Abstract

This thesis details the technical, theoretical and aesthetic concerns relating to the composition and performance of real-time audiovisual material with particular reference to the connectivity between audio and visual events. It explores the inspirations and various processes that have given rise to the attached composition portfolio. The portfolio includes the pieces *Structure Interne*, *Third World Warhol*, *Remote Control*, *Beautiful Human Canon*, *Light Speak*, *The Mudlark* and the installation piece *Aqua*.

The thesis discusses approaches to the generation of real-time audiovisual material via three distinct processes. Firstly, through audiovisual cutting, secondly through audiovisual synthesis, and thirdly, through a combination of the two. Theoretical and aesthetic goals have been approached through the use of Max/MSP/Jitter/NATO and also the Pure Data/Gem software systems. These goals incorporate an exploration of John Whitney’s differential dynamics, including the application of differential dynamics to sound synthesis, and the development of real-time cut-up techniques inspired by the work of William S. Burroughs, among others. In addition, the thesis attempts to deal with fundamental questions regarding audiovisual material, including our experience of audiovisual associations and how they impact on our perception of audiovisual art.

Finally, this thesis creates an argument for the acceptance of an audiovisual metadiscipline, that combines aspects of musical composition, film composition, experimental visual art and digital media, based on studies in perception, and Michel Chion’s theoretical approach to cinema sound.
**Table of Contents**

1. Preface ......................................................................................................................... 5  
2. Theory.............................................................................................................................. 8  
   1. Theory of audiovisual composition .................................................................................. 9  
   2. Audiovisual Relationships .......................................................................................... 14  
       The proposition - to create tools which aid in the process of audiovisual composition ... 24  
   3. Audiovisual Instrument Design: Inspiration and Architecture ........................................ 32  
       The Impact of Real-Time Computer Aided Composition .............................................. 35  
       Materials and Methods ............................................................................................... 39  
3. Audiosis and Cutting.......................................................................................................... 40  
   4. Audiovisual Sampling and Cutting ................................................................................. 41  
   5. The Political Aesthetics of WMD Audiovisual Cutting Software ..................................... 44  
3. Technical Construction: WMD Audiovisual Sampling and Cutting Software .................. 49  
   6. Audiovisual Synthesis .................................................................................................... 76  
   7. Problems of Audiovisual Synthesis: Translating Between Sounds and Images: ......... 83  
       Temporality, Translation and Scale ............................................................................. 89  
       Achieving a Congruent Translation Method ................................................................. 99  
       Light Speak ................................................................................................................ 101  
       Controlling Pattern Progress .................................................................................... 105  
   9. Audiovisual Synthesis via FFT analysis of Video Data .................................................. 110  
       Aqua ............................................................................................................................ 110  
       Technical construction ............................................................................................... 112  
10. Audiovisual Sampling Synthesis ................................................................................. 116  
   11. Structure Interne .......................................................................................................... 116  
       Technical Construction: Re-synthesising Images ......................................................... 124  
   12. Narrative Applications of Experimental Audiovisual Composition ......................... 130  
       The Mudlark ............................................................................................................... 130  
       Conclusion .................................................................................................................. 134  
   13. Bibliography ............................................................................................................... 140
Preface

This thesis is a complementary volume to the CDROM and DVD video disks entitled Audiovisual Composition. It details the inspiration and direction of the project from an aesthetic, theoretical and technical perspective. Although containing supplementary commentary and analysis, it should not be read solely as a critical review of the works on the CDROM and DVD video disks themselves. Instead, it should also be seen as an exploration of the underlying fundamentals that form the basis of the research.

The thesis comprises of two main sections. First, a theoretical section offers justification for the study of an audiovisual metadiscipline. This provides the groundwork for the entire project, including discussions on the subjects of audiovisual relationships, sound art theory, film sound theory and the impact of computer technology on audiovisual practice. In addition, it introduces a central aspect of this research - the construction of specific audiovisual instruments for the purpose of composition. This section is followed by the Materials and Methods section, which divides the study of audiovisual composition into two main areas – Audiovisual Sampling and Cutting, and Audiovisual Synthesis. Both of these areas deal with the fundamental political, philosophical and aesthetic concerns surrounding the process of audiovisual composition. With respect to Audiovisual Sampling, the philosophical debate centres around ideas of representation. With regard to Audiovisual Synthesis, there is an attempt to deal with the impact of audiovisual relationships on abstraction. Finally, this second section deals with the fusion of audiovisual representation and abstraction, and provides analysis of some of the compositions in detail.
Throughout, there is an attempt to provide equal space for both the technical and the philosophical aspects of the work. This is a direct result of the attempt to engage with both the aesthetic concerns of audiovisual composition, and also those aspects which deal with the technical, formal structuring of the material. As such, both the thesis and the works produced rise out of a convergence of interdisciplinary theory and practice, as well as an understanding of the close relationship between form and content.

The thesis references the compositional material in two main ways. First, the compositions are discussed alongside the technical and aesthetic concerns which relate to them. Secondly, where appropriate, the poetic aspects of each major composition are described in some detail. In these cases, it is important that a reader enjoys a familiarity with the works, so as to fully understand the philosophical and aesthetic arguments contained within the written material. In addition, the accompanying CDROM includes MAX/MSP/Jitter/Pure Data/GEM patches (software) which have been created in order to produce specific pieces. In some cases, these are discussed as 'compositional material', similar to the way in which one might discuss a musical score or a storyboard. Most importantly, however, the main thrust of this thesis is the development of an approach to a new type of composition. As such, the works should be viewed both in their own right, and also as part of a larger argument concerning the emerging discipline of audiovisual composition.

Finally, it is important to note that as a composer engaged in experimental image and sound creation, I have adopted a particular approach which rises out of the tradition of experimental independent and avant-garde practice. This approach, by necessity,

---

1 This would include the abstract, subjective inspirations for the material, in addition to any aesthetic visual and auditory qualities.
Mick Grierson

functions in opposition to both political and artistic aspects of conventional and mainstream culture. This is inevitable given that this project attempts to discuss and develop a new perspective on audiovisual art forms. However, it is the intention that this research should offer an approach to video, film, sound, music, and performance which will be useful in developing an effective interdisciplinary domain of knowledge that can feed directly into both the academic and industrial contexts.
Mick Grierson

Theory
Mick Grierson

1. Theory of audiovisual composition

This chapter serves to introduce the key questions and concepts which lie at the centre of audiovisual composition. It is not an attempt to describe and explain each aspect in detail, as these ideas will be developed and expanded upon throughout the text. Instead it functions as a brief survey of audiovisual composition theory in order to aid in a more explicit understanding of the theoretical groundwork and general concerns which surround it.

There are subtle but important differences between what is generally understood as audiovisual composition, and what is meant by it in the context of this thesis. Conventionally, audiovisual composition may be taken to mean the process of putting sound and music to visual material in order to create a 'soundtrack' to an existing film or animation. The study of this type of practice is often referred to as 'sound on film', 'audio for video' or 'sound design'. The theoretical debates which surround this practice are informed in particular by Michel Chion's *Audio-Vision: Sound on Screen*. However, Chion also introduces concepts which point to a more complex type of audiovisual practice. With this in mind, it is illuminating that Chion takes exception to the idea of the 'soundtrack'. According to Chion, there is no such thing as a soundtrack.

By stating that there is no soundtrack I mean first of all that the sounds of a film, taken separately from the image, do not form an internally coherent entity on equal footing with the image track. Second, I mean that each audio element enters into simultaneous vertical relationship with narrative elements contained in the image (characters, actions) and visual elements of texture and setting. These relationships are much more direct and salient than any relations the audio element could have with other sounds. (Chion 1994, p40)
Mick Grierson

Chion's justification for this view is bound up with his notion of *added value*. This is, in essence, the idea that through the combination of audio and visual elements, a third *audiovisual* element\(^2\) is generated\(^3\). This view suggests the existence of a specific audiovisual response that arises out of the synthesis of sounds and images. When experienced simultaneously, sounds and images are altered by each other, generating specific effects. Therefore, the audiovisual work is more than a combination of its component parts. As such, the practice of audiovisual composition is not simply the production of audio with video. It is the process of composing audiovisual works which exploit added value. Any meanings or effects generated by combinations of audio and visual elements (whether narrative or abstract) are explicitly audiovisual in nature, occurring as a result of specific audiovisual combinations of elements. So, it is fair to say that Audiovisual composition is a study which focuses precisely on the nature of these combinations.

This idea impacts significantly on conventional approaches to modern sound/image composition and analysis. If an audiovisual work exhibits added value, that is to say, if the specific nature of the combinations produces a separate audiovisual effect within a work, then there are grounds for assessing it on the basis of its audiovisual composition. This is very different from analysing the quality or nature of the sound or music, or similar aspects of the visual material, in isolation. In addition, it is more complex than stating that the sound or music functions 'narratively' due to cultural or other predetermined conditions. Added value depends not only on what the material is, but also where (and how) it is placed.

---

\(^2\) This could be an abstract aesthetic effect or even a meaning as a result of two intersecting sounds/images.

\(^3\) Additionally, Chion states that sound in general is not usually the main focus of an audiences attention. This plays an important role in the way in which added value functions (Chion, M, 1994, p6).
So in audiovisual composition and analysis, added value shifts the focus from separate audio and visual components, to the relationship between audio and visual components. This is particularly the case where a work has been designed from the outset with great attention to detail regarding the composition of both the sonic and visual material and their effect together, and even more so if the work attempts to exploit ideas about the formal relationships between sonic and visual material.

Computer animation and synthesis pioneer John Whitney was an artist who adopted this approach. For Whitney, audiovisual composition was the process of composing visual material using computer technology in order to exploit similarities between what he termed 'visual motion pattern' and the spectrum of sound. His theory of differential dynamics is inspired by the architecture of sound, exhibiting qualities associated with sound and music, such as harmonic resonance and noise. This idea forms the theoretical basis of almost all of Whitney's works, including his Matrix trilogy (1971-72), and Moondrum (on Twelve Works for Videodisc (1989-1995)). Through digital synthesis, Whitney was able to create an abstract visual experience which he argued was similar to the general experience of sound and music. In the case of his earliest analogue films (Five Abstract Film Exercises (1940-1945)), there is a complex compositional connection between the sound and the image. Whitney's later digital audiovisual compositions are mainly visual composition inspired by the formal structure of sounds and music. Despite this, however, it was his goal to produce a digital system for inter-connected audiovisual composition via synthesis.

For a more complete description and application of differential dynamics, see chapter 7.


Although these works have soundtracks, there is no exact formal relationship between the image and the audio. This may be a result of practical necessity rather than choice. Whitney’s animations were enormously complex projects.
Mick Grierson

As an attempt to understand and operate within the audiovisual context, Whitney's audiovisual composition sits comfortably alongside Chion's notion of added value, and together they provide a firm direction and context for this research.

Thus, audiovisual composition in the terms defined here is the development of works which are explicitly created to exhibit various degrees and aspects of added value. More specifically, the works created and cited within this thesis attempt to forge and/or exploit formal and aesthetic relationships between sonic and visual events in order to explore added value via creating connections between the two forms. The success or failure of this project could possibly be judged by the degree to which added value is produced i.e. the degree to which the artworks generated demonstrate an enhanced audiovisual experience beyond that which can be expected from a conventional film sound/music approach. Additionally, the works should not be regarded as attempts to merely visualise sound or musical material. The successful audiovisual composition is the result of a combined audiovisual composition process.

A factor which feeds into this process is the technical difficulty involved in controlling, manipulating and connecting sounds and images simultaneously. In order to facilitate the study of audiovisual relationships for the purposes of creativity, it is desirable to be able to treat both types of material in the same or a similar way. Digitisation allows for this possibility. However, through the use of digital technology, all sounds and images can be reduced to binary data. Provided that the structure of the material can be correctly interpreted, this data can then be manipulated using similar processes, and the results compared. In addition, audiovisual material can be digitally synthesised from scratch –

This alone does not mean that there will be an audiovisual connection at the perceptual level.
that is, sounds and images can be created from abstract formal and structural relationships.
2. Audiovisual Relationships

The relationship between sonic and visual material is complex. Essentially, both sonic and visual material yield effects of their own when experienced in isolation. These effects are difficult to discuss in themselves, let alone in combination. Chion states that when combined, new effects are apparent. This makes the process even more complex. Despite this, difficulties regarding the interpretation of audio and visual material do not prevent artists from exploiting that material. However, audiovisual composition which exploits structural relationships by its nature rises out of a desire to understand the combined audiovisual effect. This research offers an argument for promoting the interdisciplinary study of audiovisual composition as a super, or metadiscipline in itself that is neither Art/Film studies nor Music/Sonic arts.

Psychomusicology and psychophysics are two areas where recent research appears to back up Chion's view regarding added value. In ‘The Perception of Audio-Visual Composites: Accent Structure Alignment of Simple Stimuli’ (Lipscomb, 2004), Lipscomb’s results suggest a correlation between the synchronisation of audiovisual composites and the perceived effectiveness of the material. Where audio pulses correlate strongly with visual events, audiences (organised in ability groups according to background and training) sense a greater effectiveness.

Calculation... revealed that subject ratings of synchronization and effectiveness shared a strong positive relationship (r = .96). Therefore, AV combinations that were rated high in synchronization also tended to be rated high on effectiveness and vice versa. (Lipscomb, 2005 p60)

Defining exactly what is meant by 'effective' in this context is problematic, although for the purpose of this thesis it will be assumed that it equates to the memorability and/or subjective preference of an audiovisual event when compared to other less accurately synchronised events, as this appears to be Lipscomb’s distinction.
This research shows that audiences perceive closely synchronised material as being more effective. However, Audiovisual works which exhibit strong structural links in this way are sometimes referred to as being guilty of 'Mickey Mousing' – the exact, and by implication, simplistic synchronisation of visual and sonic events. In his paper 'Insects, Urine and Flatulence: On the Radical Potential of Mickey Mousing' (Birtwistle, 2002), Birtwistle mounts a robust response to this criticism:

The close matching of musical sound and image is seen in negative terms: "... because of the implication that exact illustration is a rather tedious and silly way to relate music and image." (Curtis, S in Altman, R 1992, p201) Mickey Mousing is poor practice. It is considered unsubtle, unnecessary and creates humour when none is required...(however) Mickey Mousing punctures the bubble in which western music has placed itself, forcing an acknowledgement of an 'outside', an other: in this case, the visual. Not only does Mickey Mousing destroy the notion of an isolated specificity, of an abstraction from all else, but it also introduces ideas of other kinds of structuration, other ways of considering structure, other ways of thinking music, and other ways of thinking about music. (Birtwistle 2002 p26)

So, according to Birtwistle, there is a desire to maintain a separation between musical and visual art which rises out of existing traditions. Criticism of closely synchronised material emanates from this perspective. In *Analysing Musical Multimedia* (Cook, 1998), Nicholas Cook argues that multimedia is 'predicated by difference' (Cook, 1998 p56), and the 'duplication of information across sensory modes' (Cook, 1998 p41) cannot be described as multimedia. If one is willing to take seriously the results of Lipscomb's experiments, it seems that it may be unwise to dismiss the effectiveness of closely synchronised material. Audiovisual composition, in fact, may rely on an understanding of this 'effectiveness' and the complexity of its operation. Perhaps Cook is right, and it is not evidence of multimedia. However, this does not mean that it is unsophisticated or lacking in value.
In the conclusion to his paper on Audiovisual relations, Lipscomb makes the point that although there is a proliferation of abundant audiovisual material in our everyday lives, very little research is being carried out to analyse the dynamics of the material.

..given the sociological significance of the cinematic experience, it is quite surprising that there is still only a small amount of research literature available addressing issues involved in the cognitive processing of ecologically valid audio-visual stimuli. (Lipscomb 2005 p65)

In addition to Lipscomb's work, recent research at the Shimojo Psychophysics Laboratory underlines the need for further research regarding audiovisual relationships. In 'Visual Illusion Induced by Sound' (Kamitani, Y, Shimojo, S, 2002), proof that audiovisual material is processed in combination, and that this combination alters the perception of the material with definite effects has been confirmed by MRI scans. In one experiment, a subject sees a black dot appear on a screen for one frame. When accompanied by two small blips, the subject sees two dots, one after the other, even though there is only one. This suggests that strongly synchronised material is effective in producing a type of experience which is distinct from the experience of images or the experience of sounds in isolation from one another. As such, it provides evidence for Chion's notion of added value, whilst underlining the fact that the structural relationship between sound and visual material is at the heart of audiovisual composition. In this way, the pejorative term 'Mickey Mousing' lacks authority with respect to the discussion of audiovisual composition, and although the process of composition may shift to and from heavily synchronised material, the effectiveness of synchronisation could be, as Birtwistle argues, truly radical.

As has been stated, this research sits well with the terminology developed by Chion. Heavily synchronised material (such as sound effects) can alter the way in which an
audience perceives an object on screen. Chion identifies the term *Synchresis* (Chion, M. 1990 p 58) to describe the fusion which occurs when sounds and images collide in close temporal proximity. He notes that a sound effect can be used to change the perception of an object’s behaviour. Chion uses the example of electronic door sound effects in the film *Star Wars*. The sound lends aspects of its amplitude, frequency and timbre to the door. It also alters the perception of the door’s movement – the door does not move, but the sound creates the impression of movement. An even simpler example would be that of a punch or gunshot. As is well known, sound effects in films can often be greatly exaggerated. Louder, dense, lower frequency sounds make punches, gunshots and explosions seem larger. Smaller, higher frequency sounds will make them seem unnaturally small with respect to the interpretation of the event, and perhaps even comic.

Early computer aided AV scratch artists, such as Emergency Broadcast Network (see *Ballistic Ordinance Beat*, EBN 1991) utilised similar strategies to great effect in their powerful polemical artworks. In this way, it can be seen that Chion’s terminology not only has uses in describing aspects of narrative cinema, but also helps to illuminate aspects of modern avant-garde audiovisual and graphic design. The title design work of Saul Bass is well known for its power and impact, often being able to generate suspense, tension and narrative contextualisation without the use of words or explanatory text. Good examples of this type of semi-abstract sonovisualisation are the title sequences for Hitchcock’s *North by Northwest* (1959), and *Vertigo* (1957 – titles in collaboration with John Whitney). In both cases, Bernard Herrmann provides music. Certainly, in the case of *Vertigo*, the audiovisual relationships are especially dense and complex. Sounds and images develop concurrently

---

9 EBN also specialised in deconstruction and re-contextualisation, see chapter 4.

Mick Grierson

and are bound together in such a way as to generate a coherent aesthetic identity. This is the additional implication of Chion’s added value. It is a technique of combining sounds and images in order to generate a third audiovisual form, which modern corporations, advertisers and media companies (such as TV stations) utilise heavily in order to create aesthetic identities. The combination of time related, or even heavily synchronised abstract/graphic (and in many cases, synthetic) material with sonic material (including music) is a common technique for creating audiovisual iconography. Film distributors, TV companies, and other media organisations have audiovisual logos which help propagate their chosen image (the U.K. Channel 4 logo being an excellent example). As Lipscomb states, the power of audiovisual relationships is being overlooked within the realm of research, while Audiovisual relationships remain at the heart of much modern communication and commerce.

Another important element in the analysis of audiovisual relationships is the idea of audiovisual congruence and its affect on memory. Research suggests that the nature of a sound can affect the memorability of a visual event (Boltz 2001). Congruous sounds and images occurring simultaneously result in greater memorability. In addition, incongruous sounds and images occurring consecutively also result in greater memorability. Simply put, moments of synchronisation have a greater impact on the viewer when images and sounds are congruent, and diversity in dynamics helps to maintain effectiveness and interest in audiovisual work. This resonates with Chion’s theory of added value, and also seems to conform to Lipscomb’s findings. Again, both Chion and Boltz are mainly concerned with the congruence of culturally recognised (non-abstract) sounds and images. However, as is the case with Lipscomb, this idea can be extended and applied to abstract

The criteria for congruence in Boltz’s article are generally cultural – e.g. dogs bark, babies cry etc. Ways in which one might identify congruence between abstract material is with reference to structural similarity – see below.
audiovisual components.

