Soma: live musical performance where congruent visual, auditory, and proprioceptive stimuli fuse to form a combined aesthetic narrative

Ilias Bergstrom[1], R. Beau Lotto[2]

[1]EventLAB, Universitat de Barcelona, Campus de Mundet - Edifici Teatre, Passeig de la Vall d'Hebron 171, 08035 Barcelona, Spain. Email: iliasbergstrom@ub.edu, onar3d@gmail.com

[2]Lottolab Studio, University College London, 11-43 Bath Street, University College London, London, EC1V 9EL. Email: lotto@ucl.ac.uk

Key-words: visual music, audiovisual composition, lumia, soma, live performance, enactive, embodied knowledge, creative coding, programming as art, new media, parameter mapping.

Abstract

Artists and scientists have long had an interest in the relationship between music and visual art, leading up to the present date art-form of correlated animation and music, 'visual music'. Current live performance tools and paradigms for visual music however are subject to several limitations. The work detailed addresses these through a transdisciplinary integration of findings from several research areas, detailing the resulting ideas and their implementation in three interconnected software applications. This culminates in the art form of Soma, where correlated auditory, visual and proprioceptive stimuli form a combined narrative.

1. Introduction

Artists and scientists have long had an interest in the relationship between music and visual art (visuals). Today, many create correlated animation and music, called 'visual music'. Established tools and paradigms for live performance however, have several limitations:

- Virtually no user interface exists, with expressivity akin to that in live musical performance where gestural controllers are used to allow for the employing of advanced enactive knowledge.
- Mappings between music and visuals are reduced to few parameters, typically static associations to the music's beat and amplitude, disallowing close audiovisual congruence, tension and release, and suspended expectation in narratives.
- Preparing or improvising performances is complicated, often requiring software development.
- Collaborative performance is difficult, in comparison to how easily it is achieved in live musical performance, due to technical limitations.

We transdisciplinarily address these, detailing the resulting ideas and their implementation in the novel Trinity system:

Musical instruments are our primary control data source, encoding all musical gestures of each performer. Musician's advanced embodied knowledge of their instruments, allows increased expressivity, the full control data bandwidth allows high mapping complexity, while musicians' collaborative performance familiarity may translate to visual music performance.

We conceived of Mutable Mapping, gradually creating, destroying and altering mappings, which facilitates a narrative in mapping during performance.

We formulated the art form of Soma, where correlated auditory, visual and proprioceptive stimuli form a combined narrative, building on knowledge that performers and audiences are more engaged in performance requiring advanced motor knowledge, and when congruent percepts across modalities coincide.

We simplified preparing and improvising, through re-adapting the Processing programming language for artists to behave as a plug-in API, thus encapsulating complexity in modules, which may be dynamically layered during performance.

Subsequent sections expand on the argument behind the identified existing limitations, and our proposal for how these are overcome.

2. Background

Ideas about the correspondence of visual and musical art have been formulated throughout history by people including Aristotle and Isaac Newton. Machines have been constructed to explore the relationship, the first known mention being the Louis-Bertrand Castel's *Clavecin oculaire* (1734), which implemented a direct note to color correspondence. Many have followed, either accompanying music with color, or providing a form of visual music - named '*Lumia*', a term coined by pioneer Thomas Wilfred, developer of the *Clavilux* color-organ (1922). Wilfred rejected absolute correspondence between sound and image, and concentrated on his art form of controlled color, form and motion, meant to stand without musical accompaniment.

Music's immediacy in communicating emotion has been envied by visual artists, most notably Wassily Kandinsky, who tried to recreate it in painting [1]. A great inspiration of his was the composer Scriabin, likely the first to write a musical composition intended to include also an evolving visual element.

Music is unique among arts in being predominantly non-representational, and self-referential. The primary music universals are: a sequence of *pitches* forming a *melody*, the *rhythm* being the duration of notes and their grouping, *tempo* being overall speed, while *timbre,* defying exact definition, loosely is what differentiates instruments playing the same score [2]. Cross' definition [3]: *"music embodies, entrains and transposably intentionalizes time in sound and action",* encompasses actions following a musical narrative, such as dance, and physical performance involving musical gestures, a central notion to present work. Music is time-based narrative affective communication, which allows for emotive content different to that represented through language or other art-forms. It thus cannot effectively be conveyed in any other medium than itself.

Pioneer John Whitney, has extensively documented his thoughts on what a harmony of visual music may entail [4]. Evans [5] identifies the fundamentals of visual musical narrative: Just as musicians and audiences tacitly understand, from upbringing and culture, harmonious and inharmonious music, and the usual sequences of musical events, so too in visual culture there are heritages of imagery: *“it is possible to resolve visual dissonance to consonance, and so move a viewer through time in a way similar to tonal harmony in music”*. He defines visual music as: “*Time-based visual imagery that establishes a temporal architecture in a way similar to absolute music. It is typically non-narrative and non-representational (although it need not be either). Visual music can be accompanied by sound but can also be silent”.* Grierson [6] adds how narrative can be constructed also vertically, between congruent visual and auditory musical events. The fusion of musical and visual art, has led to *Audiovisual Composition*, an “*artistic form which takes as its starting point the cognitive actuality of multisensory audiovisual experience*”: neither is simply accompanying the other, instead they form an inseparable whole.

In modern times, analogue video synthesizers, laser shows and computer graphics have accompanied music live and in real-time. Most frequently, there is a direct coupling between music and image, with musical input being processed to partly or wholly control live procedural computer graphics. There is a vast body of work in which performances are created by artist/engineers, using tools that they have largely developed themselves specifically to facilitate each performance at hand. This article however, concentrates on that art which does not demand that the artist is also an engineer, inventing his tools anew for each performance. It is also specifically this that we refer to as “current practice” in the text.

Under this definition, the first commercial equipment available was analog video synthesizers, primarily modular in construction. Their cost and complexity however meant that few artists could access them. Notable users include Steina and Woody Vasulka (www.vasulka.org). Using the Rutt/Etra “Scan Processor”, they produced the earliest electronic precursor to our work that we know of, with their piece “Violin Power” (1970-1978). Today, the predominant live visuals performance practice, employing readily available tools, is VJing. VJ’s (Video/Visuals Jockeys) perform by mixing pre-recorded video clips together, while manipulating their playback, and applying real-time video effects. Modern VJ software also facilitates including real-time *Procedural* computer graphics, however such features see more limited use as they are complicated to invoke in relation to performing using clips of video. The term procedural refers to graphics that are generated algorithmically, rather than having been created manually or been otherwise sampled [7]. A common practice in VJing is to control some parameters of video playback, effects and procedural graphics with values derived from stereo audio. The connection however is limited, because the data encompasses only overall beat events, amplitude and tempo.

