Paper

Peer Review:

Copyright:

Open Access:

Digital Preservation:

PAPER

Composing Visual Music: Visual Music Practice at the Intersection of Technology, Audio-visual Rhythms and Human Traces

Julie Watkins

University of Greenwich, UK

J.Watkins@greenwich.ac.uk

Creators of visual music face the challenge of retaining their own artistic impetus amidst an overwhelming choice of instruments, aesthetics, practice, techniques and technologies brought about by the impinging presence of a vast sea of data and tools. Navigating the data-driven ephemerality of artistic technology and its market-driven constraints by utilising strategies similar to composer Ron Kuivila's (1998) for getting 'under', 'over' and 'into' will be examined with the aim of elucidating methodologies for creating works that other artistic practitioners may find useful.

Leading pioneers of visual music were, of necessity, innovators of technology as well as visual musicians and artists. There is an intrinsic tension between developing new technology in order to re-imagine how music can be made visible and technological pioneers succumbing to the fascination of exploring the technology itself. Understanding aspects of perception, such as rhythm, is key to developing new technologies and processes in ways that avoid this pitfall and keep the experience of visual music central. Audio-visual synchronisation and rhythm are vital to create, in the seminal computer artist John Whitney's words: 'an art that should look like music sounds' (1980: front dust jacket).

Integrating the body, human traces and especially the human voice into visual music compositions underpins the key objective which is to create work that is non-narrative, 'abstracted animation'[1] (Watkins, 2015), and yet suffused with human presence and emotion. Visual music can be perceived as overly repetitive, cold and alienating if it seems to embody a purely mechanical alignment of music to image, or if it seems disengaged from both human emotions and natural imagery.

---

[1] 'Abstracted Animation' is my own term; it refers to animation that has texture, depth and expressive movement, without overtly representing concrete reality.

This paper is part of an on-going investigation into developing methodologies for composing new abstract visual music pieces and, ultimately, parameters for a visual musical instrument.

Keywords: Visual music; Audio-visual perception; Ryhthm; Human-traces

Introduction

As Friedmann Dahn asserts, '[visual music] should be visible music, music made visible or, to expand the term an equal and meaningful synthesis of the visible and audible, and is therefore ultimately its own art form' (Dahn in Lund & Lund, 2009: 149). Painters, artists working in light, animators, musicians and V-Js have all contributed to the long and rich history of visual music (Watkins, 2016). Within visual music the possibility of creating a synthesis of the visible and audible has been debated in terms as varied as synaesthesia and a 1:1 mapping (see **Figure 1**); new possibilities
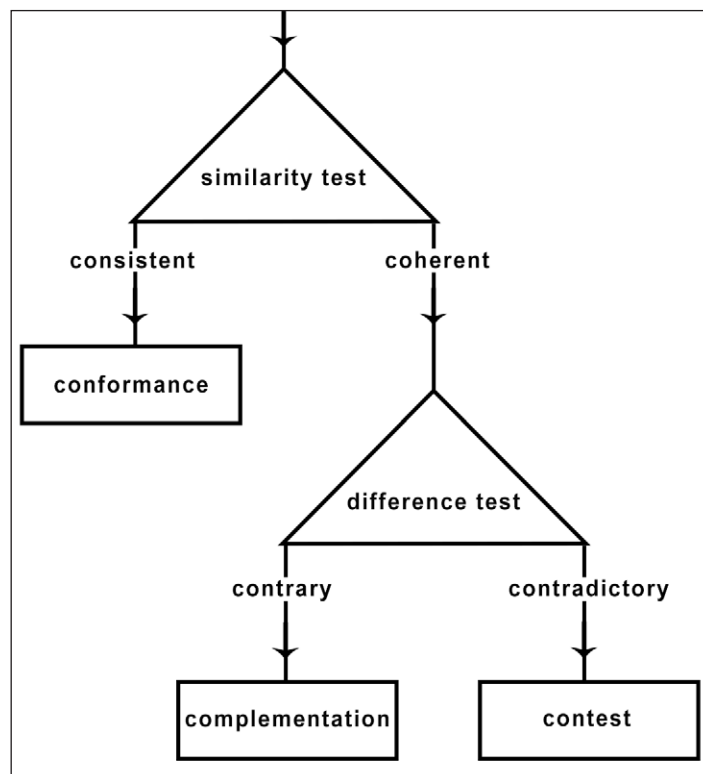
Figure 1: Diagram based on Cook (2000: 99).

afforded by current technology and new research into perception and multi-modality has given this debate new life (Gallese, 2016). Visual music will be considered in the light of: audio-visual perception, rhythm, audio-visual synchronisation, technology and human traces.

Perception of Sound, Visuals and Audio-visuals

The theorist Adrian Klein (1930: 37) posited: 'somehow or other, we have got to treat light, form and movement, as *sound* has already been treated. A satisfactory unity will never be found between these expressive media until they are reduced to the same terms.' However, our physiological, cognitive and emotional responses to visuals and sound are very different. Visual musicians, exemplified by seminal artists, such as Jordan Belson, have the ambition to communicate in a similar manner to music. Belson stated:

> I don't want there to be any ideas connected to my images, and if there are any there, if anybody sees any, those are entirely in the eyes of the beholder [...] Actually, the films are not meant to be explained, analysed, or understood. They are more experiential, more like listening to music (Belson in Brougher et al., 2005: 148).

Listening can be immersive. When music provokes emotions the listener attends to both the music and their own reaction more closely. This is a reinforcing cycle. Generally, Grewe et al., state, this process is implicit (Grewe, Nagel, Kopiez, & Altenmüller, 2007: 313): 'Our own mind seems to react automatically to music. We sense no effort; music is re-creation, but yet it is the listener's re-creation.' Emotional responses to music are influenced by: individual hinterlands, prior experiences, expectations, memories and associations.

