What is Electroacoustic Audio-Visual Music?

Nomenclature and Cognition

Andrew Hill
Music, Technology and Innovation Research Centre,
De Montfort University,

United Kingdom

andrewhill86@gmail.com

In: Motje Wolf & Andrew Hill (Eds.)
Proceedings of Sound, Sight, Space and Play 2010
Postgraduate Symposium for the Creative Sonic Arts
De Montfort University Leicester, United Kingdom, 2-4 June 2010
http://www.mti.dmu.ac.uk/events-conferences/sssp2010/

Abstract

Electroacoustic audio-visual music works explore the possibilities that the combination of their two time-based media (sound and moving image) allow. Discussion of sound and image interaction is not new. The Ancient Greeks discussed it, Newton had a theory on the subject and numerous people devoted their lives to the development of colour organs in the 18th and 19th century in an attempt to realise an art form that brought together sound and light. The most liberating technical development for the genre was the invention of tools to capture sound and image, and most importantly, to play them back again alongside one another.

The development, and recent affordability, of digital technology has encouraged this genre to expand rapidly with VJ performances in clubs becoming regular and visualisation software on music media players being almost ubiquitous. While there has been an explosion in the prevalence of audio-visual media and visuals to accompany sound there is still a fairly limited selection of theoretical texts analysing the subject field.

Many different techniques and styles are employed in the composition and association of sound and image. I will introduce the discussion surrounding the definitions and nomenclature for works that make use of sound and light and present a model for the cognition of electroacoustic audio-visual music adapted from Anabel Cohen's "Congruence-associationist framework for understanding film-music communication" (Cohen 2001: 259).

What is Electroacoustic Audio-Visual Music?

Sound with Image

For hundreds of years musicians, artists and thinkers have imagined an art form in which sound and light are united. Understandably with so many people envisaging such an art form there are many different conceptions of what this art should be.

The first practical realisations surfaced in the 18th and 19th Centuries as a variety of colour organs began to be developed. The first reported instrument of this type is that of Father Louis Bertrand Castel, Jesuit priest and mathematician. Castel modified his own harpsichord and on December 21st 1734 the Clavecin Oculaire was first played in his Paris study. Thomas Wilfred provides a description of this instrument.

> 'it had a musical keyboard of five octaves. When a key was depressed, a colored strip of paper or silk would appear above a black horizontal screen to the rear. The first octave represented the pure hues, the next the same hues "one degree lighter," and the fifth octave the highest values.
>
> Newton had once suggested that C, being the lowest note in the octave, should be red, the lowest vibration in the spectrum.
>
> Castel decided C should be blue because it sounded blue. For the same reason he made F yellow-orange, where Newton had green, and the one choice is fully as justified as the other.' (Wilfred 1947, p.248)

The seemingly arbitrary and subjective nature of these pitch colour correlations highlight the first challenge encountered by those experimenting in this field. If there **is** an absolute pitch colour correlation what exactly is it?

In subsequent years many other artists and experimenters built their own colour organs, but for each new instrument a new set of pitch colour correlations emerged too (Figure 1). Such variation can also be found when looking at the colour key mappings of composers.

> 'Skriabin considered C major to be red [while] his fellow countryman Rimsky- Korsakov considered it to be white; and while both Russians agreed that D major was yellow, this brought them into conflict with a relatively consistent tradition in early nineteenth-century Germany that the yellow key was E major' (Cook 1998, p. 35).

The arbitrary association of colour to chord and colour to pitch in the above examples is highly individualistic and possesses little logical basis, apart perhaps from some form of subconscious or cultural conditioning. One individual may perceive a certain audio-visual relationship as difference while another does so as similarity. Despite the plentiful evidence demonstrating the contradictions and its futile nature many artists and researchers in the field are still intent upon discovering a *true and correct* mapping between colour and pitch.