The process of making connections between incoming control data, and control parameters of visual/audio synthesizers, is referred to as *Mapping* [8]. Almost exclusively in established performance practice, the musical control data is derived from a stereo mix-down, processed to derive the music's overall tempo and amplitude, as well as detect individual beat events. This data is then almost exclusively statically associated to control parameters of the controlled visuals.

3. Limitations with current practice

I. The mappings between music and visuals are constrained because they are limited in complexity, and remain static over time, thus limiting the correlation between visuals and music, when in fact there is much evidence that increased correlation results in a stronger experience.

It can easily be argued, that increasing the detail of the correlation between visual and auditory musical events, to a higher level of synchronization than that achievable with current practice, will further encourage the unified experience of music and image. It is known that the human perceptual system is apt at detecting correlated stimuli across modalities, and fusing these into a single percept before their interpretation [9], [10]. Michel Chion et al. [11] argue that synchronized music and/or sound provides *"added value"* to a visual narrative, defining *Synchresis* as: *"(...) the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time"*. Experiments show that there is significant positive correlation between how closely discreet auditory and visual events are synchronized, and the perceived effectiveness of the combined audiovisual stimulus [12]. *"(...) the manner in which salient moments in the auditory and visual domains are aligned results in a significantly different perceptual response to the resulting composite"* [13].

Additionally, the benefits of allowing the mappings to be varied during the course of the performance are clear: it allows the use of suspended expectations, thus also facilitating tension and release in the aesthetic narrative developed through the mapping. Both are crucial aspects of aesthetic and visual music narratives [4], [14], and difficult to employ in the synchronization between music and visuals, if mappings remain fixed throughout the performance, as is the case in related current practice.

II. Virtually no user interface exists, that allows controlling the performance of visual music in real-time, with a level of expressivity comparable to that attainable in live musical performance in which instruments are used which allow for the employing of advanced enactive knowledge.

Performances are currently controlled using interfaces on computers and/or external hardware controllers (with knobs, sliders and buttons, etc., here referred to as *non-musical controllers*). The control is akin to that in musical *conducting gesture performance*, in that signal sources are influenced indirectly: like a conductor directs an orchestra, or an audio mixing engineer manipulates multi-track audio. This contrasts to direct instrumental performance, using controllers through which performer's musical gestures immediately translate into sound, as with traditional instruments. The term *Musical Gestures* refers to the actions carried out by a musician during performance [8]. When playing musical instruments, musicians exhibit the use of advanced *Enactive Knowledge*: knowledge that can only be acquired and manifested through action [16]. Examples include dance, painting, sports, and performing music. One can derive that non-musical controllers' reduced control complexity, does not allow performers to develop as advanced a virtuosity as musical instruments, thus affording reduced expressivity, in comparison to that musicians have with their instruments [8], [15]. While in music, conducting gesture instruments were created in a world where instruments facilitating advanced enactive knowledge were already established, in live visual music, that has not been the case, and instead the currently greatest level of expressivity attainable, is that of the conducting gesture performance interfaces. While it is not here claimed that conducting gesture performance is lesser, it is certainly different, and performance which also allows for advanced enactive knowledge is certainly also desirable.

Not employing advanced enactive knowledge also influences the audiences' experience: *"(...) the perceiver watching, listening to and experiencing another's motor performance, simulates the actions of the performance within the range of their own motor capabilities"* [17]. When performers are not using controllers that demand advanced enactive knowledge, audiences' total experience is consequently different.

III. The process of preparing or improvising live procedural visual music performances is overly complicated, in comparison to the same for live musical performance.

With current practice, artists preparing for a live procedural visual music performance necessarily engage in creating the visual instruments themselves, through software engineering at some, usually high, level of complexity. Such programs are predominantly limited in their usability, as they necessarily embody the aesthetic goals of their creator [18]. Software created by one performer is thus prevented from being used by other performers to achieve their individual aesthetic goals, without extensive modification, necessarily through software engineering.

IV. Collaborative performance, akin to the extent that this is possible in actual music performance, is difficult in live visual music performance, due to technical limitations.

With current practice, live audiovisual performance allows limited scalability in the number of performers involved. There are significant benefits to draw from collaborative performance that allows for *Mutual Engagement* [19]*: "The point at which people spark together, lose themselves in the joint action, and arrive together at a point of co-action 'where you are when you don't know where you are'* [20]*".*

4. Proposed ideas, and their implementation

A system is presented, created with the purpose of embodying our ideas towards greatly improving on and extending established practice. We increase mapping complexity, allow mappings to vary during the performance, allow greater expressivity, lower the technical barrier to creativity, allow for improvisation, and for collaborative performance.

A. Increasing input data richness

We propose that musical instruments are used as the primary source of control data. In live musical performance, much richer data is generated by the instruments than simply stereo audio. Separate audio signals emanate from each instrument, allowing the tracking of the amplitude, tempo, and detected beat events of each, rather than just of all instruments lumped together, as is used in current practice. Most importantly we use this data alongside exhaustive digital procedural control data produced separately by each suitably equipped instrument, be it either MIDI data [8] or OSC [21]. These data sets are much richer, including the onset, offset, pitch and amplitude of individual notes of each instrument, alongside a plethora of additional parameters, depending on the type of instrument in question.

As no existing software facilitated the above, we developed the Live Input Processor (LIP) (Figure 1). Briefly its main capabilities are: receiving incoming MIDI data from musical instruments, translating it into OSC, and retransmitting it in real-time; Performing low-latency pitch tracking, amplitude tracking, and beat detection on incoming audio signals, and transmitting the derived data over OSC, again in real-time. LIP serves as a pre-processor of incoming data, with no role in determining how the data is then used, that role is instead dedicated to the Mediator software, detailed in the subsequent section D.

