California Institute of the Arts

Audiovisual Mapping and Synaesthetic Music

by

Gabriel Rey-Goodlatte

A thesis submitted in partial fulfillment for
the degree of Master of Fine Arts

in the
Herb Alpert School of Music
Music Technology: Interaction, Intelligence & Design

Supervisory Committee

Dr. Jordan Hochenbaum

Mentor

Dr. Ajay Kapur

Committee

Dr. Owen Vallis

Committee

Abstract

The concept of synaesthesia–a correlation and blending of perceptual experience across multiple senses–has long been a source of inspiration in music and the arts. For centuries, philosophers, musicians, artists, and inventors have reached towards a visual analogue for–and counterpart to–music, creating theoretical frameworks, aesthetic experiments, and new technologies to correlate sound, light, and images. *Audiovisual Mapping and Synaesthetic Music* surveys historic and contemporary approaches to synaesthetic music with a focus on techniques and technology, presents new synaesthetic musical artwork, and presents new technology for crafting synaesthetic sound-image relationships in a musical context. This thesis examines how technology mediates the possible techniques and workflows in interdisciplinary and experimental work, with a focus on audiovisual techniques, technologies, and works that are specifically musical in nature. Placing visual production within the context of musical technologies, vocabularies, and methods helps facilitate audiovisual synaesthetic music. This thesis proposes the development of new technology to facilitate linking real-time musical and visual systems, presents software contributions in this vein, and examines their use in new works.

Acknowledgments

I would like to thank:

- Jordan Hochenbaum, Ajay Kapur, and Owen Vallis for teaching me how to create music technology, for their tireless dedication to their students, and for supporting my creative vision in many different ways.
- Tom Leeser, Tom Jennings, Sara Roberts, Kirsten Winter, Charlotte Pryce, and Yehuda Duenyas for informing my work at CalArts with insight from sound art, film, theater, and interdisciplinary art.
- Bob Bielecki, Brenda Hutchinson, Peggy Ahwesh, and Jennifer Reeves for supporting my initial work in synaesthetic music at Bard College.
- Erwin Helfer, Betty Ross, Lou Mallozzi, and James Rucker for being mentors who opened my eyes to diverse ways of relating to music, art, and the world.
- Dorka Keehn and Brian Goggin for creating a beautiful work of art–*Caruso's Dream*–and inviting me to help; Joshua Sophrin for generously lending his technical expertise.
- Nick Suda, Raphael Arar, Colin Honigman, David Howe, Jason Jahnke, Kameron Christopher, Jingyin He, Jeff Bryant, Jonathan Becker, and Bruce Dawson for blowing my mind with your ideas, energy, and hard work as collaborators and colleagues.
- Jacob Pritzker, for exposing me to some of the best underground music I've ever heard, and helping me hone my taste.
- Adrian Adorno, whose description of creativity in graffiti art helped me experience and appreciate music in new ways.
- Angelo Duncan, Gustav Feldman, and Thandiwe Satterwhite for all the fun we had playing in Elle Niño, and for helping me become a better musician.
- My parents, Roberto Rey and Joyce Goodlatte, for supporting me in countless ways.
- My fiancé, Erika Rand, for sharing your creative genius, being an amazing thought partner, and for everything else.
- Musicians, artists, scientists, philosophers, mystics, and technologists who have pursued, documented, and developed the idea of synaesthetic music.

Contents

List of Figures

Chapter 1
Introduction

> *Can one imagine anything in the arts which would surpass the visible rendering of sound, which would enable the eyes to partake of all the pleasures which music gives to the ears? … What could [w]e say of an art which did not only simply awaken the idea of speech, and of sound, by means of arbitrary and inanimate characters, such as the letters of the alphabet, or the notes of music; but painted it really; that is painted it with colors; in one word, rendered it felt and present to the eyes, as it is to the ears, in such a manner that a deaf person can enjoy and judge the beauty of music as well as he who can hear.* (Castel 1757)
>
> – Louis Bertrand Castel, creator of the first known color organ

This thesis discusses audiovisual, synaesthetic music, with a focus on how technology shapes and expands its aesthetic and creative possibilities. A brief historical overview is provided, as well as a survey of contemporary technology for synaesthetic music. New software is presented which helps facilitate correlations between sound and image in real-time musical systems, and new synaesthetic musical works are documented.

1.1 Defining synaesthesia

The word 'synaesthesia' (also spelled 'synesthesia') generally refers to a neurological phenomenon in which a correlation, overlap, or blending of perception across multiple senses is experienced. A synaesthete might taste colors, or see sounds.

Simon Shaw-Miller writes in his chapter "Synaesthesia" in the *Routledge Companion to Music and Visual Culture*:

> At its most general, synaesthesia can be defined as the cross-stimulation of sensory modalities. Neurologically, synaesthesia occurs when the stimulation of one sensory modality automatically, and instantly, triggers a perception in a second modality, in the absence of any direct stimulation to this second modality. To give an example from the most common case … a piece of music might automatically and instantly trigger the perception of vivid colors. The interaction of sight and sound (music and the visual) constitutes about 90 percent of neurological synaesthesia cases. (Shaw-Miller 2013)

When describing the neurological phenomenon, the word synaesthesia has a strict definition that limits its application. In "The Hidden Sense: On Becoming Aware of Synesthesia," Cretien van Campen writes that while 4% of people fit the neuroscientific definition of synaesthesia, many more are interested in "art forms that present synesthetic experiences" (Campen 2009). The word synaesthesia is often used by musicians and artists in relation to a much wider range of experience, theory, and artistic expression involving a union of experience across different senses. Van Campen's article expands the view on synaesthesia to include social and cultural interactions, calling for a definition that better accounts for the diversity of experience described by synaesthetes, and that takes into better account research which suggests that perception is primarily synaesthetic (Campen 2009) (discussed in 1.2). Similarly, Shaw-Miller writes that synaesthesia has meaning as both a neurological and cultural phenomenon, and that while science often seeks to make a very clear distinction between these two manifestations of synaesthesia–drawing clear boundaries around synaesthesia as a neurological phenomenon–the distinction between the two is overstated. The existence of synaesthesia as a consistent neurological condition which can be easily documented and defined, writes Shaw-Miller, should not distract us from the existence of synaesthetic experiences outside this framework (Shaw-Miller 2013).

This thesis deals with synaesthesia as an aesthetic in music and art. Neurological research into synaesthesia may offer profound insight into the subjects discussed here, including the artistic appeal of synaesthesia and techniques for achieving satisfying synaesthetic relationships between sound and image. This thesis uses the word synaesthesia not according to its strict neurological definition, but rather as a way of describing a particular type of cross-sensory correlation in perceptual experience which synaesthetic music and art aim to create. More specifically, this thesis focuses on audiovisual synaesthetic music (which is defined and discussed in 1.3).

1.2 Neurological synaesthesia, audiovisual perception

While only a small percentage of people fit the neurological definition of synaesthesia, the phenomenon may actually be in some ways universal to human perception. Recent research suggests that babies may have undifferentiated senses (Walker et al. 2010; Maurer and Mondloch 2004), that the connections between senses are pruned away over time, and that this process is less complete in adults who experience synaesthesia, but perhaps not fully complete in any adult.

> According to the neonatal synesthesia model, newborns fail to differentiate input from different senses–either because of connections between cortical areas that are pruned or inhibited later in development or because of the multimodal limbic system being more mature than the cortex … The remnants of this unspecialized cortex are most clearly evident in synesthetic adults who experience, for example, visual percepts in response to sound and in adults with abnormal sensory experiences, such as the congenitally blind or deaf who have unusual patterns of activation in cortical areas deprived of typical input. However, remnants also are observed in normal children and adults in their ability to match dimensions from different modalities (e.g. pitch and size) and in the prevalence of cross-modal metaphors (e.g. 'loud colors') in everyday speech. (Maurer and Mondloch 2004)

Research is mixed on whether there is universality in synaesthetic associations. Some research shows that synaesthetic associations vary from person to person and can be influenced by linguistic (Dolscheid et al. 2013) or other associations based in experience. Other research shows strong similarities in how different people perceive particular types of associations (for example, that people generally associate smallness and lightness with high pitch, while associating largeness and darkness with low pitch).

The study of multimodal perception–how sensory input from different sources is weighed and integrated into a coherent and useful perceptual experience–is also highly relevant. The integration of different sensory modalities is essential to how we perceive the world (Carterette and Friedman 1978, Hochenbaum 2013). Speech is processed in an audiovisual manner, with sensory input from the eyes (lip-reading) affecting not only the overall perception, but also integrated into the experience of hearing. The McGurk effect demonstrates how people can be fooled into perceiving phonemes not actually sounded, based on a corresponding visual image of lips speaking a different phoneme than that of the actual sound. Another example of this kind of cross-sensory integration is found in how information about low-frequency vibrations taken in by the body's sense of touch are integrated into the experience of hearing.

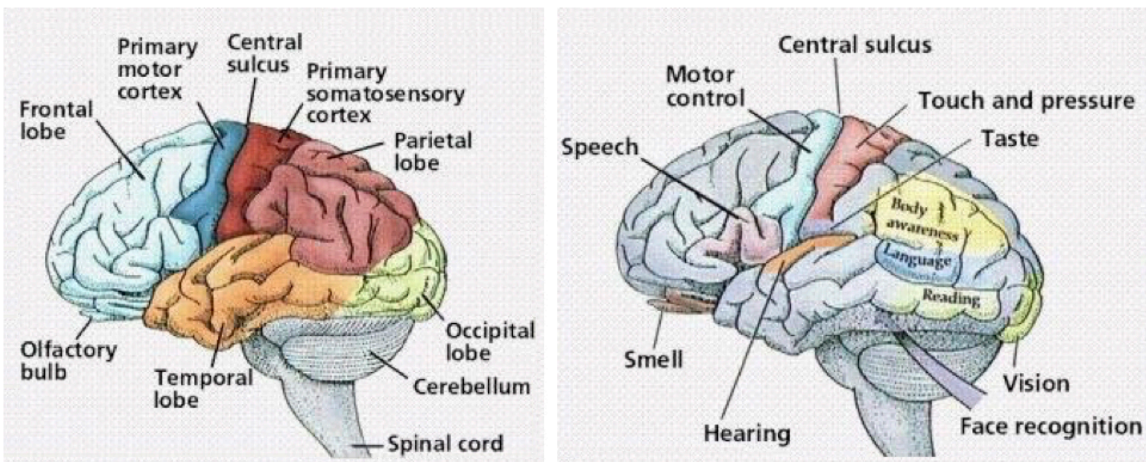

Figure 1. Diagram depicting brain regions typically associated with the five senses

Negotiation and integration between different kinds of sensory input is evident not only between the traditional five senses, but also within them (for example, vision involves integrating stimuli from different kinds of cells that respond to different kinds of visual information). Shaw-Miller writes:

> The notion that there are five senses is principally a consequence of the physical (visual) existence of sense organs: eyes, ears, nose, tongue, and skin. But of course the ear does not hear, it merely receives vibration in air (usually) before passing it on to the cochlea and translating it into a liquid medium, then further translations take place into nerve

> impulses before the brain can finally interpret and hear. Sound happens throughout this process, not as a simple unitary phenomenon, but as a mingled one. (Shaw-Miller 2013)

Multimodal perception is universal. Research suggests it may be neurologically related to synaesthesia, according to Maurer and Mondloc, who write that connections between brain regions usually associated with separate modalities (see **Figure 1**) may be the underlying basis for both synaesthesia and multimodal perception in non-synaesthetes (Maurer and Mondloch 2004). Much contemporary neurological research is focusing on mapping connections between regions of the brain in increasing detail (see **Figure 2**).