In the first instance, any abstract sound can be attached to any abstract image through a process of association, and this generates a simple cross-modal connection. That is to say, for the duration of the piece, a particular sound can be linked to an image or a set of moving images. However, according to Boltz, the memorability of the combination will depend ultimately on the level of congruence between the two types of material at the moment they are joined together synchronously. Problems arise when we consider that with abstraction, there may be little or no cultural association aiding the re-enforcement of the relationship between the sound and the image. But, if we think of abstract congruity as structural similarity, we can start to see abstract sounds and images as having congruency. The extent to which an abstract image can be understood as being similar to a sound is entirely to do with relative structural aspects of the sonic and visual components. That is to say, the structure of the sound can be shown to be similar to the structure of the image. For example, a sound which has a high ratio of unrelated overtones to harmonic ones can be described as 'noisy'. Likewise, an image which exhibits less order, less self similarity, and more incoherent complexity could also be described as noisy. In this way, we can begin to think about sounds and images in a similar way with respect to texture (or shape) and timbre. This provides us with a method for discussing the congruence of sound and visual events in the context of abstract audiovisual composition.

This criteria for discussing audiovisual relationships is also detailed in *Perceptual Correspondences of Abstract Animation and Synthetic Sound* (Abbado, A 1988). In his paper, Abbado details his work at MIT on the relationship between abstract computer generated sounds and images.
A sound can be abstracted as an aural object, as something that may be complex but has a precise identity. This identity is defined by the sound's spectrum, its energy. Similarly, visual events can also be considered as independent objects. The shape and the kind of surface define a visual object's identity. Timbre in sound, and shape-surface attributes in images, are the most powerful perceptual determinants. Their correspondence should, in my opinion, form the basis of any new sound-image language. (Abbado 1988)

This resonates with Whitney's Differential Dynamics (although there are differences relating to Whitney's interest in patterns as opposed to shapes and surfaces). However, it goes further in terms of outlining possible criteria for the identification of sound/image congruence. Here, Abbado is stating that there might be such a thing as a 'sound-image language', or a language of audiovisual composition, separate from (but informed by) the languages of sound/music and visual composition.

Abbado finds support for his hypothesis that the experience of audiovisual language can be dealt with as separate to the experience of sounds and images in isolation, by referencing experiments investigating the 'superior colliculus', a multisensory organ which exists in humans and other animals. He quotes the following:

"For example, if some multisensory cell responds to a light flash in the upper right portion of the visual field, that cell will respond to a sound only if it too comes from the same vicinity. Additionally, when visual and auditory inputs occur simultaneously, a multisensory cell responds more strongly than when either input occurs alone." (Sekuler, R. & Blake, R. 1985, p104).

In addition to more recent experimentation into cross-modality (Kamitani, Y, Shimojo, S, 2002), this research demonstrates that multisensory cells in the brain respond directly to audiovisual experience, as opposed to simply audio or visual experience, and that these
cells respond more strongly when events occur simultaneously and appear to come from the same source. This gives further support to Chion’s ideas, as well as the research carried out by Shimojo, Boltz and Lipscomb. In addition, it helps to reinforce the case for the study of the metadiscipline of audiovisual composition.

Synaesthesia has long functioned as a marker for the existence of connectivity between sonic and visual material. Nicholas Cook describes art which rises out of the interest in cross-modal perception as ‘synaesthesia by proxy’ (Cook, N. 1998 p49). Debates regarding the universality of pitch colour relationships in synaesthesia are inconclusive. Importantly, however, the structural approach to the examination of audiovisual relationships and multisensory perception provides stronger material than that inspired by pitch-colour synaesthesia. For example, V.S. Ramachandran’s Hearing Colors, Tasting Shapes (Ramachandran, V.S. and Hubbard Ed, 2003), describes many different ways in which multisensory cells in the brain can cause sensory confusion. This challenges the common conception of synaesthesia as centred around pitch-colour relationships, and offers us new ways of approaching the analysis and creation of audiovisual material. As such, this research project chooses not to focus on fixating harmony/colour relationships, favouring the congruence of structural audiovisual attributes as a means for controlling and relating to the memorability or effectiveness of audiovisual material. This approach shares similarities with the general line of research undertaken by John Whitney.

In his Complementarity of Music and Visual Art, Whitney states that the visual instrument he desires to create (which is in part the inspiration for the visual synthesiser developed as

---

12 Although there is no proof of the experience of people with synaesthesia being in any way universal, trends can be discerned between certain shapes and colours with very large samples (Harrison, J 2001),
part of this research) cannot and should not be related to a conventional scale system.

This is vital, as it is a pointer to what Whitney thought of as perhaps the most important factor in the relationship between our experience of sounds and our experience of images. In Whitney's opinion, there is no such thing as 'a certain colour for a certain note'. He saw sound and vision as fundamentally different phenomenon. However, Whitney intuitively felt that there were experiential factors binding audio and visual material together.

This is the belief which led Whitney to base his patterning system on the Pythagorean principles of harmony. In order to perceive a pattern, the elements which make up that pattern must appear in close temporal proximity. That is to say, enough of the pattern should be visible for the symmetry it creates to be apparent. So, according to Whitney, it is more accurate to say that a visual pattern is to the image, as a timbre is to a note (the timbre of an instrument having its own resonant 'harmonic' characteristics). The visual pattern (or texture) is like timbre. The implication of Whitney's conclusion is that individual note pitches are not as important as the timbre of sounds with respect to connectivity between sound and visual events. Likewise, individual colours themselves are not so important when exploring audiovisual relationships. This approach is re-enforced by Abbado's hypothesis, that there is a relationship between the texture of an image, and texture of a sound. The main difference between this view and Whitney's Differential Dynamics is that in Whitney's research, it is patterns,\textsuperscript{13} and not surface textures which are used to reflect resonance and dissonance.

However, it needs to be noted that all of the above research deals solely with the

\textsuperscript{13} Although it is fair to say that patterns can be seen to have textural qualities in general, particularly when they become more complex.
relationship between sonic and visual events in isolation from a whole work or context. It does not deal with the movement within a composition from one state to another. Nor does it deal with a composition as it is experienced in its totality. With respect to this, it is worth noting that congruence defines only one end of the spectrum with regard to audiovisual language. In order to provide space for tension and release within an audiovisual work, binary opposition or total incongruence is necessary. This opposition can be effective in pulling a sound or a visual component away from its structural direction, providing a balancing force against congruence, and allowing for movement between tension and release. Opposition in terms of rhythm, texture, volume (amplitude) etc. between sound and image can be shown to have various notable effects. Maintaining a static rhythm against a moving image, according to Chion, can have two possible results. It can become boring (and therefore unmemorable), or alternatively (perhaps depending on other aspects, such as variation in timbre), it can have a hypnotic effect, altering an audience's perception of time. Additionally, flickering images can be made more relaxing with slowly developing or rhythmically simple sonic activity. These effects are expressed in part by Chion's concept of temporal animation. Sounds can temporalise images, providing them with a sense of pace and duration.

Far from undermining the importance of congruence, binary opposition reinforces the idea of deep structural relationships between sounds and images. Incongruent aspects of audiovisual language may function as a result of an apparent lack of congruity. That is to say, although the journey from congruence to incongruence could itself generate tension, it is the movement towards binary opposites (such as harmony and noise) which could provide intensity. Lingering in between these states might be hypnotic, but may not induce the memorability or effectiveness of specific events. The lack of memorability or
effectiveness in an audiovisual composition may not be a negative however, depending on the purpose of the work (for example, material used to facilitate hypnosis). This idea bears some relation to the model proposed by Nicholas Cook. According to Cook, audio and visual material combine in one of three main ways: They either a) conform to one another ie. the relationships are congruent, b) contest one another i.e. the material is in binary opposition, or c) complement one another i.e. the materials 'are generally aligned with one another'. Cook states '..each medium elaborates the underlying structure in a different way' (Cook, 1998, p.102). The varied effectiveness of these states could be said to be useful in generating contrast within a work.

**The proposition - to create tools which aid in the process of audiovisual composition**

In order to achieve the goals set out in this research, specific tools are needed which allow for the digital manipulation of sound and image data. In addition, these tools should allow the composer to engage with the material in a way which is as flexible and responsive as possible. Developing tools which address the above requirements was one of the key focusses of this research. What follows is a brief survey of how this has been approached, the equipment which has been used, and the general philosophy behind the direction of the project.

The relationship between a task and a tool is interdependent. Instruments evolve as art forms make demands of them. Additionally, the creative task develops with the refinement and evolution of the tool. In one sense, it is true that the desire to create a set of tools for
Mick Grierson

use in composition comes about through a desire to improve both the materials available to the artist and the process of art itself. In another sense, the creation of new tools and instruments makes entirely new art possible.

In the case of audiovisual composition, rapid technological growth is providing more sophisticated resources. New real-time environments for sound, moving image multimedia and intermedia\(^{14}\) production are becoming available. Due to the fact that the data throughput of digital audio is significantly less than digital video, this revolution has been more quickly realised on a larger scale within the world of sound production. Now, software is evolving to incorporate video in the same manner. The digital sound and music platform Max/MSP (derived from Miller Puckette's *Patcher* (Puckette, M. 1988)), has recently been enhanced by the addition of 'Jitter', a set of real-time video objects, partly inspired by Netochka Nezvanova's innovatory NATO software for Max/MSP, which allows for control of Apple Quicktime video and 3D OpenGL data. However, both libraries are pre-dated by Mark Dank's GEM external for Miller Puckette's Pure Data, created as part of the Global Visual Music Project\(^{15}\). This audiovisual research project makes use of all these programming environments in one way or another for the creation of audiovisual composition tools.

Before adopting an approach to instrument development, certain questions regarding the relationship between tool and task must be considered. What is the relationship between composer and tool? How does one influence the other? How does the nature of the tool predetermine the type of work produced with it, and in what way do the choices made by

\(^{14}\) Definitions vary on the difference between Intermedia and Multimedia. Here, the term 'Intermedia' refers to inter-related media material.

\(^{15}\) A project funded by Intel and Yamaha among others, supporting the development of Pure Data/Gem. See http://visualmusic.org/gvm.htm
instrument creators reveal a desire to achieve a specific aesthetic?

It can be reasoned that throughout history, instrument development has been at the centre of all composition. In the case of art forms such as vocal music and dance, the training and physical modification of human beings is itself a type of instrument development (the castrato being an extreme case). With respect to visual media forms such as television, the development of video tape can be seen as an invention which stimulated significant aesthetic and creative activity (Williams, R 1974 p139), which aided in a redefinition of how media technology should develop. In this way, all art relies on an interaction between technological and artistic development. The interplay between the art of technology and the technique of art is central to the development of an aesthetic. This argument is advanced in more detail later with specific reference to the composition portfolio. However, it should be briefly explored here.

The construction of stringed instruments stems from the observation that agitating a length of a particular type of material causes vibrations in the air. Strung keyboard instruments, such as the piano, are good examples of design informed by this observation. They reliably produce frequencies throughout the various ranges associated with other tuned acoustic instruments – therefore they can function as a model for other instruments (as can many keyboard based polyphonic instruments). In the case of the piano, it has a highly controllable and versatile dynamic range. In addition, the piano provides subtlety in terms of timbre through the use of padded hammers. This allows for an elegant manipulation of what could be perceived as the three basic elements of sound (frequency, amplitude, timbre). The quest to generate instruments which can aid in controlling natural forces facilitates composition and performance.
However, arguments regarding the tuning of instruments such as the piano are thousands of years old, and a compromise has only been reached relatively recently. The Pythagorean comma is the mathematical ratio representing the offset from the starting note after completing a pure cycle of fifths. It is 1:2.0273. This is approximately one quarter tone (Sethares, W, 1999). Aristoxenus put forward the argument that through the use of tetrachords, one could approximate the intervals within the existing 12 note structure (Aristoxenus, c330 BC) in order to make it more pleasing aesthetically. This suggests that the Pythagorean mathematical harmonic scales do not function 'aesthetically'. There may be many people currently working within the discipline of synthesis who would disagree with this idea. Most importantly, however, the changes were necessary in order to make transposition on instruments possible, allowing for complex harmony and melody to be transposed and remain ostensibly identical in terms of frequency ratios.

In applying this idea to the development of an audiovisual instrument, several lessons can be learned. First, that a well structured instrument or set of instruments should be able to generate as complete (in terms of range) and as dynamic (in terms of variation) a manipulation of the fundamental elements of vision and sound as is currently physically possible. Secondly, that any inefficiencies regarding the relative structure and relationships between sonic and visual material are only as important as the impact of their effect both aesthetically and practically – that is to say, just as tuning systems have undergone revisions which adapt the structural relationship between the harmonic series and the scale, an audiovisual tool needs to focus on aesthetic and practical aspects as opposed to entirely structural ones in order to function more effectively. Thirdly, that an instrument can often be judged by its ability to aid in the production of effective and innovative works.
Often, the development of a tool depends on what can be seen to be done with it. The development of the synthesiser is a good example. Robert Moog, a technician employed by composer and engineer Raymond Scott, took Scott's principles of synthesis and applied them to create the modern subtractive synthesiser. In this case the central desire was to more completely manipulate the elements of music – just as in the case of the piano. The subtractive synthesis model which Scott and Moog developed is derived from the physical properties of sound. As such it comprises of waveform oscillators (for low frequency pitches and high frequency harmonic generation), filters (for generating and removing high frequency timbral content of a sound), amplifiers (for controlling volume) and transient generators (for the manipulation of all the above over time). In principle, this allows for a high degree of control, and it is for this reason that the subtractive model is one of the most widely implemented synthesis models in use today.

The composer Walter Carlos collaborated with Robert Moog in order to create 'Switched on Bach' (1968), an attempt to show that synthesisers can be an adequate replacement for acoustic instruments. However, it was not until much later, when the subtractive synthesis model was combined with sampling technology, that synthesisers could effectively model the timbre of real sounds with the aid of digitisation. Even so, the modern popularity of this approach is not solely due to its ability to emulate acoustic instruments. It is also due to its ability to create sounds which have not been experienced before. As opposed to the conventional interpretation of the word 'synthetic', whereby an existing 'thing' is emulated through technical means, synthesis in this case is synonymous with the generation of new sounds (and visuals). As a product of the desire to exploit the elements of audiovisual experience, the audiovisual instrument draws inspiration from this
interpretation of the term 'synthesis'. That is to say, its output is not just imitative or representational. The images and sounds which it produces can be entirely new.

So, audiovisual composition involves the generation of essentially abstract sounds and images via synthesis, as well as the re-production of existing sounds and images via sampling. An efficient and reasonable way of achieving this goal is through the use of modern computer music and video tools. The relationship between sound and visual synthesis coupled with the development of an appropriate control mechanism for composition and/or performance is vital in the development of the instrument, as is the ability to perform, record and replay with relative levels of congruence between sound and image. This can only be achieved efficiently through the use of computer technology. These factors in themselves suggest that there may be a general aesthetic effect apparent within the work due in some part to its exploratory nature and the use of computer technology in order to facilitate synthesis.

It is already the case, and may be more so in the future, that software manufacturers and developers of composition tools are themselves seen as artists, as the effect which they have on aspects of composition becomes more complex and intertwined. This is already evident with certain types of software such as Propellerhead's *Reason* and *Reason 2*, and even more so with software such as *Reactor*\(^\text{16}\). The desire to synthesise the physical elements of sound have expanded the palette of possible tones available to the composer, just as the palette of possible movements made available by computer animation have revealed new territory in terms of representation. In this way, the more definitive a technological technique becomes, the more it could be said to be representative and

\(^{16}\)From Native Instruments – the real time time-stretching software from Native Instruments is called *Kontakt* – one can be safe in the assumption that references to Stockhausen are intended.
perhaps even responsible for a certain manner of composition. The extent to which
computer technology influences creative decisions can only be revealed by the process of
computer composition. Interestingly, artists/programmers such as Miller Puckette, David
Zicarelli and Netochka Nezvanova may have always been aware of the potential impact of
software design on the aesthetic of the composer, in terms of the restrictions that software
may impose on a user's creative practice. At the recent IRCAM Resonances festival
(October 2003), both David Zicarelli and Miller Puckette agreed that their programming
environment was intentionally aesthetically bland so as not to enforce a particular
composition style\textsuperscript{17}. In support of this, it is fair to say that Max/MSP/Jitter and Pure
Data/GEM are good examples of software which impose minimally on the aesthetic of the
composer. This is because they are graphical representations of high level programming
languages, having being designed in order to simplify the development of custom tools by
each user.

PD/Max\textsuperscript{18} has been designed to most closely resemble the C family of programming
languages. This facilitates its use as an application development environment, and allows
it to offer the same or similar functionality as command based languages. PD/Max can
also execute commands and statements like a traditional command based language. In
addition, PD/Max users are not immediately provided with a functional environment within
which to create musical or visual works. This environment must be developed entirely by
the user. This is the same 'blank canvas' approach offered by development environments
traditionally associated with computer programming. It is extremely flexible, in that
theoretically, most computer music processes of which a composer can conceive can be

\textsuperscript{17} As far as I am aware this debate is not documented, and I can only report the statements from my own notes.
\textsuperscript{18} The term 'PD/Max' is used to describe both Pure Data (and related libraries), and Max/Msp/Jitter when dealing with
elements which both platforms share, such as the fundamental paradigm which governs their design.
Mick Grierson

achieved\(^{19}\). However, this flexibility comes at the expense of usability. Pd/Max platforms do not do anything 'out of the box'. At the Resonances festival a commentator described feeling 'raped\(^{20}\)' by the user interface (presumably a reflection on the 'blank canvas' approach of PD/Max). It could be argued that this is an inevitable factor of its enormous functionality, and that if one attempts to make software 'transparent' and 'simple' to use, then inevitably the software itself is less versatile. Beyond this, the process of designing and building a synthesiser brings with it the inevitable relationship between the technical and the aesthetic, i.e. the output of a device is defined by the technical construction of the device. An instrument designer can then dictate the aesthetic by controlling the output type. This explains the main motivation for adopting PD/Max as the only environment suitable for the development of this research – complete control of all factors which relate to the aesthetic of the material.

With this approach, the landscape is open for the development of tools which continually question and redefine the relationship between composer/performer and instrument/technology in both the sonic and visual artistic worlds.

---

\(^{19}\) Given the range of possibilities provided by the hardware. The software gives the user a degree of control similar to that which one would have when using any other application programming interface such as C or C\(^++\).

\(^{20}\) A rather subjective response brought on by the confusion of having too many options, none of which were readily acceptable to the subject.

With respect to audiovisual instrument design, there are three main ways in which this research draws inspiration from the world of real-time sound technology. First, digital sampling provides a model for the playback of material. Sampling synthesisers allow for the real time manipulation of this material, providing an interface whereby the 'real' can become 'the abstract' through real-time synthesis. Secondly, within the field of real-time sound synthesis, instruments emulate the actual physical properties of the sound in an attempt to generate material. That is to say, the devices themselves are capable of producing sounds which have no origin in the natural world. Instead, they manipulate the parameters of the natural world in order to create sound. Thirdly, real-time control allows a composer or performer to develop ideas as they occur. The operator can interact in a more dynamic and immediate way with the instrument. Comparatively, this approach is somewhat of a novelty in the world of moving visual art, although it has always been a feature of music\(^21\).

The most obvious use for sampling is playback on demand. Once a sound has been digitised, it can be replayed, but with some, if only minimal delay. This is useful for both the immediate and controllable replay of complex sounds which are difficult to reproduce/synchronise, and also for collage. Schaeffer's early *Music Concrète* works make use of collage techniques. Material is acquired, cut up, manipulated and replaced in order to generate new material. Real-time sampling takes this idea one stage further. Through digitisation, the techniques employed by Schaeffer can be used to generate similar works in real time. Sections can be replayed at various speeds in any order in a

\(^{21}\) Although historically, there have been many attempts at ‘colour organs’, including Louis-Bertrand Castel’s Occular Harpsichord, originally constructed in 1734.
non-linear fashion. In addition, material can be immediately reversed, and subjected to complex Digital Signal Processing (DSP) techniques, such as Granular Synthesis. None of these functions are available to the composer with tape.