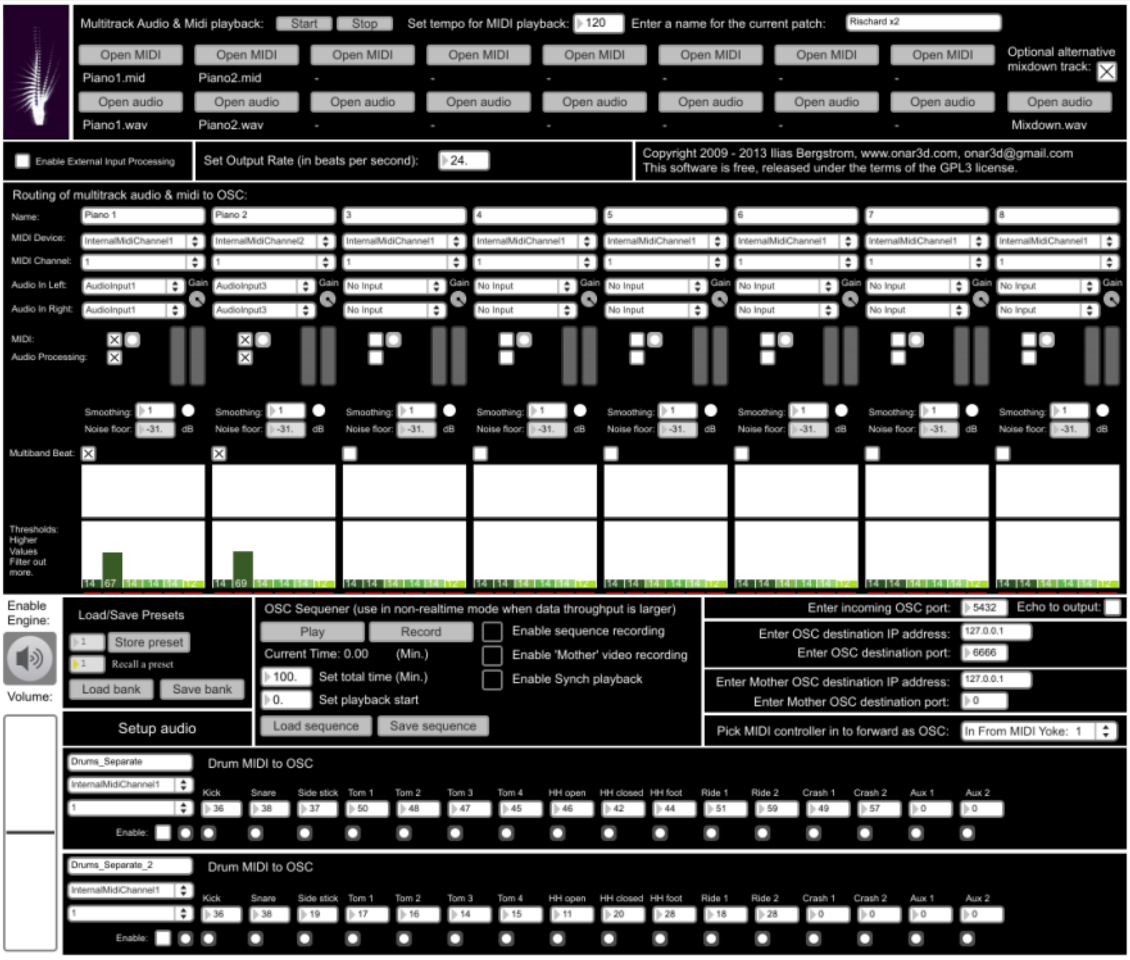

Figure 1: The Live Input Processor software

B. Increased expressivity

Through using much richer data, the musical gestures of performers are more accurately encoded. Musicians are, from this increase in control data bandwidth, expected to better transmit their intent, and thus to a greater extent usefully take advantage of the advanced embodied knowledge they have of their instruments, achieving an increase in expressivity.

As we saw in section 3, the level of control complexity that a new digital musical instrument provides is considered a precondition to its expressivity [15], although high complexity does not guarantee expressivity. Established musical instruments however can be assumed to have reached maturity, their high control complexity thus guaranteeing correspondingly high expressivity.

When live visuals are controlled using signals from live musical performance, the musician(s) performing are controlling the visuals with their instruments. We hypothesize that taking advantage of the full complexity of data generated by musical instruments, gives rise to a considerable increase in the expressivity of said instruments also for performing live visuals. While not guaranteed, it is made possible, depending on the suitability of the mappings made to the visual synthesizers at hand.

The subjective and highly elusive experience is desired, of the musicians to some extent controlling the visuals, as opposed to the visuals reacting to their playing.

C. Simplifying preparation for performance

In existing practice artists are often required to engage in software development to prepare their performance, due to the lack of suitable software. We addressed this by developing a new application, dubbed *Mother* [22]. We retained the paradigm of mixing multiple layers of moving graphics; however, these are not pre-rendered videos, but the output of real-time *Visual Synthesizers* (*synths*) that run in parallel within the main host application. Each is a program rendering a particular visual effect, whose control parameters are all accessible during a performance, allowing its appearance to be controllably altered over time.