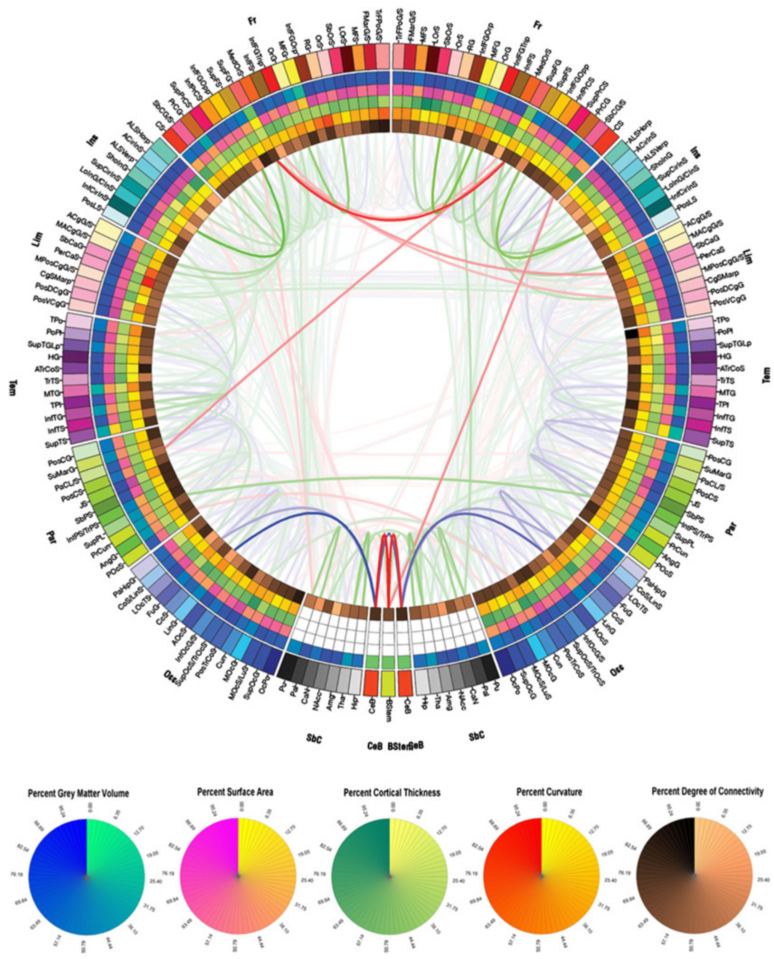

Figure 2. A visualization of connections between brain regions

Another striking example of how auditory information can result in a visual image is in human echolocation, practiced by some blind people who learn to use information from their ears about how sound is reflecting in the space around them, to inform a visual sense. Bats do this with an auditory system that is uniquely geared towards spatial sensing and that evolved as a system for vision, while human hearing is more spectrally and temporally oriented. Human practitioners of echolocation, nevertheless, have learned how to sense visually with their ears, and say that it produces visual images. It has been demonstrated that during echolocation, activity occurs in parts of the brain normally associated with visual perceptions. (Thaler, Arnott, and Goodale 2011) This phenomenon and the McGurk effect both demonstrate that sensory signals originating from the ears and eyes are, in the brain, in some sense interoperable and translatable from one domain to another.

1.3 The concept of synaesthesia in music and art

This thesis uses the term "synaesthetic music" to refer to music (or a musical art) which aims to be both sonic and visual in nature, with a synaesthetic relationship between sound and image. Attempts to correlate musical sound and sight have a very long history, which is discussed in Chapter 2. When technology began to meaningfully make possible the technical coordination of sound and image in the late 1800s and early 1900s, a shared vision of an audiovisual, synaesthetic, and musical art form spread among some musicians, artists, scientists and technologists. The art forms associated with this concept have often been described as color music or visual music, terms which refer to the idea of music in a visual medium. The concept of visual music is related to synaesthesia, at least in its inspiration (given that music is normally considered a sonic art). Many works often described as visual music can also be described as synaesthetic music, but visual music is a broader category because it may or may not involve representing synaesthetic relationships between sound and image. Visual music also includes silent film works inspired by the form and structure of music. Some of these silent films can have a synaesthetic effect (creating a corresponding sonic experience) for people with some synaesthetic sensitivity. However, without sound, such films do not in themselves contain synaesthetic sound-image relationships. Additionally, visual music can also refer to audiovisual work in which the emphasis is on a musical interplay and perhaps counterpoint between sound and image, but where sound and image are not directly linked in a synaesthetic manner.

This thesis is interested primarily in synaesthetic musical work that involves a consistent and strong correlation between sound and image, especially a strong temporal correlation (time is music's medium, and should be central to the audiovisual relationship in synaesthetic music). While acknowledging that this is closely related to, and in some ways a subset of what is more generally referred to as visual music, this thesis uses the term "synaesthetic music" to refer specifically to such work. Specifically, this thesis investigates how technology has shaped, expanded, and can continue to expand the possibilities for expression in synaesthetic music.

1.4 Overview

The new creative work and software presented in this thesis is informed by a survey of historical and contemporary synaesthetic music and its relationship to technology. Chapter 2 traces the history of synaesthetic music, with an emphasis on how technology has mediated and expanded its possibilities, beginning with ancient views on the relationship between sound and form, and moving through the development and impact of color organs, abstract film, video, and computers.

The remainder of this thesis presents new synaesthetic musical works and new software to facilitate the production of synaesthetic music. Chapter 3 documents my work on *Caruso's Dream*, a permanent site-specific public art installation in San Francisco that integrates lighting sequences with original recordings of Enrico Caruso. Chapter 4 documents several installations and performances that approach the idea of synaesthetic music, and which informed the development of the new software presented in this thesis. Chapter 5 and Chapter 6 present two new software applications–ControlMaps and FFT Control–which help facilitate the correlation of sound and image in a musical context.

Appendix A defines several types of methods for linking sound and image, and discusses their application to synaesthetic music, citing examples of each method. Appendix B surveys contemporary technology for synaesthetic music, focusing on how technology can interconnect different disciplines and systems, work with the relationship between sound and image, and place the creation of such synaesthetic relationships in a musical context, enabling synaesthetic musical expression.

1.5 Summary of Contributions

The list below provides an overview of the technical and creative contributions to the practice of synaesthetic music documented in this thesis (organized by order of appearance).

1. *Caruso's Dream* (a collaboration with artists Brian Goggin and Dorka Keehn)
    a. Uses custom software for generating and shaping light according to its format in the piece (13 illuminated glass piano sculptures).
    b. Explores an approach to synaesthetic music involving a combination and integration of both audio analysis based mapping techniques, and manual composition.
    c. Investigates synaesthetic music by representing and embodying music in light.
    d. Situates the production of lighting sequences in a musical technical context.

2. Installations and performances
    a. *Emulsion Juice #1*: Creates a live, interactive synaesthetic musical experience by visualizing sound and inviting the audience to musically explore the resulting audiovisual synaesthetic relationship using instruments.
    b. *Ces Balons de Lumière*: A collaboration with David Howe and Raphael Arar. Visualizes sound in colored light using audio analysis in a musical performance. Sound from a modular synthesizer generates light which flickers from floating helium balloons.
    c. *Audiovisual Ouroboros #1*: Explores audiovisual feedback through the use of bi-directional mapping between sound and image involving both visualization and sonification. Sound from a microphone is visualized in a particle system, attributes of the particle system are sonified, and the resulting sound is played from speakers which point into the microphone, completing the loop. Participants are invited to "play" the feedback by positioning the microphone.

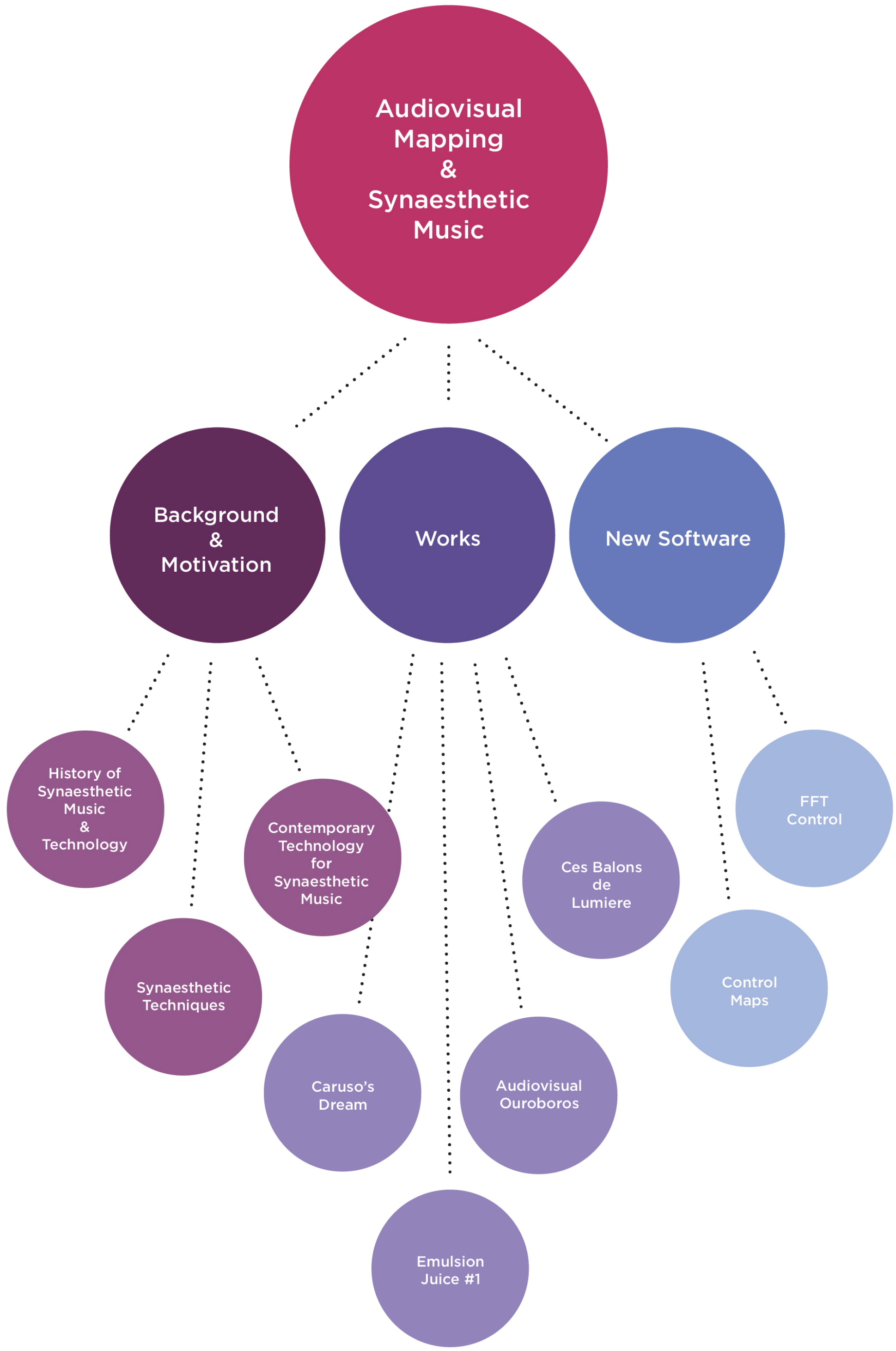

Figure 3. Overview

3. ControlMaps
    a. One of the first software applications aimed specifically at nonlinear mapping of parameters between sonic and visual systems.
    b. Allows for quick and intuitive adjustment of audiovisual relationships based on parameter mapping.
    c. Allows one parameter to control several others according to different mappings of incoming to outgoing values.
    d. Communicates via MIDI and Open Sound Control for integrating real-time parameter mapping between most audio and visual software, and hardware interfaces.

4. FFT Control
    a. Generates real-time control signals using FFT-based spectral audio analysis.
    b. Offers an extensive set of audio-based feature extraction methods, and provides a simple interface for selecting different kinds of feature parameters for synaesthetic mappings.
    c. Packaged as a VST plugin for easy integration with most audio software.
    d. Allows for the modular use of audio analysis on a variety of sound sources within music production environments.

Chapter 2
History of synaesthetic music and its technology

This chapter provides an overview of synaesthetic music and the technology used to create it. It is not intended to be comprehensive; rather, this chapter focuses on the ways in which science and technology mediate the possibilities for experimentation and expression in synaesthetic music, and thus has served as a direct motivator to the artistic and technological contributions presented in this thesis.

This chapter demonstrates that attempts to correlate sound and image according to musical concepts, and to create synaesthetic music, are integral to the history of the arts and their relationship to technology. Scientific understanding and technological development have always been limiting and defining factors in attempts to create synaesthetic music. The desire to create synaesthetic music has motivated the development of interdisciplinary technology and techniques that have expanded what is possible in synaesthetic music while also having a profound impact on the relationship between technology and more traditional, well-defined artistic disciplines.

2.1 Ancient roots

The idea of a unity or correlation between sound and the visible is very old.

Dieter Daniels and Sandra Naumman write in the introduction to *Audiovisuology Compendium* that over the course of history, many models postulating a correspondence between colors and sounds have been constructed, and that such models were often a part of systems that sought a holistic model of the universe.