Interestingly, both collage and pastiche (in terms defined by Jameson, F. 1991) are generally referred to and understood as 'sampling' within popular culture. In this context, sampling refers to material being captured and juxtaposed to create new material rather than a simple description of the technique of sampling. This shows the immense effect that the technology of sampling has had on culture. Its technical term has become a description for a phenomenon of culture itself. This interpretation of 'sampling' is related directly to the idea of Collage, and related to the concept of montage in the work of Sergei Eisenstein and Lev Kuleshov. Aesthetically, however, there is an essential difference in that Eisenstein, like Schaeffer, generated the content used within the montage specifically (as is the case with the Kuleshov Experiment). This is where a distinction can be drawn aesthetically between pastiche and collage. Even in cases where material for composition comes from other sources, the extent to which the material can be seen as derivative sets it apart as pastiche. Therefore, where works are derived from material which already has cultural currency, significant manipulation is necessary for it to function as collage, as opposed to pastiche. One way this can be achieved is through techniques of synthesis. The audiovisual instrument created as part of this research attempts to create a form of audiovisual montage/collage through a combination of manipulation and synthesis that moves beyond the realm of pastiche (see 'Audiovisual Sampling Methods').

The paradigm of synthesis is applied in the audiovisual instrument through both the

---

Mick Grierson

generation of material from scratch, and also the application of synthesis techniques to
sampled material. In a common real-time audio sampler for performance, a sound can be
treated in the same way as an oscillator within a subtractive synthesiser. That is to say,
Transient generators and filters can be applied to the sound sample. The sample becomes
a wave table for use in synthesis. Taking advantage of this idea, the audiovisual
instrument takes both sound and visual material and applies synthesis techniques to it.
Both audio and visual synthesis techniques are applied to both types of material, with the
visual synthesis engine utilising the video sample as a texture. Likewise any audio material
can be treated in the same way at the same time through audio synthesis methods.
Aspects of the visual texture can be used to create new sounds or interact with existing
sounds and vice versa. Patterns generated by Whitney's Differential Dynamics can be
used as a map for various implementations of synthesis, including subtractive, additive
and physical modelling. This generates a high degree of congruity in what can be referred
to as the spectral movement of a sound, and the visual patterns being generated. These
patterns can also be used as a basis for re-synthesis23. In addition, the shape of an object
can be directly related to the structure of a sound. For a further description of this method
see Chapters 6 and 9.

Thus, the practice of audiovisual composition as proposed in this research combines two
distinct approaches inspired by both cinema and sound theory. First, there is the concept
of audiovisual sampling and cutting. Real world sounds and images are captured,
processed, and re-constructed in real-time in order to generate new material. Secondly,
there is the concept of audiovisual synthesis. Sounds and images are generated in a way
which seeks to exploit possible structural relationships on a formal level. These two

23 See chapters 7 and 8, synthesis based on visual patterns
approaches are then combined to form the final performance instrument, where audiovisual material (samples) can be treated via a process of audiovisual synthesis.

**The Impact of Real-Time Computer Aided Composition.**

Composing with computers has become a real-time activity. That is to say, a computer can now be treated as an interactive instrument, like a piano or a guitar. The impact of this is considerable when one considers the way in which computer composition is achieved. Computer aided composition functions through the use of algorithms. Sounds (and images) are synthesised and developed over time according to a specific set of computational rules. The success or failure of a particular piece may depend on the extent to which the algorithm matches the composer's desire for a particular aesthetic. As is discussed later, an audiovisual cutting mechanism which produces extremely noisy output due to the calibration of its algorithm is successful in so far as it is useful to a composer wishing to produce that particular type of effect. Sometimes these algorithmic rules can be extremely simple, governing the form and content of an entire piece. Likewise, through visual synthesis and cutting/re-synthesis techniques, images can be generated alongside sounds, developing stylistically and in a compositional relationship with sound signals. This is a useful aspect of composition achievable via computer technology. Systems can be developed which encompass both sonic and visual development through patterns.

One aspect of this which is important is the use of loops, either of sampled material or of controller data generating iterations and mathematical functions. Pieces can be made up of various looping sections which have specific sound and visual components. In the case
of synthesis, pieces can rely on algorithms for the generation of their entire content. In certain types of visual synthesis, such as fractal generation, a loop is simply a mathematical iteration, placing a single dot on a field of blank pixels\textsuperscript{24}. Fundamentally, however, now that an algorithm can be performed by a computer in real time, the composer's approach to the algorithm can change. An algorithm need not be calculated prior to performance. The analysis of an algorithm need not wait for the entire program to run. On the contrary, an algorithm can be instantiated and manipulated by a performer as a piece unfolds. If a composer doesn't like the output of an algorithm, it can be changed whilst it is being calculated to achieve a different effect. Therefore, algorithms can now offer a palette of real-time possibilities for the composer, in the same way that a scale offers a palette of possibilities to Jazz saxophonists such as John Coltrane or Ornette Coleman.

Technology now offers digital artists the opportunity to place themselves within the flux of the creative moment in order to draw out those hidden opportunities that only occur in that moment. This revolution follows on from the revolution of recording. With the creation and subsequent adoption of the tape machine by both the music industry and composers alike, composers and audiences were able to review compositions again and again. In addition, Pierre Schaeffer found that the tape machine was an instrument in itself, with functions which could be programmed and replayed. The need for both instrument and score was dissolved – the tape machine could be both at the same time. This is one of the most important realisations in electronic music, and has led to an enormous revolution in the way we perceive musical and sonic arts.

\textsuperscript{24}In terms of audiovisual synthesis, there are occasions when neither the sonic nor the visual aspects of a piece can be generated on its own, as without the other they would fail to exist at all (see Aqua p96).
Mick Grierson

Film technology did not revolutionise image-based art forms in this way. Instead, it created a temporal image-based art form where there had been none before. The way in which Film and Video artists compose their works is varied, but within the conventionally accepted methods the following elements exist: A work is written as a script or outline; the script is turned into a story board; it is broken down into a variety of shots; the shots are taken; the work is 'cut together' to resemble the story board; and finally the work is shown. Here we have a modern compositional process which is echoed in music and sound composition. The work is scored, arranged, performed (and perhaps recorded). However, if one is to attempt a progressive composition technique, then one could proceed to edit the work in order to improve upon it. Also, one could dispense with the score entirely, proceed to the performance stage, and then edit in order to produce a work. This is the technique that Schaeffer employed. It is also a technique used by Stockhausen. Interestingly, Francis Ford Coppola adopted a similar compositional approach when creating *Apocalypse Now*, and this technique of composition (or film making) is central to the film works of the Dogma 95 group (see Trier, L V *The Idiots* 1998).

So, technology offers new aesthetic possibilities, and forces us to address new compositional methods. An understanding and engagement with the implications of real-time composition is essential to the development of Audiovisual composition, most importantly because the very nature of combining interconnected sound and visual algorithms with human performance characteristics often demands that they are executed and recorded in real-time. Sounds which depend upon highly detailed developmental visual processing cannot be predetermined in the traditional way. This process requires a new perspective which embraces the world of real-time digital technology.
Materials and Methods
As has been discussed, many tools exist which allow for the creation of audiovisual work, but only a few are suitable for real-time experimentation and development. Of these, only two offer the ability to interrelate audio and visual data in a flexible and interesting way suitable for this research. Almost all other existing software platforms are restrictive in that visual elements can not be sufficiently inter-related or cross-synthesised\(^{25}\). Additionally, existing non-linear editing techniques do not provide either the real-time interoperability (between sonic and visual elements) or the absolute control necessary in order to develop new techniques. This research has therefore explored audiovisual cutting techniques using both the Pure Data/Gem platform and Max/MSP/Jitter. After reflection on the results, these techniques have been redeveloped and applied in composition. The finished applications form the basis of two software suites, one entitled 'Ariel' (generated specifically for the composition *Remote Control*), and the other 'Weapons of Mass Deconstruction' (WMD used to create the work *Third World Warhol*). Both are tools for creating specific types of audiovisual material, being crafted in such a way as to generate work that manifests a particular type of aesthetic.

\(^{25}\) Conventional video editing software such as Avid and Final Cut Pro has limited audio editing capabilities. Likewise, Pro-Tools has flexible audio editing, but no control over the visual information whatsoever. In addition, no conventional software allows for sound and image data to cross-modulate, cross-synthesise, or co-develop.
Remote Control

Remote Control was composed for live performance as part of the Kent International Film Festival in 2002. It comprises of eight main sections, and is over an hour long in total. The software created in order to generate the piece is entitled Ariel. It is taken from the 16mm film project of the same name, completed in 1999. This earlier project is an experimental narrative inspired by the films of William Burroughs and Anthony Balch. Of particular interest is the fictional idea that a tape machine could be used to control or even destroy the world (Towers Open Fire, 1963). The film Ariel is, in many ways, a re-working of this idea. A tape machine with the ability to effect time and space (labelled ‘Ariel’) falls into the hands of a couple of East-Asian teenagers, who accidentally use it to change the world. In the real-time experimental documentary artwork Remote Control, the bespoke software that is used to control the material is also called ‘Ariel’.

This personal reference highlights some general artistic inspirations of this research. First, that the ability to create and control software has an emancipatory effect on digital art, by freeing it from the functional constraints of mainstream software. Secondly, that by taking a radical approach to audiovisual cutting, one can change the way in which an audience views representation. Thirdly, that the control of art and representation is in some way related to, or acts as a sign for the control of society. This view is inspired in part by The Cut-Ups (Balch, A 1966), in which seemingly redundant words are systematically cut together over a sequence of apparently random and unrelated shots. The effect of the film is sometimes noted as rather disorienting:
Mick Grierson

The Cut Ups was first shown in the mid-1960's at the Cinephone on Oxford Street, where the manager begged for a change of programme on account of the keys, coats, bags, underwear, and other strange items left behind by the disorientated audience. (Cumming 2004)

*Remote Control* is born out of a convergence of this and other related impulses. The majority of the visual source material is documentary in style, and the task of the performer is to edit the material in real time direct from hard disk onto the projection screen. The material is controlled and developed through the use of several algorithmic processes which are then used to treat the material, in order to abstract it. At the beginning of each section, the material is allowed to play normally, with only minor editing, in one full screen frame. As each section progresses, the material becomes more abstract until it collapses into the next section.

The majority of the sections involve some sort of interview. In many cases, random people are approached and asked to describe the nature of art. This is deliberately vague, resulting in descriptions which are relative and non-specific. The abstraction of the material is ironic in this sense. The inability to describe the nature and quality of art with words is highlighted by the re-creation of the material as art. This is most adequately demonstrated by the extract on the DVD entitled 'remoteControl excerpt 2'. Just prior to the start of this extract, a retired academic has been asked to define artistic quality in terms which do not involve capital. He states that this cannot be done. This eventually leads him to the conclusion that the pianist Vladimir Ashkenazy is of no greater quality than Cliff Richard. At this point, the material veers off towards abstraction.
In addition to the abstraction and re-cutting of interview material, *Remote Control* consists of aesthetic material which functions to demonstrate the creative potential of the software. The DVD contains a long extract, wherein the source material is re-iterated and re-structured in time to a regular beat. The main rhythm begins after a short introduction, and it is this rhythm which in part makes this extract more comfortable to view than other sections. On performing the piece, I was conscious that the audience’s attention could be galvanised by the use of familiar and repetitive material. This view is in part supported by the research regarding effectiveness and memorability documented in the theory section. The long extract demonstrates an attempt to create a section that relates more directly to conventional music video. It is important to stress, however, that except for the source material, this extract was generated in real-time.

One final aspect of *Remote Control* which it is important to mention is the continual use of text. The piece makes use of a large text file, generated by myself, which makes various references to abstract art, avant garde film, experimental sound and music. This text file is displayed on screen one word at a time in a semi random order, with variations in tempo, size, and three dimensional rotation. In addition, the display font is continually altered. This can be seen in the extract entitled 'remoteControl excerpt 3', and is again inspired by *The Cut-Up’s* (Balch, A, 1966), and by cut-up technique in general (see below). In many ways, the cut-up is a technique for removing received meaning from patterns of communication, and attempting to re-establish them. As such, the application of this technique is as valid to text as it is to any other form of representation. Indeed, it is a technique which forms the basis of many of Burrough’s books, such as *The Naked Lunch* (1963).
Mick Grierson

**Third World Warhol**

*Third World Warhol* is an artwork created through the development and use of the Max/MSP patch 'Weapons of Mass Deconstruction'. The WMD software suite is designed in order to generate a particular type of output, and is in many ways an extension of the earlier real-time audiovisual cut-up machine 'Ariel'. It has also been designed with the need for a certain type of input, although it functions well with any type of audiovisual material in the correct format. It is flexible in that the user can force the software to act in a wide variety of ways. Indeed its variability is key to its usage in performance. However, instantiations of this software can be utilised to produce any number of separate works with the same or similar aesthetic focus and effect. One example of this is *Third World Warhol*.

*Third World Warhol* deals with themes of terrorism, appropriation, representation and art. An important factor in the way in which it does this is the choice of source material. It uses footage of the World Trade Centre collapsing, as shot by a member of the American public (whose identity has been difficult to establish). The footage was sourced from a BBC television program, *Correspondent Special: One Day of Terror* (eds Murch, F & Durrani F, 2001), broadcast on Tuesday the 18th of September 2001.

This small section of audiovisual material is the basis for *Third World Warhol*. The material is transformed through live performance and manipulation, in order to foreground two key inspirations. First, the piece seeks to emphasise the horror of the event. No attempt is

---

26 As the software is an audiovisual instrument, it is capable of producing a variety of different pieces.
Mick Grierson

made, however, to create a sense of realism. Instead, the piece seeks to reinforce, represent and re-contextualise the horror that we associate with the September 11th terror attacks. Through treatments and abstraction of the source material, the diegetic sounds and images become re-interpreted as non-diegetic material. In this way, the representational material (i.e. the ‘realist’ video footage) is used to create its own abstracted context, whilst generating an aesthetic of disorientation and noise.

Secondly, the piece is inspired by a desire to explore a cultural and political viewpoint regarding the nature of representation. This is evidenced by a number of factors, but chiefly by the choice of video footage itself. The person recording the event is viewing it through the lens of a portable video camera. When passers by take her into a nearby shop, she seems annoyed with them. She is unaware that her life has just been saved. As smoke and dust fly past the window, she starts to scream, and the camera she is holding shakes violently. However, her panic soon subsides and is overtaken by her preoccupation with the practical task of filming the event. She states ‘I hope I’m getting audio on this’, only moments after she has recorded an event which marks the potential death of thousands of people.

The piece of footage is used as an attempt to pose questions regarding the way in which we as a society produce, consume and consider representations. We view her experience through a representation, and must make our own judgements regarding the nature of the situation and its morality. It seems unlikely that this representation adequately conveys her experience. We experience a simulation of the event – a recorded version, as framed by the camera. We are thankfully removed from the experience, and respond to it through the
filter provided by the screen, framed by the contemporary context surrounding the terrorist attacks. It could be said that there is an irony to this particular piece of video footage – the camera person is focussed on the screen in a way which puts her own life at risk. It is the fact that she is determined to record this event without considering the danger to herself that is the key reason for choosing this clip\textsuperscript{27}.

Third World Warhol attempts to ask the following question. Does the camera person find it difficult to negotiate between the value of the real and the value of the artifice? That is to say, does her preoccupation with creating and experiencing simulations of the real, impact negatively on the way she deals with her material situation? If so, then what effect do these simulations have on the viewer? Should we reconsider how we address this type of spectacle?

With respect to the overall inspirations of the piece, there is some attempt to engage with certain aspects of post-structuralist and related approaches to postmodern critical theory. One theoretical approach which examines similar issues surrounding representations is Baudrillard’s notion of the precession of simulacra (Baudrillard, J 1994). Baudrillard argues that in the postmodern age, we\textsuperscript{28} find ourselves unable to distinguish between representation and reality, due to the substitution of the real with the representation i.e. we experience a chain (precession) of representations which fail to lead back to the represented. Whilst we attach significance to representation, the real cannot be detected. Baudrillard states that this occurs as a result of a number of interrelated factors. First, we

\textsuperscript{27}It is important to stress that this is the only piece of source material in the entire portfolio which does not come out of the research itself.

\textsuperscript{28}This perhaps should be read as ‘many people who live within capitalist culture’. However, Baudrillard makes no such concessions to our sensibilities.
Mick Grierson

no longer acquire goods due to our needs, but instead are encouraged to fulfil our desires by the proliferation of images available to us as a result of increasing mediatisation. Secondly, the use value of goods has been superseded by their capital value (i.e. their exchange value – a central inspiration for Remote Control). Therefore, we no longer attribute function to form. Instead of use value, we have capital value. Thirdly, global multinational capitalism creates a distance between the means of production (labour) and consumption. As a result, we have no knowledge of the production process, and understand little of what we consume. Fourthly, the Earth has become increasingly urbanised, to the extent that we often experience nature and indigenous culture through the mediation of signposts and managed environments. This removes us from the actual experience of nature. As such, it is difficult to say exactly what nature is. Finally, Baudrillard sees the postmodern condition as one where all ideas concerning reality are ideologically based through necessity. Whereas in the past, representations and ideologies were mainly interpreted as illusions or constructions, the postmodern condition makes no distinction between ideology and truth.

Baudrillard's position, although certainly complex, is useful when considering political aspects of representation. Third World Warhol is inspired by a desire to engage with debates surrounding the cultural, personal and political aspects of image making. The initial hope was that the particular piece of footage chosen could metaphorically represent aspects of simulation and its effects. A secondary goal was the extrapolation of these ideas to a level whereby the effect of global capitalism could be considered by a viewer during the experience of the work.
The resolution of the piece, where square fragments of the material converge to create a pixelated image of Osama Bin Laden, generates a synthetic portrait of the aggressor out of a representation of the products of his aggression. This image was inspired by two main factors. First, it is a reaction to Stockhausen's comments regarding the nature of the September 11th 2001 attacks as a form of art. Stockhausen made the following clarification of his remarks:

On the occasion of a press conference on Sunday, September 16th 2001 for the Hamburg Music Festival I was asked if Michael, Eve and Lucifer were historical figures. I answered that they are always present, for example Lucifer in New York. I called to mind the role of destruction in art. All other words outside of this context have nothing to do with my opinion. (Stockhausen, K. 2001)

Stockhausen here makes reference to his calling to mind '..the role of destruction in art'. Third World Warhol explores this role via the WMD approach to processing. In addition, the final sequence echoes the desire to highlight the relationship between destruction and re-construction. Additionally, the image of Osama Bin Laden has become a significant aspect of mainstream culture. Bin Laden's image is reproduced in Third World Warhol in a way that treats the image as a pop culture reference. This is an ironic comment on the commodification of images, and also the existence of commodified pop culture references in art, such as in the art of Andy Warhol. Warhol took popular culture artefacts and attempted to recast them as art. The desire behind Third World Warhol is to take the horrific media spectacle (itself a form of popular culture) and comment on it as a sign of the distaste with which some cultures view the western obsession with commodification.

29 This should not be read as a flippant comment. It is a very complex argument, archived by eyewitnesses at http://www.stockhausen.org/eyewitness.html.
The Political Aesthetics of WMD Audiovisual Cutting Software

The WMD software attempts to elaborate practically on ideas of deconstruction, and also reconstruction, in real-time, whilst additionally allowing for the exploration of radical political and cultural ideas relating to the urge to destroy and reshape. Although a direct relationship between deconstruction and destruction is not explicit, it is certainly true that the analysis of material through a process inspired by philosophical deconstruction reveals elements (or events) which may not be obvious in the primary (or intended) context. By applying this process to audiovisual material in real time, one can uncover events which can be used in re-construction. As such, deconstruction forms part of a compositional approach. Additionally, WMD exploits the aesthetic effects that result from such compositional approaches. These approaches have a great deal in common with the philosophy and technique of William S Burroughs, and, like Remote Control are directly inspired by aspects of Burroughs’ film and sound work (The Cut-ups, Balch A (1966) and also Towers Open Fire Balch, A (1963)).