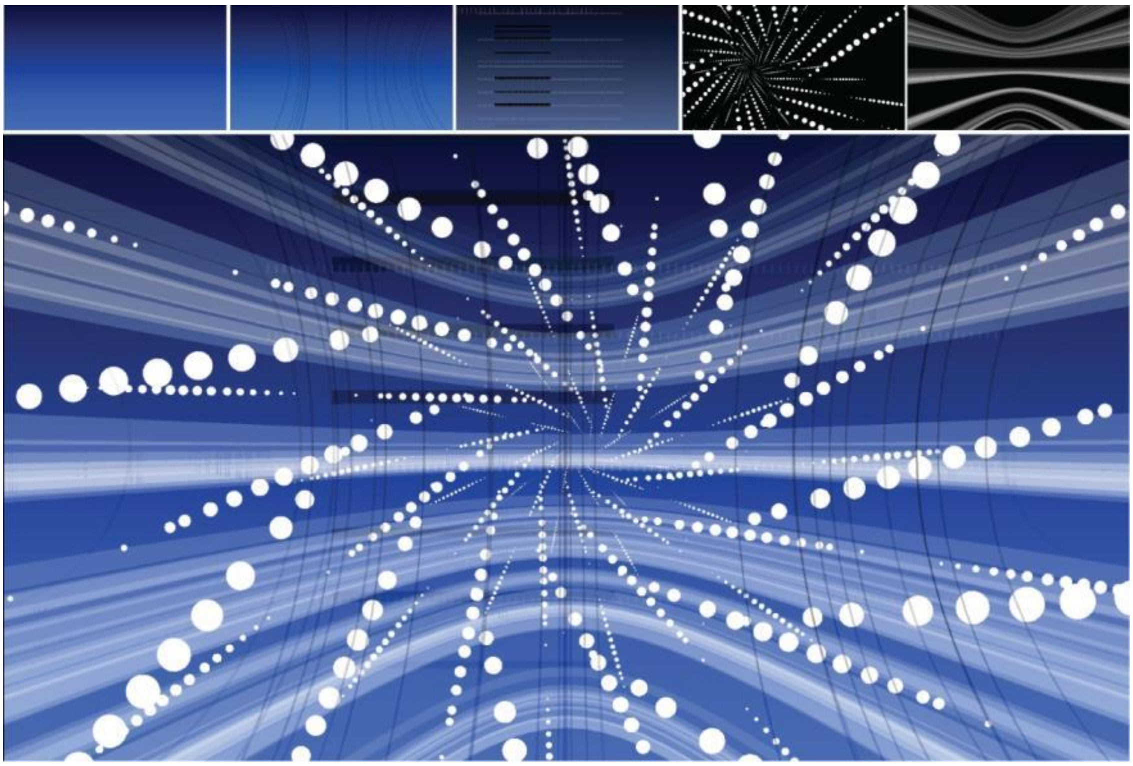

Figure 2: Illustration of how several processing sketches (each a small image at the top) are within Mother layered to produce a single complex output (all images and programs involved created by I. Bergstrom).

Synths can either be created anew as Processing *"sketches"* [23], or come from a pre-existing library. Mother further allows artists to forward digital control data to synths so as to finely control what each displays, and to dynamically rearrange, add and remove synths during a performance. The parameter space of control input for the visuals therefore varies constantly, as synths are added and removed. See Figure 2 for a real world example of the layering of several synths, and Figure 3 for the code drawing the first of these layers.

This simplifies preparation for performance on two levels: firstly, artists may employ a combination of pre-existing synths from other artists. Through selecting and combining these, they achieve a higher level of artistic control over the visual content than in current practice, without engaging in any programming. Secondly, if they do not have such pre-existing synths available to them, or if artists choose to program, to achieve greater control still, we facilitate using the Processing programming language in a modular manner, thus greatly reducing the effort necessary. Note that the choice of the Processing language also signifies a considerable reduction in complexity, given it was created with artists as its intended end users. Processing already has a very large user base of such artists, who have shared thousands of programs generating visuals as free examples. These can all be converted into synths for Mother with little effort, by anyone familiar with Processing.

What is conceptually novel about the above approach is the re-adaptation a programming language intended for artists to instead behave as a plug-in API, thus further increasing the language's usefulness to a context for which it was originally not intended; a specific application of a conduct we have christened code-bending, and detail in a dedicated article [24].

Interested readers are encouraged to download Processing (www.processing.org/download), and, following the "Sketch->Import Library->Add Library" option, select and install Mother. Mother comes with examples, and a pdf with documentation on its use.

```
public Foetus f; // This field needs to be public for Mother to access it.

FoetusParameter m_TopR, m_TopG, m_TopB;
FoetusParameter m_BotR, m_BotG, m_BotB;

void setup(){
 // When run as a synth, "setup()" is never called!
 // put the necessary initialization code in "initializeFoetus()".
 // The Processing initialization call "size(x, y, nnn)" is called by Mother
 // and so should be left out from initializeFoetus().
 // For the synth to work as a processing sketch within the PDE,
 // call initializeFoetus() from within setup().
 size(1280, 720, OPENGL);
 frameRate(30);
 initializeFoetus();
}

void initializeFoetus() {
 noStroke( );
 f = newFoetus(this);// Instantiate foetus object here

 // The below maps incoming OSC messages to values.
 m_TopR = new FoetusParameter(f, 0.0, "/TopRed   ",    "f" );
 m_TopG = new FoetusParameter(f, 1.0, "/TopGreen",     "f" );
 m_TopB = new FoetusParameter(f, 0.0, "/TopBlue" ,     "f" );
 m_BotR =  new FoetusParameter(f, 0.0, "/BotRed   ",   "f" );
 m_BotG =  new FoetusParameter(f, 0.0, "/BotGreen",    "f" );
 m_BotB =  new FoetusParameter(f, 0.0, "/BotBlue" ,    "f" );
}

void draw() {
 f.startDrawing(); // Notify foetus that this synth starts drawing.

 clear(); // Clear this synth's canvas (not Mother's canvas).

 // Draw a gradient, using standard Processing commands.
 pushMatrix();
 beginShape(QUADS);
 fill (m_TopR.getValue(), m_TopG.getValue(), m_TopB.getValue());
 vertex (0, 0);
 vertex(width, 0);
 fill(m_BotR.getValue(), m_BotG.getValue(), m_BotB.getValue());
 vertex(width, height);
 vertex(0, height );
 endShape();
 popMatrix();

 f.endDrawing(); // Notify foetus that this synth stopped drawing.
}
```

Figure 3: Simple real-world visual synth example for drawing a gradient. It can receive OSC parameters setting the top and bottom color.

D. Increasing achievable mapping complexity and variability

In current practice, given the limited input control data, defining mapping to control parameters of live visuals is simple, and has virtually always been hardcoded. With the approach presented here however it is neither feasible nor desirable to hardcode mappings, because of the significantly greater amount of available input and output control data.

To address these issues we have devised the artistic conduct of *Mutable Mapping*: gradually creating, destroying and altering mappings between the two parameter spaces of input and output control data, before and during the course of a performance. Furthermore a software application has been developed for this conduct, *Mediator* [25]. The main screen for Mutable Mapping is presented in Figure 4.