Many spiritual and religious traditions have described sound as a force of creation which gives rise to form. The Vedas of ancient India describe four levels of sound corresponding to four levels of existence and consciousness, ranging from the most subtle and undifferentiated (para-vak) to the externally manifested, physical form of sound, vaikhari-vak (which corresponds roughly to the scientific definition of sound as a physical phenomena). In pashyanti-vak, the second of these levels (one level less subtle than para-vak), sound has form and color. Pashyanti means 'that which can be seen or visualized.' Advanced spiritual practitioners (yogis) in this tradition can perceive these qualities in sound. (Nitai Das 2014)

In *Egyptian Harmony: The Visual Music*, Moustafa Gadalla writes that ancient Egyptian civilization saw a unity between sound and form and pursued principles of harmony in both music and architecture, because sound and form were seen as two aspects of the same reality, their relationship conforming to the same harmonic principles that governed the entire physical and metaphysical universe. (Gadalla 2000)

According to Gadalla, there was also a kind of synaesthetic connection between sound and form in Egypt's written language:

> Tehuti (Thoth) set the principle of the written language—letters—as the graphic representation (image/picture) of the spoken/sound vibrations. This point illustrates the intimate relationship between the sound and the visual (written/illustrated) forms. The image of each Egyptian symbol contained its specific vibrational pattern. Words were constructed of these symbols in a manner incorporating and amplifying the meaning of the individual symbols, so that the meaning of a word emerged from the interplay of symbols, as the meaning of a chord or a musical phrase results from the combination of notes. (Gadalla 2000)

In *The Secret Teachings of All Ages*, Manley P. Hall discusses the Pythagorean roots of the philosophy of music, first noting that the Greeks likely learned from the Egyptian philosophical approach to music. Hall describes how Pythagoras' study of harmonic ratios in sound led to an extension of this harmonic theory to other aspects of nature and reality, and within this system were various musical structures and analogies:

> Pythagoras applied his newly found law of harmonic intervals to all the phenomena of Nature, even going so far as to demonstrate the harmonic relationship of the planets, constellations, and elements to each other … The Greek Mysteries included in their doctrines a magnificent concept of the relationship existing between music and form. The elements of architecture, for example, were considered as comparable to musical modes and notes, or as having a musical counterpart. Consequently when a building was erected in which a number of these elements were combined, the structure was then likened to a musical chord, which was harmonic only when it fully satisfied the mathematical requirements of harmonic intervals … In the Pythagorean *tetractys*–the supreme symbol of universal forces and processes–are set forth the theories of the Greeks concerning color and music. The first three dots represent the threefold White Light, which is the Godhead containing potentially all sound and color. The remaining seven dots are the colors of the spectrum and the notes of the musical scale. The colors and tones are the active creative powers which, emanating from the First Cause, establish the universe.

In his history of color in Western art, *Colour and Culture*, art historian John Gage examines how analogies between sound and color were an important part of how artists, scientists, and philosophers in the ancient and medieval European world sought to understand and develop theories about the nature of color.

> The experience of colour in the West has always been closely interwoven with the experience of music … Some Greek theorists considered 'colour' to be a quality of sound itself, together with pitch and duration; it may have been thought akin to what we now describe as timbre. What most impressed the Greeks, it seems, was the capacity of colour, like sound, to be articulated in a series of regularly changing stages whose differences were perceptible in an equally regular way–for Aristotle and his school light and dark appear to have been cognate with clear and muffled sound or even high and

> low pitch … The analogy between musical sound and color seems to have been most compelling to the Greeks because both could be organized in more or less regularly stepped scales …. (Gage 1993)

In ancient Greece, the physical nature of sound and light were not yet well understood, and there was speculation about possible physical similarities between color and sound (which continued until the late 1800s, when scientific and technological advances allowed light and sound to be quantified in new ways). Attempts to create systems to organize color along a scale, and to create a theory of color harmony often sought guidance from established knowledge about sound and music. Analogies between color and sound generally focused on potential correspondences between colors and musical pitches, and used this analogy to look at musical scales and musical harmony as a guide for creating color scales and color harmony. Gage writes that such analogies were speculative and never bore the kind of fruit that ancient Greek philosophers hypothesized they might, but they were the inspiration for subsequent attempts to correlate color and pitch in medieval Europe.

2.2 Color Organs

In Renaissance Europe, analogies between color and musical pitch continued to be a significant thread in how artists and scientists grappled with color theory, and efforts to create mappings between pitch and color continued (see **Figure 4**). Color organs have their roots in these attempts to map pitch to color, and are some of the earliest known examples of an audiovisual music technology designed to correlate light and sound.

The first color organ, called the Ocular Harpsichord, was built by the Jesuit mathematician Louis Bertrand Castel:

> [It] consisted of a 6-foot square frame above a normal harpsichord; the frame contained 60 small windows each with a different colored-glass pane and a small curtain attached by pulleys to one specific key, so that each time that key would be struck, that curtain would lift briefly to show a flash of corresponding color … a second, improved model in

1754 used some 500 candles with reflecting mirrors to provide enough light for a larger audience, and must have been hot, smelly and awkward, with considerable chance of noise and malfunction between the pulleys, curtains and candles. (Moritz 1997)

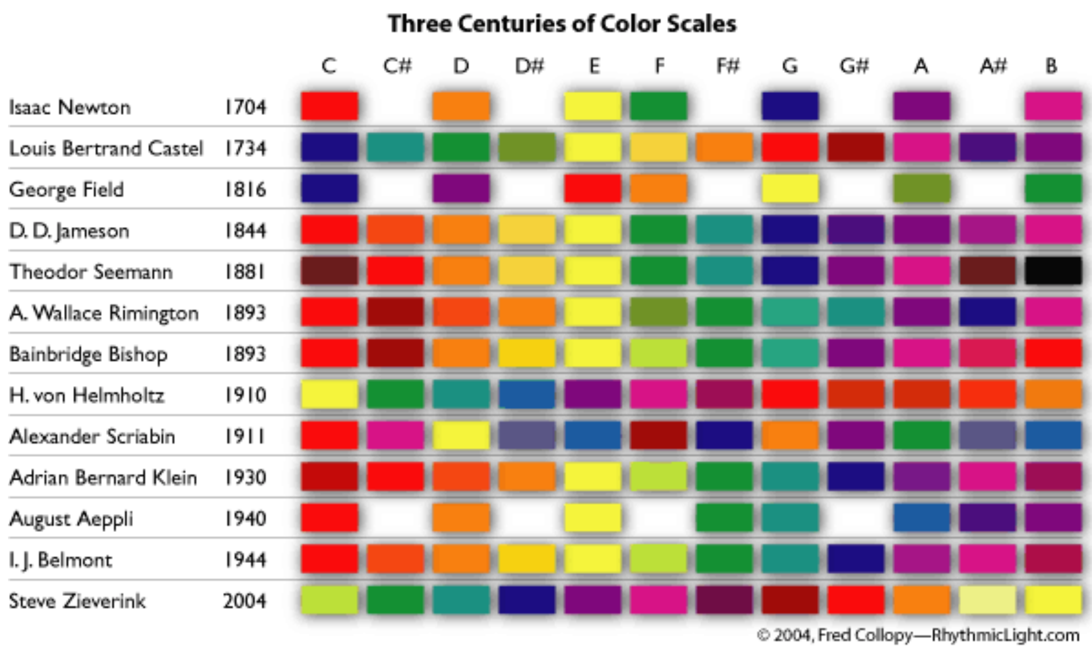

Figure 4. A diagram depicting historical pitch to color mappings

In the late 19th and early 20th century, with more advanced technology (including electricity) at their disposal, many others invented color organs, which usually used a musical keyboard to control the projection of colored lights, and often also simultaneously produced sound like a traditional instrument. Others did not produce sound, but aspired to facilitate the manipulation of light and color in a manner analogous to how sound is manipulated in music, with the goal of helping bring about a new art of "color music" or "visual music."

Notable inventors of color organs include Bainbridge Bishop, Alexander Wallace Rimington, Mary Hallock-Greenewalt, and (later) the experimental filmmaker Oskar Fischinger. *Visual Music Instrument Patents*, edited by Michael Betancourt, collects 27 patents, from 18 inventors, for such instruments–and for systems for linking color and music–issued from 1876-1950.

William Moritz writes that Mary Hallock-Greenewalt, in the process of creating her color organ, "invented the rheostat in order to make smooth fade-ups and fade-outs of light, and the liquid-mercury switch, both of which have become standard electric tools." (Moritz 1997)

Most of the devices documented in Betancourt's collection are color organs controlled by a musical keyboard, linking color and pitch by triggering each simultaneously, in response to input from a player. Richard M. Craig's 1929 patent for a "Radio Color Organ," however, is worth noting here for the different technical approach it took to linking sound and light: it was a system designed to visualize, in the form of fluctuating colored light, the sound played back by a radio. It achieved this by linking switches to tuned musical reeds that would vibrate in sympathy with frequencies produced by the radio. There were many of these reeds, each tuned to a different frequency, which, when vibrated, would complete a circuit, varying the amount of electricity flowing to various colored lights. This was a primitive but ingenious form of electro-mechanical audio analysis and real-time translation of spectral content to color, not dissimilar in function to FFT analysis (with each musical reed functioning like a bin). Craig was interested in creating a relatively inexpensive, automated system which could translate sound into color without needing a skilled player to be involved, and which could be installed in any radio cabinet–a form of automated music visualization, for entertainment purposes, which anticipated software music visualizers by decades.

Color organs are an early example of situating visual production in a musical context, and giving its tools a musical interface. This thesis discusses and explores the possibilities afforded by software to do the same.

2.3 Film

Another nascent technology in the late 19^{th} and early 20^{th} century with a clear and profound impact on the possibilities for visual and synaesthetic music was film. Some of the earliest abstract filmmakers explored the medium because of their interest in representing music visually.

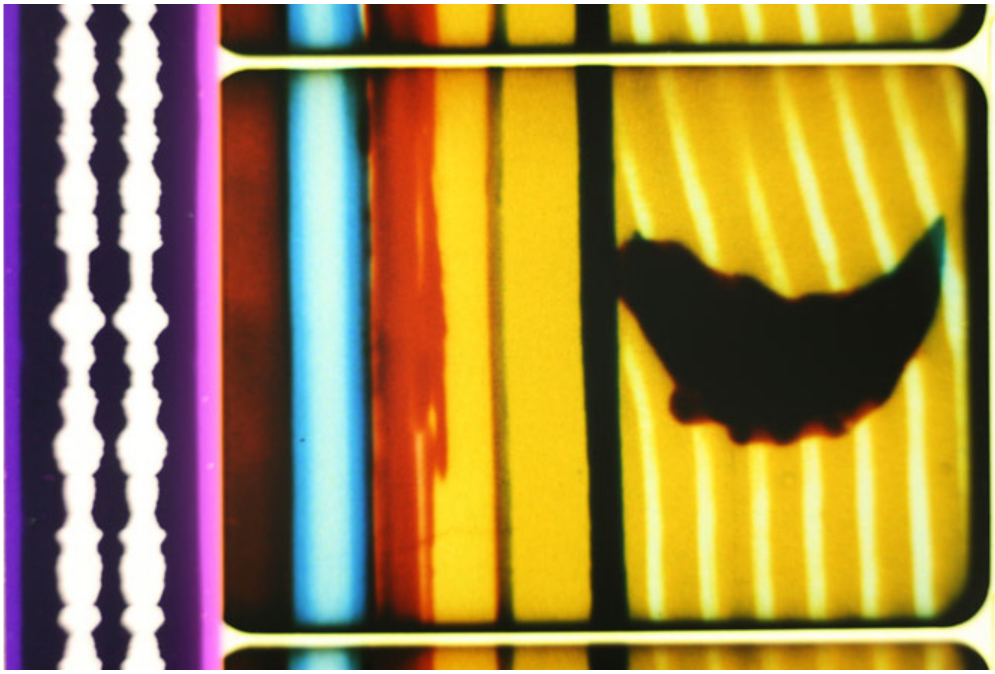

Figure 5. Still from Len Lye's "Colour Flight"

The first experiments in using film as a medium for abstract, musical images may have been conducted by Mary Hallock-Greenewalt, who was not primarily a filmmaker and is better known for her color organs, but nonetheless looked to film technology as a means of realizing her vision of correlated music and light. The Historical Society of Pennsylvania's *Mary Elizabeth Hallock Greenewalt Papers* mentions her "initial experiments in 1905 with coloring photographic film," and says that in 1916, she gave a demonstration to the Illuminating Engineering Society of Philadelphia in which she showed an invention involving rotating rolls of film as the medium for compositions of colored light composed and synced to music.