This central inspiration operates in conjunction with other significant theoretical, political and cultural influences, most notably Fredric Jameson. Jameson's take on postmodernism as detailed within the essay Postmodernism, or, The Cultural Logic of Late Capitalism (Jameson 1984) details the Postmodern condition as a preoccupation with modes of production in the Marxist sense, whereby capitalism is taken for granted and goes unchallenged. All Postmodern culture (including theory) can therefore, in Jameson's view, be 'read' as a statement or element of capitalism itself. This idea, that all culture is

---

30 In this case the term re-construction refers to real time collage or pastiche of material which has been 'deconstructed'.

49
Mick Grierson

sublimated and becomes commodity, resonates with Karl Marx’s theory of ‘Cultural Fetishism’.

The image is the commodity today, and that is why it is vain to expect a negation of the logic of commodity production from it, that is why, finally, all beauty today is meretricious and the appeal to it by contemporary pseudo-aestheticism is an ideological manoeuvre and not a creative resource. (Jameson, F p135)

The appropriation of cultural material including theory, film, sound, colloquial dialects and all forms of art, and their inevitable sublimation into the dominant cultural aesthetic is key to all Jamesons' writings31. His view could be said to concur with and relate to aspects of Richard Dawkins Memetics, which is detailed in the final chapter of Dawkins’ book The Selfish Gene (Dawkins, R 1976). In this case however, Dawkins only discusses the concept of ideological transmission via cultural forms. He proposes that scientific observation regarding viral transmission, evolution and propagation can be employed as a metaphor for the development and transmission of ideas, including religion, politics and culture.

Interestingly, William Burroughs’ concept of the idea as virus forms part of his writing regarding language as ‘mass ventriloquism’.

My basic theory was that the…word (is) actually a virus…The word has not been recognised as a virus because it has achieved a state of stable symbiosis with the host… (Burroughs, W.S. 1971 in Miles, B, 1992 p150)

31 A similar, but not identical critical approach is offered in Storey, J. (1996), Cultural Studies and the Study of Popular Culture: Edinburgh University Press, although this analysis is far less pessimistic with respect to the contemporary cultural climate. Jameson's particular analysis is itself absent from this text.
Mick Grierson

If Burroughs is to be believed, his 'word virus' theory pre-dates Dawkins' development of memetics by almost a decade. In perhaps what is an attempt to counter that which he sees as transmission of didactic cultural ideas, Burroughs attempts to wage war on the word virus through the use of 'the cut-up' technique. The cut up as a technique of radical artistic protest dates back to Dada and the early Surrealists, and in many ways is a characteristically controversial product of radical modernism (see http://en.wikipedia.org/wiki/Cut-up_technique). It pre-echoes elements of Derrida's deconstruction, in as much as it deals with the breaking down of an image from its context, and its inevitable re-contextualisation.

This concept of the 'Cut-up' and its application feeds directly in to the development of the software 'Weapons of Mass Deconstruction'. The software is created from the premise that in order to metaphorically deconstruct and re-construct political and cultural signs of Postmodernism (as capitalism), it must itself function as a tool for deconstruction and re-appropriation in its earliest and most original sense – the Cut-up. As such it embodies a combination of technical and aesthetic foci, inspired by theoretical, political, and cultural concepts.

Most importantly, WMD is a radical reaction against cultural misappropriation and information distribution. As such, it exists not just as a compositional tool (as its output is consistent with a particular aesthetic approach), but also as a work of art which embodies the desire for the deconstruction and re-appropriation of media-based contemporary culture. Through WMD processing, cultural forms can have their meaning de/re-
contextualised in real-time through the control of a performer\textsuperscript{32}. Metaphorically, it attempts to operate as a \textit{Vaccine} (a vaccine for a meme: Dawkins, 1982) against cultural transmission by deconstructing formal aspects and given meanings from cultural media material as it is produced, using any particular input as the method of its own deconstruction, rendering it impotent as regards its intention as a memetic ‘virus’, and revealing aspects within it which are not immediately apparent in its original form.

Partly in an effort to avoid re-appropriation itself, WMD attempts to produce an aesthetic effect that is fundamentally tied to its production methodology. In this way, it has much in common with the work of structural materialists\textsuperscript{33} such as Malcolm Le Grice, Peter Gidal, and also Michael Snow. In Michael Snow’s \textit{Wavelength} (1964), a faux narrative is presented within the film in order to confound (and distract) an audience’s expectations, drawing away from those aspects normally associated with traditional ‘cinema’, to those features which deal with the material. The film is a statement about the material itself, focusing on the nature of waves as they are recorded on film and on tape. The camera performs a slow zoom accompanied by an electro-acoustic soundtrack until it focuses on the image of waves on the far side of a room. The volume of the oscillating sine waves increases towards the end of the film, pointing towards the material of audiovisual production. However, as a result of the general lack of movement within the image (discounting the use of gels and other forms of manipulation), and the slowly developing soundtrack of pure tones (excluding the occasional synchronised sound element), the piece has a transcendental quality. The structural materialist ‘narrative’ is accompanied by

\textsuperscript{32} One could also locate the control with the composer, or even the programmer. See chapter 1.

\textsuperscript{33} For a review of structural materialism, see Peter Gidal (ed.), \textit{Structural Film Anthology} (London: British Film Institute, 1978), and www.LFMC.org.
an aesthetic effect which is a direct result of the nature and speed of the processes utilised in its production.

This sort of structural materialist focus and its resulting aesthetic is also demonstrated in the works of composers as apparently different as Stockhausen and Steve Reich. Reich’s desire to emulate the behaviour of tape machines in works such as ‘Piano Phase’ points directly to the process of composition.

I am interested in perceptible processes. I want to be able to hear the process happening throughout the sounding music. (Reich, S. 1974 p9).

The process is the means by which the aesthetic effect is produced. Coincidentally, in this case, the work could also be described as having a transcendental effect, which is perhaps due to the use of repetition of both harmonic musical material and rhythm, alongside the gradually phasing piano texture. Although Reich’s philosophical position appears incompatible with acousmatic approaches, it is interesting to note that much that is acousmatic (without identifiable source) is by nature process driven. In terms of the relationship between the sound and the material, the understanding of process is by its very nature linked to knowledge of how sound is produced. Without this knowledge, the sound can be easily characterised as acousmatic.

Stockhausen’s process driven methodology is also echoed within the structural materialist tradition. Works such as Mikrophonie I (1964) and Mikrophonie II (1965) focus on the material and process of production. The explicit use of experimental microphone
techniques and recording processes demonstrates significant commonalities with works of the structural materialists. However, in the case of both of these works, there is a fundamentally unsettling aesthetic effect generated by the experimental nature of the processes themselves. The process of manipulating the material is directly exposed in an effort to exploit the significantly more complex sonic aspects of modulation and timbre offered by technology (Smalley, D, 1986). The inharmonic and rhythmic complexities produced by this process are by their very nature potentially disturbing (see below), and in this way, in addition to the concepts of Burroughs, the work of Stockhausen has been a direct inspiration on the development of Weapons of Mass Deconstruction.

The process of production in this case results in the aesthetic effect of the Cut-up, combined with theoretical inspiration from Derrida and Jameson (among others), and also the concept of culture as virus. These elements are synthesised in the hope that the work will remain politically and aesthetically radical regardless of the type of input material. The parameters of the software are calibrated to allow for a variety of different intensities and audiovisual relationships. However, a conscious choice has been made regarding the treatment of the input material inasmuch as the material will always be violently abstracted from its initial form. The degree and manner of abstraction is within the control of the composer/performer. A separate version of the software could be produced without the degree of manipulation exhibited within WMD, and the aesthetic difference between the two could be described as similar to the relationship between low frequency amplitude modulation, and violent ring modulation. Metaphorically speaking, the software should function in an offensive way - as an actual aesthetic weapon. As such the results should be similarly offensive.
Due to its philosophical and aesthetic ambition, it is important that when treatments are applied aurally, visual treatments are also applied to the same degree and in the same fashion. This could be seen as contrasting with the views of Nicholas Cook, who suggests that overt synchronisation can overwhelm either the visual or sonic aspects of a work, and that sensitivity to the concept, power and design of each aspect of the media is key to producing multimedia work (Cook, N 1998 p41). This view seems only to be applicable to a composer writing music for existing material from other media, and may well be influenced by commercial factors rather than artistic ones. Additionally, the concept of WMD precludes by its very nature its usage as an *accompaniment* to existing cultural environments. Its purpose is to devolve meaning from existing culture (including TV and club culture) to a new radical form as a method of protest, based on a metaphor of ‘cultural terrorism’.

This type of experimental cultural protest is prevalent in many forms of ‘industrial culture’, such as the work of Throbbing Gristle (*20 Jazz Funk Greats, Zyklon B Zombie*) who are early pioneers in experimental audiovisual work. It could also be said to resonate with the ideology behind the provocative work of John Cage. For example, Cage’s *Cartridge Music* (1960) takes the technical instrumentation utilised for reproduction of recorded music, and attempts to use it as an instrument. It functions as both a statement about the process of reproduction, and also as an attempt to provoke contemporary music audiences. Cage’s music was also a great influence on structural materialism, as it aims to emphasise the relationship between the process of production and sound making material. Beyond this, it stimulates radical thought by drawing parallels between radical musical and sonic activity,
Mick Grierson

and cultural relationships. This is revealed by Cages' statements on social philosophy:

For the anarchist, every being is a ruler...My music liberates because I give people the chance to change their minds...We need first of all a music in which not only are sounds just sounds, but in which people are just people, not subject, that is, to laws established by any one of them, even if he is 'the composer' or 'the conductor' (Cage, J in Kostelanetz, R 1992 p257)

This radical sound world metaphor (or what could be described as such) is recognised by Richard Kostelanetz in Conversing with Cage (1992). Kostelanetz refers to it as Cage's 'Anarchic harmony'.

The thrust of Cage's Anarchic Harmony is a desire to create space where sounds can exist on their own terms. As such they become sonic events which jar with the conventional musical aesthetic of the period. This has two main effects. First, the 'anarchic' experimentation of treating sound in a similar way to music leads us to question what we define as musical artworks. The lack of conventional harmony, melody and rhythm forces us to address music in other ways, such as through analysis of texture and timbre. The common framework or language of western music is subverted by this lack of harmony in particular, instead placing the sounds themselves at the root of the audience's experience. Secondly, through opening up the musical world to the entire sonic spectrum, audiences are forced to encounter sounds which as yet have not been incorporated into a cultural framework in this way, highlighting the strangeness of the sonic musical experience. This

34 Having had contact with the author personally, I am still unable to find the specific paper reference for this description. However, Kostelanetz directed me to this internet site: http://www.sterneck.net/john-cage/richard-kostelanetz/index.php
strangeness is compounded when we realise that the material has its roots in the everyday (such as the process of sound reproduction in *Cartridge Music*). It seems from Cage’s own words that he seeks to exploit a listener’s lack of familiarity with the sound of a familiar object in order to exploit creatively the uncertainty it causes. Again, such techniques resonate strongly with the Dadaist tradition.

The generation of experiences which are by their experimental nature unusual, can result in an audience feeling discomfort, shock, fear, terror, and even generate laughter. These are the types of reactions which WMD is primarily designed to produce. Importantly, this approach has been adopted so as to allow abstract material to surface in a way that appears alien to the majority of audiences. An attempt is made to remove the received meaning which one may normally attach to the representation by a process of seemingly random arrangement, thereby making the material appear more abstract. This is similar to the way in which Burroughs employs the Cut-up technique. Burroughs uses the Cut-up as a direct attack on the desired or imposed meaning of the sentence. A simple example of this is repetition. When a person repeats a word over and over again, its context is removed (as explored in *The Cut-Ups*, Balch, A 1963). Therefore, the ability for it to function as part of an existing message is diminished or removed. This can be absurd for a viewer or listener, and as such it can stimulate laughter or become unsettling. The Cut-up itself, although sometimes restricted to a re-arrangement or re-ordering of words, can also become a re-arrangement of word fragments. When word fragments are used as raw material, the reconstruction can be even more flexible, abstract and effective.

The relationship between abstraction and the shock of protest has historical precedents.
As we have seen, in the case of structural materialist and experimental film, electro-acoustic and avant garde music, abstraction through process may stem from a radical desire to reveal the form of the material, which can be interpreted as, and directly relates to protest. The desire to abstract the everyday functions as a protest against the socio-political and economic narratives of the time stems from Dada. In the case of Dada, the initial stimulation for the movement may be interpreted as World War I. In the case of WMD, the stimulation is postmodern capitalism (as analysed and critiqued by Jameson and Baudrillard) and the negative aspects of its global effect.

Finally, with respect to WMD, and the audiovisual works generated with it, there is a strong and vital link between the technical, the theoretical, and the aesthetic. The technical process and construction is the means by which the aesthetic effect is produced. Beyond this, the theoretical and cultural concerns which inspire the technical approach are fundamentally related to both the aesthetic and the process, and thus the process attempts to reflect cultural, political and aesthetic arguments. The radical treatment of the material functions so as to immediately re-construct the material in real time at the precise moment that it appears at the input (so no previous capturing of material is required). The device is constructed with maximum possible radical potential in mind. The deconstruction of the material is preformed in such a way as to create new, more abstract forms from within it, which are re-constructed 'by the roots' in order to generate material with potential for protest and absurdity.
Mick Grierson

5. Technical Construction: WMD Audiovisual Sampling and Cutting Software

_Ariel_ was a first attempt at creating audiovisual cutting software for use in composition and performance. Many of the ideas developed in this application were reviewed and redesigned in order to create WMD\textsuperscript{35}. With regard to the technical construction of WMD, the process of development began by adopting an approach to the treatment of audio material (live or pre-recorded). These approaches were then applied to visual material. The final aspects of the development dealt with the interplay between audiovisual material as informed by Chion, and the honing of the aesthetic focus in relation to the poetic inspiration. Early versions of the software were developed in PD/GEM. Some of the examples given here deal with PD/GEM, others relate more specifically to MAX. In all cases the work now exists as a MAX/MSP application.

The processing necessary to cut and re-sequence an audio stream is time based. Before any processing can take place, the data must be written to a buffer. At first the process was non-real time, in that the audio was pre-recorded and loaded into buffers. After testing and development, the design was modified to allow for real-time sampling and re-sequencing. This approach facilitated the development of the re-sequencing technology.

In Pure Data, audio information is written into a table and then accessed at audio rates to reproduce the sound.

\textsuperscript{35} The differences between the two pieces of software are discussed later in this chapter.
Mick Grierson

At this stage the goal is to produce a technique of sequencing and repetition of audio material that utilises carefully controlled aspects of randomness and can be controlled in real time, through a process similar to granular synthesis, with the capacity for the manipulation of grains of variable sizes up to several seconds. In addition the processor needs to be ‘aware’ of the material and its contents, so that re-sequencing can occur in an intelligent way i.e. the material can be re-ordered with reference to the type of material contained in the buffer. Of course, it is impossible for the machine to determine any real meaning from the material. However, accurate information regarding the amplitude of the signal can be analysed and used to create regions for re-sequencing.

The analysis is performed by first simplifying the material. Any real-world sound signal is generally very noisy. That is to say, there is a great deal of high frequency material which may well make analysis more difficult. The only information required for WMD to operate at this level is the start location of any discrete events within the sound signal. A low pass filter is used in order to facilitate the event detection process. In this case, the LPF is

---

36. This does not mean that the re-sequencing is random. Randomness is used similarly in Granular Synthesis  
38. In the sense that the material is automatically analysed and the analysis is then used in re-sequencing
Mick Grierson

implemented by calculating the average amplitude value within a given number of consecutive samples. This reduces the volume of high frequency content in the sound signal by effectively ‘smoothing’ the waveform. Therefore, only the lowest frequencies contained within the material are sent to the event detection system. The actual frequency threshold depends on the nature of material being utilised by the system. In most cases, a filter which averages a second's worth of samples is sufficient in detecting any significant events in a stream of speech or music (the event detection system can be calibrated in real time to achieve the desired effect). In the case of the non real-time example, the audio signal is played through the event detection system prior to re-sequencing.
The detection of an increase in the amplitude of a signal can be easily set to produce a 'bang' event in both PD and Max. In order to reduce the effect of anomalous 'bang' repetitions, bangs which occur in close proximity (over time) are lumped together into a single event. The bang is then used to set the start position for a region. As bangs are received, the number of regions increases, and the location points are stored in milliseconds. The end of the region is taken to be the beginning of the following region. In this way, region start and end points can be stored and accessed from a table.

There are several ways in which the re-sequencing can take place. If the sound signal consists of a definite rhythmic loop, then each region can be played out in any order once, with each region commencing as the other finishes. With this form of re-sequencing, a loop can retain its length and tempo, providing it is played accurately. In addition, if the re-
Mick Grierson

sequencing is fed random region points, it will exhibit constant variation and development. This process produces highly ordered results from highly ordered material. It also makes new sentences out of old ones with reasonable accuracy, provided that the signal is clear and direct. Additionally, the regions are mapped to midi notes as they are detected, and this allows for instant performance control.

With reference to WMD it is required that the material is processed with the maximum potential for de/reconstructive effect. With this in mind, several real-time controllable functions are available. The regions can be repeated up to a set number of times. This means that a region can be instantiated, and allowed to repeat for a given number of times before moving on to the next region. A controller is used to set the number of times the loop repeats, which can then be altered by a random value (calculated before the beginning of each repetition), the magnitude of which is set by another controller. In addition to this, the length of the loop can be adjusted to within the nearest millisecond (standard for Max/MSP). This allows for the repetition of small grains of material, the length of which can gradually be altered whilst repetition is occurring until the maximum
loop time is reached and the next region begins. The start position of a loop can also be altered. A controller allows for the region start position to be changed whilst playback is in operation. The loop start time can then be adjusted via a random value of variable magnitude, which is also calculated before each repetition. Finally, the playback speed of the audio can be adjusted. If the sections of audio are significantly small, then the frequency (perceived pitch) of the material can be perceptually shifted and/or stretched. In this way, the material can be manipulated to an extreme degree in real time.

As has already been described, this type and manner of audio processing is based on granular synthesis methods. In order to reduce audible clicks in the re-synthesis, however, windowing functions ideally need to be implemented (Roads, C. 1988). This allows for tiny sections of audio to be overlapped, creating smoother playback of small grains of sound. This also provides more convincing pitch manipulation. The main difficulty with this functionality is the processor overhead. This audio cutting system must form part of a larger video processing suite, so the audio processing must consume the minimum amount of CPU power.

Instead of utilising overlapping windowing systems to reduce clicks, the audio signal is automatically enveloped. The length of each event is calculated at the beginning of each repetition, and the start and end points are treated with a fade of only a few milliseconds. The length of the fade is variable. As the next loop will not start until the previous one has completed, the likelihood of clicks in the audio stream is reduced.

---

39 The material is looped beyond the perceptual threshold whereby it can be perceived as discrete event. This allows the material to merge with material immediately following (also looping at these rates), creating a time stretch.
There are cases where this system is insufficient to produce click free re-sequencing. When the audio regions are triggered by a separate source (such as a midi event, or gate of any kind) and not when the previous loop ends, a form of windowing is required. The simplest method is to use a dual playback system, whereby consecutive repetitions or region changes are sent to separate playback algorithms. In this case, the playback algorithm does not ‘loop’ as such. Playback functions swap between the two systems allowing each loop portion to complete. In order to prevent unwanted overlapping, the fade out is triggered as the next loop portion begins, creating a minimal overlap. This does not allow for the same degree of smoothing offered by traditional granular synthesis, but is efficient in performing the basic manipulation required for the re-sequencer to function.
To allow for the creation of a real-time version of the audio re-sequencing algorithm, data has to be written into the buffer as it is received by the audio inputs of the computer. However, it is important that the machine does not try to read from the buffer as it is being written to. If this is allowed to happen, two problems occur. First, a buffer cannot be infinite in length. If the end of the buffer is reached, recording has to re-commence from the beginning of the buffer. The region definitions stored by the event detection system will then become inaccurate, as they will refer to a section of the buffer that has been overwritten. Secondly, as playback can be entirely random, the computer may be asked to write into the same memory space that it is currently reading from. The effect of this sort of error can be extreme.