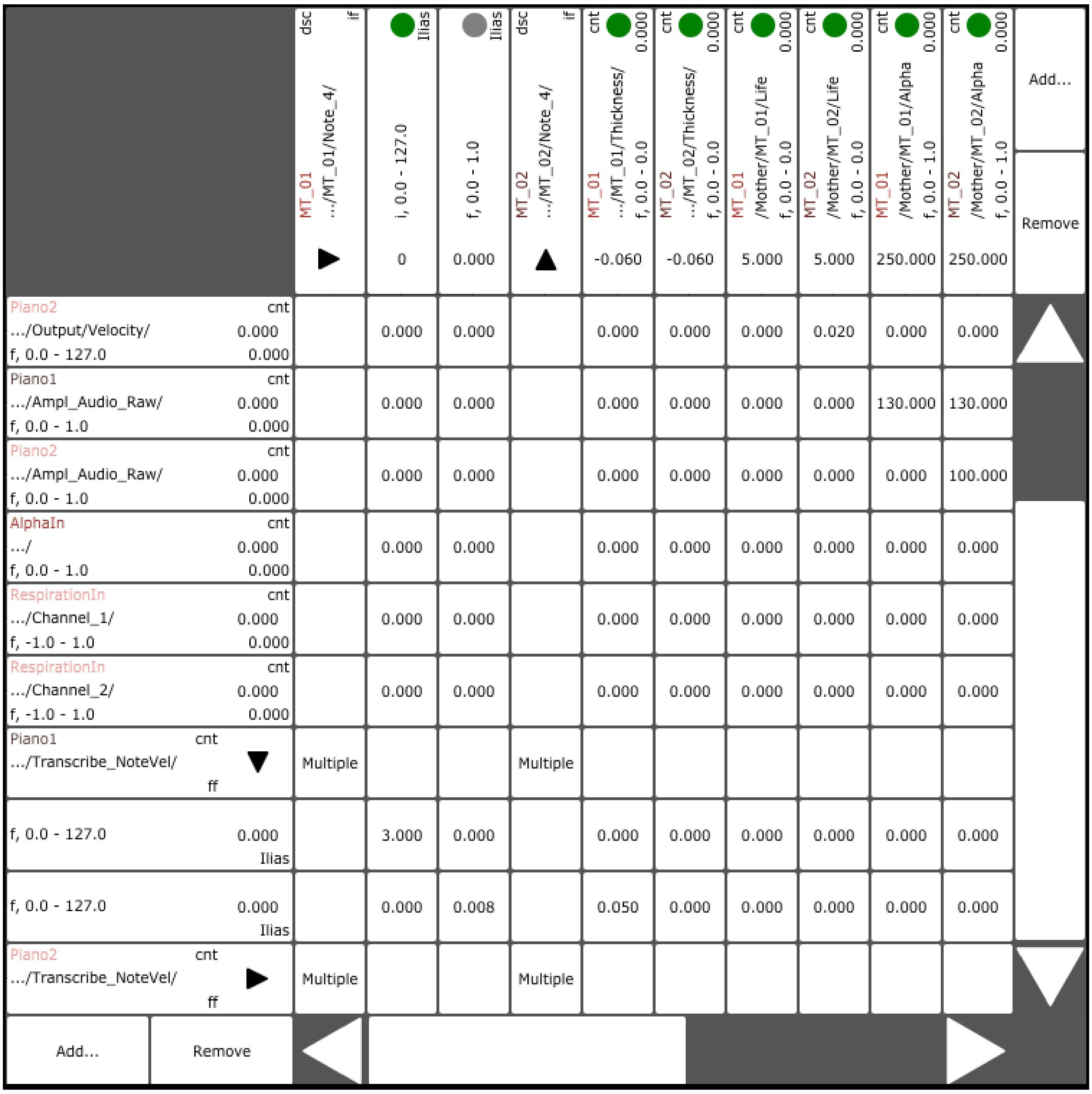

Figure 4: An interactive routing matrix within Mediator, for performing Mutable Mapping. In the left column, sources can be added, while the top row is an analogous list of destinations. A new value is sent whenever one is received and the pairs corresponding cell is non-zero. Because the incoming value is multiplied by the cell value, connections can be gradually manipulated.

The mediator software does not make decisions about the mappings implemented. Instead it presents a detailed user interface which gives access to all necessary controls for manually performing mappings. With mutable mapping, the constraints of current practice are lifted as both high mapping complexity can be achieved, and altering the mappings over time is made possible. A narrative in the mapping between music and visuals, using tension and release, and suspected expectation, is therefore made possible.

With all elements of the Trinity system introduced, readers may now view the signal flow diagram between these, as employed in our practice (Figure 5). Figure 6 presents examples of live visual output.