Around the same time, Arnaldo Ginna and Bruno Corra experimented with both color organs and film. After they were disappointed by their initial experiments with a color keyboard, they began experimenting with film, creating films to pieces of orchestral music. They were excited by film's technical properties–the brightness of the light produced, and the phenomenon of persistence of vision. (Gage 1993)

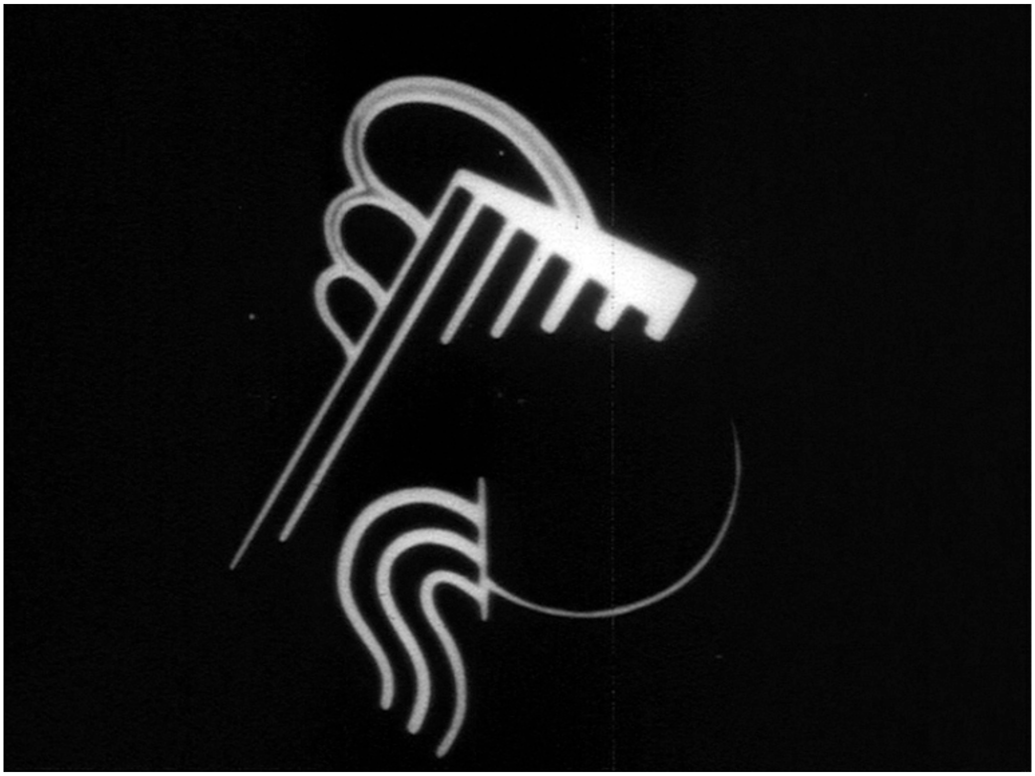

Figure 6. Still from Viking Eggeling's “Symphonie Diagonale”

There are no surviving copies of their films, and it’s unclear whether they were shown publicly, but they appear to be some of the earliest known abstract films, and also some of the earliest known instances of using film as a medium for synaesthetic music.

Films by Walter Ruttman, Viking Eggeling, and Hans Richter are the earliest surviving abstract films, and they all dealt with the idea of synaesthetic music. Some of these films were silent and some were shown with music, but even the silent films were directly inspired by musical principles in the way they explored temporal developments of abstract visual form. The titles of their pieces–for example, Ruttman’s “Lichtspiel Opus 1” (Moving Picture Opus 1) and Eggeling’s “Symphonie Diagonale”–often made this intention clear. (Schwierin and Naumann)

While these early abstract film works were sometimes accompanied by live music, the invention and commercialization of the optical film soundtrack in the 1920s allowed, for the first time, for film’s moving images to be perfectly synchronized with sound, in a unified audiovisual format (it

Figure 7. Oscar Fischinger with his optical film soundtrack

also preceded tape as the first format that allowed sound to be extensively edited by cutting and splicing).

Optical sound not only allowed for the composition of films with a perfect and consistent synchronization between sound and image–it also "facilitated for the first time direct inter-transformation of acoustic and optical signals." (Daniels and Naumann 2010) Sound and moving images both, in a sense, became data signals that, while originating in different sensory phenomena, and played back in different ways, could be recorded and edited on the same physical medium. Oskar Fischinger was among the first to experiment with using the medium to generate both sound and image from the same source:

> In ca. 1931, he realized that the optical soundtrack on film consists of abstract patterns. In the subsequent months, he systematically worked on the possibilities of producing sound by means of abstract drawings or ornaments on the film's soundtrack … In the films he produced, sound and image at least theoretically comprise a unit; one hears the same ornaments coming from the soundtrack as one sees on the image track. (Daniels and Naumann 2010)

The way optical soundtracks allowed for the generation of sound–as opposed to the recording and playback of sound–is also notable in the history of music technology as one of the first examples of sound synthesis.

Another notable abstract film work with an optical soundtrack which used a similar technique is Norman McClaren's 1940 film "Dots" (McLaren 1940). McLaren's work with optical sound, does not, like Fischinger's, use the same visual pattern as the source of both the film image and the soundtrack; rather, it appears that the relationship between sound and image, while synaesthetic and tightly synchronized, is constructed in a more intuitive manner. While both the sound and image in "Dots" were created using a manual direct-on-film technique (a pen) McClaren went on to develop advanced techniques for "animated sound," which he describes in "Notes on Animated Sound." (McLaren and Jordan 1953)

Len Lye and Mary Ellen Bute were two other early abstract filmmakers who used direct-on-film techniques and whose work involved a synaesthetic correlation with music. They both composed abstract films meticulously synchronized to popular music by scratching, drawing, and painting on film.

Brothers John and James Whitney were motivated by a desire to create synaesthetic music. With their *Five Film Exercises* (1943-1944),

> They wanted to create an 'audiovisual music' by not only laying out a comprehensive structure based on fundamental musical forms, but also translating these into image and sound by means of comparable production processes. For this purpose they developed revolutionary techniques for the generation of sound and images. They shot direct light for the first time, which they modulated with the aid of stencils. Based on a limited set of geometric forms, they thereby produced serial permutations. For the production of the sound they constructed an instrument that consisted of a series of individually controllable pendulums that could record the oscillations directly onto the soundtrack. They succeeded in this way in precisely controlling synthetic sounds and assembling them into more complex oscillation patterns, thus not only creating the equivalent to

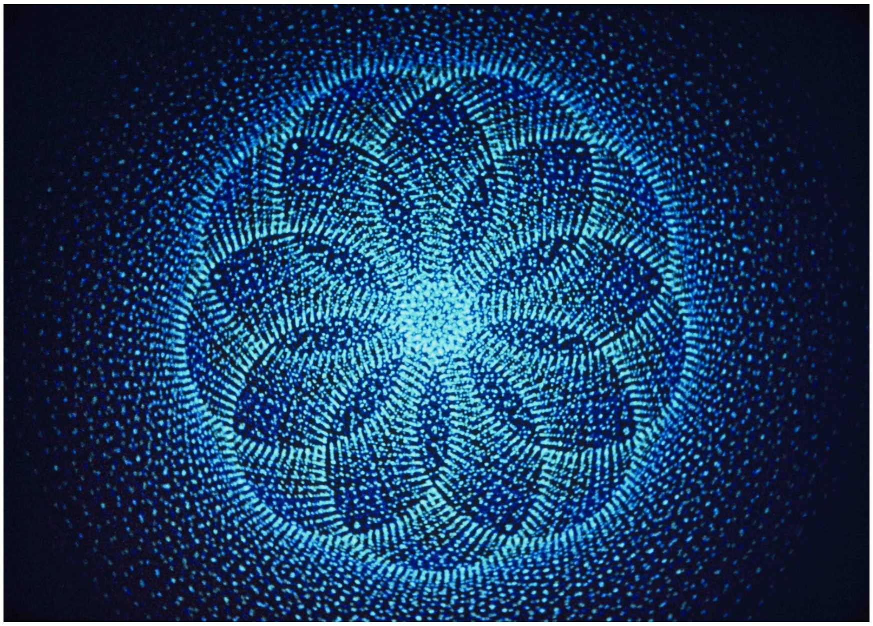

Figure 8. Still from John Whitney's "Lapis"

> visual design, but also anticipating developments in electronic music. (Schwierin and Naumann)

John and James Whitney went on to become pioneers in the use of early computer technology in film.

Other abstract filmmakers notable for either synaesthetic correlations between music and image or for silent films that acted as purely visual music include Harry Smith, Stan Brakhage, Jordan Belson, and Stan Vanderbeek.

I explored abstract film techniques at Bard College under Jennifer Reeves, and at CalArts under Kirsten Winter and Charlotte Pryce. Such techniques are an expressive way to create abstract, musical moving images by hand, and they were an inspiration for some of the works discussed in this thesis. I learned from exploring these techniques that streams of fast visual changes can create a sense of texture in time that feels like sound. When the images change every frame, without visual registration between frames, the experience becomes less about shapes and objects, and more about other qualities of vision like color, brightness, and texture. This abstraction of the visual experience can aid in creating synaesthetic connections with music.

In *Caruso's Dream*, the lights–driven by pops, clicks, and hiss in the recordings–evoked the flickering quality of film, and of direct-on-film techniques in particular. The particle systems used in *Emulsion Juice* and *Audiovisual Ouroboros* could be configured in a way that evoked painting on film, with its lack of visual registration–but continuity of other elements like color palette and texture–between frames.

2.4 Sound visualization techniques before video and computers

Optical sound (which allowed sound to be recorded as images on film), television and analog video (which allowed moving images to be represented as a fluctuating electrical signal, just as sound could–discussed in section 2.5), and the advent of computers and digital media (which allow both sound and image to be represented as digital data–discussed in section 2.6) might be considered the three most important technological developments that expanded the technical possibilities for correlating sound and image. However, several other developments in scientific research and technology from the Renaissance through the 20th century should be noted here, given how they allowed sound to be visualized in powerful new ways.

Cymatics

Cymatics (a term coined by Swedish scientist Hans Jenny in the 1960s) is the study of visible sound and vibration. It involves vibrating a surface with sand, a powder, or a liquid on it. According to Cymascope.com (citing *Encyclopedia of Religion Volume 4* by Mircea Eliade and *The History of Musical Instruments* by Curtis Sachs), this practice has very old roots:

> The provenance of Cymatics can be traced back at least 1000 years to African tribes who used the taut skin of drums sprinkled with small grains to divine future events. The drum is one of oldest known musical instruments and the effects of sand on a vibrating drumhead have probably been known for millennia. (Cymascope.com 2014)

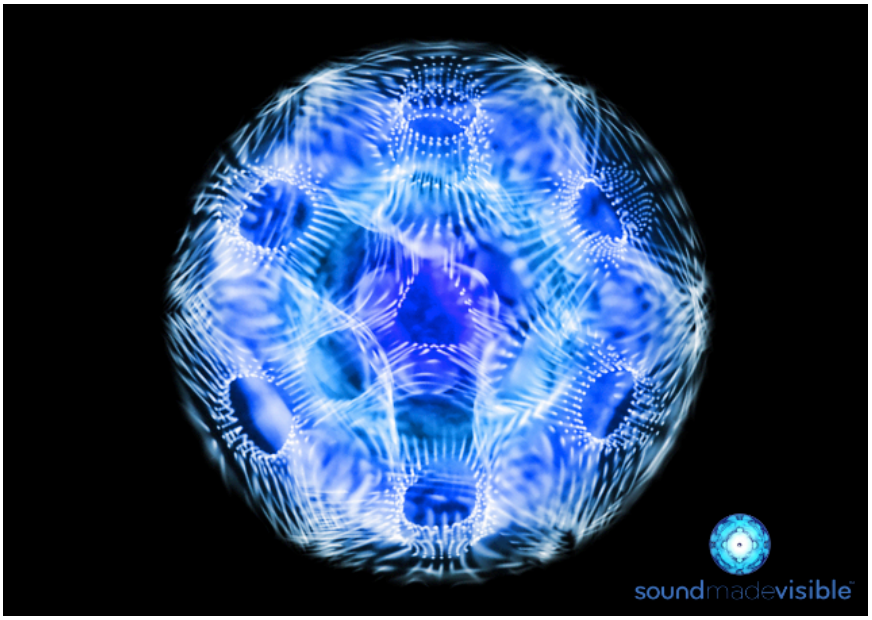

Figure 9. An image from John Stuart Reid's Cymascope

Leonardo da Vinci and Galileo Galilei both made notes that indicated they may have noticed the phenomena, but Robert Hooke, an English scientist, seems to be the first to have documented experiments with the phenomena: "Hooke devised a simple apparatus in 1680 consisting of a glass plate covered with flour that he 'played' with a violin bow." (Cymascope.com 2014) Hooke found that when he bowed the glass plate, patterns emerged in the flour. Ernst Chladni, described by some as the father of acoustics, took these experiments further in the late 18th century. (Rees 2009) His "Chladni plates" were metal plates that, like Hooke's glass plate, could be made to vibrate with a bow, generating patterns in sand on these plates. Chladni plates resonate at particular resonant frequencies (depending on the size and shape of the plate), and the patterns in the sand arise from how sound vibrations take place across the surface of the plate: sand is pushed off the portions of the plate that are vibrating, and settles in the nodes of the sound waves. Higher frequencies create more complex and intricate patterns (smaller wavelengths will create more nodes, and less distance between them, across a given plate). Others continued this kind of research with vibrating plates and membranes, including Michael Faraday, Lord Rayleigh, Mary Desiree Waller, and Margaret Watts-Hughes.