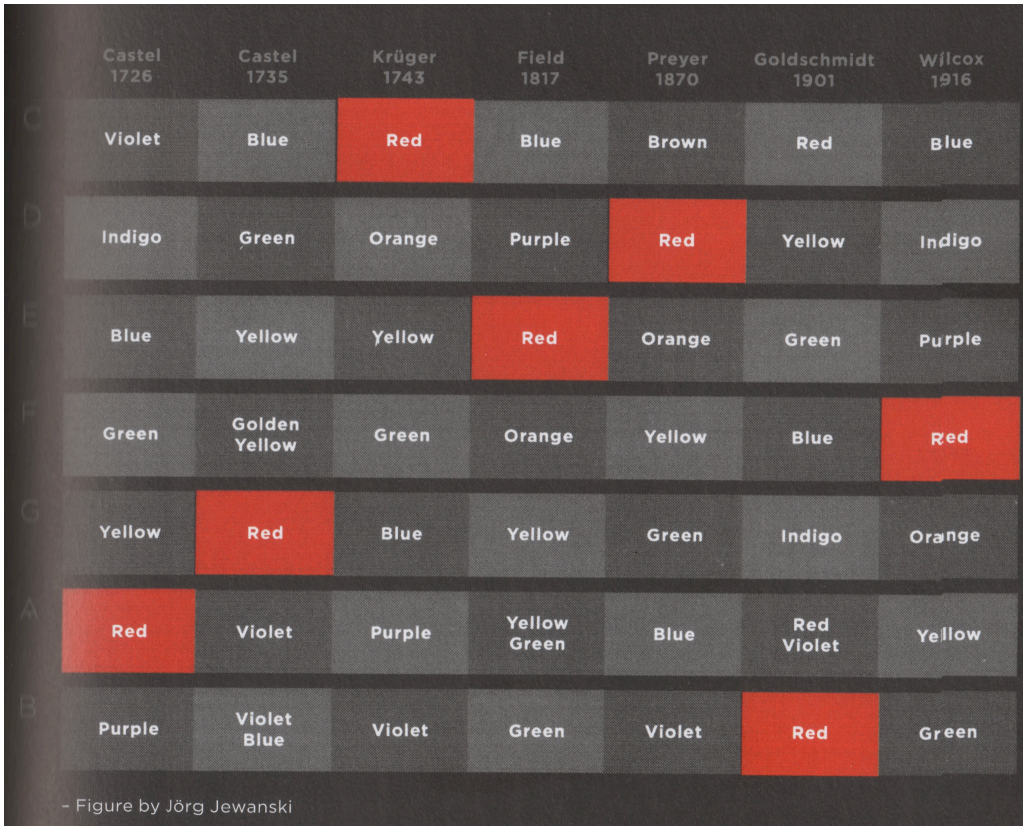
Image with Sound

With the rise of abstract painting many artists turned to music as a source of inspiration. The most famous of these artists is probably Wassily Kandinsky who was greatly inspired by music and named many of his abstract paintings composition or improvisation. The shift from still to moving image facilitated by the development of film led to the inception of "Absolute film" an artform that allowed visuals to develop and change over time.

> 'The term "Absolute Film" was coined by analogy with the expression "Absolute Music," referring to music like Bach's Brandenburg Concertos which had no reference to a story, poetry, dance, ceremony or any other thing besides the essential elements - harmonies, rhythms, melodies, counterpoints, etc. - of music itself.' (Moritz 1999).

Filmmakers like Hans Richter and Walther Ruttmann created silent animations in which visual forms evolved and flowed on screen.[1]

The next step was facilitated by the development of a medium that could store and re-present both sonic and



	Castel 1726	Castel 1735	Krüger 1743	Field 1817	Preyer 1870	Goldschmidt 1901	Wilcox 1916
C	Violet	Blue	Red	Blue	Brown	Red	Blue
D	Indigo	Green	Orange	Purple	Red	Yellow	Indigo
E	Blue	Yellow	Yellow	Red	Orange	Green	Purple
F	Green	Golden Yellow	Green	Orange	Yellow	Blue	Red
G	Yellow	Red	Blue	Yellow	Green	Indigo	Orange
A	Red	Violet	Purple	Yellow Green	Blue	Red Violet	Yellow
B	Purple	Violet Blue	Violet	Green	Violet	Red	Green