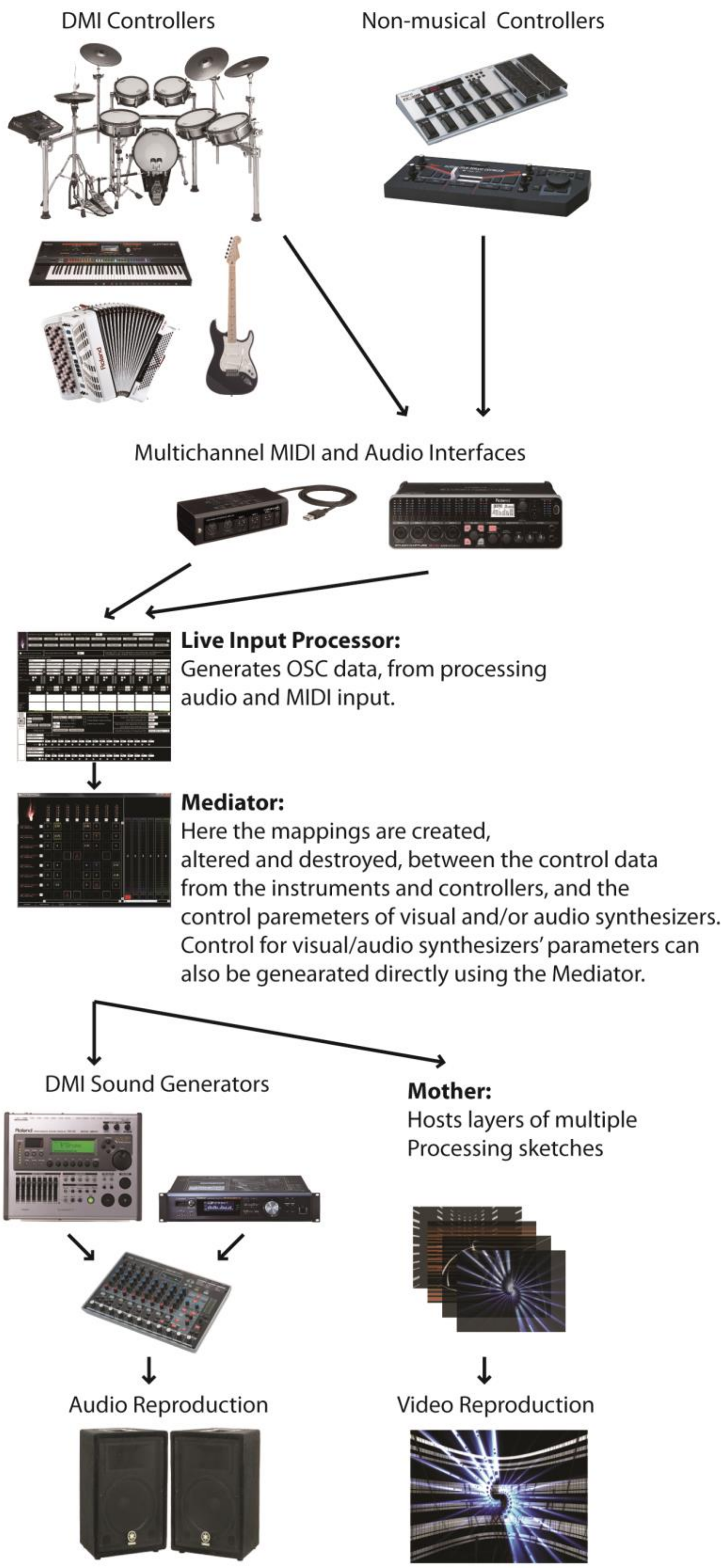

Figure 5: Illustration of signal flow. Note how audio and video are independent, both being controlled from the gestural control data, rather than one from the other.

E. Allowing for collaborative performance in a manner directly comparable to how musicians perform together

In live musical performance (and live performance art) it is common practice that groups, sometimes with conductors or live mixing engineers, perform together. While it is true that in current practice live visuals artists can and do regularly perform together, this is not with the same ease or scalability with which musicians may do so. Musicians who have never before met, may stand on stage together and improvise there and then. This is the point we wish to reach for live visuals. In present work, musical instruments are employed as sources of control data, and mutable mapping is a practice much akin both to live music mixing and conducting. Our hope is that by harnessing these already established roles from musical practice, the immediacy and spontaneity with which collaborative musical performance is attainable, will also translate to the context of visual music. The same benefits that allow collaborative live musical performance may thus translate to also making collaborative audiovisual performance easier, scalable also to larger numbers of participants, rendering the benefits of collaborative performance more easily achievable, and allowing for a heightened experience for performers and audience.

F. Soma: a new artistic practice

In Audiovisual Composition music and visuals are experienced as an inseparable whole. However, it is known that humans fuse more percepts than just auditory and visual; these are processed alongside tactile and other sensory stimulation, depending on the level of congruence between the stimuli [26]. Furthermore *"(...) the perceiver watching, listening to and experiencing another's motor performance, simulates the actions of the performance within the range of their own motor capabilities"* [17]. Consequently, when performers employ advanced enactive knowledge, audiences perceive a richer sensory experience. Many music researchers assert that a direct consequence of richer sensory experience is that audiences are more engaged in the performance [27].

From the above, the theoretical foundation for combined tri-modal sensory stimulus emerges: music is performed alongside congruent visuals, involving performers employing advanced embodied motor knowledge. Thus audiences perceive the performance on three congruent modalities: audiences subconsciously simulate perceived actions, activating their own brains' motor capabilities, while simultaneously experiencing congruent musical and visual stimulus. The experience is similarly strengthened for the performers enacting the physical gestures.

Although the premise is tri-modal congruence, it is recognized that for tension and release and suspended expectation to be possible, it is also necessary to allow for narratives that can transition between states of high congruence and *"Binary opposition or total incongruence"* as also Grierson recognizes. Such narratives are here achieved through employing mutable mapping. Grierson defines congruence between abstract stimuli to multiple senses both as temporal congruence, where events co-occur in multiple stimuli, and as structural similarity. Following the tradition established by our works predecessors, of inventing and naming new artistic practices, the art-form has been given the name *Soma*: the ancient Greek word for body, and the state sponsored drug administered to citizens in Aldus Huxley's *"Brave New World"*.