Hearing and sight function very differently; we process acoustic waves into a perception of sound in a completely different way to how we process electromagnetic waves into vision. 'Synaesthesia is an involuntary response in one sense, such as sight, triggered by the stimulation of another sense, such as hearing' (Watkins, 2016). Creating a meaningful synthesis of visuals and sound would be greatly simplified if we were all synesthetes and, additionally, we all experienced sound-as-colour and colour-as-sound in a similar way. Clearly this is not part of normative cognition but

the idea of synaesthesia has given impetus to visual music. I would argue that the idea of synesthesia, or pseudo-synaesthesia, is an expression of the cross-modal integration of the senses.

When viewing audio-visual works physiological, perceptual, cognitive and emotional effects are intertwined. Audio-visual pieces have a different effect from audio or visual pieces alone, as the seminal American editor Walter Murch points out: 'We never see the same thing when we also hear; we don't hear the same thing when we see as well' (Chion, Gorbman, & Murch, 1994: xxii). Audio greatly affects what we see, it changes our understanding of what we see and how what we see resonates emotionally. We watch audio-visual works and appreciate their sound, for the most part, visually.

Rudolf Arnheim, the German-born perceptual psychologist and visual theorist suggests: 'the ear is the tool of reasoning; it is best suited to receive material that has been given shape by man already – whereas seeing is a direct experience, the gathering of sensory raw material' (2007: 195). Experientially I find 'shaped' audio much easier to listen to. However, as French theorist Michel Chion observes 'sound more than image has the ability to saturate and short-circuit our perception' (Chion et al., 1994: 33). Sound generally has more of a direct physiological effect than vision, for example film viewers' breathing can be changed by the breathing noises on the sound track of a film. The direct effect of sound may be due to how sound is experienced; sound is in the air, surrounding viewers. In contrast screen-based images are localised to the screen. The audience will remember the images more readily and understand the images more rapidly if the sounds support the images, see the British musicologist Nicholas Cook (see **Figure 1**); this enables a faster and deeper immersion in the work.

As Chion (1994) argues the immersive effect of audio-visuals is not due to synesthesia but is trans-sensorial in nature. Some perceptions are unique to eye or ear, for example, whereas colour is only experienced visually, pitches and the inter-relationships between pitches are only experienced auditorially. However, the majority of perceptions, including perception of rhythm, texture, and material affect both senses. Forms such as music, radio and silent film are less sensorially complete than

audio-visuals and so allow the audience to engage their imaginations to fill the sensory gaps. The equivalent mode of engaging the audience in an audio-visual work is the metaphoric use of sound, i.e. reassociating less expected sounds with images to enrich their relationship by adding a measure of ambiguity. As Murch argues:

> The metaphoric use of sound is one of the most fruitful, flexible, and inexpensive means: by choosing carefully what to eliminate, and then reassociating different sounds that seem at first hearing to be somewhat at odds with the accompanying image, the filmmaker can open up a perceptual vacuum into which the mind of the audience must inevitably rush (Chion, 1994: xx).

To make this metaphoric use of sound possible the viewer must accept that, as Chion defines his 'audio-visual contract' (1994: 222): 'the elements of sound and image to be participating in one and the same entity or world'. Though a limited use of acousmatic sounds will not break audience immersion, viewers generally expect to see the causes of the sound in the images on screen; this is how synchronisation and synthesis, in Chion's terms 'synchresis', occurs and the images gain 'added value', added emotion or information, from sound. Clearly this is predicated on the viewer being able to bridge the gap between the 'reassociated' sounds and images. This leads to an examination of the relationship between similarity and difference that is at the core of audio-visual media. Nicholas Cook analyses the relationship of vision and audio in this way: 'The pre-condition of metaphor – and if I am right, of cross-media interaction – is what I shall call an enabling similarity…Rather than simply representing or reproducing an existing meaning, it participates in the creation of a new one' (2000: 70). What this means is that the images and audio are not so similar as to be redundant nor so different as to be contradictory and in contest with each other; rather the images and audio dynamically complement each other and thus a new meaning is constructed (see **Figure 1**).

In non-representative work, such as 'abstracted animation' the role of rhythm is crucial in determining whether the result of the difference test is 'contrary' or

'contradictory'. Completely arrhythmic, asynchronous audio-visuals are likely to be 'contradictory' and in contest with each other.

Rhythm and Audio-Visual Synchronisation

Rhythm is a vital component in creating a meaningful synthesis of vision and audio. The visual music instrument designer and composer Fred Collopy concludes that:

> Rhythm has played a particularly important role in the thinking of painters who have been interested in the relationship of music to their work. There is a rhythmic element to each of the three dimensions. The changing of colors is rhythmic, the ways in which forms are arranged (even in static images) is often described in terms of rhythm, and movement in time is inherently rhythmic. This suggests that rhythm constitutes a particularly rich point of entry for the design of instruments and for the development of technique for playing visuals in performance with music (2000: 360).

Perception of musical rhythm[2] has been extensively researched. Timing and tempo rely on the individual listener's perception and cognition: the listener organises their understanding of a rhythm. As Henkjan Honing the Dutch theorist of music cognition concludes: 'A listener does not perceive rhythm as an abstract unity, as is notated in a score, nor as a continuum in the way that physicists describe time' (2013, 380). The Flemish musicologist Mark Leman (2008) has found physiological correlations; looking at the effects of embodied phenomena, such as walking speed and heart rate on the perception of pulse and tempo. The pulse is identified by Smalley (1997) as the smallest rhythmic structure in tonal music. We have an innate skill to find a musical pulse when listening to a varying rhythm (Honing, 2012). We are particularly attuned to listening for the onset of beats, as, in evolutionary terms,

---

[2] Musical rhythm consists of meter (a beat, either single or compound), rhythmical structure (shorter groups of sequential patterns of emphasised beats that are grouped into a long hierarchically based grouping of groups), tempo (the impression of speed or change of speed), and timing (nuances of when notes are played, slightly 'early' or 'late' or mechanically regular).

prediction is a powerful tool (Huron, 2007), and so the onset, the very start of the beat, garners most attention.