In order to circumnavigate this problem, two buffers are used. The two buffers are identical in length. Each buffer records for a set amount of time. When the buffer is full, a signal is sent in order to tell the second buffer to commence recording. The first buffer is then used for re-sequencing. The event detection system is identical in construction. It detects when an event of a set amplitude has occurred and reports the value in milliseconds to a table. When a buffer becomes available for playback, it uses the table as a list of region definitions. Simultaneously, it begins storing events in a separate table. When the buffer is swapped, the corresponding table of region definitions is also swapped, maintaining coherence between region definitions and the currently active audio stream. This means that the system can deconstruct and re-sequence in real time with a delay identical in length to the size of each buffer.
There are two significant drawbacks with this system. First, sections of material can only be used for re-sequencing whilst the buffer is being used for playback. If the buffer is ten seconds long (10000 ms), material recorded within the ten second time frame can the be re-sequenced. However, if radical re-sequencing is being performed (i.e. random repetitions of extremely small sections of material) then it is inevitable that not all of the material will be used, as the material will continue to swap at ten second intervals. This places restrictions on the amount of material available to the composer/performer at any time. The problem can be avoided by pausing the buffer swapping algorithm temporarily – however this means that new material arriving at the input will not be re-sequenced at all. The second problem is that material residing at the end of the first buffer cannot be re-sequenced alongside material at the beginning of the second buffer. Both of these issues
Mick Grierson

arise from the fact that one simply cannot have an infinite amount of memory. Increasing the size of the buffers succeeds in disguising the first of these drawbacks, but creates a greater delay, and additionally requires more memory. As we are going to apply these techniques to video, the amount of memory required is a significant factor.

Once the process has been set in motion, the material can then not only be re-sequenced, but also re-synthesised. As we have seen, information within a buffer can be treated in a manner not unlike a form of semi-intelligent\textsuperscript{40} granular synthesis. As a result of this, it can be pitch shifted, giving the impression that it has been stretched. In addition, the audio can have transient generators applied to it, such as oscillators and envelopes, in order to control relative frequency and amplitude. This allows for quite complex treatments such as ring modulation, frequency modulation and amplitude/pitch enveloping (useful, for example, for creating synthetic percussive sounds). As is well understood, using random waveform high frequency modulators can result in any signal being reduced to complex noise.

On its own, the audio re-sequencing and re-synthesis aspects of the software allow for any real-time recorded sound of any type to be used as a basis for wavetable synthesis. Specific preset modulations can be applied to a particular region number. Consequently, when the region material is swapped, the modulations can remain the same, instantly being applied to the new region. Rhythmic and harmonic data transmitted over MIDI can be interpreted via the regions defined in the buffers. With the implementation of simple polyphonic playback algorithms, whole electro-acoustic works can be composed and

\textsuperscript{40} With respect to the analysis of the signal and the use of this analysis in processing.
Mick Grierson

performed wherein all raw material is generated by one acoustic performer (such as a singer), and all treatments are performed in real time through the use of a MIDI interface with only minor variation of buffer material from performance to performance depending on the accuracy of the acoustic performer (see below).

With regard to polyphony, the software used to create Third World Warhol does not implement polyphonic buffers at this stage (unlike other software developed as part of this research) for a number of reasons, both technical and aesthetic. First, the audio material itself is linked almost perpetually to the accompanying visual material. Conventional video display could be described as ‘mono-visual’. That is to say, there is often only one ‘playback channel’ for video or film material. When mixing multiple audio signals with similar frequency content in a single channel, sounds will easily mask one another. Through the use of multiple channels, the coherence of the frequency content can be maintained via the process of spatialisation.

In terms of video, although many images can be placed within the same video frame, their size needs to be reduced in order to prevent masking. In addition it is difficult to disguise the straight edges of the frame without the use of blending, which is processor intensive. Multiple images can be displayed together through the use of simple mixing and keying. The colour content (relative to the frequency content) and the contrast ratio (relative to the dynamic range) need to be carefully controlled, however. Aspects of this problem are addressed in more depth later, particularly in relation to video synthesis and the use of multiple textures (see below).
Another suitable solution may be to use multiple displays in different spaces (such as in the case of Bill Viola’s *Five Angels for the Millennium*, 2000), employing a form of visual ‘spatialisation’. However, this produces problems similar to those encountered when utilising multichannel surround formats during performance in large spaces i.e. a single member of the audience cannot perceive the whole artistic experience in one single viewing. Some audience members may not see this as an issue, preferring the opportunity to enjoy the work in different ways on multiple occasions. In the case of WMD, aspects of its output depend entirely on input, and variability is an inherent factor of performance.

With respect to cinema presentation, Chion’s concept of *magnetisation* states that audio material which is closely connected temporally to on screen material will appear to emanate from the screen, as it becomes perceptually *magnetised* (Chion, 1994) towards the central viewing focus. This may mean that there is little point in spatialising large amounts of sound material which relates directly to on screen activity. This view definitely requires further research and discussion with relation to the use of multi-screen moving image installations. Additionally, with regard to the development of the software and the possible use of multiple screen-space, it makes more sense technically to use multiple instantiations of WMD in order to produce a ‘poly-visual spatialised’ performance.

The application of the re-sequencing mechanism to the video material is theoretically identical to that which has already been described with respect to the audio material. There are a number of significant factors which make the process more difficult, however, and many attempts have been made to address the issue in this research. An appropriate model has been developed and decided upon, and it is detailed below.
Two versions of the software have been developed. One version, part of *Ariel*, is designed to operate on a pre-recorded signal stored on a hard disk. This version was used to create *Remote Control*. The second version of the software has been designed to operate on an incoming 'live' signal (such as a Television signal). This version was used to create WMD, and the associated work. With regard to both versions of the software, video and audio information is captured simultaneously, and then split into separate data streams. It is necessary to treat the data separately due to the fact that the video and audio data are fundamentally different types of data, operating at different resolutions, each having specific technical requirements with relation to processing. Additionally, in Pure Data, video processing is treated separately to audio processing. The same is true in Max/MSP/Jitter, and although the latter allows for rudimentary playback and control of audio and video data from a single data source, adequate processing of the audio data is far more difficult if the data is linked in this way.

Once the data has been separated and loaded into a playback algorithm, it must be re-synchronised so that when a region is triggered from the event list, the correct frame is recalled. In the case of the early versions of the software, this is done by sending the region start time in milliseconds to the video playback algorithm. This value is then used to calculate the current frame number\(^{41}\). Providing that both streams are identical in length and overall start position, playback of both audio and visual material should commence at the same point. Further to this, however, audio and/or video playback can suffer from what in general terms is described as system ‘latency’, the amount of delay between the sending of a signal, and the receiving of a response. Computer systems suffer from variable latency of up to several hundred milliseconds. The actual latency is unimportant;

---

\(^{41}\)By dividing the current time in ms by 1000, and multiplying by the current frame rate.
the only important factor is that whatever the latency, it should be identical for both the video and audio streams. When this is not the case, the difference in audio/video latency needs to be compensated for. WMD provides a real time controller for adjusting the start frame position and synchronisation tracking of the source material.

This control has to be set prior to performance or installation. Presuming that the hardware functions correctly and is of sufficient quality, it should not need to be altered again. Factors which may require the value of this controller to be altered include changes in video resolution, audio resolution, hardware drivers, hardware efficiency and processor/memory usage.

Additionally, playback synchronisation must be efficient so as to prevent the separate streams from becoming desynchronised. Rapid alterations in playback rate and direction can cause the stream to slow down and lose synchronisation, so an accurate tracking system is required in order to maintain overall system stability. There are two main ways that stable synchronisation can be attempted - by either constantly reporting the time value of one data stream to the other, or by synchronising the data streams to an external clock source.

In the case of the version of the software developed for the composition Remote Control, the video stream is used as the timecode reference for the audio stream. The current frame number is reported, and its value in milliseconds calculated with regard to the beginning of the data streams.
Mick Grierson

This system results in the software maintaining reliable synch throughout long performances of rapid real-time treatment. The major difficulty with this solution is that when variations in the timecode do occur, or one stream loses its position due to inefficiency, it can produce a loud click as the audio stream is repositioned.

In order to rectify this problem, the algorithm has been redesigned for the development of WMD in Max/MSP/Jitter. In the revised design, the position of the audio stream in milliseconds is used to check the frame number held within the video buffer (the video buffer in this case is the Max/Jitter object jit.matrixset). When the frame number value is updated, it requests the relevant frame from the video stream. This means that any inconsistencies within the video stream are eliminated, preventing the audio stream from clicking unnecessarily. In addition to this, the playback of the video can be entirely...
controlled by reporting the buffer position in milliseconds, removing the need for the start value to be reported to the video buffer. As in the case of the audio re-sequencing algorithm, there are two video buffers that record for an identical amount of time, which in turn are the same size in milliseconds as the audio buffers. All four buffers swap at the same precise moment, with one pair in playback and the other pair in record at all times (unless paused).

This level of control over synchronisation is necessary in order to make possible the extreme methods of treatment required by the WMD research. Changes in pitch are reflected in the rate of playback of the video material. Additionally, if the audio material is reversed, the video playback can also be reversed, maintaining synchronisation.

Video material can be processed via the same transient generators used for audio processing. Transient generators and frequency modulators can be routed to a number of inputs, allowing seamless integration without assigning specific visual processes to fixed audio signal processors. This means that, for example, an LFO or envelope modulating the amplitude of the audio signal can easily be patched to modulate any number of video processing parameters. These include brightness, contrast, hue, individual RGBA colour values, frame width, frame height, zooming, frame rotation, position, and spatial deformation. In addition, a form of simple visual feedback has been implemented. The extent to which these values are affected by the transient generators can be set by a controller. This allows the operator to limit the ‘depth’ of the visual effect in real time.
There are problems with aspects of audiovisual cutting and re-synthesis which relate to visual perception. For example, a looping grain of sound ceases to be perceived as a string of separate events once it exceeds 20 repetitions per second (the Haas effect). This represents a frequency of 20hz, or a period of 50ms. This is similar to the function of a rapidly changing stream of still images. There exists a perceptual threshold whereby the smoothness of a piece of video or film material reaches critical flicker fusion – the point whereby it ceases to be seen as a string of single images and becomes a ‘moving image’. However, there are also significant perceptual differences. The ear is faster than the eye (Chion, 1994 p 10-11). For a sound to be sampled and reproduced with acceptable fidelity, over 44100 samples per second need to be processed. This creates a barrier in the application of sonic processing techniques to video.

As a result, applying visual modulators which operate at frequencies beyond the frame rate results in a form of audiovisual aliasing. If the frame rate is 20fps, and the modulator applied to the rotation factor is functioning at a speed of 20hz, then the frame will not appear to rotate. Each rotation will take exactly one twentieth of a second. The frame itself is displayed every twentieth of a second (50ms), so in each displayed frame, no movement will be shown. As a result of this issue, visual modulation beyond the current frame rate is ignored.

Most importantly, the looping of a small section of audio above the perceptual event threshold will produce a timbre directly related to the frequency content represented by the amplitude of the different samples (consider a wavetable oscillator for example). Increasing the frequency will also change the timbre. However, with an image, there is no
Mick Grierson

benefit in increasing the rate of image playback, as the effect will not be perceived due to the limits of distribution technology (generally 24-30 fps, although speeds of 60 fps and greater have been used in film installation to increase perceptual accuracy). Also, repetition of one frame will not be perceptually altered by the speed of repetition whatsoever.

Therefore, in order to more gracefully integrate timbres resulting from granular and wavetable synthesis into visual processing, video processes are linked to high speed audio processing at slower rates, relative to the perceived movement of the timbre. This encourages greater audiovisual connectivity. It is important that a suitable manner of video processing is utilised so as to relate directly to the aesthetic effect of the audio processing. For example, granular synthesis produces a smearing effect, as does FFT\(^42\) based processing. In WMD, the ‘smearing’ is emulated by visual smearing and spatial repositioning processes. This reflects a perceived relationship between the timbre of sound, and the texture of the image.

**Beautiful Human Canon**

The relationship between timbral evolution and image texture was a feature revealed through working on the piece *Beautiful Human Canon*. In many ways, the piece fails to convey coherent sound-image relationships evident in the other pieces. However, there are several aspects of the work which are interesting, and as such deserve to be mentioned here.

\(^{42}\) Fast Fourier Transform.
Beautiful Human Canon was initially conceived as a surround sound piece. The sounds were generated through the application of bespoke real-time FFT based polyphonic time-stretching algorithms to a number of sound sources. The initial aesthetic inspiration for the work was the perception of memory, particularly childhood memories, and the effect that certain experiences can have on a child. The intense treatment of the sound sources, coupled with the violent movement from speaker to speaker, are intended to generate disturbance and uncertainty in the audience. The sounds themselves are treated so as to appear recognisable for a short length of time, and are then quickly distorted via the time-stretching algorithm.

These treatments are accompanied by a video track of a child at play which repeats on a short loop. The video is heavily treated in terms of colour and texture. These textures vary throughout the piece, so that each image is re-visited with a different textural effect. The intention is to highlight minor differences throughout the multiple viewings of the material, in order to show how the memory of an event can be altered over time. In addition, a white filter is used in order to create audiovisual accents and crescendos. This technique is used heavily throughout the piece, and adds to the intensity of the viewing experience. The white filter interacts with the intensity of the more prominent distorted sounds, providing a white 'fade up', and also various sudden white flashes, at times functioning like a cymbal.

The main flaw in this piece is the lack of coherent audiovisual interaction with regard to the complexity of texture evident in both the audio and visual material. This is in part a technical issue. As Max/MSP requires huge amounts of processing power in order to
generate simultaneous audio and video streams, it is very difficult to reserve resources so that the material can be effectively rendered (recorded to disk) in a way which was representative of the piece as it progresses. It is possible for the material to be rendered to disk in non-real time. However, this type of rendering process does not allow for the same level of connectivity between real-time sound and video streams (as the audio and video materials are not being processed simultaneously with reference to each other)\(^\text{43}\). This means that the developing audiovisual textures are rendered incorrectly. The only means of ensuring that the material is recorded as intended is to output the work directly to a separate video recording device through a scan converter. Unfortunately, at the time of making *Beautiful Human Canon*, a scan converter was unavailable.

Recent works are much more successful in exploiting the relationship between audio and visual texture. However, *Beautiful Human Canon* was a key moment in the exploration of texture with respect to audio and visual connectivity in composition.

\(^\text{43}\) Although current versions of Jitter allow for this, it was a very difficult problem to solve in version 1.
6. Audiovisual Synthesis

There are a variety of ways in which visual material can be synthesised. Random values can be used to vary the brightness of an array of pixels. Algorithms can be executed in order to generate any number of given shapes by altering differentiated pixel values. Beyond this, pixels can be manipulated in order to resemble objects and environments.

The virtual 3D world is the most common form of visual synthesis. In the same way that sound synthesis attempts to reproduce real-world sounds, visual synthesis attempts to generate depictions of existing things. In cinema, computer graphics provides a means of generating visual material which is either difficult or sometimes impossible to effectively manufacture in the real world. Interestingly, the argument as to whether computer graphics succeeds in generating realistic objects and environments parallels arguments surrounding audio synthesis.

3D graphics standards such as OpenGL and Direct3D are a product of the desire to synthesise virtual images. They allow for the shaping, positioning, lighting and animation of objects in virtual space. As a by-product of the desire to generate virtual real-world objects, the 3D OpenGL graphics standard also provides an environment for abstract animation. Elements which are commonly found in the outside world (such as lighting and shadow) can be manipulated in order to generate experiences which are not normally found in our everyday existence. As such, abstract visual synthesis has direct parallels with sound synthesis in its development and applications. It makes sense that the two should be combined in order to help create an instrument designed to explore the phenomenon of added value.
 There are huge differences between composing with synthetic material, and composing with real-world material. Real world material is inherently noisy. That is to say, it has elements of randomness and incoherence from the start. However, synthesis does not necessarily exhibit this incoherence. When we discuss synthesis of any type (especially within the digital domain), we almost always start from the perspective of synthesising simple structures, such as shapes or waveforms. These structures are then modulated to make them behave more like real-world phenomena, such as sounds. Conversely, sampled material can never be 'pure'. It is always noisy. For example, the tuning fork is one of the closest acoustic models for the basic elements of all sound – the sine wave. However, it is still not a pure tone. So, it could be said that the real and the synthetic are at opposite ends of a modern creative continuum. In the real world, we start with complexity (noise) and attempt simplicity (harmony). In the virtual world, we start with simplicity and attempt complexity.

John Whitney said of his friend and contemporary, John Cage, that his disregard for conventional harmony showed a lack of musical appreciation (Whitney, J, 1980 p10). To Whitney, the organisation of frequencies into harmonic patterns is a fundamental defining characteristic of music, and harmony is at the root of musical composition. Whitney sought to develop a framework for visual synthesis which exhibited formal similarities to these fundamental aspects of music. To achieve this, he had to organise pixels into recognisable patterns utilising algorithms. In this situation, the exact mathematical nature of an algorithm is vital in generating the type of ordered complexity normally apparent in non-digital creative work. It may be the fact that he was engaged in this process that made it difficult for Whitney to engage with the revolutionary aspects of Cage’s work. In many ways, Whitney was at one end of the continuum, and Cage was at the other, placed there
Mick Grierson

by their chosen field of speciality. Whitney was one of the first composers ever to utilise synthesis, and although some of Cage's work deals with reproduction, the material is in no way synthetic in the same sense.

Whitney is well respected for his considerable contribution to science, art and the media industry. As well as being renowned as an artist in his own right, he was a key figure in optical film effects production, being responsible for the technology which made possible Douglas Trumbull's Stargate Sequence in Stanley Kubrick's *2001 a Space Odyssey* (1969). He had previously used this technology to create advertising for Coca Cola\textsuperscript{44}. In 1957, he had collaborated with Saul Bass to create the title sequence for Alfred Hitchcock's *Vertigo* - regarded as the first ever computer-generated title sequence. An algorithmic pattern was photographed one frame at a time from a high resolution monochrome monitor, and then colourised. In addition, he had by this time committed a considerable amount of experimental computer graphics to film. Whitney\textsuperscript{45} was also an avant garde computer animation pioneer, winning first prize at the First International Experimental Film Competition in Belgium in 1949. He is revered amongst abstract video and film artist for his work in the field, and his contribution to the canon of twentieth century abstract animation.

As such, it is interesting to note that visual art was not necessarily his first love. He spent a year in Paris studying serialist music with Rene Liebowitz in 1937. In addition, he worked on his own 'electronic compositions' (about which not much is known), before becoming

\textsuperscript{44} Slit Scan Photography: A camera technique where a matte box with a thin slit is placed over the camera lens. The slit can be moved either by hand or using a motor, allowing only certain sections of the film to be exposed at any one time.

\textsuperscript{45} Brother James Whitney collaborated with John Whitney during earlier works, but their techniques diverged later, James preferring to animate by hand.
Mick Grierson

interested in the relationship between the perception of sound and the perception of light. What seems to have attracted him was the notion of 'movement' in music, and the comparative experience of motion in visual art forms, such as film. It appears that through analysis of the mechanics of both art forms, he was able to inform and develop his theory of 'Differential Dynamics', a theory which he asserts is an application of the mathematical basis of music to visual art.

It could be said that in the earliest Whitney works (*Five abstract Film Exercises*, 1944), Whitney's vision appears more completely *audiovisual* than in his later films. The use of pantographs in conjunction with synchronised synthetic sound (see below) is an exercise in developing congruent abstract audiovisual motion. Sound does not predicate the image in the early exercises, and neither does the image generate the sound. The creation of both sound and image is a simultaneous process, painstakingly plotted out on graph paper and then shot frame by frame. As such, John Whitney is one of the primogenitors of audiovisual composition.