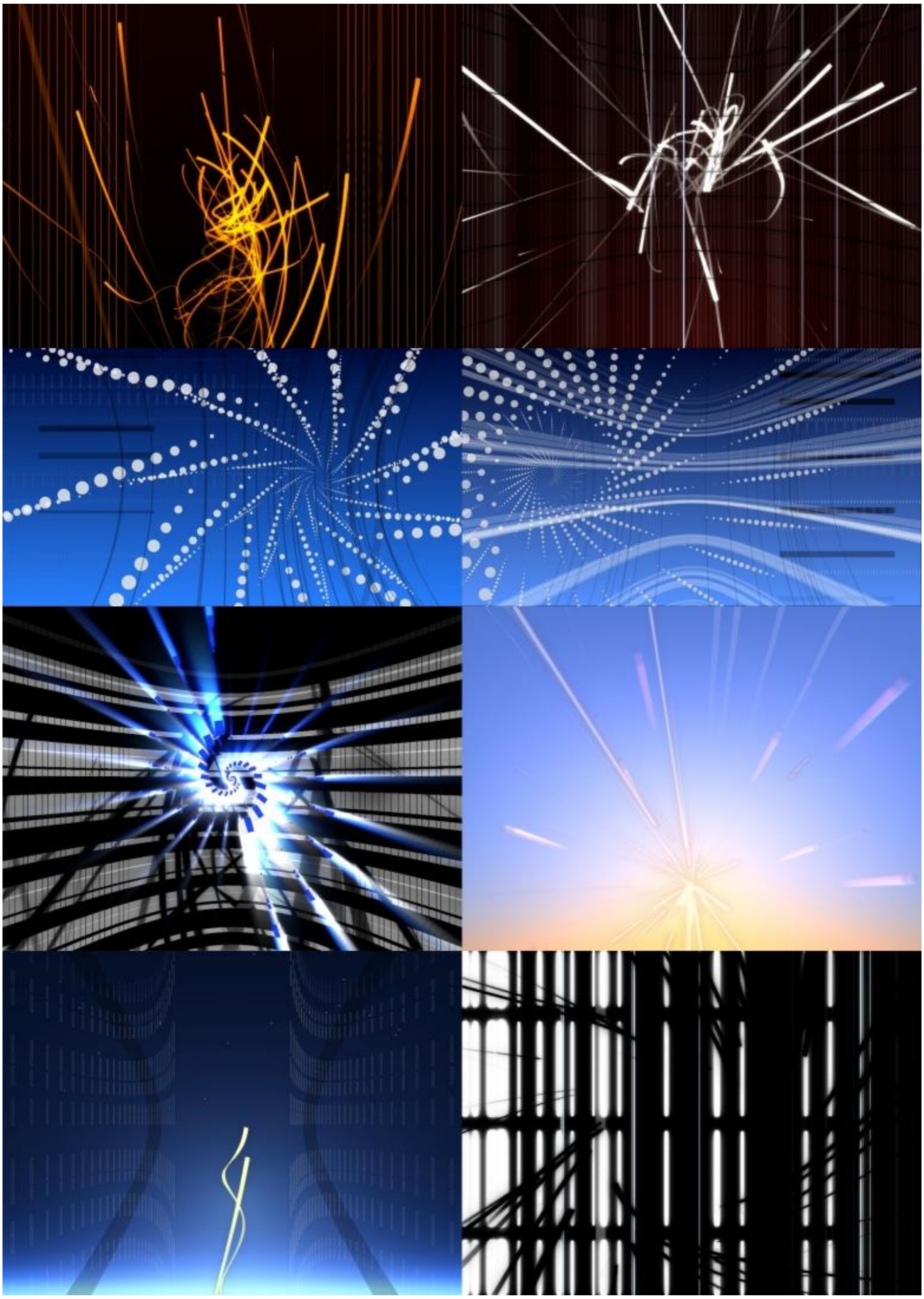

Figure 6: Examples of the visual output possible when using the Trinity system.

5. How our contributions address described limitations

Our complete software system, Trinity, consists of three applications, the Live Input Processor, Mother, and Mediator. There is no one-to-one correspondence between our contributions and described limitations. Instead, our ideas and applications address these in combination. LIP (A) allows the real-time access to a rich dataset from a group of musical performers, thus helping in addressing limitations I, II, and IV. The conduct of mutable mapping with its implementation in the Mediator software (D), forming the central hub of our system, is central in addressing all limitations I-IV. The Mother application simplifies preparation and improvisation as described in (C), thus addressing limitation III. By exposing the entirety of its controllable parameter space, it is receptive to a vast variation of rich mappings, this addressing limitations I and II. Mother has been available as free open source software since 2008, and LIP will at the time of publication also be released as such, to engage the live visuals community with the ideas they engender. Mediator too is under active development towards release in the near future.

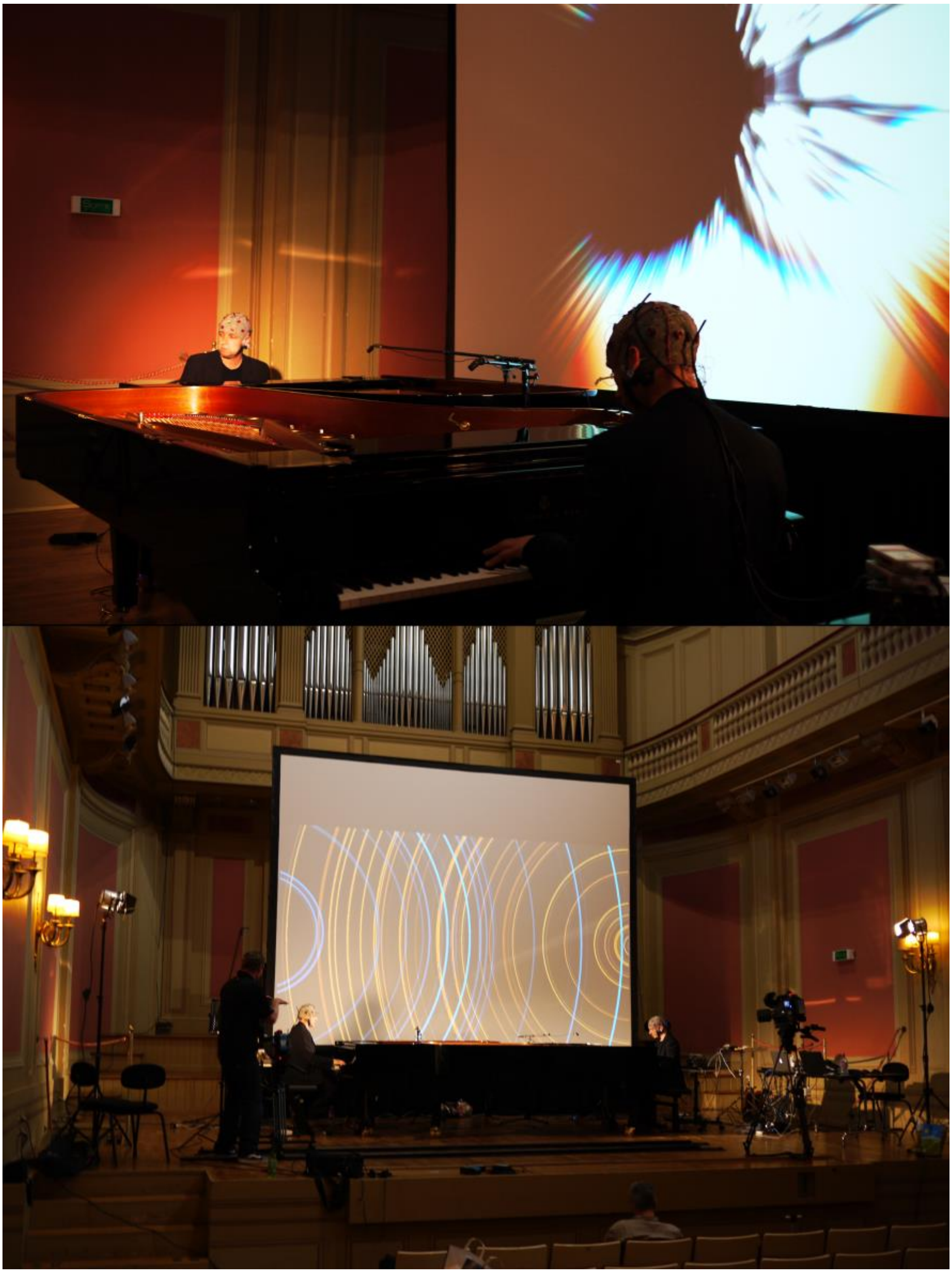

Figure 7: Images from rehearsing the *Music and Brain in Unconscious Waves performance*. Pianists Richard Rentsch and Orazio Sciortino improvise, while the first author accompanies them performing Mutable Mapping for the projections, driven by data from their pianos, electrophysiological and EEG measurements. Joan Llobera and Nathan Evans are involved as scientific collaborators.