Audio-visual rhythm has some similar effects. As Chion reminds us:

Rhythm is an element of film vocabulary that is neither specifically auditory nor visual…the phenomenon strikes us in some region of the brain connected to the motor functions and it is solely at this level that it is decoded as rhythm (1994: 136).

Audio-visual synchronisation relies on the coincidence of action in the image with an auditory emphasis such as a beat. Chion defines these coincidences as 'sync points', an 'audio-visually salient synchronous meeting of a sound event and a sight event' (1994: 233). Synchronous points are similar to a musical chord in that they vertically divide the audio-visual flow, shaping it and creating phrases. Moreover, each sync point emphasises a point in time and imprints an audio-visual moment more heavily in our memories.

There are different types of synchronisation. The most obvious is at the level of a pulse, an image event coinciding with a short duration audio event. An 'absolute synchronisation point'[3] is most impactful and the most percussive; usually the audio coincidence is the onset of an accented beat. Visual coincidences include: a flash frame or a cut, or the movement of the subject in the frame, especially at the height of the action (for example a punch making contact), or used metaphorically (for example the gun shots exactly on the beat in Edgar Wright's *Baby Driver* (2017)), or movement of the camera (whether the camera is real or virtual). I would argue that we see so much moving image constructed with 'absolute synchronisation points' that we also have statistically learned expectations that are consistently being fulfilled. This fulfilment of expectations does not become boring because we are given very different opportunities of association with the audio-visuals, there are

---

[3] 'Absolute synchronisation point' is my own term; it refers to animation that starts absolutely on the onset of the first note of a musical phrase.

endless nuances in the execution and there are many possible variations. There is a continuum of synchronised points from 'absolute' to 'metaphorical'.

The degree of realism, the stretch of 'en creux' in moments of 'synchresis' also affects our sense of audio-visual synchronicity. If the audio appears naturalistic we only pay attention to the visuals. If the audio creates, in Chion's terms, a gap, and if we can bridge the metaphor, this bridging emphasises the moment. If the gap is too wide the audio-visuals become asynchronous. Asynchronicity emphasises the distinct media within the audio-visual work; it gives a much greater recognition of audio appealing to our auditory senses and images appealing to our visual sense as the two media pull apart. As Honing (2012) argues, to a great extent synchronicity is subjective. We favour synchronicity over asynchronicity; we prefer synchronous works and we often perceive non or randomly synchronised stimuli as synchronous. For example, when turning on music and windscreen wipers in a car and feeling that the music and motion of the wipers coincide, or that raindrops running down a window coincide with randomly chosen music. We are wired for apophenia,[4] wired to see patterns and create connections from unconnected events. As Chion (1994: 211) asserts: 'disorder with no apparent goal is intolerable for human beings. We cannot resist giving it structure and form, a teleology, a shape and direction, even when it itself has none'.[5]

Listeners categorise metres and rhythmic genres from their remembered experiences (Snyder, 2000) and form expectations of both. This applies to periodic temporal structures and changing temporal structures such as a bouncing ball, or speech. Honing states: 'We actually tend to hear rhythm and timing in what one might call "clumps"' (2013: 380). Putting the beat into the hierarchy of a rhythm structure may be (statistically) learned. We favour the rhythms we know the best. As

---

[4] The psychiatrist Klaus Conrad initially coined the term 'apophany' in the 1950s, from the Greek apo [away from] and phaenein [to show] to emphasise that delusion can appear to be revelatory to schizophrenics. Over time the meaning has changed to the propensity for seeing connections between phenomena that are not related.

[5] Psychology is beyond the scope of this paper, but this echoes the Gestalt psychologist Max Wertheimer's assertion that the perception and interpretation of incomplete or contradictory images is always into the simplest form, the 'Law of Pragnanz' (1938: 71–88).

argues: 'It is easier to process, code, or manipulate representations when they are mentally attached to events or objects' (Huron, 2007: 124). 'Event-related binding' (Huron's term) refers to how we unify phenomenal experiences; in vision we bind shape, colour and object recognition, in audio we bind timbre, pitch, loudness and location. Additionally we seek to lighten our cognitive task by tackling the relationship between a small number of elements, discerning neighbouring relationships (this uses less short-term memory than distant relationships), and discerning the amount of change (rather than a meta-level change in the rate of change). All these perceptual tendencies inform our appreciation of visual music.

It is possible to consider the 'purest' strand in the topology of Visual Music to be literally what you see and what you hear being one and the same, a 1:1 mapping; which the visual music artist and historian Jack Ox and the director of the Center for Visual Music, Cindy Keefer (2008) term: 'a direct translation'.

Many artists have sought absolute audio-visual synchronisation. This is in sympathy with our liking for synchronous events and 'event-related binding'. Laszlo Moholy-Nagy claimed that: 'to develop creative possibilities of the sound film the acoustic alphabet of sound writing will have to be mastered; in other words, we must learn to write acoustic sequences on the sound track without having to record real sound' (1947: 277). Seminal visual musicians, such as Norman McLaren, realised Moholy-Nagy's creative vision by producing a visual optical soundtrack in the soundtrack area of celluloid film. He used several means, including creating an optical soundtrack by photographing shapes or by manually painting or scratching individual frames. His *Synchromy* is close to a 1:1 mapping of sound and shape; he used the same shapes to create the soundtrack as to create the visuals. The piece starts by introducing each note-shape singly. He did, however, add colour variation and visual repetition, with the intention of making the visuals more interesting. The piece is engaging at the start but then the visuals become overly predictable and repetitive and the variations weaken the absolute synchronisation without adding interest. When it was made in 1971 it was a technical feat; it required creating each pitch optically and filming them in sequence on to the sound track and a great number of optical passes to achieve the multi-layered visuals. Today's digital

processes offer much greater speed, ease and flexibility. A 1:1 mapping of data can be cold and mechanical (Watkins, 2015). At first one experiences 'pure' visual music with pleasure as the 1:1 mapping fulfills one's prediction of the relationship between music and image, but soon the very predictability of this relationship dulls the pleasure of the experience.