Whitney's later works deal specifically with the attempt to create a metaphorical model for the 'harmonic motion' of patterns. As such they do not exhibit the close sound-image relationships apparent in his early abstract films, and are not as focussed on the *fusion* of sonic and visual creativity. However, Whitney's later, purely visual works are tremendously sophisticated. Throughout the sixties, seventies and eighties, Whitney chose to focus on the technology, concepts and processes governing elemental forms of pure visual synthesis. As such, his later films are an attempt at generating a *musical* visual experience.

---

46 This may be because they were created in collaboration with his brother, James - although it is clear that James Whitney's artistic motivations were very different, centring on metaphysical/spiritual themes

47 Except in as much as the audio is printed onto the optical track.
Mick Grierson

using early digital means. The sonic aspects of his composition methodology were quite possibly sidelined by his desire to perfect a model for visual synthesis as a first step in generating audiovisual computer art. As well as being of special interest to the development of audiovisual composition, it also plays an important role in the history of algorithmic computer graphics.

Whitney's Differential Dynamics successfully models the behaviour of sound and our appreciation of it as music in three main ways. First, it takes as a basis the idea of harmonic resonance and relates it to the idea of visual symmetry. Secondly, it takes the idea of the harmonic series and applies it to a series of nodes in a visual pattern. Thirdly, it takes the idea of tension and release, and applies it to the process of moving towards and away from symmetrical forms in a system of patterns. As such it is a system for creating visual composition from a framework based on our experience of music. This means that it is highly compatible with the desire to develop a system for composition whereby both the sound and the image generation process are calibrated on the same plane with a similar degree of temporal relation. This particular technique demonstrates significant similarity to our general experience of music.

Problems of Audiovisual Synthesis: Translating Between Sounds and Images: Temporality, Translation and Scale

The oscilloscope is a device for displaying electrical or electronic representations of sound waves on a screen. The signal flow from negative to positive represents the compression and rarefaction of vibrations as they interact with the molecules of a gas, generating sound. Oscilloscopes are reasonably effective in displaying the amplitude of a signal in
Mick Grierson

real time, and as such are often used in order to create generative visual art. Technically, the oscilloscope provides a means for the visual analysis of sound (among other things). However, if the frequency of a sound is too complex (i.e. If there are frequencies within the material which are faster than the refresh rate of the screen), it naturally becomes harder to discern a similarity between the image and the sound. The fundamental aspects of a sound are still apparent, but the internal harmonic relationships can be difficult to perceive. This is due to the speed and complexity of the variation exhibited by complex sounds. Simple sounds with clearly defined, obvious transients (such as repetitive dance music) can produce acceptable results. This is an in-built limitation of direct audio visualisation, and clearly demonstrates the problem of scale.

It could be said that the way in which artists such as Oskar Fischinger have attempted to generate audiovisual work takes greater advantage of multi-sensory perception than an oscilloscope representation does. His work, and the work of others such as Vibeke Sorensen attempts to match the experience of music through images via the use of texture, rhythm, variation and complexity. This approach is less scientific, but perhaps more successful than a simple oscilloscope or oscilloscope-inspired interpretation of sound. This implies that their approach to the process of composition allows for a more appropriate translation between the sonic and visual material. Vibeke Sorensen is transparent about the fact that she is attempting to deal with the what she refers to as 'the binding problem' - a term for describing the phenomenon that sounds and images (as well as other sensory input), seem to form an entire unit in the brain, as if they are a single definite and intended event when they occur over a similar time frame (see Chapter 1.

---

48 See Cthuga, or more recently, Winamp, Windows Media Player and itunes for examples of this type 'Audiovisualisation'. Again, this is not audiovisual composition – it is algorithmic visualisation.
49 Hence the focus shift towards less sonically related generative visual effects within audiovisualisation software, such as colour cycling, and random particle generation (Cthuga etc).
Mick Grierson

Theory of Audiovisual Composition). In effect, these artists have been, and still are interested, at least in part, in the process and dynamics of translation between aural and visual events.

As is discussed in chapter 2, any critique centreing around 'Mickey Mousing' of sound and image relationships cannot be applied to this kind of art for two reasons. First, the art is abstract – so its interpretation is inherently more complex than conventional narrative material. Secondly, the work attempts to exploit the experiential binding of audio and visual material directly. It is this very phenomenon which leads us towards our point of discussion (Birtwistle, A 2002). Whitney, Sorensen and countless others have been trying to explore 'audiovisual relationships', to discover congruities and their opposites in terms of the way that we, as human beings perceive sound and visual events, not necessarily the way that they actually are as 'things-in-themselves'.

So, in order to take the visual synthesis work of John Whitney into the audiovisual realm, the visual motion patterning exhibited by Differential Dynamics must be 'mapped' into sound synthesis using an existing synthesis method. As part of this research, several attempts were made before settling on a suitable model. The main difficulty in establishing connectivity between Whitney's complex pattern synthesis and the synthesis of sounds is the difference in scale. In Whitney's earliest work, his sound synthesiser functioned at the same scale as the animation system. A system for controlling the position and movement of cell animated pantographs was designed and developed alongside what could be described as an early (perhaps even the first) optical sound synthesiser. By using a series of adjustable weight pendulums, waveforms and patterns were 'drawn' onto the optical sound track of the film, effectively in 'slow motion'.
This is similar to the technique used by Oskar Fischinger in his early pixelation works, and additionally by Daphne Oram in her real time optical synthesis. However in the case of the Whitneys' invention, the waveforms were not drawn by hand on to the cells, but instead controlled mechanically. The 'score' was a list of states, with each harmonic or overtone in the sound being represented by the speed of each pendulum; the exact frequency being dictated by the position of the weight on the pendulum, and the overall speed of the film. The result was a sophisticated additive synthesiser able to produce a wide variety of timbres from reasonably pure components. The synthesis was not 'real time'. Effectively, the soundtrack was animated i.e. the temporal framework of sound production was very similar to the process of film animation. Both aspects of the production process functioned on a similar scale, and could be manipulated together on a similar level. In this way, the Whitney brothers were able to deal with sound and image synthesis, exploiting the relative novelty of both the forms whilst controlling the output with the same level of detail simultaneously.

This reveals Whitney's approach for combining audio and visual material. When one deals with the creation of absolutely abstract sounds and images through technology, one has the opportunity, just as Whitney did, to connect the generation of sonic and visual elements through similar processes. The important aspect is that both the sonic and visual components must be calibrated so that their production technique is properly translated. The more accurate one makes the translation process, the more highly congruent the audiovisual material will be. The fundamental difficulty with respect to this is translating the material in such a way as to generate an effective result.

50 http://www.web-dictionary.org/encyclopedia/or/Oramics.html
A two-dimensional spectrograph only displays one unique time-slice of the timbre of a sound. Recording and synthesis technologies make it possible to analyse sound as a time line of modulating timbres. Likewise, Whitney's visual synthesis generates a kind of moving visual timbre (texture). In this way, it is the timbral quality (harmonic complexity) of a sound which relates more directly to the process of visual synthesis. This perspective is echoed in Abbado's thesis also (See Chapter 1).

That is not to say that changes in pitch do not have a symbolic impact on the perception of visual material. For example, Oskar Fischinger would often illustrate sections of melody by animating objects or shapes which were comparatively close together on the screen. In addition, the direction of a particular sequence of notes can have significant congruence with a visual technique or effect. For example, in Hitchcock's *Vertigo*, Herrmann's music oscillates in thirds, with both the upper and lower registers moving towards and away from the tonal centre in opposite directions. In addition to providing an eerie soundtrack to the film, it also acts as a sonic representation of the contrazoom effect employed by Hitchcock at crucial moments throughout the film. So it is not the case that notes have no relevance in audiovisual composition. It is simply that in synthesis, the timbre and texture relationships may initially be more significant.

The timbre of a sound is directly related to its overtones. A sound is considered 'noisy' when it has a large number of loud overtones which are not mathematically related. In

---

51 There are many examples of this type of shape-melody relationship in Fischinger's work. These include *Koloraturen* (Coloratura - 1932), and *Ornament Sound Experiments* (1932), among many others. See Nicholas Cook’s excellent analysis of *Fantasia* (1940) in *Analysing Musical Multimedia* (1998).

52 The contrazoom – the process of zooming in whilst tracking out – or vice versa, resulting in a gradual revealing or obstructing aspects of the on screen space, due to the actual distance between the camera and the space being altered.
Mick Grierson

some fields, these sounds are referred to as 'complex' tones. To extend this idea, we could also say that a sound's complexity can relate to the complexity of an image. We can then use this relationship (a texture/timbre relationship) as a basis for generating abstract patterns or shapes within reasonable limits in terms of their complexity. When trying to achieve this in real-time, aspects of a sound's complexity can be mapped automatically to an image or pattern. This is where the problems of scale become most apparent when dealing with Whitney's visual synthesis. It does not matter whether the movement in the shape is an accurate representation of the complexity of the sound (as this is impossible anyway – see previous discussion regarding the speed of sound perception). What matters is that the relative texture of the material appears congruent (or stands out due to total incongruence – see Cook's binary opposition, p 21).

If one follows Abbado's approach, that synthetic sound and image can share structural relationships which exhibit congruence, then one finds that these ideas sit more comfortably with the notion of surface and shape than they do with the notion of Whitney's dynamic patterns. Whitney's patterning theories rely on the relative spacing of objects in order to function as a metaphor for harmonic motion. This being the case, the texture of a shape or the surface of the shape is not affected. The only important factors, according to the fundamentals of Whitney's theory, are the positioning of the points in space. However, it seems logical that once the relationship between visual pattern and sound synthesis is solidified, the structural aspects relating to shape and surface can be applied secondarily. Before this can be done, however, the problem of the relationship between the nature of abstract visual material and the nature of sound needs to be addressed practically. As it happens, addressing this issue reveals solutions which make the application of Abbado's theory of audiovisual structural relationships more effective.
Whitney's perspective on complementarity between music and visual art asserts that our responses to sets of moving visual patterns can be shown to echo aspects of our experience of music. That is to say, through convergence and divergence towards and away from recognisable patterns, visual material can be made to behave like musical material. This approach can be supported by the research of John Sloboda. Sloboda states that human responses to music can be predicted in terms of intensity through an analysis of pattern usage within the work. In *Music and Emotion: Theory and Research* (Sloboda, J. A. and Juslin, P. (Eds.) 2001) Sloboda supports the theory that even though it is impossible to generalise regarding the emotions people experience given a certain piece of music, the *intensity* with which those emotions are felt can be reliably predicted and measured. The use of patterns can cause a build up of expectation through repetition and variation. When expectations are not met, this generates tension. When they are fulfilled, this creates a release. It is this aspect of our experience of artistic material which Whitney is referring to when he talks of *complementarity* between sound and vision. Whitney effectively *translated* aspects of musical experience to a visual format through experimentation and analysis.

Most of his work deals with manipulating a matrix of points. In his model, the points are interconnected, being part of the same system (like a complex particle system), and each point in the system is a representation of the motion of the system at that particular phase. This is a useful analogy - it translates well to certain aspects of synthesis (such as...
Frequency Modulation, Phase Modulation and Physical Modelling). The difference between the speed of oscillation and the speed of shape or surface movement is the main barrier when connecting audio and visual components. However, if one restricts the connections to the development of patterns in both musical and visual material, both audiovisual components are experienced in a very similar way. The process of recognising the movement from one pattern to another occurs at similar speeds in each medium. So perceptible changes in patterns provide a means for establishing an abstract audiovisual relationship.

Patterns are not only a major part of our experience of art. They are also a major part of our existence. We make sense of the world through the use of patterns. Our concept of time is regulated through the application of patterns - we divide up our time so that we can organise our lives successfully. Time effectively gives us a unique identifier for every moment of our lives. However, the notion of time progressing in equal steps is an illusion brought on by our familiarity with the concept of time. Our perception of time is known to be flexible. Commonly, the composer plays with notions of time through establishing patterns, and then restructuring them in order to play with the audience's expectation. The way that this is achieved is often purely intuitive. However, the patterns almost always have a mathematical basis: patterns are a form of order.

Maths offers us a way to understand the world. The use of mathematics in music and art succeeds in providing familiarity – a type of understanding. In composition, the use of pattern satisfies a human need to establish order (Sloboda, J.A. 2001). The belief in a deity, for example, could be said to spring out of the human desire to see the universe as coherent and ordered. Beyond this, however, the universe exhibits strong evidence of
coherence (such as life). Mathematics is developed through observation. The basis of mathematics and pattern making is essentially repetition and similarity. As repetition and similarity occurs naturally, its inclusion in art can be seen as representative of coherence in the universe. Equally, the use of noise and incoherence in art could be seen as representative of disorder and chaos in the universe. In this way, abstract art can use order and disorder to resonate with our perception of the universe. This is certainly the case with the abstract artworks submitted as part of this research.

In life, as in art, the divergence from harmony, resonance or pattern, towards noise, destruction and chaos is a fundamental and defining factor of existence. Entropy is the loss of coherence and structure. It is the entropy of noise and chaos which creates tension. It is harmony and order which creates resolution. This relationship between our desire to order the natural world, and the belief that the universe is in constant flux between order and chaos functions as the inspiration for James Whitney's *Yantra* (1957).

Curves generated by mathematical functions are given names which take their inspiration from nature, such as the 'Roses of Grandi'.\(^{53}\) Additionally, algorithmic computer graphics research engages with the various practical and aesthetic functions of fractal geometry - which is similar to the territory that Whitney's work occupies. However, Whitney's work also deals primarily with the movement from one pattern to the next. It is the journey between states, the speed of this journey, and the dissonance moving towards resonance, which drew him towards that field of work.

\(^{53}\) The 'Roses of Grandi', created at some time between 1723 and 1728 by Guido Grandi, is the curve used in many of Whitney's works, although he is a little oblique about this in his writings. It is so called due to the fact that it generates a variety of multi-petalled rose shaped curves.
Whitney explains differential dynamics by providing the example of a length of rubber with holes spaced at regular intervals. If one wraps the rubber round a stick, at certain tensions, the holes within the rubber will line up to create patterns. The manner in which these 'holes' or points in the system are created and manipulated is key. It is both highly complex and highly ordered - much like a fractal.

Many fractals are generated as filled colour images, with the actual intensity of the colour being related to the length of time it takes for a point to iterate through a function beyond a set boundary. The amount of processing power required to generate an image is prohibitive in terms of showing the fractal moving from one state to another\textsuperscript{54}. Fractal self-similarities exist in one frame, and they are generated and explored by iterating through a set function (in most cases). In this way, it could be said that many fractals produce a two-dimensional map of how functions alter over time.

With Whitney it is very different. A field of points within a matrix defines the resolution at which the fractal, or function curve, is rendered. Also, unlike the majority of modern fractal generation techniques, the function is simple enough to allow it to be manipulated over time with the use of computers. It is the change in the values defining the function that creates the 'pattern system'. It is these pattern systems that express self-similarity. A pattern at one time may be a small echo, or large iteration of a previously arranged pattern. It is the movement from one function to the next which creates resonance and dissonance (basically, symmetry and noise in the image). Whitney's specific visual synthesis technique explores fractal-like relationships across the dimension of time - a dimension which places the change in pattern, self-similarity and resonance upon the

\textsuperscript{54} This is not necessarily the case when one uses the less efficient 'Choas game' to generate the fractal.
same plane as that of music. The changes in pattern behaviour echo the pattern behaviour of sound and music\textsuperscript{55}.

\textsuperscript{55} The exploration of the above concept using traditional fractals, by altering the equation during real-time rendering may lead to a specific type of fractal audiovisual synthesis.
7. Technical Construction: Reproducing Whitney's Differential Dynamics

In order to put Whitney's differential dynamics into practice, a system had to be devised which allowed independent and/or relative control of a number of points in a matrix. To begin with, sixty points were used. If the first point in the system moved 1 unit (or pixel), then the second point would move 2 units, and so on, the sixtieth point moving sixty units. This process produces a dynamic complex system based on a relatively small amount of input data. Additionally, this bears some relation to the physics of sound - or at least the patterns of complex tones. The way in which the points in the system move is governed by a function rule, or similar, such as \( r = n \sin \theta \) - producing a set of polar co-ordinates. This can be plotted as either polar co-ordinates or converted to Cartesian form by taking the angle of rotation to produce \( x \) and \( y \) (\( \cos \theta = x \), \( \sin \theta = y \)), and then multiplying by the radius. Although Whitney's actual 'Roses of Grandi' program was never published, he does make a passing comment on it. It was possible to combine his published Fortran software with the details provided in Digital Harmony in order to re-create his method.

(See DVD extract 'Video Synthesis 2002' – (mute) for a record of the initial implementation of Whitney's Differential Dynamics generated by this research. In addition, please see CDROM extract entitled 'Visual Synthesis1.mpg' for a video presentation).

---

56 The Roses of Grandi
57 The final chapter of Whitney's 'Complementarity of Music and Visual Art' contains a few basic notes, programs and diagrams which were used as a starting point for research.
Mick Grierson

These are the two numbers you give to the equation.

$\exp(\text{amp}, \text{phase} \cdot \#1)$

This is the equation. It is set up to produce 36 pixels.

We end up with the polar coordinates (amplitude and phase).

This algorithm is needed for each pixel. We want to render, so we need all of these:

$\text{amp}$, $\text{phase}$, $\#1$

This takes the polar coordinates and multiplies them by a number unique to the pixel.

$\sin$  

The coordinates are translated to Cartesian XY screen space.

$\#0$, $\#5$, $\#6$

$\#0$, $\#5$, $\#6$

$\text{setPixel}(0,0,\#4)$

This cell in the matrix is drawn by this command.
Further to the application of the standard 2D curve functions defined by Whitney, the system has been refined and improved in a variety of ways. First, the algorithm has been altered to allow it to function with virtual objects in 3 dimensions. The library being used to allow particles to be drawn to the screen (Silicon Graphics OpenGL) is mainly intended for the position, translation and lighting of objects in virtual 3D space. In order to exploit this factor, Whitney's original program has been changed so that each particle can be represented by a geometric shape such as a sphere, cube, tube or even a complex model. Additionally, basic lighting has been implemented, providing shading and depth. In addition, lights can be 'moved' in real time using the same algorithm. Most importantly, the particles which make up the system can move in three dimensions, across a Z plane. This diagram shows the matrix of points (represented by cubes) being translated in 3d space.

The second major improvement to the system concerns the use and manipulation of
algorithms (functions). As a result of being able to manipulate the device in real time, the range of creative possibilities is considerably wider. At one stage, Whitney would have spent hours waiting for a section to render before he could see the result. As has been already stated, the fact that the machine performs the algorithm in real time allows for the composer/performer to alter the nature of the algorithm as it is being produced. In addition to the fact that the variables within a function can be controlled in real time, the actual function can be altered as it is processing the input. This is a major benefit when attempting to create varied works. Below can be seen the section of the user interface which controls the synthesis algorithm. The mathematical formula is constructed by choosing a combination of mathematical operators. Each object in the system has a unique number, which is then used to differentiate each object's movement (usually by multiplying the result of the equation by each object's unique number). The user can select the desired function whilst data is being processed. In addition, combinations of mathematical functions can be selected, stored and recalled as preset messages.
The third major development is in terms of colour. Colour can be defined and manipulated in a number of complex ways. Colour and transparency information can be generated through algorithmic functions; that is to say, just as the position of an object can be selected with reference to its position in the system, the colour of an object can be determined in the same way. For example: In modern computer graphics and video processing, colour values are specified through the value of 4 different 8 bit values – one for each of the three primary colours, red, green and blue, and another for the alpha value. These four values can be set by an algorithm whereby the current object’s position is taken into account. This can be either the object’s position relative to the screen, or the

---

58 In terms of RGB colourspace.
59 The alpha value refers to the transparency, in situations where transparency is being used.
object's position relative to the system i.e. the 10th object in the system can use its unique number to generate a unique colour value, which can then be altered in ratio to the rest of the objects in the system.

This allows for the generation of patterns with a huge amount of colour variation. In addition to this, the colour values can be manipulated with reference to the audio material as it is produced. Furthermore, video images can be textured onto the 3d objects. This is the process which I have referred to as audiovisual sampling synthesis, and combines the techniques described in Chapter 5 with those discussed here. For a more complete analysis, including the discussion of audiovisual performances using this technique, see Chapter 9.