6. Conclusion

We expect there can only be an increase in interest towards live visual music and audiovisual art, using systems such as Trinity in live performance. The expensive and complex system that is a visual synthesizer has very recently become far more accessible both to afford owning, and to learn using. Although there still is no software that can currently have an impact analogous to that observed when for example 3D animation software appeared on desktop workstations, a similar development does not seem too distant. The emergence of analogous software for live visuals performance will surely also augment the relevance of ideas such as ours.

Since the detailed account of the research process for present work cannot be summarized in this paper, interested readers can find the full report freely available online, in the form of the first author's PhD thesis [28].

New versions of the LIP, Mother and Mediator applications are all under development, and new performances involving live musicians are at the time of writing in advanced stages of preparation (see Figure 7 for images from a recent rehearsal), so we are certain this article is not the conclusion of the work detailed, only an introduction to a series of improvements to it, and performances employing it, hopefully both by ourselves and others!

Acknowledgements

This work has been supported by an EU Presenccia grant, the Spanish INNPACTO Melomics project reference code IPT-300000-2010-010, and through a grant by the Agalma foundation.

References

1. M. Dabrowski, *Kandinsky: Compositions*. (New York: Harry N. Abrams, 1997).
2. D. J. Levitin, *This is your brain on music*. (London: Dutton Publishing, 2006).
3. I. Cross, "Music and evolution: Consequences and causes," *Contemporary music review*, vol. 22, no. 3, pp. 79–89, 2003.
4. J. Whitney, *"Digital Harmony: on the complementarity of Music and Visual Art"*, (New York: McGraw-Hill, 1980).
5. B. Evans, "Foundations of a Visual Music," *Computer Music Journal*, vol. 29, no. 4, pp. 11–24, 2005.
6. M. S. Grierson, "Audiovisual composition," PhD Thesis, University of Kent at Canterbury, 2005.
7. D. S. Ebert, *Texturing & Modeling: A Procedural Approach*. (Massachusetts: Morgan Kaufmann Pub, 2003).
8. E. R. Miranda and M. M. Wanderley, *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard*. AR Editions, 2006.
9. L. E. Marks, *The unity of the senses: Interrelations among the modalities*. (Massachusetts Academic Press, 1978).
10. P. Larsson, "Virtually hearing, seeing, and being: room acoustics, presence, and audiovisual environments," PhD Thesis, 2005.
11. M. Chion, C. Gorbman, and W. Murch, *Audio-vision: sound on screen*. Columbia University Press, 1994.
12. L. Shams, Y. Kamitani, and S. Shimojo, "Visual illusion induced by sound," *Cognitive Brain Research*, vol. 14, no. 1, pp. 147–152, 2002.

13. S. D. Lipscomb, “The perception of audio-visual composites: Accent structure alignment of simple stimuli,” *Selected Reports in Ethnomusicology*, vol. 12, pp. 37–67, 2005.
14. N. Steinbeis, S. Koelsch, and J. A. Sloboda, “The Role of Harmonic Expectancy Violations in Musical Emotions: Evidence from Subjective, Physiological, and Neural Responses,” *Journal of Cognitive Neuroscience*, vol. 18, no. 8, pp. 1380–1393, 2006.
15. C. Dobrian and D. Koppelman, “The ‘E’ in NIME: musical expression with new computer interfaces,” in *Proceedings of the 2006 conference on New interfaces for musical expression*, Paris, France, 2006, pp. 277–282.
16. F. J. Varela, E. Thompson, and E. Rosch, *The embodied mind: Cognitive science and human experience*. (Cambridge, MA: The MIT Press, 1992).
17. M. Rodger, J. Issartel, and S. O’Modhrein, “Performer as perceiver: Perceiver as performer,” *Proceedings of the ENACTIVE conference*, vol. 7, pp. 237–240, 2007.
18. F. Collopy, D. Jameson, and R. Fuhrer, “Visual music in a visual programming language”, in *1999 IEEE Symposium on Visual Languages*, 1999, pp. 111–118.
19. N. Bryan-Kinns and F. Hamilton, “Identifying mutual engagement,” *Behaviour & Information Technology*, no. 1, pp. 1–25, 2009.
20. M. Tufnell and C. Crickmay, *Body space image*. (London: Virago, 1990).
21. M. Wright, “OpenSound Control: A New Protocol for Communicating with Sound Synthesizers,” *Proceedings of 1997 International Computer Music Conference, ICMA*, pp. 101–104, 1997.
22. I. Bergstrom and B. Lotto, “Mother: Making the Performance of Real-Time Computer Graphics Accessible to Non-programmers,” in *(re)Actor3: The Third International Conference on Digital Live Art Proceedings*, 2008, pp. 11–12.
23. C. Reas and B. Fry, *Processing: A Programming Handbook for Visual Designers and Artists*. (Cambridge, MA: The MIT Press, 2007).
24. I. Bergstrom and B. Lotto, “Code-bending: a new creative coding practice,” *Leonardo*, In Press, 2014.
25. I. Bergstrom, A. Steed, and B. Lotto, “Mutable Mapping: gradual re-routing of OSC control data as a form of artistic performance,” in *Proceedings of the international conference on Advances in computer entertainment technology*, Athens, Greece, 2009, pp. 290–293.
26. D. R. Wozny, U. R. Beierholm, and L. Shams, “Human trimodal perception follows optimal statistical inference,” *Journal of Vision*, vol. 8, no. 3, p. 24, 2008.
27. N. Armstrong, “An Enactive Approach to Digital Musical Instrument Design,” PhD Thesis, 2006.
28. I. Bergstrom, “Soma: live performance where congruent musical, visual, and proprioceptive stimuli fuse to form a combined aesthetic narrative,” PhD Thesis, 2011.