John Whitney did not use 1:1 mappings, but developed 'differential dynamics', i.e. linked nested, or interrelated motion paths, the result of which is that shapes are overlaid, creating harmonic visual patterns via computer algorithms.[6] This was a result of noting that rhythm in music and rhythm in vision are very different:

> Often referred to as the drive of a piece of music, is almost automatically enhanced with metrical or cyclical consistency and repetition. Rock musicians know this-perhaps too well. On the other hand, the most difficult visual quality to compose into a composition, as every abstract filmmaker may know, is the same driving propulsive thrust with a visually rhythmic metrical cycle (1980: 69).

As technology has exponentially increased in power it has allowed composers of visual music such as Bret Battey to create much more complex patterns and more complex links between audio and vision. Composers such as Battey have achieved this by using the same algorithms to create both sound and image. Current technology, coding and processes allow much greater artistic freedom and flexibility than the original analogue, optical methods used by animators and film-makers such as McLaren and, more recently, Guy Sherwin.

Song Serie (see **Figure 2**), are studies of audio-visual synchronisation ranging from the 1:1 mapping of *Variation 6*, inspired by McLaren's *Synchromy*, to *Variation 5* with 'metaphorical synchronicity' (Watkins, 2015).

---

6 In contrast Thomas Wilfred's Lumia and James Turrell's artworks eschew audio-visual synchronisation; they are mute. This increases their sense of being timeless and having open-ended (perhaps vast) scale. Light is the key character and is worked with directly.
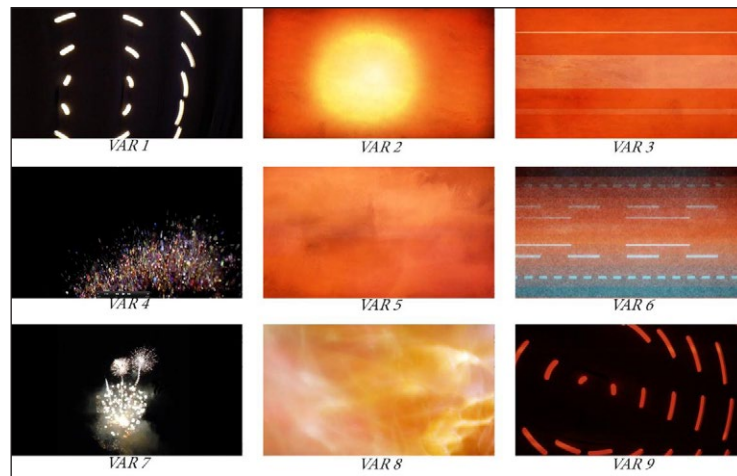
Figure 2: *Song Series Animacy Variations* 1 to 9. Images retrieved from https://vimeo.com/143060494. Copyright 2015 by Julie Watkins.

Shadow Sounds (see **Figure 3**); is a test of creating and composing with 'audio-image units'.[7] Non-verbal vocalisations such as 'ooh' 'ah' 'eeh' and 'pah' are not mapped but visualised using an animator's skills. Thomas Wilfred's *Lumia* is an inspiration for the flowing animations. Each sound and animation is consistently used together, as one 'audio-image unit'. 'It is built from individual vocal gestures that are analogous to notes in tonal music' (Watkins, 2016). The work has 'absolute synchronisation points' but ultimately the amorphous animations needed to be more nuanced to reflect the audio shapes more clearly and so create a more meaningful synthesis between the visual and the audio.

A Continuum of Audio-visual Synchronisation

There are other types of audio-visual synchronisation beyond matching audio beats, for example the widely perceived feeling that higher pitched notes with brighter tones, and lower notes with darker tones match each other better. Similarly, higher in screen correlates with higher notes, lower in screen correlates with lower notes, ascending motion matches ascending musical pitches, and descending motion

---

[7] 'Audio-image unit' is my own term; it refers to instances of audio and animation that always appear together synchronised in the same way, these may be combined in any number of combinations.
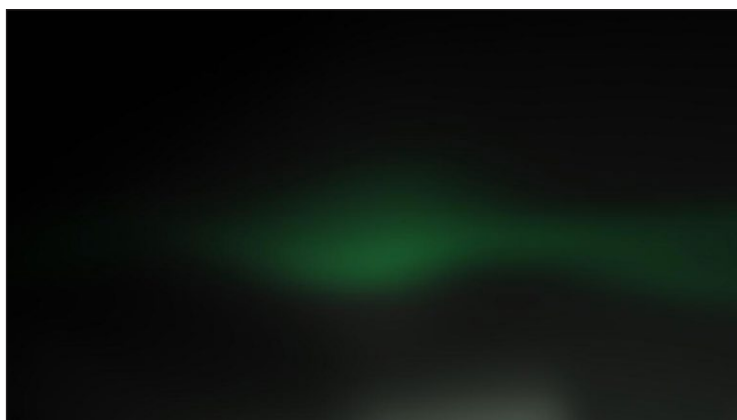
Figure 3: *Shadow Sounds.* Copyright 2015 by Julie Watkins.

matches descending musical pitches. The sonic artist Diego Garro (University & Garro, 2005) scrutinised audio visual mappings in Visual Music, including between bright sounds and bright colours, sound shape and geometrical shape, between how sounds evolve and motion of shapes, and more. Unsurprisingly, he found that redundancy is caused by too great a degree of correspondence. As Bret Battey and the conductor and composer Rajmil Fischman elucidate, there is a continuum of methodologies that range from artistic interpretation to a close mapping of the audio and visual; neither extremity makes a pleasing work (Kaduri, 2016). If the audio and visual elements have no mapping and remain separate the work will fail to be a coherent piece of Visual Music, but if the audio-visual mapping is too close the piece will generate uninteresting perceptual relationships and at best mediocre aesthetical relationships. Battey and Fischman recommend using both clearly recognisable local mapping and (2016: 73) 'higher order intuitive alignment' [that is underpinned by] 'emergent affective properties'. Garro (2005) also recommends using a mix of mapping methodologies.