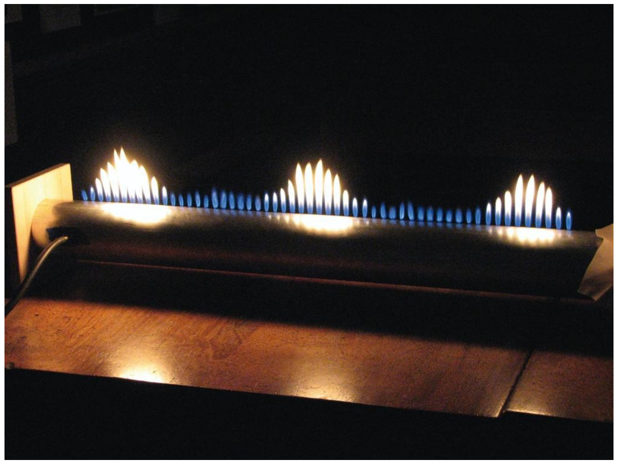

Figure 10. A Rubens' Tube

In the 1960s, Hans Jenny studied the phenomenon of visible sound intensely, coining the term cymatics. He expanded on the work of Chladni and others, using electrical devices and speakers to vibrate different kinds of membranes, with various materials on them. "His two volumes [*Kymatic* (Cymatics) Volumes 1 and 2] are rich sources of cymatic imagery, which he observed and described in great detail, although leaving scientific and mathematical explanations to scientists who would come after him." (Cymascope.com 2014)

Others since Jenny have continued experiments and research with cymatics. The Cymascope, created by John Stuart Reid, is a device that produces richly detailed, highly complex images by vibrating "ultra pure water" with sound.

Another notable experimental device for visualizing sound (somewhat related to cymatics) is the Rubens' Tube, created by Heinrich Rubens in 1904, which visualizes sound in a row of flames. It involves a long tube, oriented parallel with the ground, with evenly spaced holes drilled in the top, filled with a flammable gas. Both ends of the tube are capped, one of them by a speaker which plays sound into the tube. The openings at the top of the tube are lit with a match, and as the sound causes pressure in different parts of the tube to vary, the height of each flame of burning gas changes, visualizing the waveforms.

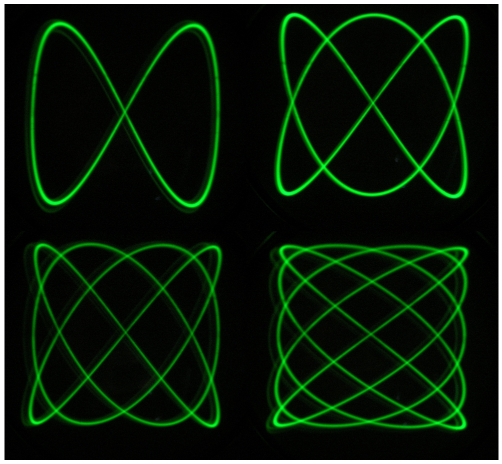

Figure 11. Lissajous curves on an oscilloscope

Lissajous figures

In 1905, Jules Antoine Lissajous conducted an experiment in which he reflected beams of light off of two mirrors in sequence, with each mirror attached to vibrating tuning forks. These tuning forks were positioned perpendicular to each other, in such a way that the ratio between the frequencies of the tuning forks was visualized as a geometric figure with corresponding harmonic proportions.

These figures became known as Lissajous figures or Lissajous curves. Lissajous curves can also be graphed on an oscilloscope by inputting sine waves of differing frequencies on each axis. A harmonograph is a similar device, inspired by Lissajous figures and producing similar images, but driven by the kinetic motion of pendulums.

2.5 Television and video

Television and video technology allowed moving images to be captured, broadcast, transmitted, and recorded as an electrical signal, making it even easier to correlate sound and image, and transform them into each other. Sound and image were transmitted and recorded on the same medium, and using the same kind of signal (a fluctuating electrical signal). It became technically

Figure 12. Charlotte Moorman playing Nam Jun Paik's *TV Cello*

much easier to transform sound into an image, or vice versa, and early video art pioneers experimented with this. (Spielmann)

In the late 1960s and 70s, artists and engineers like Nam June Paik, Shuya Abe, Steina and Woody Vasulka, Skip Sweeney, Steve Rutt, Bill Etra, and Stephen Beck began experiments with video signals that explored various technical means (beginning with video feedback) of distorting them, generating them, and translating them from sound to image or vice versa. Many of these early works achieve a striking degree of correlation between sound and image by taking advantage of the technical interoperability of the sound and image signals.

Spielmann writes that some of these artists were musicians who wanted to use video in musical ways: "In particular the trained composer Nam June Paik and the trained violinist Steina Vasulka dealt in their videos with issues related to structural correspondences between music and video and considered video an extension of their musical practice." (Spielmann)

Analog video synthesizers grew out of these experiments, and some of them could be connected to modular audio synthesizer systems. These video synthesizers created a whole range of new visual effects, and made it relatively easy to link them closely to sound in a variety of ways.

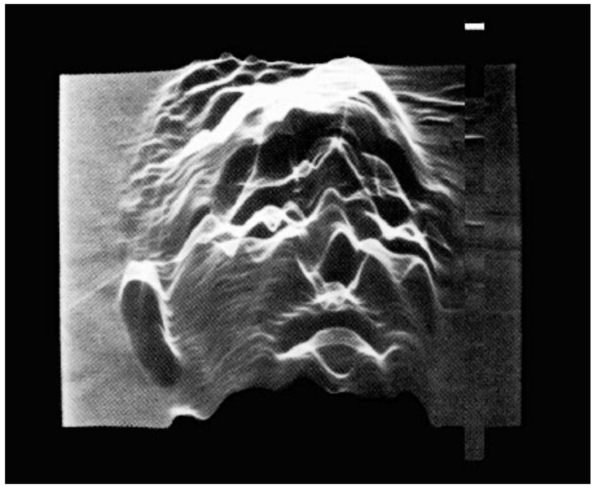

Figure 13. A still from the Rutt-Etra video synthesizer

Spielmann writes that as digital computers began to become more widespread in the late 70s and early 80s, there was a hybrid use of analog and digital technology in video: “Overall, the methods and concepts of analog video processing were continued with increased complexity with the digital programmability and greater storage capacity of the digital computer.” (Spielmann) This parallels the development of sound synthesizers during the early to mid 1980s, when many generated and shaped sound via analog electronics, but under digital control, bringing the advantages of patch memory and MIDI.

2.6 Computers

Similarly to how analog sound and video technology came under digital control, some film technology, while still using optical technology as the recording and playback medium, used digital computers to control the optical equipment–optical printers used in both commercial and experimental film were sometimes computer controlled, which allowed for processes to be automated in new ways.

However, the real impact of digital computers would come when digital technology began to be used to create and manipulate audio and visual media directly in digital formats. There were experiments in using computers to digitally generate imagery (John Whitney was a pioneer in

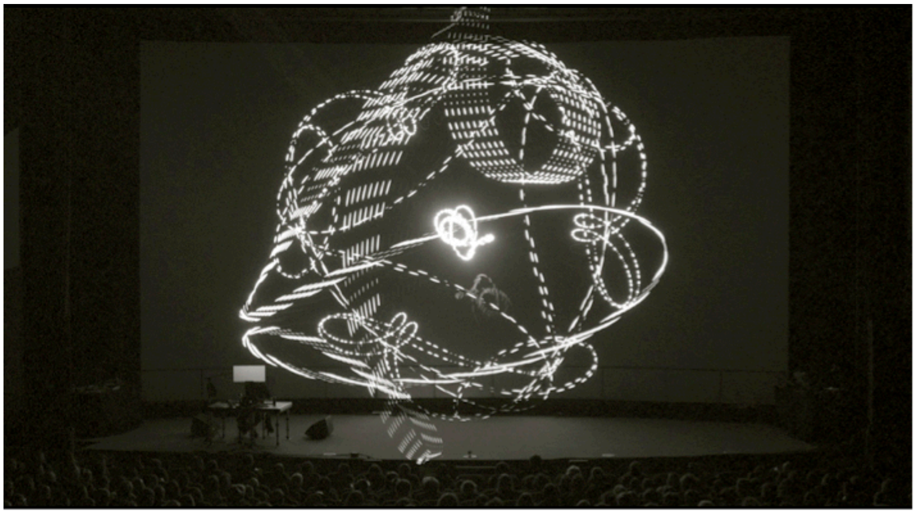

Figure 14. Robert Henke's *Lumière*

this) as early as the 1960s, and some artists had access to supercomputers that would allow for this. But it was not until the 80s and 90s that computers, including personal computers, began to be able to record and play back sounds and images with enough speed and fidelity to become a widespread format. Since then, the digital medium has made it possible for sounds and images–their recording, manipulation, transformation, generation, sequencing in time, and playback–to be correlated in increasingly diverse and powerful ways.

Daniels and Naumann write in *Audiovisuology Compendium*:

> Through digitalization, electronics integrate all current media formats. All the devices that once led separate lives in photography, film, video, radio, television, and audiotape now run as emulations in the universal machine of the computer, so that audiovisuality does not have to generated by the combination of separate media, but is implicitly and explicitly already given. (Daniels and Naumann 2010)

Appendix A describes several categories of techniques for correlating sound and image, with a focus on contemporary techniques using contemporary technology. Appendix B surveys contemporary technology (from the digital age) for correlating sound and image. I believe we are still in the early stages of discovering and inventing computer techniques and technology that create synaesthetic connections between sound and image. Computers and digital media allow for powerful translation and correlation between sound and image, but they don't achieve it on their own; synaesthetic techniques and technologies need to be designed and shaped by creative vision. Such techniques are a major focus of the artistic work and software presented in this thesis.

Figure 15. Brian Goggin and Dorka Keehn's Caruso's Dream

Chapter 3
And My Room Still Rocks Like a Boat on the Sea (Caruso's Dream)

And My Room Still Rocks Like a Boat on the Sea (Caruso's Dream) is a permanent public art installation in downtown San Francisco by artists Brian Goggin and Dorka Keehn. The piece consists of thirteen slightly larger-than-life piano sculptures constructed of glass and steel, suspended from the side of a new residential high-rise, above the sidewalk. Each of the piano sculptures has a cluster of LED lights inside. I was commissioned by the artists to create lighting sequences for the sculpture in sync with Enrico Caruso's original recordings, and to create a system to control the lights inside the pianos.

3.1 Background

Caruso's Dream pays homage to Enrico Caruso, one of history's greatest and most widely known opera singers, and an early adopter of recording technology. On the first night Caruso spent in San Francisco, the great earthquake of 1906 struck as he slept in the Palace Hotel (after which he vowed never to return). He awoke to experience his room rocking "like a boat on the sea," witnessed from his window the surreal sight of the city crumbling around him, and at first assumed that he was dreaming. *Caruso's Dream* imagines his experience that night. The pianos appear to be tumbling or flying through the air as if time has been frozen in the moments after the earthquake. The materials used in the sculpture evoke the industrial economy of the early 1900s that shaped the sights Caruso saw on his trip to San Francisco: the pianos are constructed of recycled factory window glass, their steel is painted a rust color, and they appear to be held up by wooden trusses and suspended by hemp ropes. The sculpture was draped in ship sails before its unveiling. While the lights inside the sculpture are clusters of LEDs, they emit a warm white light; the LED clusters (a straight line of LEDs inside each piano) are diffused by the glass, but still visible, lending each piano the resemblance of a giant incandescent bulb.

The technical and formal parameters of my work on the piece were as follows: the light clusters inside each piano would be treated as one unit (for a total of 13 independently controllable lights); the lighting sequences should have their basis in Caruso's original recordings (which are locally broadcasted over FM radio), but should also work without music (the way most people will experience the sculpture most of the time); and the lighting sequences would play from dusk until dawn each day.

3.1.1 Synaesthesia in a work of multimedia public art

Goggin and Keehn's vision for the light compositions was very in line with my own ideas about audiovisual musical works of art: the aim was toward an abstract, synaesthetic relationship between the music and the lights. The music was fixed and determined (pre-existing recordings), so a primary goal was for the lights to represent and embody the music.

The parameters of the piece presented some aesthetic questions and challenges. First, the sculpture itself, as a medium for the light sequences, is fairly simple and "low-resolution": there are only 13 individually controllable units (the pianos), their position is fixed, and with just one color of light, each of these can only change and be controlled in one dimension (the brightness of the light). Sound in general (like most sensory experience), and music in particular, tends to feel vastly complex and textured, multi-dimensional, spacious, like it opens up world or a landscape for us. One of the challenges posed by this piece was how to create compelling, satisfying, and natural-feeling relationships between music and light, given the technically limited palette and canvas of lighting in the piece.