- Figure by Jörg Jewanski

Figure 1.

visual information. Since the development of Edison's Phonograph in 1877, which facilitated the recording and subsequent reproduction of sound, designers and scientists sought to realise the goal of reproducing sound and image simultaneously. Between 1880 and the 1920's many different machines were developed by competing researchers and scientists (see: Geduld 1975). Many of the early ventures utilised independent machines that separately dealt with the reproduction of sound and image but were synchronised either manually or mechanically. These were often cumbersome and unreliable and were eventually surpassed in the early 1920's by techniques facilitating the recording of sound onto film, and most importantly alongside the captured image (Ulano undated). The optical soundtrack made use of a photoelectric cell to transform the patterns of light and dark running along the edge of the film into electrical signals that were then amplified and connected to loudspeakers. Finally both the sound and image information was recorded onto a single medium and could be reproduced in synchrony.

These advances in technology made it possible for artists, such as Oskar Fishinger to compose animations frame by frame alongside Orchestral and Jazz recordings. Possibly the most famous example of this is Walt Disnesy *Fantasia* completed in 1940. Oskar Fishinger was employed to animate sequences accompanying Bach's Tocatta and Fuge in D minor (the first feature in the film). Unfortunately Fishinger's entirely abstract animations were too extreme for Walt Disney and so Fishinger was fired and his sequences adapted by subsequent artists so as to be more representational (Moritz 1999, p.84).

Around the same period (late 1930's into the 1940's) a number of artists (including Fishinger) began to experiment with composing both the visual animations their own audio tracks. The Whitney brothers developed complex machinery that translated the sine-wave motion of a series of pendulums into sound while other artists, such as Norman McLaren, began to draw sounds directly onto the optical soundtrack (James 1986).[2]

Continued developments in both sound and image media increased the possibilities of working in and across both domains and with the digitisation of both sound and image workflows the affordability of working in the audio-visual field has been greatly increased, while the impracticalities have been greatly reduced. Using modern digital technologies it is now possible to manipulate and assemble sound and image into an audio-visual composition on a single computer using almost identical workflows.

What's in a name?

Unsurprisingly, due to its diverse and interesting history, there are many varying forms of works that explore the interaction, and possibilities of combining, sound and image. With techniques and creative people being brought together from many different disciplines and backgrounds it is not surprising that the vocabulary and nomenclature quite rapidly becomes pretty confusing. This is further compounded by the fact that 'the speed of audiovisual praxis today far outstrips that of theory formation' (Daniels and Naumann 2010).

Visual music, absolute film, cinedance, light shows, lumia, videomusique and audio-visual music are all terms used by practitioners to describe their work. Each of these terms can refer to various historical or technical compositional methodologies yet all describe work exploring visual materials structured in a musical way. As with any art form practitioners often seek to associate their work in a historical context, appropriating names or terminology that are then applied to new styles, in some cases quite different form the original. This is one of the factors that leads to a confusing nomenclature in this field.

In an attempt to clearly distinguish between the different types of fixed media compositions in this field all works have been separated into the following four categories:

A) The purely visual approach to Visual music, for example Thomas Wilfred's *Lumia*, or the works of Kandinsky or Klee. Works that aim to emulate music, or contain structures and forms inspired by those within music but contain no sonic content themselves.

B) Visual composition to pre-existing musics such as in some of the early works of Oskar Fischinger, the artistic interpretations of Walt Disney's animators in the 1940 film *Fantasia* or music videos of the type found on MTV.

C) The composition of sound and image informed by traditions of music in which materials are structured within time. This form is here defined as audio-visual music because works contain both sonic and image elements. The sound and image are regarded as equal components joined in the context of a work and are both structured musically. A work itself would be an audio-visual composition.