As visual music is time-based, correlations need to be considered across a range of time from absolute simultaneous synchronisation, to events, to gestures, to phrases, to sections, to the whole work. Seminal early visual music pieces by Walter Ruttmann, Hans Richter and Oskar Fischinger seem to have been composed in this way. These findings would appear to support Rudolf Arnheim's view (2007) that audio-visuals

are more effective when either the visuals or audio are complex, than when both visuals and audio are complex. I would argue that when both imagery and audio are complex the viewer tends to discern some occasions of local synchronisation or some phrasing, but that the patterns quickly become too complex to enjoy and the piece tends towards fragmentation or audio-visual dissonance in the mind of the audience.

My feeling and intuition for combining audio and visual elements comes from working for many years as an animator and (mainly) timing animation and live action to audio. I would argue that audio-visual synchronisation also corresponds to the motion embodied in sounds; most sound has a forward impetus, a vectorisation, which means the sound cannot be reversed without changing. The composer Denis Smalley states that sound-making gestures create, in his term, 'spectromorphological life' giving sounds a strong forward impetus (1997: 111). This is in contrast to artificial sounds, for example white noise, which can be reversed without changing; these sounds lack impetus.

I have used this research in my practice. When creating *Variation 7* for *Song Series* (see **Figure 4**), live action of fireworks was edited to the onset of sung phrases, resulting in 'absolute synchronisation points'. The sung phrases are vectorised and have a strong forward impetus. The fireworks begin as if caused by the beginning of the wordless sung phrases and are edited and speed-ramped to echo the arc of the phrases to form 'gestural animation'.[8] The detailed development of the fireworks is not planned to the music. Perceptually the viewer adds sync to the visual detail, demonstrating our liking for, and ability to create, synchronous events. This combination of the human voice as impetus to 'gestural animation', initiated with 'absolute synchronisation point' is key to my works creating a meaningful synthesis between the visual and the auditory.

In contrast *Reservoir* (see **Figure 5**) has a much more metaphorical audio-visual synchronisation. Many diegetic sounds were combined with impressionistic images.

---

[8] 'Gestural animation' refers to Kimon Nicolaides' notion of 'gesture': 'Gesture has no precise edges, no exact shape, no jelled form. The forms are in the act of changing. Gesture is movement in space. To be able to see gesture, you must be able to feel it in your own body' (1988: 15).

The sound data for *Reservoir* was captured at the same time as the point-of-view footage, whilst circling the reservoir on foot. All the sounds are acousmatic; the viewer sees the effect of the sounds on the camera movement and not the makers of the sounds. The diegetic audio increases the sense of place and time; using these sounds in the order they were captured keeps the original acoustical geography of the circular walk intact. 'I abstracted and re-timed the imagery and created a layered time montage through re-synching visual and audio components' (Watkins, 2016). The audio events play in real-time; they are not tied to the images in the realist manner of synchronous sound but form audio-visual chords with the step-framed images, which synthesise the impressionistic visual and the diegetic audio in a meaningful way.

Audio-visual synchronisation is linked to anticipation. As Chion describes: 'the listener's anticipation of the cadence come to subtend his/her perception. Likewise, a camera movement, a sound rhythm, or a change in an actor's behaviour can put the spectator in a state of anticipation' (1994: 58). There is a tension around anticipation of audio, visual and audio-visual events; we derive pleasure from predicting events. As Huron states, in relation to music: 'Pleasantness is directly correlated with predictability' (2007: 173). But we also like some surprise.

Repeated listening changes the experience; the listener expects to hear the surprises of the first listening repeated. Huron states that 'repeated listening makes the music more predictable. Veridical memories for music hold an extraordinarily

Figure 4: *Song Series* – Variation 7 Image retrieved from https://vimeo.com/143060494. Copyright 2015 by Julie Watkins.
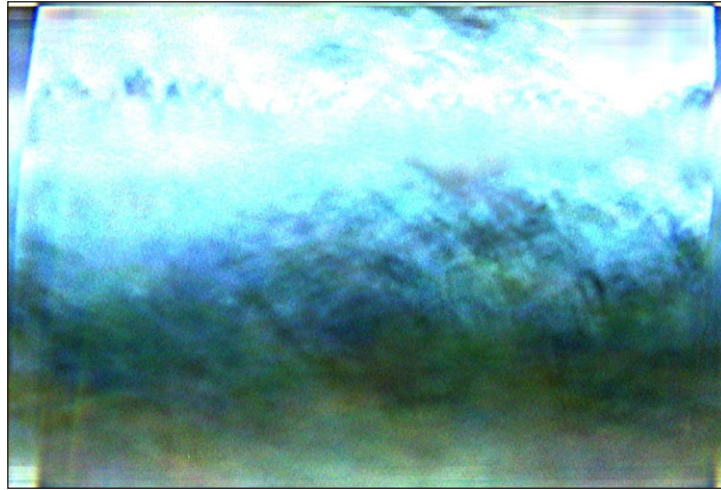
Figure 5: *Reservoir* Image retrieved from https://vimeo.com/181309546. Copyright 2014 by Julie Watkins.

refined level of detail. Listeners are highly sensitive to the slightest changes from familiar renditions' (2007: 241). Chion (1994) argues that, because we are wired for speech, the ear processes faster than the eye, therefore, replaying a rapid image sequence will not allow the viewer to distinguish more. However, this does not take into account an animator's intensive viewing. When I am working as an animator I view sequences that I am working on numerous times, mute and with audio, in real-time and frame-by-frame. I view sequences just looking at the foreground or subject, or concentrating on the background, or just transitions, fragmenting the sequence to see the details ever more clearly. This intensive repeated viewing has a similar effect on me as the repeated listening cited above. I build an extraordinarily in-depth, detailed memory of the audio-visual piece; when it is played I anticipate every moment and if even one frame is altered it jumps out, even though the piece is playing in real-time at 25 frames per second. This ability to re-mix and review is very much a product of our digital technology.