**Achieving a Congruent Translation Method**

Once the basics of Whitney's synthesis had been understood, replicated and improved, a decision had to be made. Should one try to make the visual synthesis match a separately generated audio signal (as is the case with audiovisualisation\(^60\)), or should one attempt synthesis based on analysis or ‘mapping’ of the visual system\(^61\)? I have chosen to attempt the latter, believing it to be a more suitable method of progress. It provides a great degree of control and flexibility over the output, and while this may make producing audiovisual works more complex, it provides a greater potential for creativity. The movement of a particle system on a screen which is being generated algorithmically is far easier to translate upwards in terms of velocity, than it is to take an audio stream and scale it

---
\(^{60}\) Audiovisualisation: A generic term which describes sound to light devices and software. Normally generative and non-specific. See Cthugha, Winamp, Windows Media Player, i-tunes etc.

\(^{61}\) This is actually the opposite of audiovisualisation – the visual pattern generates the sound.
downwards, towards an existing set of patterns. As has been discussed at length earlier in this chapter, sound is full of complex inharmonic spectra, making visual analysis an unreliable process, for reasons already explained.

Because each particle is essentially sending a value to an oscillator, the number of particles in the system is relative to the number of modulations that can occur within the sound. Additionally, self-similarity and patterned behaviour is inevitable, making the process of synthesis naturally resonant (provided that the instrument is calibrated correctly), with the sound approaching noise between periods of resolution (resolution occurring at stages which exhibit symmetry). This is accented due to the fact that Whitney's system is based on the principle of Pythagorean musical harmony, with harmonic resonance occurring at integer intervals. Echoes and reflections of patterns are also revealed within the sound - depending on the calibration of the modulating waves.
Light Speak was designed as a technical study in Audiovisual Synthesis. In the first successful set of examples, FM was used to generate the sound signals. For the initial studies, a drone signal was produced, and the pattern distribution was mapped to the modulation index of phase modulated signals. Each particle in the system represents a signal being used as part of an FM algorithm. This highlights an earlier discussion regarding audiovisual synthesis. It is the movement within the particle system which echoes the progression from musical dissonance and resonance to noise and dissonance.

Effectively, when the position of each particle in space is related to a particular timbre, the
instrument can function as a visual representation of those timbres. Because of the way the particle system is organised, specific timbres can be 'locked in' to specific pattern types. Just as the space between the particles can be shifted, so can the distance between each signal in the system. However, the spectral distance is always shifted in ratio to the whole matrix. As such, the audiovisual synthesiser can be used for creating and controlling specific and repeatable audiovisual timbres.

The FM aspect of the instrument works through phase modulation, which is an efficient way of achieving frequency modulation with multiple operators. Each particle in the system is basically an operator helping to modulate a carrier frequency. The carrier frequency could be any signal. A Sine wave is the usual starting point for FM synthesis, so the mapping used in Light Speak is sine wave based. In other works, especially where full

---

62 This could be incorrectly interpreted as a type of audiovisualisation. However, audiovisualisation does not have access to the actual frequencies discreetly. This could be achieved through the use of a continuous and discreet Fourier transform, but this is currently almost impossible to achieve in real-time.
motion video files are being modulated by the Differential Dynamics particle synthesis engine, simultaneous modulation of an audio track is not performed by FM\textsuperscript{63}.

FM operators have two main parameters\textsuperscript{64}: modulation frequency and modulation index. If the modulation frequency is mathematically related (i.e. a factor or multiple) of the carrier frequency, harmonic sidebands are produced. This is the fundamental concept behind FM synthesis. The number of harmonics is dependent on the modulation index - basically the amplitude of the modulation frequency. If the modulation index is zero, then there is no effect. As the amplitude increases, so does the frequency and number of the overtones/harmonics (if the modulation frequency is not related, the sidebands will be inharmonic partials). By altering the modulation index, the operator changes the bandwidth of the signal.

There are several ways the system can be connected, with each exhibiting varying degrees of congruency. In the case used for Light Speak, the modulation index is dictated by the distance from the central particle. The greater the distance from zero, the wider the bandwidth of the signal. Essentially, this is achieved by calculating the radius of the particle from zero, and using that value to control the amplitude of the modulating frequency relative to the particle number in the system. As the particle passes through zero, the number of sidebands is reduced to zero (or a set number defined as part of the calibration process). The particles move at different velocities, following the set trajectory. This produces highly complex and congruent modulations which are all generated and

\textsuperscript{63} Other types of audiovisual synthesis which have been attempted include FFT resynthesis, and polyphonic digital filtering algorithms driven by differential dynamics.

\textsuperscript{64} J. Chowning "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation" J. Audio Eng. Soc. 21, no. 7 (1973):526-534
controlled by the particle system.

A key drawback of FM synthesis is that it is difficult to control. If one intends to create a specific timbre, then it can take a long time to determine the algorithm which should be used (depending on how complex the timbre is). When only one modulator is used, it is relatively simple to calculate the harmonics which will be produced given the frequency and amplitude of the modulator. When using multiple operators, controlling the modulation index of each operator in real time can be taxing, and the output difficult to predict with accuracy. Using Differential Dynamics, however, a high degree of control and accuracy can be achieved, and the nature and effect of the multiple modulations can be altered by changing the algorithm which controls the particles. A sixty operator FM synthesiser can produce a highly complex and constantly shifting sound through the manipulation of only one value. Additionally, the visual feedback produced by the instrument provides an engineer or operator with a very effective way of predicting and reproducing a complex set of multi-operator FM algorithms. The Timbre can be identically reproduced by recalling the shape created by numbers fed into the input\textsuperscript{65}. The two control numbers (plus two or three optional calibration factors) are all that is required. Alternatively, a keyboard map can be produced which provides a set of keys for certain shapes, enabling an engineer to recall an image and its related FM timbre instantly by pattern recognition.

\textsuperscript{65} Provided that the spectral width of the particle system is also recalled
Controlling Pattern Progress

The system implements low frequency modulators in order to control aspects of pattern development. These act much like a common LFO, only they relate directly to the numbers being fed into the input stream. It was initially implemented as a simple way of alternating back and forth through modulations (pattern systems) automatically. It can also be used as a simple sequencer. Upper and lower limits of the pattern progression can be altered dynamically, allowing an engineer or composer to move from one set of patterns through to another. The speed of the LFO progression is directly related to the alterations in timbre.

In the case of the synthesiser developed for the piece Light Speak, the FM model has been incorporated into a conventional subtractive synthesis model. The FM oscillator is fed into the mixer section of the synthesiser, where the signal is combined with other oscillators. In addition, there are several important enhancements to the overall operation of the instrument. The particle system and synthesis engine can be controlled by transient generators (envelopes). The alpha value (transparency) of the system, the relative size of the particles, the overall rotation, position and scale can all be controlled by envelopes and low frequency oscillators. Additionally, the position of the Whitney particles can be set via the user interface, so that when a key is pressed, the whole system moves to the desired position.
In this diagram, the user interface can be seen. The conventional functions of subtractive synthesis have been applied to the visual patterning system. In the above example, the filter cut-off frequency has been mapped to the transparency of the objects in the system. When the filter is closed, the object is completely transparent. When the filter is fully open, the object is opaque. The extent to which the filter affects the alpha value can be set via a calibration slider.

The particle system is moved through Y (or X) and Z directions by the pitch envelope and
amplitude envelope respectively. The amplitude envelopes control the movement of the object system in the Z plane, that is - towards and away from the viewer. The pitch envelopes translate the object system vertically or horizontally. When the pitch is modulated, this translates the object system in either of these directions around a central point (usually the point in the XY plane where the system originated). The relative positioning of the objects can be set separately, and the degree to which the envelopes affect the object system can be set by a separate controller. In addition, the object system’s rotation can be controlled via an LFO, allowing the entire object system to spin with relation to its centre as it is being translated through the screen space. The size of the particles in the system can also be set differentially. That is to say, the first particle in the system will be of a particular size, and the size of all of the other particles in the system will be set by this in ratio relative to their position.
Mick Grierson

In order to control the various functions of the synthesiser (including the pattern progress of the object system), a MIDI control surface can be used. This allows for all of the functions available via the GUI to be mapped onto the physical controller. With reference to the position of the particles in the system, it is calibrated so that during operation via the control surface, the system interpolates integer numbers between 0-127. Therefore, when the machine is left alone, it will resolve to a harmonic pattern. This is true for the alterations in both radius and angle of the particles in the system, as well as three dimensional transformations. This makes control via MIDI devices elementary. However, if the system is left in the same state for extended periods of time (without any spectral movement whatsoever) the result can be a little dull.

Thus, Differential Dynamics is used here in conjunction with conventional subtractive synthesis to generate a system whereby highly complex 3 dimensional visual patterns are created and manipulated alongside definite and specific audio synthesis. Repeating a given function always results in the same visual and sonic effect. The instrument used in Light Speak is the first step in creating a complete audiovisual synthesiser which is both complex and highly ordered, in addition to being completely repeatable with predictable results. Additionally, the process of algorithmic computer graphics composition which has been used allows for sophisticated multi-operator real-time FM synthesis with direct visual feedback of the timbre being generated.

As is the case with audiovisual sampling, when modulations in an audio track go beyond frequencies which can be discerned as separate events, there is no longer any point in modulating the visual material at that rate. In these cases, visual modulations can be

---

66 Graphic User Interface.
restricted in speed. For example, when a tone is frequency modulated beyond 20hz, it begins to alter the frequency content of the sound. This can be represented visually in a number of ways. The important factor is that the visual representation of this modulation maintains the audiovisual link in a stimulating way. There has been some attempt to do this with later pieces, such as *Structure Interne*, but the idea in general may require more research at a later period.
8. Audiovisual Synthesis via FFT analysis of Video Data

(Note: For a video presentation of this software please see the file 'Aqua Demonstration.mpg' on the CDROM)

Aqua

As part of this research, there have been several attempts to synthesise audiovisual material in different ways. In addition to the application of Frequency and Phase modulation, a form a FFT resynthesis has been deployed in order to generate the material for the installation work Aqua.

The piece functions both as a study in synthesis, and also as a creative work. The installation itself consists of a single screen display and speakers arranged at one end of a room, with the speakers in a stereo configuration. There are two modes of operation. First, in cases where a touch screen is available, users can enter the space and touch the screen. This stimulates a fluid dynamics algorithm which manipulates the image on the screen, with the audiovisual effect emanating from the point where it is being touched. In the case where there is no touch screen available, the user can use a mouse to trigger the algorithm, generating the same effect. Ideally, the installation screen should be as large as possible so that the audience can be more completely submerged in the piece and its effect. The algorithm generates a wide range of results, from delicate waves, to enormous ‘storms’. Depending on how it is stimulated, the piece is capable of being incredibly fierce. The image synthesis also controls an FFT audio filter, so that when the waves are at rest,
there is no sound. When the algorithm is stimulated, the filter opens up set frequencies, depending on the position of the pixels on the screen. The effect should be dynamic enough to draw the audience in, and then to shock them with a seemingly brutal reaction.

The behaviour of water is extremely dynamic, varying from small drops to great tidal waves. Many people find the sound of a small waterfall relaxing and comforting. However, water can also be violent. This is the idea behind *Aqua*. Water can be deceptive. We perceive water, like most things in nature, in different ways, lying in the balance between beauty and horror, resonance and dissonance. In this way the piece serves as an analogy for the unpredictability of natural forces.

Water is also a useful subject for composition due to its physical and environmental properties, or more specifically, its formal qualities. Changes in atmospheric pressure cause waves of varying speeds to be produced. As such it is similar to the way in which both sound and light behave. Indeed, in very dynamic situations (such as a waterfall), water can generate a huge variety of frequencies, resulting in a sound close to white noise. So as a theme for electronic composition, it offers material from a personal perspective, which is then interpreted through an interaction with technology. Of course, synthesising the sound of huge amounts of water is fairly simple, being similar in sound to white noise. However, as has been detailed in the previous two chapters, the key factor in producing a coherent and highly congruent audiovisual illusion is in matching and correctly scaling the different behaviour of the sonic and visual aspects of the work.
**Technical construction**

As part of the standard MAX/MSP/Jitter distribution, there is an example which utilises a fluid dynamics equation\(^\text{67}\). The technical thrust of *Aqua* has been to apply Whitney's differential dynamics to the given fluid dynamics example, and then to enhance the appearance of the display. Upon completion, the process of developing a congruent sound synthesis method begins.

The fluid dynamics model must be applied to a method of sound synthesis for the process

\[^{67}\text{../examples/jitter-examples/other/pool. The fluid dynamics equation is stated thus: h(x, y, t+1) = damping * (1/2 * ( h(x+1, y, t) + h(x-1, y, t) + h(x, y+1, t) + h(x, y-1, t)) - h(x, y, t-1))}\]
Mick Grierson

to work. In order to achieve this, the decision was made to use the technique of FFT filtering. An existing sound source is used as a starting point, and then the signal is analysed by FFT. Once the signal has been transformed from the time domain to the frequency domain (which is the purpose of the FFT analysis), the amplitude or phase of any frequency component in the signal can be altered in real time with a high degree of accuracy. The main difficulty with this theory is again the process of translation. The analysis of digital video data normally results in a two dimensional output. However, FFT analysis is one dimensional\(^68\). More specifically, an FFT analysis results in a graph of frequency against amplitude. This is essentially expressed as a one dimensional array of numbers, with each number representing the amplitude of the frequencies that may possibly be contained in the signal. If the signal does not contain a particular frequency, then it will not appear in the analysis, resulting in an amplitude of zero. In contrast, video information is a two dimensional array (such as a 320 x 240 matrix of RGBA colour values).

In order to solve these translation difficulties, two solutions were adopted. First, the two dimensional video data had to be transformed into a set of single dimensional data. Secondly the data had to be transformed from RGBA colour values (four sets of data) to a single data set. In the system pictured below, First, the data set is reduced by converting the RGBA data to luminance data. That means that the colour values are dropped, and the overall brightness values are used instead. This effectively turns a colour image into a black and white image. Secondly, and most significantly, the screen is divided up into multiple sections or 'slices'. Each slice of information is generated by averaging the pixel values from a specific number of lines (in this case, ten lines), therefore converting a range

\(^{68}\) In the sense that the results are expressed in only one array. Normally, video data is a 2D matrix of changing values. FFT data is a 1D array. There is only one amplitude value per frequency.
Mick Grierson

of slightly different values to a single array of values. Each slice of video information is then sent to one of a number of FFT bins. Therefore, the screen is divided up into multiple arrays of movement, with each array being processed simultaneously. The incoming audio signal is then filtered by each FFT process simultaneously, producing a polyphonic multi-band filter controlled by the visual data.

An array of FFT filters take input from a 160 by 120 matrix of values.

In the piece Aqua, the congruency between the visual and audio material is generated by effective audiovisual modelling of fluid dynamics. Just as in the case of Whitney’s differential dynamics, because the synthesised screen material is reasonably well
organised, it is far easier to develop audiovisual interaction than with more random 'real world' material. Below is a still from a study in the effect of the *Aqua* filter on a piece of real world footage. The lack of simplicity in the image makes the effect of the filter difficult to perceive. However, it is possible to restrict the filters so that only the important elements of motion within the image are used as data for the FFT arrays.
9. Audiovisual Sampling Synthesis

Structure Interne

Through a process of applying existing real-time sound production methods to visual material, and the application of informed translation between audio and visual material of both recorded and synthetic types, two distinct models for audiovisual composition have arisen: Audiovisual Sampling and Audiovisual Synthesis. Both of these models carry with them unique aesthetic baggage which flows into and out of the process of composition. Due to factors inherent in the process of recording and reproduction, the sampling methodology bears a heavy cultural weight. Regardless of whether video material is originally created or not, video processing carries with it the possibility of the abstraction of real-world representations. This lends itself to particular types of aesthetic uses, such as the cut-up. As all art could be described as a desire to bend the world to one's internal vision, the process of acquiring and manipulating representational footage places images of the world, with all its noise and texture, at the mercy of deconstruction. In stark contrast, the process of synthesis is almost the exact opposite. Forms originate in the abstract. Through mathematical processing and manipulation, the artist attempts to speak of experience from a starting point of very basic materials which might appear to bear no relation to the real nature of experience in the first instance. Yet, through a process of development, the synthetic is made to function as a metaphor for various types of experience due to the type and manner of control to which it can be subjected.

In the world of electronic sound production, the most effective model for the generation of
complete works is the sampling synthesiser. This refers to the process whereby a piece of real-world material is acquired, and then subjected to a process of synthesis. By effectively combining the two concepts of sampling and synthesis, their defining aspects as described in the previous chapters can be fundamentally altered. In the realms of synthesis, we attempt to produce complexity via the combination of simple building blocks. However, if we introduce real-world sound into the synthetic world as one of the building blocks, we can jump start the process, injecting the noise and complexity of the real world straight into the heart of the synthesiser. The synthesiser becomes capable of producing extremely complex and realistic sounding audio material. However, the synthetic world also has a dramatic and useful impact on the recorded representation of the real world. Whereas before, sampled material lent itself to the process of deconstruction, almost never completely removing itself from its primary function as a real world representation, after it has been combined with synthesis, it is transformed into a more abstract, less culturally specific form, less direct in its referentiality. It becomes partially synthetic, masking aspects of its reality, and in turn revealing new elements within itself which were not previously apparent, such as its density and texture.

When this process is applied to visual material, we can see that the same thing occurs. For example, the piece *Structure Interne* is a study in the application of synthetic visual and audiovisual processes to audiovisual sampled material and vice versa. The piece is inspired by a combination of factors. First, it is a piece for the piano. This grows out of traditional approaches to composition practice. The central process began by recording video and audio footage of a piano in a particular environment. This was then used as the sole material for producing the piece, intentionally echoing the early music concrète work
Mick Grierson

of Pierre Schaeffer\textsuperscript{69}. As this is the first piece written for performance with the audiovisual sampling synthesiser, it seemed fitting to make reference to Schaeffer’s work, and the beginnings of music concrète in this way. Secondly, in this work, the piano acts as a symbol for conventional conservative musical values. As a part of the process of generating the video material, the piano was separated into several pieces. The recording of the piano's destruction was achieved through collaboration with sculptor Siobhan Timoney, who both sourced the instruments, and provided the space.

The piano was filmed whilst being beaten, struck, bent, and dropped. Its internal workings were used as a means of beating, striking, and breaking the instrument. The hammers were removed and used as beating implements. Strings were woven into other strings, then dragged back and forth in a rhythmic motion. In addition, the piano was swung from the ceiling, and struck whilst it rotated. The manner of performance was generally violent and aggressive, and the treatment of the piano was filmed in such as way as to produce a violent effect. That is not to say that this was the only effect achieved by the process. It was purely my intention to record the activity in such a way as to accent the violent nature of the piano's deconstruction and destruction.

When capturing the video footage, two types of shot were used. Wide angle shots, and close-ups. Stationary wide angle shots were taken with the camera sitting on or next to the piano, with the remains of the instrument stretching away from the camera. The internal workings of the piano are not commonly seen from this angle. As a result, when one views the footage it may not be immediately apparent what the object is. The surface of the piano from foreground to background produces the illusion of landscape. This is as a result

\textsuperscript{69} Etude au piano I and II Etude au piano II (1948)
of the heightened sense of perspective. When the piano was struck, the striking implement generally appeared in the frame, enhancing the sense of movement, whilst adding a visceral quality to the material. In contrast, the close up shots were generally taken in such a way as to highlight the movement of the strings. When the strings are beaten, they can be seen oscillating. Finally, there was an aesthetic need to remove any reference to human interaction from the material. This was done by ensuring that no part of any person ever entered the frame

The desire to destroy the piano, and to record its ‘suffering’, rises from an impulse to attack the dominant conventionality and conservatism inherent in western art music. This includes attitudes towards technology, sound as music, performance, culture, improvisation and experimental composition. As such, it is an impulse which exhibits a close relationship to the aesthetic of experimental sound art, such as the work of John Cage. In addition, it also resonates with the general aesthetic of Punk. It is a reference to the radical impulses which can be located within many 20th century art movements, such as Dada, Surrealism, the Beatniks, Situationism, abstract art, and experimental electronic music. The idea that these movements are connected to a general culture of resistance is detailed in the book Lipstick Traces (Marcus, G. 1989). However, fifty years after Cage and Stockhausen came to prominence, the dominant ideologies within western art music remain the preservation of existing and well understood hegemonic cultural forms (which do not include the further exploration of experimentalism initiated by Cage, Schaeffer or Stockhausen), or else the production of art for the sole purposes of financial gain. So, like other works in the portfolio, significant aspects of the composition rise out of radicalism, in

---

70 The sound of voices can be heard at times during the piece. This was due to interference in the studio space. However, rather than eliminating this material, I chose to keep the soundtrack as it was to preserve the resulting unsettling effect.
response to conservative values which can be linked to the problems inherent in postmodern culture, or the cultural logic of late capitalism (Jameson, F. 1991).