As is often the case, the limitation here helped focus the aesthetics of the lighting sequences. In some ways, the "lo-fi" nature of the medium for light in *Caruso's Dream* echoes the limitations in the nascent recording technologies that Caruso's recordings helped popularize. These recordings are technically primitive by all modern standards–they are monophonic, have a low signal-to-noise ratio, and a lot of artifacts from the physical medium–yet despite the limited amount of information they can provide to one's senses, and the limited clarity of that information, they still open up a whole world for the listener. They achieve this partly by packing in as much information as technically feasible: though the format (by modern standards) is monophonic, noisy, and has a low dynamic range, there is a lot of change in the signal over time, allowing for much of the richness of sound to be reproduced despite the limitations. Additionally, experienced from a modern perspective, the poor quality of the recordings adds another element–a sense of nostalgia, of peering back into the past and remembering an earlier time in history. In exploring the possible synaesthetic relationships, these qualities in the original medium and recordings were considered as clues for how to achieve a powerful visual interpretation of the music.

Spectral analysis was used to generate a control signal that could be mapped to changes in light intensity over time, which proved to be a powerful way of translating the dynamics and sonic character of the original recordings into the visual realm. When applied to lights, this signal created flickering patterns that reflected both the dynamics of recorded sound as well as the artifacts–hiss, pops, and clicks–from the recording medium. The flickering quality also evokes early film technology, connecting with the theme of Caruso's early embrace of media technology.

Sound changes quickly and subtly over time. Using audio analysis to track nuanced changes in volume–and echo them in the lights–created a natural and compelling relationship between sound and light, and also helped achieve an organic and rich visual experience despite the limited palette of visual parameters under my control. Recorded sound itself, in a basic technical sense, is essentially volume changing over time, but what arises from this perceptually is not just a sense of rhythm, but also pitch, timbre and every other aspect of sound. Light is physically different from sound. But the quickly fluctuating light patterns that result from mapping volume to brightness (with a high time resolution) begin to impart a sense not just of brightness changing over time, but of visual texture and pattern. When experienced with the sound that generated them, such light patterns feel naturally and intuitively linked to the sound, allowing sound and image to begin to blend into a less-differentiated perceptual experience. The degree to which this synaesthetic blending happens is directly linked to the degree of correlation between the changes in intensity of sound and light. Because sound changes quickly, therefore, an effective implementation of this technique will result in fast visual changes.

In the context of this project, since the only dimension of control for each light was over intensity of brightness, the fast changes in sound could not be represented in a variety of ways like with more complex visual media (changes in color, texture, shape, etc.). This presented a challenge: lights flickering quickly and intensely could be jarring and invasive as a piece of public art on the front of a residential building, especially without the accompanying sound to contextualize the lights and provide an audiovisual synaesthetic experience. And without hearing the music simultaneously, if the light patterns were too frenetic, they might not be a satisfying visual experience on their own, or feel like they reflect the mood and pace of the music they are based on. The lighting compositions needed to function as both a casual experience (for many drivers, passersby, and building residents) and an immersive experience (for those choosing to tune in); a visual experience as well as an audiovisual experience. The need to maintain a balance between close audiovisual correlation, standalone visual appeal, and suitability as public art became an important consideration.

3.2 Integrating aesthetics, composition process, and hardware implementation

Realizing the compositions required an integrated approach to developing both a compositional process and a technical implementation that could achieve the aesthetic goals of the piece within the technical requirements, and within a relatively small budget.

Because *Caruso's Dream* is a permanent installation, stability, longevity and minimal maintenance were important requirements of the technical system that would control the lights. For this reason, while computers were ultimately a central part of the composition process, they were ruled out early on as the final technical means of playing back lighting compositions.

I developed a software composition environment that could create fixed compositions for each song, and integrated a hardware system that could play back these compositions in sync with

Enrico Caruso's music. This allowed for manual composition while leaving open the possibility of incorporating audio analysis and custom composition tools.

The composition environment was structured around the use of MIDI data (specifically 14-bit pitch bend messages, to harness their increased resolution as compared to standard MIDI note and CC messages) that would eventually be output and recorded as DMX lighting data. Hardware was selected that could play back DMX lighting data in sync with chased SMPTE timecode. Using MIDI and music software to compose the light sequences made it relatively easy to integrate SMPTE timecode (which can be recorded and played back as audio when using balanced connections) into the composition environment.

The requirements of *Caruso's Dream* (lights synced to recorded music in a permanent installation), and the compositional process used to realize it were unique, interdisciplinary, and specific to the piece. This meant that it was not possible to simply pick from tools geared towards established workflows in a well-defined discipline. It also required there to be a conversation between the process of designing and integrating a hardware system for the final installation, and the process of experimenting with and iterating on different combinations of software tools for composition. These design processes necessarily informed each other and were developed side by side.

3.3 Techniques for composing light to music

This section discusses various techniques used in the creative process for the lighting sequences, and their technical dimensions.

3.3.1 Visualizing sequences using a virtual prototype

The pianos were still being constructed when I began my work, and were not installed on the building until the project was almost complete, so it was not possible to view the lighting sequences on the actual sculpture as I was creating them; thus, it was necessary to visualize the light sequences in software.

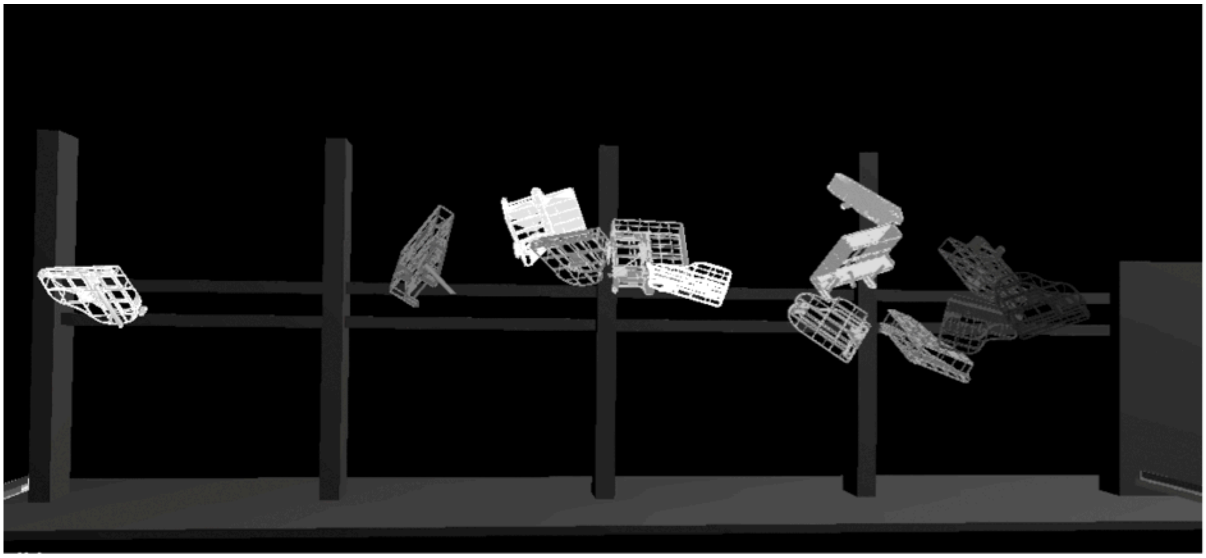

Figure 16. Visualizer developed in Quartz Composer by Joshua Sophrin

Collaborating artist and technologist Joshua Sophrin developed a visualizer in Quartz Composer[1] (shown in **Figure 16**). Sophrin used 3D models of the piano sculpture (created as part of Goggin and Keehn's design process, and to plan for the fabrication process), as the basis for the visualizer, importing them into Quartz Composer and applying virtual lighting with brightness individually controllable for each piano. The entire collection of pianos (along with a skeleton representing the façade of the building) could be rotated in 3D space and viewed from any angle, making it feasible to judge the aesthetics of the light compositions from various angles that a member of the public might experience (**Figure 17** shows the visualizer displaying a side view). I made a small modification to the patch to accept MIDI pitch bend as input on 13 separate channels (each controlling one piano).

The visualizer was a very accurate model in terms of the size and shape of different pieces of the sculpture, and their relationship to each other and to the building in space. Its accuracy in this regard proved invaluable, as it made it almost entirely unnecessary to imagine or guess at the way the composition of the light sequences would relate to the spatial composition of the sculpture.

[1] http://en.wikipedia.org/wiki/Quartz_Composer

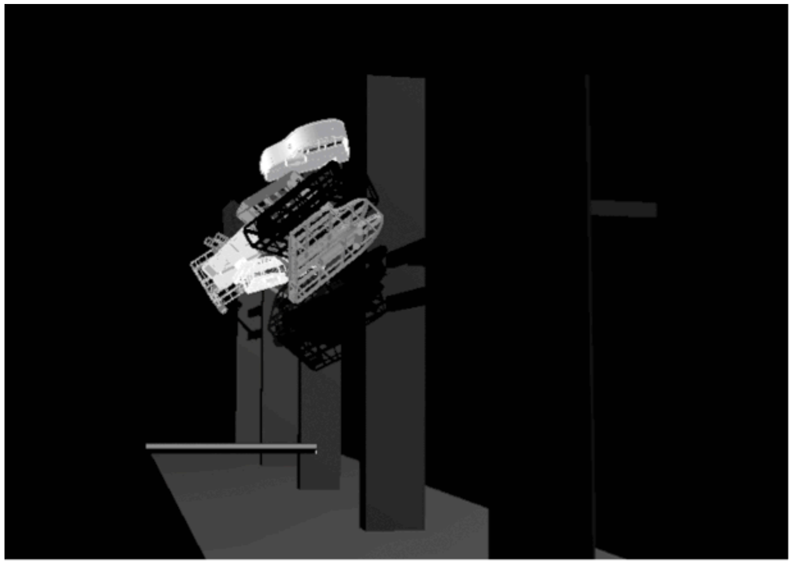

Figure 17. Quartz Composer visualizer displays the piece from a side angle

The visualizer also performed accurately in terms of its response to the lighting control data. Its response in this regard was not identical to the actual lights–we ultimately decided to subtly re-scale the lighting data through a logarithmic curve–but it was close enough for the visualizer to be a reasonably accurate reference of lighting levels for creating the compositions. Before seeing the completed sculpture, a sense of the similarity of response between the visualizer and the lights was achieved by testing parts of the lighting sequences on small test strips of the LEDs used in the sculpture (provided by Mike Hollibaugh, who constructed the LED clusters for the artists).

3.3.2 Audio analysis: generating a control signal as raw material

FFT Control is software I developed to generate control signals from spectral analysis of an audio signal (FFT Control is discussed in detail in Chapter 6). The first prototype of the software, which was used in *Caruso's Dream*, features an interface that visualizes the spectrum of incoming audio, and allows the user to graphically select a frequency range whose magnitude will be calculated, multiplied by a scaling factor, and output as a stream of control events via Open Sound Control (see **Figure 18**).

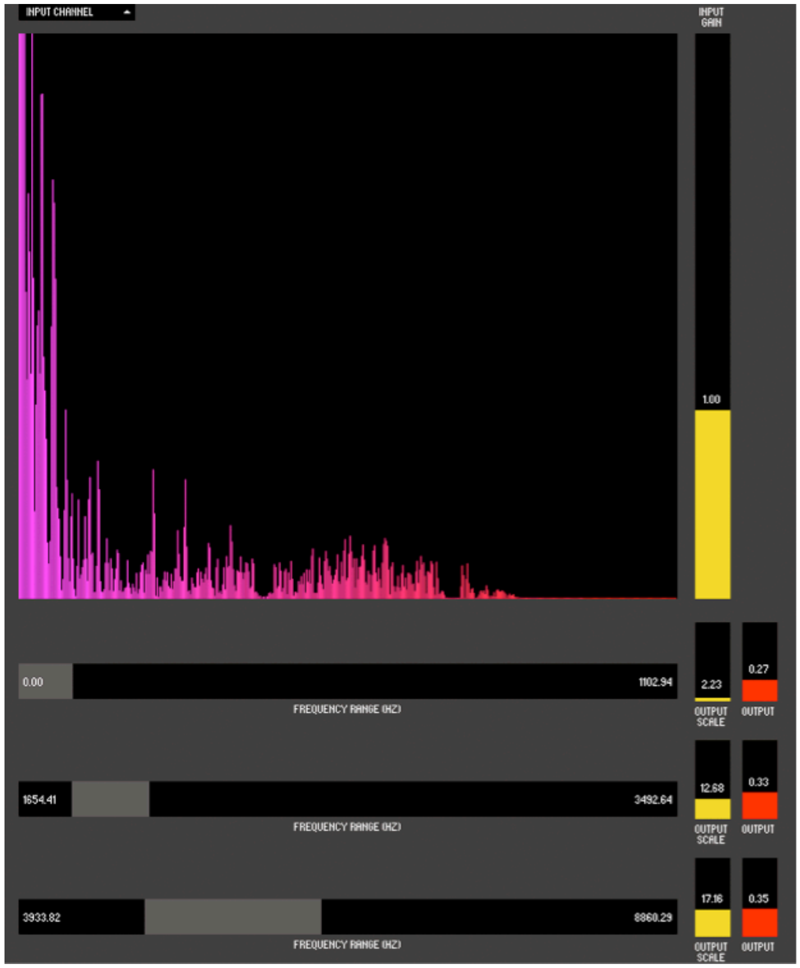

Figure 18. FFT Control software
(ChucK with a Processing sketch, communicating via OSC)

The OSC messages sent by FFT Control were received by Osculator[1] to be converted into MIDI for recording in Digital Performer.[2] Osculator can graph an incoming OSC signal, a useful feature for understanding and adjusting the parameters of the signal before recording it (see **Figure 19**).