Technology and Data

Creating works using current technology and processes affords opportunities (see above) and poses potential problems and challenges. Technology is ephemeral. It can be superseded and then be unavailable or it can be ubiquitous and clichéd. New

technologies seem to offer new creative potential, the energy of the pioneer is felt, but when they become ubiquitous they quickly become clichéd, for example using data derived from volume to control lighting at an event; making lights brighter as the music is louder in real time. Additionally, as Professor of Digital Creativity at the University of Greenwich Gregory Sporton (2015) elucidates: technology, both hardware and software, can also be a costly trap that limits creativity, as all too often the artist seeks to create something new by pioneering new techniques but ends up illustrating the technology's affordances and constraints rather than creating a new form. Ron Kuivila advises getting 'under' technology, by working directly with physical principles; staying 'over' technology, by working with abstract principles; or by diving 'into' obsolete or banal technologies (Kuivila & Behrman, 1998: 13).

Golan Levin's Opto-Isolator, from 2007, demonstrates how an interactive artwork could stay 'under' the technology. The installation represents a human eye that reacts to being looked at. Eye movements, including blinking, that mimic psychosocial eye-contact behaviours, are triggered based solely on the physical measurements (direction, duration and blink) of the viewer's gaze. The direct correlation between the viewer's gaze and interactive response makes the technology feel transparent. In the early 1960s John Cage's *Variations V* demonstrated how performed electronic music could stay 'over' the technology; how a composer could create the parameters, the framework allowing the performer freedom to explore their own sensibilities in the moment, rather than follow a composer's instructions. This mode requires a refined sensibility but greatly expands the possibilities of composition and adaption of new technologies. Practicing music not preserving music. Computers, via algorithms, are consummate preservers of presets and samples, and of transposing data into sound and sound into image. The sensibility of the composer is either obviated or, as if in aspic, coded into an algorithm.

Another way of aiding engagement is to create a piece of music that is also an instrument. Laurie Spiegel's computer program *Music Mouse* (1985) was simultaneously a piece of music and an instrument. Eno + Chilvers' *Bloom* app (2008) is advertised as being a combination of an instrument, composition and artwork. The parameters for the visual music instrument were defined through composition:

by creating a composition that is also an instrument and additionally provides parameters or 'rules' for visual music composition and so stays 'over' technology. Bjork's *Biophilia* (2011) demonstrates the expanding possibilities for multiple outputs providing many levels of engagement. *Biophilia* is a multi-disciplinary, cross-platform release, encompassing an album, live shows, website, an iPad application for each track, and a film documenting the project.

Other technological influences on the work include the Mellotron (1963); a pre-synthesizer instrument used for the beginning of *Strawberry Fields* (The Beatles, 1967). It had a keyboard that played tape loops; one key played one sound. The pitch of the sound could be altered, by varying the speed at which it was played. Additionally there was control over tone and volume. The concept of linking one motion, the pressing of one key, to one sound, within a process that allows both pitch and volume to be altered fed into the design and process for Watkins' *Sky* (2017) (see **Figure 8**).

Data was used to humanise abstract animation: data from images of landscape and data of human traces. Given that the material is digital video, which lacks the tangible physicality of film, this humanising data is especially important. As Guy Sherwin, the pre-eminent British film artist, points out, when talking about the materiality and processes of celluloid film: 'For an artist materials matter, they become important' (Lumière, 2011). *Reservoir* (see above) explores the materiality of digital video, using an old format and resolution of digital image, layering it and colourising it until the image almost disintegrates.

The gathered data becomes the artistic material. To create *Horizon* (2014) (see **Figure 6**), inspired by the artist William Turner, footage and stills of a seascape at sunset, audio of the sea were captured and brushstrokes were digitally collaged to create an abstracted seascape (Watkins 2016).

Inspired by visual musicians such as Jordan Belson, the aim is for the works to be experiential. The work is as much about creating a shared experience, as it is about self-expression. To this end the gathered data was formed into 'abstracted animation', to lose any association of a specific real landscape and evoke a more meditative experience in which motion is the most important visual element. This

is supported by the canon of visual music. The great Lumia artist Thomas Wilfred (1947) identified form and motion as the two most important elements in his work. 'Abstracted animation', also embodies Malcolm Le Grice's philosophical concept of the resulting work allowing elements to be: '"raw material" available for "retrieval" in ways which construct a new experiential model of the world' (2009: 317). Allowing the audience to participate in this way, to use 'abstracted animation' to form their own associations, their own experiences is at the heart of my practice. This supports the concept of the work as an immersive experience, similar to the psychologist Stephen Kaplan's description of 'soft fascination' in nature (1992: 139):

> Many of the fascinations afforded by the natural setting might be called
> "soft fascination". Clouds, sunsets, snow patterns, the motion of the leaves
> in a breeze – these readily hold the attention, but in an undramatic fashion.
> Attending to these patterns is effortless, and they leave ample opportunity
> for thinking about other things.

A fruitful way forward in the face of the challenges and opportunities of technology and data is to combine Kuivila's staying 'over' technology, with the use of data as artistic material that can be found in Le Grice's term 'retrieved' to form new experiences, moments of in Kaplan's term 'soft fascination'.
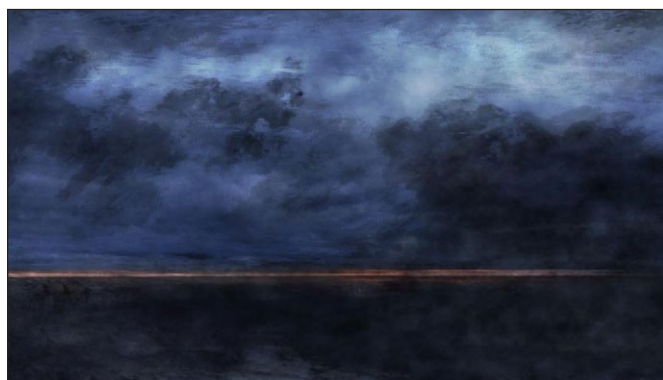
Figure 6: *Horizon* Image retrieved from https://vimeo.com/188289141. Copyright 2014 by Julie Watkins.