With regard to Structure Interne, if the completed piece was simply an edited version of the untreated material, it would have been fair to criticise the work as contrived and powerless. The power within the work does not come from the destructive aspects of its making. The work is intended to reveal the true structural nature of the piano i.e. its materiality. The images and sounds of the piano are re-synthesised into a new form which attempts to capture the essence of the piano's structure, and also the essence of the movement within the video material. The goal of the piece is the exposition of the internal structure of both the piano, and also the structural relationship between its images and sounds. With respect to the visual aesthetic, it is inspired directly by Stan Brakhage. The re-synthesised images are an attempt to create a metaphor for the way in which movement is synthesised in the mind through collections of images. The images represent distorted memories, twisted by the process of remembering and imagining. They are dreamlike representations of experience, existing as an attempt to visualise the essence of internal visual experience – the pure abstract movement of colours and shapes in mental space (see Brakhage, S. Night Music 1986).

Structure Interne is different to almost all of the other pieces in the portfolio in that it has a much more considered pace than many of the other works. This is because it attempts to model different aspects of both audiovisual experience and memory. In Remote Control, Beautiful Human Canon and WMD (Third World Warhol), the speed of the editing process in part is an attempt to represent the freneticism of the process of remembering (although it also reflects many other things). When one has an extreme experience, for example, the
memory of that experience can resurface sporadically and when least expected. The mind can easily be destabilised, bombarded with memories and emotions. The memory revisits the moment again and again, slowing it down, speeding it up, replaying it, perhaps in an attempt to better understand or alter it – to see it in a different light. If the memory is particularly nice, the frenetic back and forth of the memory can be irritating, as one tries to replay the event in sequence, within a realistic time frame. If the memory is negative, each visitation is like torture, making one flinch with fear or even pain. Memories can be relentless and unforgiving. Works such as Third World Warhol attempt to model this relentless behaviour, in an attempt to reach into the mind, and to speak about its processes. In contrast, Structure Interne, is an attempt to capture the shape and texture of distorted or half-dreamt memories. It is dreamlike. Both approaches deal with the externalisation of mental visual processes. As such, they are both heavily inspired by Brakhage (as are all of the pieces in the portfolio). However, they deal with different aspects of our internal visual (and auditory) worlds.

Structure Interne is in three main sections. It begins with an exposition, where the main themes are expressed, both audiovisually, and visually alone. The lack of sound during certain parts of the exposition functions in two main ways. It is an attempt to allow the visual images to resonate in isolation, so that when they are revisited, they have an increased power and memorability. In addition to this, it allows the visual material to function as a voice on its own, in the hope that this will generate a greater degree of counterpoint during the development section. Once the main visual and sonic ideas have been expressed, the piece continues through a process of repetition with increased velocity, leading up to the first red noise climax. After this point, there is a blank section. This reflects the silence which occurs at the start of the piece. However, in addition it
Mick Grierson

allows the sound image moment to resonate in the mind of the audience, like a dramatic pause. These audiovisual silences function in such a way as to enhance what has been seen, whilst simultaneously generating tension through an audience's anticipation of what is to come. Like many of the other pieces in the portfolio, relentlessness is key to its aesthetic, functioning as a metaphor for violence and suffering. However, in the case of Structure Interne, the violence is intended to appear premeditated and deliberate, and not merely reactive or instinctual. The blank sections emphasise the deliberate harshness of the piece, by reinforcing what has already appeared, and what is inevitably to come.

The development section explores themes already introduced in more detail. Material is reversed, slowed down and repeated, centring on moments of audiovisual impact. The real time modulating delay functions as a sonic counterpoint to the existing audio material, providing both a sense of space, and also a second voice, generating high frequencies in opposition to the low and deliberate striking of the piano. Visually, the material is re-interpreted in a number of ways. First, the material is organised via the Whitney formula so as to produce an echoing symmetrical matrix. This matrix of video material resonates when similar colours pass in front of each other. This resonance affects the spatial positioning of the sounds, on occasion generating four channel spatialisation which comes close to panning modulation synthesis. Secondly, the visual material is stretched and re-positioned to fill the screen at moments of high volume. This is to create the illusion of the piece bearing down on the audience.

Following on from the main development section, the piece enters a second red noise stage. This recapitulation signals that the piece is nearing completion. Before the final moments, however, there is a revisiting of the initial opening theme, which again resolves
Mick Grierson

to a series of still and noiseless images. These images resonate in the silence as a final closing statement, like a black and white memory.
Technical Construction: Re-synthesising Images

(See the DVD extract entitled 'Visual Synthesis 2004' (mute) for a demonstration of the visual techniques expressed in this section. In addition, please see CDROM extract entitled 'Visual Synthesis2.mpg' for a video presentation of the technology)

As has been discussed, Structure Interne is the first piece to be completed using real-time audiovisual sampling synthesis. The piece was performed live at IRCAM at the 2003 Resonances festival using an Apple Macintosh G4 dual 800 with a 64MB NVIDIA graphics card. There are several aspects of the technical construction which are important with respect to using the instrument in performance, and as such they are detailed here.

The most important factor with regard to the combination of audiovisual sampling technology and audiovisual synthesis is in applying the video information, which is the output of the audiovisual sampler, to the surfaces of the geometry used by the audiovisual synthesiser. This is done by a process called texturing. The video material is used as a texture for the colour values which are applied to the surface of the geometry. For example, if one of the shapes in the Whitney particle system is a square, then when the video image is piped into the rendering engine as a texture, each side of the square has the image applied to it. The exact nature of the texture's appearance depends on specific mapping co-ordinates which can be defined separately.
In the case of *Structure Interne*, each object in the system has the same video image source as its texture. Under normal circumstances, if the particle system is at rest with all objects in the central position, only the main video image will be seen. If the remaining particles are extrapolated in the Z plane either towards or away from the viewer, the object nearest the viewer will probably be the only one to be seen, unless the object’s size is relative to its position in the system in such a way as to make it appear smaller (see below). In both these cases, there is no visible effect of applying the texture to the objects, save for the possibility of mapping a rectangular video image onto a spherical object.
Mick Grierson

In order for the texturing process to take advantage of the object system, alpha\textsuperscript{71} blending must be enabled. This allows for the transparency of each particle in the system to have an effect on the colours within the texture. There are a variety of different blend modes available through the use of different video cards, each resulting in a different kind of blend. However, it is through calibration of the alpha blend modes that the texturing can be allowed to function in an effective way. Once blending is enabled, the relative colour values also have a more distinct and interesting effect. When some blend modes are in operation, the intensity of the colours within the image will combine when objects pass in front and behind one another.

\textsuperscript{71} This is the technique of combining colour information through the use of an Alpha ‘channel’, or more specifically, a value which controls transparency/opacity.
This process generates a form of colour feedback which gives the impression of oscillating through virtual 3 dimensional space. This is quite effective in abstracting the image, revealing the more intense shapes and colours, whilst nullifying the less intense ones, depending on the blend mode being used. In this way, the straight edges of each texture become less noticeable, giving the simulated 3d objects more of a noisy, real-world appearance. Thus, the application of textures to the geometry of the visual synthesis method can be demonstrated to have a similar effect to the application of samples in sound synthesis. The output of the software is infected with the noisiness of real life, allowing it to produce much more complex and compelling material more quickly. The
processor overhead of this operation can be controlled by reducing the number of particles in the system. This does have an effect on the intensity and detail of the material, but is a compromise worth making. In order to perform *Structure Interne* at IRCAM, the number of particles had to be reduced from 30 to 15 to allow the audio processing to function effectively. As technology becomes more powerful and efficient, these restrictions will allow for real-time performance at a higher resolution.

The connectivity between the sonic and visual material is of greatest importance in this model. There are several important factors with respect to the way audio synthesis and sampling have been combined. First, the objects in the system normally have the same texture applied to them. If a second texture is desired, another system would be necessary. As such, there is only one audio stream being generated by the visual material, despite the fact that there are up to 60 copies of the visual material in the object system. Just as in the audiovisual synthesiser described in the previous chapter, the object's overall parameters are controlled via envelopes and modulators. In addition to this, the position of the objects in the system is controlled by the amplitude envelope. In the case of *Structure Interne*, the following sound image mappings are used. The filter and alpha values are controlled via the same envelope, just as in the previous model. This means that as the filter is opened up, the image slowly appears. As the image dies away, the filter is gradually closed until no sound is heard. In addition to this, the intensity of the colour is calibrated to relate to two functions of the audio material. The amplitude of the audio is used to control the saturation of the image. Due to the violent changes in dynamics, and slowly vibrating bass frequencies, the image saturation oscillates wildly at moments of high impact and noise. This also means that when the video footage is slowed

---

72 This could be seen as a form of polyphony.

73 This is expediency, as only one copy is actually required.
Mick Grierson

down, the oscillations vary in ratio. This is effective in generating congruence with the resonating frequencies. Also, the colour values of the image are related to the speed of video playback. When the video footage is slowed, the colour values change. When the video stops, the image becomes black and white.

These mappings are not set in stone, however. Just like any synthesiser, elements of each generator can be applied to any aspect of either the image or sound through a simple routing system. This allows for a wide variety of different architectures and relationships to be instated.
10. Narrative Applications of Experimental Audiovisual Composition.

The Mudlark

The Mudlark is a short narrative film which springs out of an attempt to employ the aesthetics and techniques of audiovisual composition within the framework of conventional storytelling. In a way, it functions as a test bed for the applications of audiovisual sampling and synthesis, so as to demonstrate the practical implications of audiovisual composition theory. The film utilises a simple story structure inspired by the same themes which drive the more abstract work within the portfolio. However, the techniques which have been developed and discussed within the thesis are secondary within the film to the storytelling process. The work has been shot in a conventional way, with a central character on a journey to a conclusion. In this sense, its narrative is classical. This is an attempt to engage with an audience in a more accessible way. However, the material is generated and treated in a way identical to the rest of the work in the portfolio, with the possible exception of Third World Warhol, which utilises the technique of appropriation in order to discuss themes of appropriation and representation.

The video material itself was acquired in the usual way for narrative film making practice. A storyboard was developed and circulated to a small production team, including a central actor. Under close direction, the visual material was shot over a period of one weekend. This material was then assembled into a rough edit using Max/Msp. Following this, specific effects were developed and rendered with close reference to the audio material.

The treatments in The Mudlark are mainly a mixture of textural cross-synthesis and noise
Mick Grierson

synthesis. Almost all of the sounds within the piece are generated via subtractive and additive synthesis, with the exception of a small number of vocal sounds, and a piano/guitar theme. The custom audiovisual synthesiser is responsible for generating mud, water, noise and visual distortions which interact with the soundtrack. In addition, the whole work is framed by a version of John Conway's 'Game of Life' equation\textsuperscript{74}, which has been modified so as to represent a muddy growth of mould on the screen. This is done in order to place the viewer in the mud, giving the impression that the work has risen out of the same material which dominates the narrative. In addition, the game of life equation is linked directly to the sound synthesis engine, so that as the velocity of the sound effects increase, the mould effectively grows faster.

With respect to the compositional structure, the piece is in three main sections. There is a prologue, where the main character is introduced and placed in a situation. Following this, the character goes on a journey into the mud in search of coins. Finally, the epilogue shows the character buried in the mud. Sound cues place the main protagonist in a direct relationship with the crabs he encounters on the mudflats. In addition, visual cues relate his arm to other inanimate objects in the surrounding area.

Just as Whitney, Brakhage, Burroughs, Cage, Stockhausen and Schaefer (among many other experimental practitioners) contributed to the development of 'popular' culture through radicalism and experimentation, contemporary arts practice and conventional film making must embrace the complex, new and 'unpopular', in order to uncover the forms of culture which we will be making in the future. Audiovisual Composition allows us to progress to a more complete understanding of this new art of added value, an art form

\textsuperscript{74} Gardener, M 1970 Scientific American 223 pp120-123
Mick Grierson

which currently dominates western society at a point when our understanding of it is particularly weak. A proper engagement with it in an experimental and independent context will result in a better understanding of our relationship to our world, and the production of ever finer works of audiovisual art.


Conclusion

The purpose of this interdisciplinary research project has been fourfold. First of all, it has sought to extend and develop the vocabulary of contemporary multisensory practice by synthesising elements from sonic and visual disciplines. This approach has laid the groundwork for a combined metadiscipline, referred to here as Audiovisual Composition – an artistic form which takes as its starting point the cognitive actuality of multisensory audiovisual experience. Secondly, the project has successfully focussed on the task of developing the technical tools necessary to make this aesthetic extension possible. Through a process of experimentation and innovation, instruments have been created for the precise purpose of exploring audiovisual relationships in the context of contemporary arts practice. Thirdly, the project has positioned itself within the context of self conscious experimental audiovisual practice (including with respect to its theoretical elements), so that its developing vocabulary and tools can remain relevant and useful to the practices from which it has sprung. Finally, the project has attempted to demonstrate some links between cultural / philosophical discourse and specific audiovisual creative activity. Fundamentally, this project has successfully produced a variety of experimental and developmental practical works and tools that are emblematic of a convergence between two disciplines that have traditionally been seen as separate.

The newly emerging debates surrounding the multisensory perception of audiovisual material is of huge significance, and its exploration within this thesis has been a parallel
project to the composition work. There are some important aspects of this debate which have been realised in practice. The highlighting and use of structural similarities between our experience of sound and our experience of moving images has resulted in an approach to audiovisual sampling synthesis which reflects a desire to both reinforce and develop ideas that have been central to the process of avant-garde and experimental practice throughout the twentieth century. This project's engagement with contemporary research into cognition and perception within the context of creative work has been an attempt to both solidify existing aesthetic elements, whilst simultaneously developing them in a new and useful way. These approaches are highly relevant to the development and aesthetics of audiovisual art. Most importantly, it may be the case that understanding the way in which we perceive audiovisual art is crucial to understanding the value of that art within a variety of contexts. This is even more important when one considers the degree to which sounds and images are an embedded feature of our environment.

The compositions in this portfolio offer a vision of artistic creation which is neither traditional music composition, nor filmmaking practice. They are an example of self-conscious experimentalist work, and are an attempt at creating challenging and stimulating material that rises directly from the tradition of avant-garde creative activity. As such, they demonstrate an engagement with the ideas, techniques and inspirations of the many practitioners who have attempted to explore the world of audiovision. Now that science has confirmed that there is indeed a significant multisensory level to experience, it seems appropriate that the study of this discipline should be encouraged for its own sake, and not simply as part of the debates about the nature of music, sound or moving images. These debates are certainly important, but not key to an understanding of what Chion has called
The audiovisual instruments created as part of this project can create a wide variety of textures and sounds. Fragments can be incorporated from anywhere, either entirely synthetically or through a combination of synthesis and sampling. In addition, the textures generated by the visual sampling instruments allow for the same velocity and intensity of cutting which is exhibited in works such as Third World Warhol and RemoteControl. As such, the representation of audiovisual connectivity as it may occur in the mind draws inspiration from the way we piece together fragments from our existence. Beyond the purely aesthetic applications of the instruments, they can also function as a metaphorical model for the way in which audiovisual material is dealt with in memory. The combination of audiovisual sampling and synthesis has in some ways been used as an attempt to create an analogue of the way we synthesise real-world material into abstract, dreamlike internal spaces. As such, it is a comprehensive and successful model which attempts to allow us to represent our existence through truly combined audiovisual means, as informed by the confirmation of the experience of added value.

Life is a cut-up. As soon as you walk down a street your consciousness is being cut by random factors. The cut-up is closer to the facts of human perception than linear narrative. (Burroughs, W, in Maeck, K, Commissioner of the Sewers, 1991, Video).

This project has used the concept of the cut-up as one of its main aesthetic starting points. However, the transferring of this cognitive concept over to the real-time digital audiovisual
domain is perhaps one of the projects most significant achievements. The incorporation of these concepts within the context a synthetic continuum offers a great deal of creative potential. Fundamentally, it sends a signal that there are indeed structural similarities between the way we both create and experience, and that these are the basic tools of the artist.

There are many areas where this research offers opportunities for further investigation. Several works in the portfolio attempt to explore the relationship between sonic timbre and visual texture. This idea has great compositional and exploratory potential. This potential would be significantly increased if one had the opportunity to generate and exploit congruence between sonic and visual events at higher frame rates (for example, 100 fps). This would greatly decrease the problems of scale. That is not to say that any effects would be perceived in the same way as material playing at 25 fps. Much of the illusion of movement generated above 60 fps may only be perceived subliminally. However, this is of great interest in itself. The obvious obstructions to this approach are twofold. First, there is a difficulty in generating complex real-time audiovisual material at high frame rates due to restrictions in both computing power and projection technology. Secondly, any work made in this way would not be easily distributable. However, it is possible that both of these issues may be overcome in the near future.

In addition to higher frame rates, increased efficiency in real-time rendering will undoubtedly bring about a huge improvement in both the stability of real-time systems, and also the available resolution. Much of the work that has been produced as part of the
Mick Grierson

composition portfolio is of a relatively low resolution when compared to High Definition Television. The opportunity to produce works of art at the improved 1080 resolution will undoubtedly revolutionise the audiovisual experience, one day allowing real-time work to rival the quality of traditional cinema.

Finally, one exciting potential area of future research may be in the exploration of cognitive and structural approaches to Audiovisual Composition. Through further analysis of the perceptual and cognitive aspects of aesthetic experience, experimental artists may develop entirely new ways of structuring audiovisual compositions that more directly and effectively interface with audiences and their experience of art. Interactive audiovisual artworks, where the viewer has the impression of personal agency, may result in more profound aesthetic experiences, reaching far beyond that which we currently imagine possible.
Mick Grierson

**Bibliography**


Mick Grierson

Modulation" Audio Eng. Soc. 21, no. 7 pp. 526-534.


Eco, U. (1997), Kant and the Platypus, Milan: Bompiani R.C.S. Libri SPA.


Gidal, P (ed. 1978), Structural Film Anthology, London: British Film Institute.


Jameson, F (1991), The Geopolitical Aesthetic: Cinema and Space in the World System,
Mick Grierson

Bloomington: Indiana University Press.


Ramachandran, V.S. and Hubbard Ed (2003), Hearing Colors, Tasting Shapes, Scientific
Mick Grierson


Schaeffer, P. (1967), La Musique Concrète, Presses Universitaires de France.


Mick Grierson


Mick Grierson

Film/Videography:


Delicacies of Molten Horror Synapse (1991), Stan Brakhage, Film, Independent.

Five Abstract Film Exercises (1944), Directed by Whitney, J & Whitney, J, Film, Independent.


In Absentia (2001), Stockhausen, K & ‘Brothers Quay’, Film, Arts Council of England/Canadian Film Fund, UK.

Matrix I (1971), John Whitney, Music: Antonio Soler sonatas arranged and performed on piano by Delores Stevens, Film, Independent.


Night Music (1986), Stan Brakhage, Film, Independent.

Towers Open Fire (1963), Burroughs, W.S. Directed by Balch, A. Film, Independent.

The Cut-ups (1966), Burroughs, W.S. Directed by Balch, A. Film, Independent.

Wavelength (1966), Michael Snow, Film, Independent.

Yantra (1950-55), James Whitney, Film, Independent.