D) The synthesis of visual materials from sound and the representation of sound visually. This includes visualization software such as those within media players, oscilloscopes and computer algorithms that render visually spectral and waveform images of sonic material. (Hill 2010a)

These distinctions strive to operate outside of historical contexts, focusing instead upon the properties of the work in question. The current research investigates what can be defined define as electroacoustic audio-visual music[3] classified within category C.

These works could also be described more succinctly as works of organised sound and image (Hill 2010a) (Landy

In: Motje Wolf & Andrew Hill (Eds.)
Proceedings of Sound, Sight, Space and Play 2010
Postgraduate Symposium for the Creative Sonic Arts
De Montfort University Leicester, United Kingdom, 2-4 June 2010
http://www.mti.dmu.ac.uk/events-conferences/sssp2010/

2007). Due to this the theories discussed within this paper have been developed with category C works as their focus and so while in some cases it may be possible to transfer the same theories to the other three categories this may not be universally true.

It is important at this point to note that while this form of composition is related to, it is not a subversion or rejection of Acousmatic tradition but something else. Quite often audio-visual works are met with resistance or suspicion by those in the acousmatic field as they are viewed as being in complete opposition to acousmatic ideals. However, as in the entire history of its development, a culture of cross fertilisation and exchange has always been vital to the development of audio-visual music and there is no reason why acousmatic and audio-visual music's should not continue to flourish in parallel. After all as Michel Chion writes 'We never see the same thing when we also hear; we don't hear the same thing when we see as well' (Chion 1994, p. xxvi).

A Model of Communication

Where audio and visual materials are combined as within an audio-visual discourse signification is formed in the mind of the listener/viewer. In composition the composer, either consciously or unconsciously, selects the most appropriate materials for the task at hand and combines them in a way that will fulfil their intent for the work. In some cases this compositional process may take into account:

- Individual material properties,
- The type of sound/image association within individual audio visual events,
- The impact of these individual events within the context of the entire work.

Anabel Cohen rationalises this same division within audience perception and suggests that the 'brain operates both through innate grouping principles and by learned connections' (Cohen 2001, p. 259). To describe how sound and image interactions operate within film she presents a model called the 'Congruence-associationist framework for understanding film-music communication' (Cohen 2001, p.259).

This model proposes that various channels representing the communicative elements of the film operate in parallel across four processing levels. Interaction occurs between channels and both ways along the channel (Figure 2).

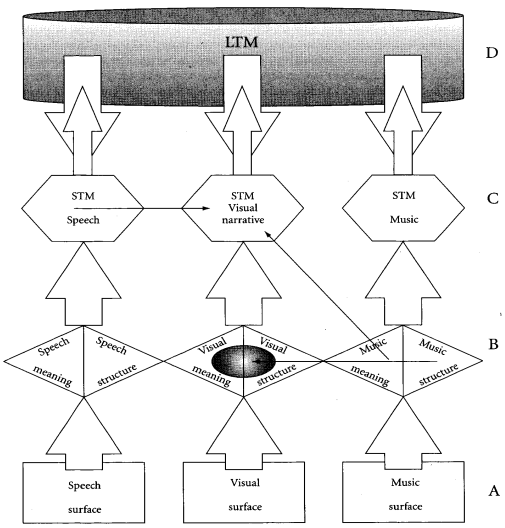


Figure 2.

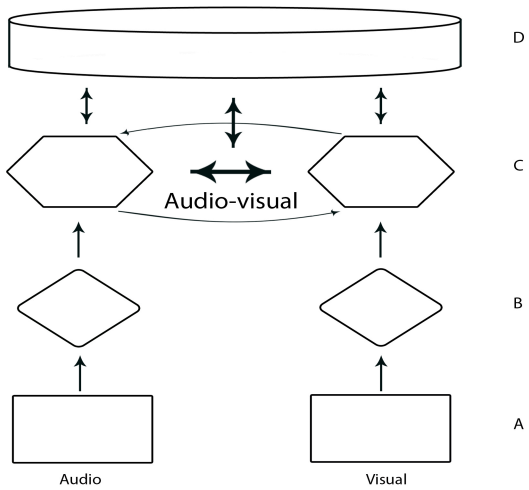


Figure 3.