Human traces

In this age of burgeoning artificial intelligence in the arts, human experience and input seems ever more crucial. My ideal visual music starts with the human voice. Watkins (2016): 'The special qualities of the human voice as an instrument has been recognised ever since Darwin brought attention to the primal nature of voice both conveying and affecting emotions'. My visual music is about communication, not about language and is designed around non-verbal communication. The importance of non-verbal or emotional signals is well documented and there is much research around the area of making this visual, for example the *Moodies* app (2013) from Beyond Verbal that claims to analyse raw human vocal intonations in real time and visually indicate the speaker's underlying emotion with a cartoon face. Additionally we have a social awareness and we recognise and sometimes identify with the emotions of the performer (Thorn, 2016). Sharing a liking for a song or style of music bonds social groups and reinforces the emotional communication at a wider level beyond the individual listener.

For *Ambience* (see **Figure 7**), audio was used that embodies emotions in the form of traditional songs, sung on vowels only: 'to underpin the abstract movement of light and colour with human motivation and emotion' (Watkins 2016). Particle systems were used to gain more detailed control of the flowing shapes. Nuance was added by directly translating some human traces, for example turning tracking data from the singer's head movement into the movement of a particle-emitter and the data from the singer's mouth movement into circles of colour.

Sky (2017) combines new elements with processes from *Shadow Sounds* (audio-visual units based with non-verbal vocalisations) and *Ambience* (particle flows) and *Horizon* (gathering data to create an 'abstracted animation). The process was initiated by creating a library of unique animated shapes driven by vowels and consonants. 12 vowels that moved from the front to the back of the mouth, including ih uh aw oo, were used. The consonants B D G V L Z M were chosen to give a distinct range of sounds. These 12 vowels each had 15 variations: the vowel is sung by itself and 7 consonants were placed both before and after the vowel. Clearly vowels and consonants cannot be cut together but must be sung individually, i.e. Z and ih do not sound the same as Zih.

Informed by my research (see above) the onset of the sound and the visuals are absolutely synchronised and the development has a looser correlation that is predicated on my choices as an animator. Using parameters such as sound shapes, motion paths in 3-dimensional space, velocity, density, textures, particle shapes, fine lines and motion blur, a library of 'audio-image units' was created. The animations are inspired by the impetus and 'spectromorphological life' of the sounds. The singer, Martin Nelson, asked if the shapes were programmed from the sound data, as they seemed to fit so well. It was pleasing that using the sensibilities of an animator resulted in animations that felt so right to him. Given that rhythm is vital to create meaningful audio-visual synchronisation, the rhythmic nature of the 'audio-image units' is emphasised, chiefly by using sounds with little pitch variation. The frequencies are contained within about two pitches, allowing a rich, reverberant, resonant human dissonance. This is distinct from the many data-driven transpositions of pitch/frequency and or volume-to-image that are common within visual music.

Isolating each 'audio-image unit' and then blending them together into a visual music composition allows great flexibility and the possibility for other composers to use these animations as an instrument. The background of *Sky* is created from footage of clouds treated as data: re-timed, layered, colourised and revealed through particle animations. The *Sky* is under three minutes long. It is the first part of 24 parts that will make up an hour-long piece. Like Richter's *Rhythmus 21* and later McLaren's
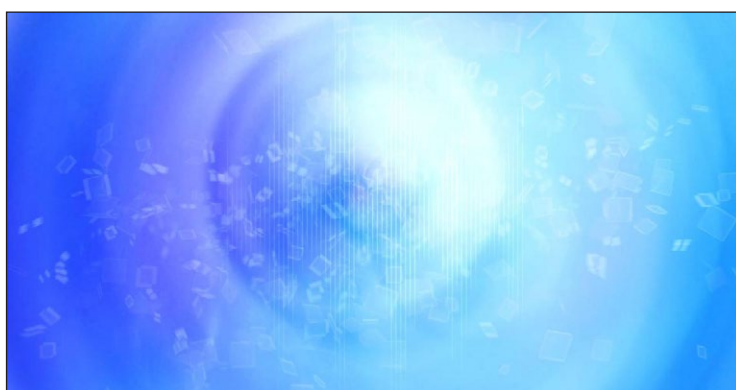
Figure 7: *Ambience* Image retrieved from https://vimeo.com/181305587. Copyright 2016 by Julie Watkins.

Synchromy it is purposefully simple at the beginning in order to aid the viewers' understanding of the relationship between visuals and sung sounds.

Conclusion

This paper delineates an evolving visual music practice, underpinned by an animator's fervour, current research into audio-visual perception and the canon of visual music. In response to work that has a more mechanical mapping the aim is to use current technology to create new visual music that affords 'soft fascination'; works of 'abstracted animation' that are suffused with human presence and emotion. This work starts with the audio, audio that has the emotion of the human voice in the sung non-verbal sounds. Visual music depends upon a meaningful synthesis of visuals with audio. These visuals depend on an animator's sensibility; they are not mechanistically or algorithmically produced from the audio. 'Audio-image units' are created that are initiated by an 'absolute synchronisation point' and develop to embody a diversity of audio-visual synchronisation. A nuanced use of audio-visual rhythm is developing through using these 'audio-image units' to simultaneously compose light, form and movement and sound into longer phrases, sections and pieces. Thus the two very distinct perceptions of light and sound have been synthesised in a meaningful way, without reducing them to limited, cold, mechanical terms. I hope this approach will be useful to others practicing in this area.

Figure 8: *Sky* Image retrieved from https://vimeo.com/237235704. Copyright 2017 by Julie Watkins.