[1] http://www.osculator.net/
[2] http://www.motu.com/products/software/dp

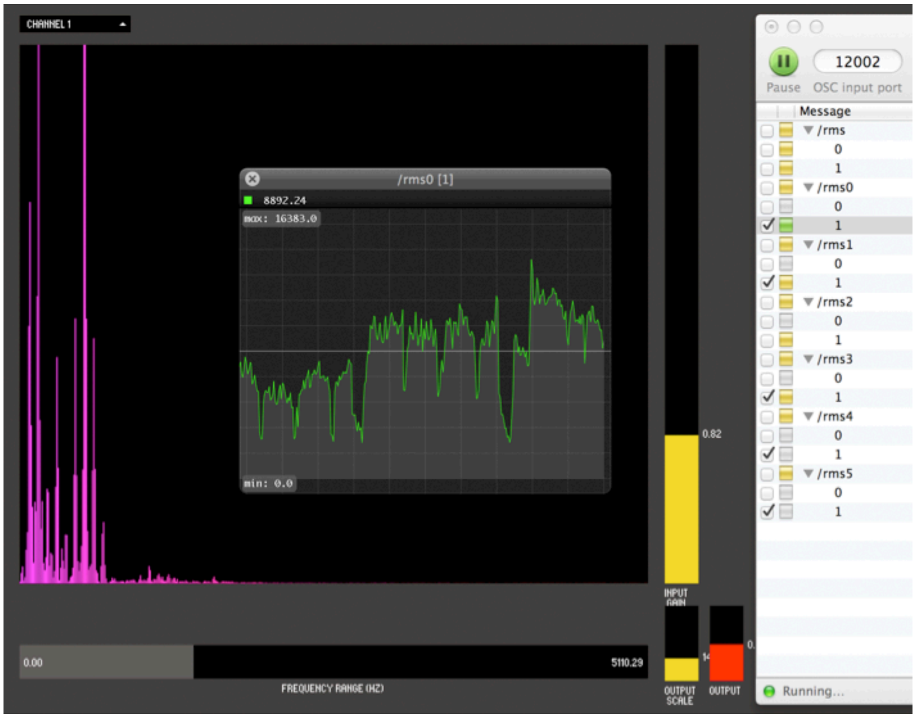

Figure 19. Osculator's "Quick Look" window floats above FFT Control, graphing the OSC signal being sent as a result of audio analysis

With this moving graph as a guide, the audio signal and the parameters of the audio analysis were shaped with the goal of generating a signal with the following characteristics:

- A high dynamic range. Just as with audio recordings, the loudest point in the song should be close to the highest possible value in the signal.

- Responding primarily to the voice and instruments in the music (as opposed to artifacts).

- Enough subtle fluctuation in the signal for it to feel sensitive and responsive without feeling excessively dynamic in comparison to the sound.

Audio compression was used to shape the dynamics of the audio signal and EQ was used to emphasize and de-emphasize certain parts of the spectrum of the audio signal entering FFT Control. Spectral analysis was focused on the frequency range of Caruso's voice (while making sure it also captured the sound of the instruments where desired).

Some techniques that varied between different songs:

- In many songs, I would manipulate input gain to FFT Control, and/or configure Osculator in a way such that the floor of the OSC signal was just above any signal produced by the noise floor of the original audio recordings. This meant that the noise of the recording itself–and with most of these recordings about a century old, the noise was significant–did not produce light; light would not be generated until a voice or instrument made sound above the threshold. In these cases, I was interested in maintaining a synaesthetic audiovisual relationship focused on the musical content. This became especially important in pieces where silence is used significantly for dramatic and rhythmic effect.

- In other songs, the threshold was set slightly below the noise floor in order to capture some of the noise of the recording as a baseline flicker in the lights. In these cases, the focus broadened a bit to also include a synaesthetic audiovisual relationship connected to the artifacts of the medium (which are, after all, a part of the perceptual experience of the sound, and which add to the emotional and psychological experience of the music).

- In one case, separate recordings of control signals for the vocals and instruments were used, focusing the parameters of the FFT analysis on each specifically.

After conditioning the signal in these ways, it was converted by Osculator to MIDI pitch bend messages and recorded in Digital Performer in sync with the audio from which it was generated.

These recorded signals became the basic raw material that would form the basis for all the light compositions. The result aimed to achieve a compelling synaesthetic relationship, facilitated by the richness of subtle variation in brightness, perfectly in sync with the sound, and which closely reflected the dynamics of the vocal performance. The result felt like it had movement and depth despite the fact that it essentially stayed still, in one place.

3.3.3 Sculpting the source: moving a control signal with a musical interface

Another approach employed was to dynamically move this signal around the composition, spreading it out or focusing it into a small area. This was facilitated using an application written in Max[1] that takes one channel of incoming pitch bend data on a virtual MIDI bus, and scales and routes that data to each of thirteen output channels, according to two parameters: position and spread (see **Figure 20**). The basic idea was to have a controllable position that would be the "center" of where the signal (or in aesthetic terms, the visual representation of the sound) is in the composition. The signal would be spread from this center point, "fading out" the further it was from the center; the spread parameter would control how far the signal would spread before fading out completely. This application was dubbed the "panner" because its positioning behavior, driven in real-time by a knob or fader, was analogous to the way in which one might pan an audio signal on a mixing console or in a DAW.

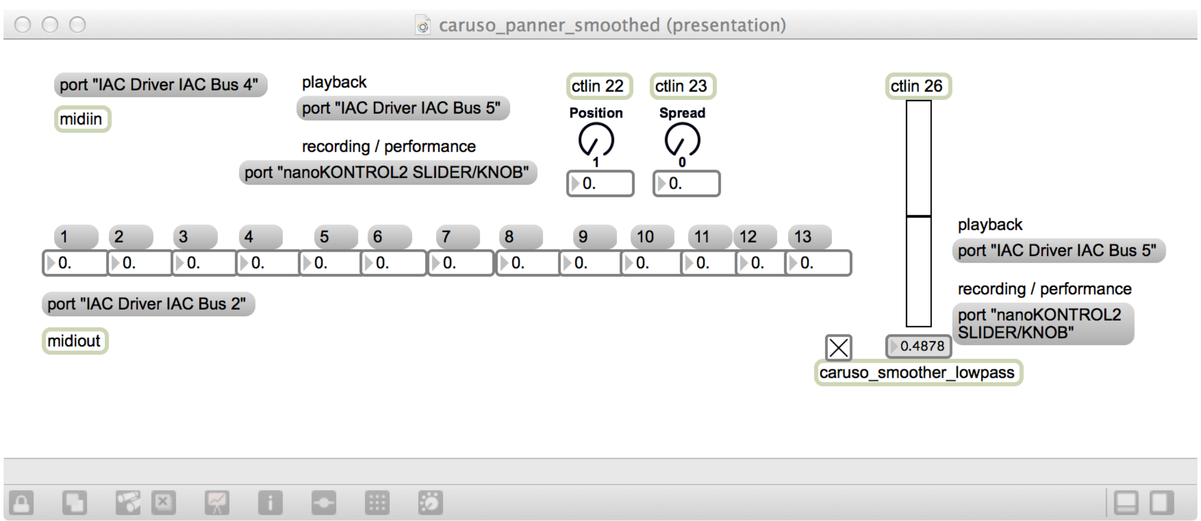

Figure 20. Max instrument for spreading and moving light through sculpture

This technique enabled the representation of the music generated by audio analysis to become almost like an image or physical object which extended across multiple pianos, moving through them, with a center, but with its boundaries not necessarily defined by the boundaries of individual lights and pianos.

[1] https://cycling74.com/products/max/

The position and spread parameters were manipulated by MIDI CC messages sent from knobs on a MIDI controller, allowing for tactile, real-time control over the how the light moved, spread, grew, and shrank over the course of the light compositions. This allowed for a new dimension of dynamics and motion in the light compositions driven by performance of control input. A MIDI controller–a tool designed for a musical context, and familiar in that context as a way of manipulating parameters of effects and synthesizers in real-time–could now be used. This made it possible to begin composing in a more intuitive, musical way; results could now be driven not only by timeline-based editing but also by performance. This allowed me to utilize my skills as a musician: the ability to listen to, memorize, and follow along with a performance; the ability to anticipate and react to musical changes; and the ability to perform physical actions with my hands, with musical timing.

Performing the position and spread parameters with physical controls (and recording and editing those performances, as is common in electronic music production) would become the primary composition technique for the *Caruso's Dream* lighting sequences.

3.3.4 Envelope generator driven by MIDI notes

A way to generate short sequences of control data in response to MIDI notes was desired, so that individual instrument hits and stabs could be represented. Essentially, what was required was an envelope generator, which, rather than shaping an audio signal (the traditional role of envelope generators in synthesizers and samplers), would output a control signal when triggered. As pictured in **Figure 21**, a velocity-sensitive ADSR envelope generator was developed in Max which responded to MIDI notes, choosing its output MIDI channel (and hence which piano sculpture it "triggered") based on the incoming MIDI note. This made it possible to generate envelopes for light levels for each of the piece's 13 piano sculptures using an octave (plus one) of MIDI notes.

Similarly to the "panner," the envelope generator was a tool that facilitated crafting the light compositions using musical tools and processes. It was now possible to listen to a recording, and represent certain instrumental notes in the song by playing along with those notes on a MIDI keyboard. In this case, the process was entirely compositional (manually replicating musical structures in the visual realm, alongside the audio), versus being partly driven by a mapping/visualization technique. However, because timing was directly driven by real-time human input, a greater number of "takes" and more editing were required in order to achieve rhythmically accurate and satisfying results.

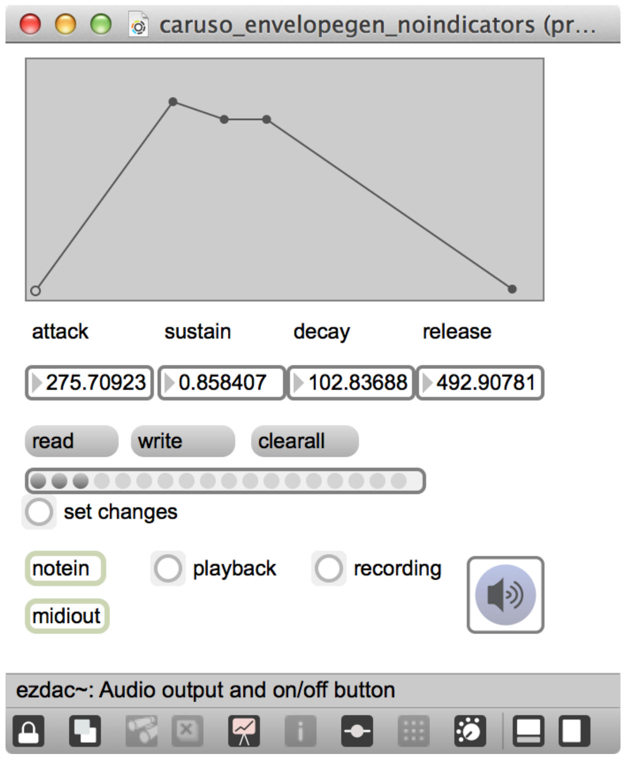

Figure 21. MIDI-controlled envelope generator

The decisions about which piano sculptures would represent which notes were made intuitively, not analytically. In general, I aimed to convey the feeling of melodic and harmonic movement in the musical elements I was representing. This often translated to a low/left to high/right correlation between musical pitches and spatial position, but this relationship was not, and could not be strict: while there were 13 piano sculptures, the musical content often moved across more than one octave, making it impossible to preserve a low-to-high/left-to-right mapping and also maintain a one-to-one relationship between musical pitches and piano sculptures. So while such relationships were an aesthetic guide and starting point, the decisions were made intuitively. Because of the musical processes used in constructing these relationships, the "feel" of things was explored in a musical fashion, trying out different ideas in real time and seeing how they felt.

Such intuitive, improvisational jamming is essential to many music-making processes, and I believe it also has an important role in crafting synaesthetic musical experiences.