The four processing levels begin with analysis of physical features of the sound/image element (A), through the analysis and construction of these elements into structural (gestalt) and semantic (associationist) information (B). The outcome of level B allows association between channels and can direct forms of pre-attention at higher levels. Short-term memory (STM) exists at level C where narrative information can be constructed from multiple sources, and their immediate contexts, in order for the audience member to make sense of the stimuli. The top level (D) represents the long-term memory (LTM) which receives information from all levels and exerts a top down inference with the goal of constructing a narrative across a wider time frame than that of the short-term memory and involving the integration of experiential knowledge (Cohen 2001, p. 260). This model has been developed and designed for understanding mainstream Hollywood film and so cannot be directly applied to electroacoustic audio-visual music. However it can serve as a very useful basis for the development of a suitable framework model.

To make a model appropriate for electroacoustic audio-visual music it is essential to first remove any focus upon the visual narrative. In Cohen's model both the speech and music channels feed ultimately into supporting the visual channel. Due to the understanding of an equal reliance upon sound and image for electroacoustic audio-visual works this visual centric approach must be removed.

Secondly the speech and music channels should be amalgamated into a single channel, in which there is no distinction between music, sound and utterance. In the first reworking all of these are recognised as the collective audio materials which make up the audio component of the electroacoustic audio-visual work.[4]

The new diagram that we have now created contains two channels, one for the visual element of an audio-visual work and another for the audio. (Figure 3) We have removed the visual centric interactions occurring at lower levels but have retained the interactions, both up and down individual channels. As it stands within this new model (Figure 3) visual and sonic information will be processed at each of the four levels within their discrete channels until level C and D. In Cohen's model it is only the short and long-term memories which infer between one another.

The new model proposes that the attack, sustain and decay of a sound will be perceived at level A. Level A then infers upwards to level B where the individual perceived properties allow for the construction of a perceived audio or visual event. This event from level B will then infer up to level C where, in relation to other similar events, a short-term relationship is perceived. Once again this is inferred upwards into the long-term memory at level D in which the individuals experiential knowledge is utilised to make sense of the combinations of events occurring at level C.

In order to test the suitability of this model (and to better explain it) it is most useful to work an example through it.

- At level A. Sonically: a sharp attack followed by a sustained section containing in-harmonic resonant pitches and a slow decay is perceived. Visually: a hollow, cup shaped, metal object is struck by a smaller metal object hanging in its hollow.
- At level B the sonic attack is perceived as a result of the smaller internal metal object striking the larger exterior object, and the sustain and decay are a result of the resonance of this larger cup shaped metal object.
- At level C this event is related to similar or nearby events in order to build up a picture of event relationships.

- At level D previous experience allows the perceiver to infer that this sound has come from an object commonly referred to as a bell.

The context in which this bell sound is located (the relationships between events at level C) allows the perceiver to infer its purpose through reference within the perceiver's experiential knowledge (level D). If it is accompanied by the sound of clinking glasses, a frantic scramble and the vocal phrase 'last orders', then the perceiver is able to infer that this bell is signalling closing time at the bar.

At first inspection the model appears to function appropriately. However we can soon begin to see that there are a great number of other possible interactions between hierarchical levels as yet unaccounted for. Knowledge from level D may be used to recognise the causes of stimuli present at level B. For example we do not need to know anything about the context (level C) if we recognise the event itself. Using the previous example to explain: our experiential (level D) knowledge of what a bell previously sounded and looked like can allow us to decode that the event perceived at level B is also a bell. If we have previously experienced an event we can discern what it is before we know anything about its context (level C).

In a similar way information from level D can also be utilised to interpret audio and visual elements at level A. Elements such as similar envelopes or spectral makeup can be recognised based upon past experience. A sharp attack can be recognised as a sharp attack, before it is conceived as either part of an event or within a context. Both of these possible interactions can be easily represented within the model through the addition of extra interaction arrows between level D and level B, and between level D and level A.