Acknowledgements

Competing Interests

Author Information

Julie Watkins is a senior lecturer in Film and Television and affiliated to the University of Greenwich. She worked as lead creative in prestigious Post-Production facilities in Soho and Manhattan. She designed concepts, led Technical Direction, Animation, Motion Graphic and Visual Effects Teams, for Commercials, Broadcast Graphics and Films. She taught at New York University. She joined the University of Greenwich in 2006, initiated a Film and Television degree and partnership with the BBC. She has MA (distinction) in Graphic Design from University of the Arts London. She has presented papers and shown work at DRHA 2014, 2015, 2016 and 2017 and Sound and Image 2015, 2016 and 2017 is now completing a PhD.

References

Arnheim, R 2007 *Film as art* (2. Dr). Berkeley, Calif.: Univ. of California Press.

Brougher, K, Mattis, O, Museum of Contemporary Art (Los Angeles, Calif.) and **Hirshhorn Museum and Sculpture Garden** (Eds.) 2005 *Visual music: synaesthesia in art and music since 1900.* [London]: Washington, D.C.: Los Angeles: Thames & Hudson; Hirshhorn Museum; Museum of Contemporary Art.

Chion, M, Gorbman, C and **Murch, W** 1994 *Audio-vision: sound on screen.* New York: Columbia University Press.

Collopy, F 2000 Color, Form, and Motion: Dimensions of a Musical Art of Light. *Leonardo*, 33(5): 355–360. DOI: https://doi.org/10.1162/002409400552829

Cook, N 2000 *Analysing musical multimedia.* Oxford: Oxford University Press.

Gallese, V 2016 The Multimodal Nature of Visual Perception: Facts and Speculations. *Gestalt Theory*, 38(2/3): 127–140.

Grewe, O, Nagel, F, Kopiez, R and **Altenmüller, E** 2007 Listening To Music As A Re-Creative Process: Physiological, Psychological, And Psychoacoustical

Correlates Of Chills And Strong Emotions. *Music Perception: An Interdisciplinary Journal*, 24(3): 297–314. DOI: https://doi.org/10.1525/mp.2007.24.3.297

Honing, H 2012 Without it no music: beat induction as a fundamental musical trait: Honing. *Annals of the New York Academy of Sciences*, 1252(1): 85–91. DOI: https://doi.org/10.1111/j.1749-6632.2011.06402.x

Honing, H 2013 Structure and Interpretation of Rhythm in Music. In: *The Psychology of Music*, 369–404. Elsevier. DOI: https://doi.org/10.1016/B978-0-12-381460-9.00009-2

Kaduri, Y (Ed.) 2016 *The Oxford handbook of sound and image in Western art.* New York: Oxford University Press.

Kaplan, S 1992 The Restorative Environment: Nature and Human Experience. In: Relf, D (Ed.), *The Role of horticulture in human well-being and social development: a national symposium 19–21 April 1990, Arlington, Virginia*, 134–142. Portland, Or: Timber Press. Retrieved from: http://www.mairibudreau.com/uploads/1/2/2/4/12240077/nature_and_the_human_experience_the_restorative_environmnetkaplan-1992.pdf.

Klein, A B 1930 *Colour-Music: The Art of Light* (Second Edition). London: The Technical Press Ltd.

Kuivila, R and **Behrman, D** 1998 Composing with Shifting Sand: A Conversation between Ron Kuivila and David Behrman on Electronic Music and the Ephemerality of Technology. *Leonardo Music Journal*, 8: 13. DOI: https://doi.org/10.2307/1513392

Le Grice, M 2009 *Experimental cinema in the digital age* (1. publ., reprinted). London: bfi Publ.

Leman, M 2008 *Embodied music cognition and mediation technology.* Cambridge, Mass: MIT Press.

Lumière, R 2011 *Encuentro Con Guy Sherwin: Short Film Series.* Retrieved from: https://vimeo.com/28717998.

Lund, C and **Lund, H** (Eds.) 2009 *Audio visual: on visual music and related media.* Stuttgart: Arnoldsche Art Publishers.

Moholy-Nagy, L 1947 *Problems of the Modern Film* (Third). Chicago: Paul Theobald.

Ox, J and **Keefer, C** 2008 On Curating Recent Digital Abstract Visual Music. *Abstract Visual Music*, (2006–2008).

Smalley, D 1997 Spectromorphology: explaining sound-shapes. *Organised Sound*, 2(2): 107–126. DOI: https://doi.org/10.1017/S1355771897009059

Snyder, B 2000 *Music and memory: an introduction.* Cambridge, Mass: MIT Press.

Sporton, G 2015 *Digital creativity: something from nothing.* Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan. DOI: https://doi.org/10.1057/9781137486417

The Beatles 1967 *Strawberry Fields Forever.*

Thorn, T 2016 *Naked at the Albert Hall: the inside story of singing.*

University, K and **Garro, D** 2005 Research, Keele University. In: *http://www.ems-network.org/spip.php?rubrique34*. Montreal: Keele University. Retrieved from: http://www.ems-network.org/spip.php?article169.

Watkins, J 2015 Animacy, Motion, Emotion and Empathy in Visual Music: Enhancing appreciation of abstracted animation through wordless song. *Body, Space & Technology Journal, 16.* Retrieved from: http://people.brunel.ac.uk/bst/vol16/cover.html.

Watkins, J 2016 An Investigation into Composing Visual Music Today. *Body, Space & Technology Journal*, 17. Retrieved from: http://people.brunel.ac.uk/bst/vol17/cover.html.

Wertheimer, M 1938 *A Source Book of Gestalt Psychology.* London: Routledge & Kegan Paul Ltd.

Whitney, J 1980 *Digital harmony: on the complementarity of music and visual art.* Peterborough, N.H: Byte Books.

Wilfred, T 1947 Light and The Artist. *Journal of Aesthetics and Art Criticism*, 5(4): 247–255. DOI: https://doi.org/10.2307/426131