Recalling the Chion quote above[5] it becomes clear that there must also be interaction between the audio and visual channels in the construction of an audio-visual composite. Sonic material coinciding with visual material will radically alter the final interpretation and vice versa so it is necessary for our model to recognise this link. The hierarchical nature of sound and image interaction means that associations may be drawn between related audio and visual events (level B), and the combination of audio-visual events within a short-term context (level C). Such interactions will also obviously be related to the experiential knowledge at level D. This audio-visual interaction therefore requires the addition of a third audio-visual channel that exists between the individual sonic and visual channels at levels B, C and D. (Figure 4).

This audio-visual channel does not assume a position of primacy over the individual audio and visual channels but rather operates as a synthesis between them. The audience member might direct their focus of attention towards the visual narrative (as might be the case were an audio-visual composition to be shown at a film festival) or to the audio narrative (as might be the case if the piece were projected within an electroacoustic music concert). In such situations the focus of audience members would be shifted closer towards either of the appropriate individual channels. But, so long as they are not observing one of the individual channels in isolation, the perceptual focus will always remain within the audio-visual channel.

We can think of the audio-visual narrative as being the main component at level C. Between that of the audio and visual channels, receiving information from both at level B, but which might be pushed either towards the sonic or the visual depending on the context of a works presentation or the audience members perceptual focus. (Figure 4.)

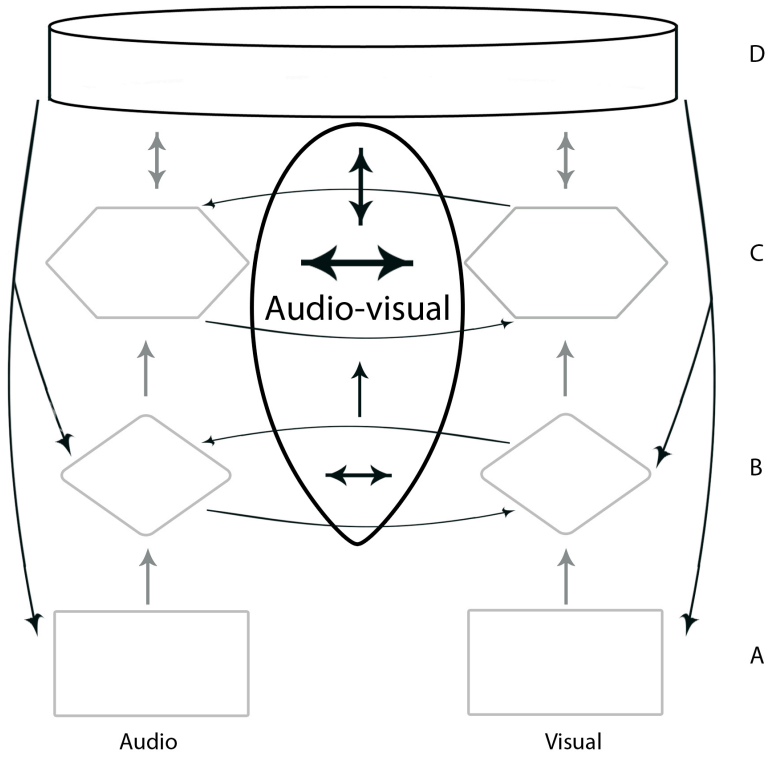


Figure 4.

Conclusion

With the availability of digital technologies for manipulating and assembling sound and image, more and more practitioners are beginning to experiment with the composition of audio-visual works. While conceived of for hundreds of years it has only been with the advent of recording technology that it has been possible to create audio-visual works exploring interactions that are not limited mechanically (as with colour organs). The collision of practitioners from the sonic and visual arts domains have left a confusing landscape of conflicting and duplicate terminology which it is essential to pick apart and clarify before it can be applied to the discussion of audio-visual music. The nomenclature and terms used here might not be universally accepted by those working in the field but are a necessary step in clarifying the difference between works allowing the progression to more complex analysis. One such analysis has been the investigation of audience perception for electroacoustic audio-visual music through a congruence associationist framework. This framework might hopefully provide an insight into how audio-visual works interact and might be used as a tool for composers wishing to understand how electroacoustic audio-visual works operate and how to create compelling associations between sound and image.

Bibliography

CHION, M (1994) *Audio-Vision*. New York: Columbia University Press.

COHEN, A. J. (2001) Music as a source of emotion in film. In Juslin and Sloboda eds. *Music and Emotion Theory and Research*. New York: Oxford University Press. pp. 249-272

COOK, N. (1998) *Analysing Musical Multimedia*. New York: Oxford University Press.

GEDULD, H. (1975) *The Birth of the Talkies: From Edison to Jolson*. London: Indiana University Press.

HILL, A. (2010a) Investigating Audience Reception of Electroacoustic Audio-visual Compositions: Developing an Effective Methodology. *eContact!* 12(4). Available from: http://cec.concordia.ca/econtact/12_4/hill_reception.html

HILL, A. (2010b) Desarrollo de un lenguaje para la música audiovisual electroacústica: investigación sobre su comunicación y clasificación. *En el Límite* — Escritos Sobre Sonido, Música, Imagen y Tecnología, pp. 144–165. Editado por Universidad Nacional de Lanús, 2010; compilado por Raúl Minsburg.

JAMES, R. S. (1986) Avant-Garde Sound-on-Film Techniques and Their Relationship to Electro-Acoustic Music. *The Musical Quaterly*. 72(1) pp. 74-89.

JEWANSKI, J. (2010) Colour-Tone Analogies in Dieter Daniels and Sandra Neumann Eds. *Audiovisuology: See this Sound, An interdisciplinary Compendium of Audiovisual Culture.* Köln (Cologne) : Walther König. pp. 345-

LANDY, L. (2007) *Understanding the art of Sound Organization*. Cambridge, Mas: MIT Press.

MORITZ, W. (1999) Lecture notes, *WRO99, Media Art Biennale*. Wrodaw, Poland: 1999.

MORITZ, W. (2004) *Optical Poetry: The Life and Work of Oskar Fishinger*. Bloomington: Indiana University Press

DANIELS, D. NAUMANN, S. (2010) *Audiovisuology: See This Sound, An interdisciplinary Compendium of Audiovisual Culture*. Köln (Cologne) : Walther König.

WILFRED, T. (1947) Light and The Artist. *Journal of Aesthetics and Art Criticism* 5(4), pp.247-255.

ULANO, (Undated) *Moving Pictures That Talk*. [WWW] FilmSound. Available from: http://www.filmsound.org/ulano/talkies.htm [Accessed 15/03/11]

Media

Pen Point Percussion (1950) Film. Directed by DON PETERS, LORNE BATCHELOR, NORMAN MCLAREN. Canada: National Film Board of Canada.

Weekend (1930) Audio-film. WALTHER RUTTMANN. Germany: Berlin Radio Hour.

[1] Interestingly Walther Ruttmann also experimented with the possibilities of sound montage and editing in the late 1920's using film with optical soundtracks. The most famous example of this experimentation is the composition *Weekend* (1930) commissioned by Berlin Radio Hour in 1928).

[2] The realities of this are eloquently explained in the 1950 film *Pen Point Percussion* demonstrating the compositional techniques of Norman McLarren).

[3] An electroacoustic audio-visual music work could be defined as a cohesive entity in which audio and visual materials are accessed, generated, explored and configured, primarily currently with the use of computer-based electronic technology, in the creation of a musically informed audio-visual expression. Electroacoustic audio-visual music works explore the possibilities that the combination of their two time-based media (sound and moving image) allow. (Hill 2010b)

[4] For works that make use of a spoken word narrative it may be desirable to retain the speech channel however its temporary removal greatly simplifies the ongoing discussion.

[5] We never see the same thing when we also hear; we don't hear the same thing when we see as well" (Chion 1994, p. xxvi).