3.3.5 Smoothing

As described earlier, while the audio analysis techniques used resulted in rich, fluctuating light patterns which were synaesthetically compelling, the technique also presented a challenge: the quick changes in light intensity, while essential to the synaesthetic effect, could be visually jarring and intense if too extreme and omnipresent. While the flickering sometimes felt soft and subtle, like a flame, at times it also felt overly dynamic in comparison to the mood of the music, and visually fatiguing. This was especially the case when the light compositions were viewed without music: with the relationship between sound and image no longer apparent, fast visual changes could sometimes feel faster, and more intense.

To address this problem, smoothing techniques were used on the control signal generated by the audio analysis, as a way of subtly dialing back the degree of change over time while still preserving much of the nuance and subtlety in the signal. A low-pass filter smoothing method (in Tristan Jehan's "smoother" external[1] for Max) was used, connecting the degree of smoothing to a MIDI CC input so that it could be dynamically adjusted in real-time. This allowed for a very wide and continuously variable range of smoothing intensity, making it easy to experiment with extremes (useful, as in audio mixing, for gaining a clear understanding of the effect applied), and hone in on a more subtle setting.

While the smoothing capability was developed in order to "rein in" a few songs with too much flickering, it was eventually applied to almost every light composition (sometimes very subtly). It worked well to reduce the intensity of quick changes in brightness where necessary, while preserving the changes and their relationship to the sound from which they arose.

Final decisions about what degree of smoothing to use were made while viewing the light compositions on the sculpture itself. The brightness of the lights in the sculpture, degree of

[1] http://web.media.mit.edu/~tristan/maxmsp.html

ambient light (which affected the perceived contrast ratio between light and darkness in the sculpture), diffusion through the glass, and reflection off of nearby surfaces were all important factors in the perceived intensity of flickering–the visualizer alone would not have been a sufficient basis on which to make final decisions about smoothing (or the lack thereof).

3.4 Reflections

Crafting the light compositions for *Caruso's Dream* required an experimental process in which technological and aesthetic experimentation were linked and integrated. This experimentation and innovation had to be balanced with a deadline, and the need for a final system that would be reliable and function without intervention, for years. The technical and creative aspects of the project informed and inspired each other, which was an enjoyable and fruitful way to work. The experimental nature of the techniques used in the creative process, and of the final hardware implementation, meant that problem solving and troubleshooting were necessary to get everything working properly together. The final system was successful in supporting the creative vision, and has proved reliable.

3.4.1 Process

Given the current state of technology for such interdisciplinary works, technical and aesthetic experimentation is often a very important component of the creation process. This type of experimentation is necessary in order to lay down foundations (both technical and aesthetic) for future work. As interdisciplinary work that involves technology continues to grow and evolve, tools will evolve to support it. Prioritizing integrated technical and aesthetic experimentation in interdisciplinary work will help inform the development of these tools (with the tools, in turn, shaping the aesthetic possibilities and tendencies of future generations of work).

3.4.2 Technical

MIDI was not designed to control lights, but using it as a proxy for lighting control data allowed the use of musical tools, and for the composition of light sequences to take place in a musical context. This made perfect sense given the nature of the project: representing music in lights. Other tools are better suited to working with lighting in more traditional ways–designing lighting for a scene in theater or film, or lighting a band on a stage, for example. In *Caruso's Dream*, however, light was the material itself (as opposed to playing more of a supporting role for other visual and experiential content), and the goal was for it to represent music and extend it into the visual realm. Though it presented technical challenges, using tools designed for musical composition facilitated a creative process that was well suited to the aesthetic goals. Incorporating expressive, performative, musical control into the visual composition process was a good idea, and worked well.

3.4.3 Aesthetics

The aesthetic approach taken was aimed at representing the central instrument of the music–the voice–as the most prominent figure in the light compositions. Caruso's voice was consistently the loudest and most prominent element of the music. This made it relatively straightforward for audio analysis to closely track fluctuations in his voice without interference from other elements of the music. Because Caruso's voice is the musical and sonic focus of the recordings, it made synaesthetic sense for it to be the most prominent visual element.

With the soloist's voice prioritized as the focal point of opera music, a great deal of the expressiveness in the music is up to the soloist. Caruso's voice in particular is considered by many to be incredibly expressive with its very high dynamic range. By achieving an accurate and nuanced tracking of the dynamics in Caruso's voice, the light compositions captured much of the musical expression in the recordings.

Opera is a musical form in which emotional intensity is often linked to volume. This made the music used in this case especially well suited to the type of mapping used between sound and image (volume to brightness). The most intense and dramatic parts of a song were usually the loudest, so mapping volume to brightness naturally resulted in the compositions feeling their fullest, most active, and most intense at the musical climax of each song.

Caruso's Dream was unveiled on February 23rd, 2014 in a public art event attended by thousands of people, which incorporated live musical performances, acrobats, and the first public showing of the light compositions. The installation is on permanent display at 55 Ninth Street, San Francisco, and the light compositions can be seen every night from 4pm to 9:30am, in sync with Caruso's music, by tuning in to 90.9 FM.

Chapter 4
Synaesthetic installations & performances

To most of us music suggests definite mental images of form and color. The picture you are about to see is a novel scientific experiment–its object is to convey these mental images in visual form.

–Oskar Fischinger in "An Optical Poem"

The installations and performances documented below explore synaesthetic audiovisual relationships in a musical context. Work on these projects is part of what motivated the development of software documented in Chapter 5 and Chapter 6. These projects used early prototypes of the software and informed its further development.

4.1 Emulsion Juice #1 (Digital Arts Expo 2013)

Emulsion Juice #1 (pictured in **Figure 22**) is an interactive audiovisual installation that explored the use of real-time spectral audio analysis to affect a generative visual particle system. The first full working prototype of FFT Control (discussed in Chapter 6) was created for use in this installation. Using this software, values representing the energy in three selected frequency ranges were sent over OSC as a control signal to a particle system created in Processing.[1] In this manner, the frequency content of the audio was mapped to color in the particle system: low frequencies controlled the amount of red in the particles, midrange frequencies controlled the amount of green, and treble frequencies controlled the amount of blue.

[1] https://processing.org/

Figure 22. *Emulsion Juice* on display at Digital Arts Expo 2013, CalArts

Because the spectral content of the audio dictated the color of particles, pitch and timbre both became correlated with color. Because the overall brightness of the particles was determined by the intensity of each of these red, green, and blue values, this also resulted in a correlation between volume and brightness. FFT Control enabled easy adjustment of the audio analysis parameters, in order to achieve a good balance of different colors, and a satisfyingly rich relationship between color and sound. When properly adjusted, the system produced a unique palette of colors for each different sound it visualized, with the same sound producing the same color palette, and similar sounds producing similar color palettes. The particles left trails as they moved, allowing the recent history of color in the particles to mix with the current color. This frequency-to-color mapping, although based on only a few data points about the sound, resulted in a surprising amount of variation and nuance in the visualization of sound.

Visitors were invited to make sounds into microphones using provided percussion instruments, and discover a synaesthetic experience as the sound generated colorful patterns on a large screen

Figure 23. David Howe's light balloons in *Ces Balons De Lumière*

in front of them. The immediacy of the sound visualization and the high degree of correlation between timbre, pitch and color illustrated the concept of synaesthetic sound in an intuitive and interactive way. The use of percussive musical instruments was an invitation for participants to play rhythms, and those who did discovered that it created periodic patterns in the particle system on the screen.

4.2 Ces Balons de Lumière

Ces Balons de Lumière is a piece made in collaboration with David Howe and Raphael Arar which visualized sound in colored light balloons. Howe's light balloons were controlled via live audio input from Arar's modular synthesizer, using FFT Control (software described in Chapter 6). A similar mapping of frequency to color as that described for *Emulsion Juice* was employed.

This also resulted in a close synaesthetic relationship between sound and light. A modular synthesizer was a good source of sound to show the capabilities of the visualization system, because of its fast changes and wide range of spectral content. The balloons were filled with helium and strung across the concert hall 10-15 feet high. In their audio-reactive mode, all of the balloon lights changed color together. The simplicity of the visual representation of sound helped focus attention on the relationship between sound and image. Arar's synthesizer playing created tones that swept smoothly across a wide range of frequencies, which helped reveal to viewers the mapping between frequency content and color. At other moments, the sound pulsed with loud clicks and pops, quickly alternating between silence and a high volume–this made clear the link between volume and brightness, and the precise temporal relationship between sound and image that arises when visual content is driven directly by real-time audio analysis. The audiovisual relationships created were obvious and easily observable because of the totality of the mapping between sound and image: all aspects of the visual content displayed by the balloons (which were few: color, brightness, and time) were directly determined by the audio analysis.

The visual content of the piece was much less complex than what might be displayed on a screen–it involved only a single color changing over time (in contrast to the millions of pixels in a high definition screen or projection). But the piece was in some ways more visually immersive than pieces that encourage the audience to focus their attention on a screen within a frame. Because the visual medium in the piece was a string of floating balloons, lit from inside, it had a physical, three-dimensional presence. Light from the piece illuminated the concert hall, surrounding audience members, reflecting off of surfaces, interacting with the space. As a result, light was on more of an equal footing with sound. Light became closer to something that used the physical space as a medium, the way sound exists as vibrations in the air. This enhanced the synaesthetic effect of the piece, making it easier to perceive sound and light as physically or ontologically linked (without having to first suspend disbelief and enter the imaginary world of the screen).

4.3 Audiovisual Ouroboros #1 (Digital Arts Expo 2014)

This installation expands on the concepts explored in *Emulsion Juice #1*. The key difference is that the project involves sonification of a visual system in addition to visualization of sound, and connects these processes together into an audiovisual feedback loop. The title references the ancient symbol of the Ouroboros, a serpent or dragon eating its own tail.

The format of the installation was as follows. A microphone on a stand was positioned in front of a screen onto which moving images were projected. Nearby speakers were pointed towards

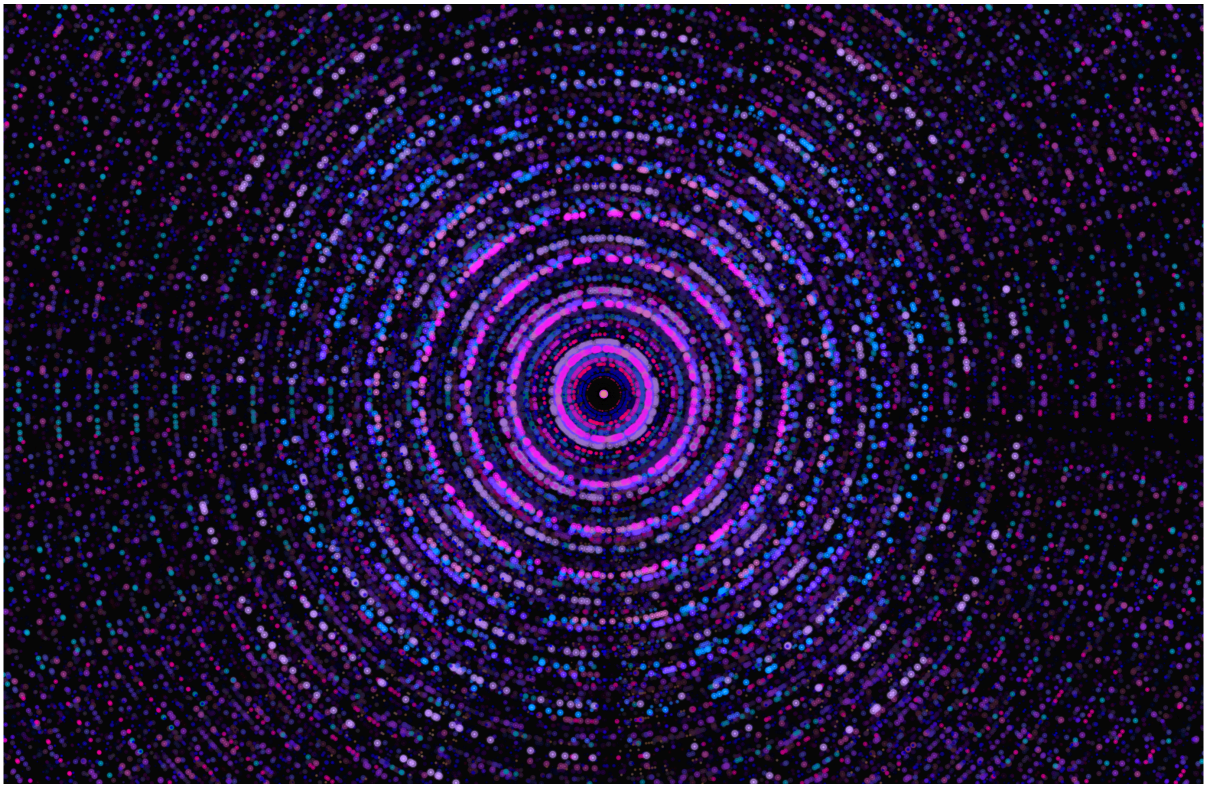

Figure 24. A still from *Audiovisual Ouroboros*

the microphone and whoever was standing in front of it. The speakers played back the sound component of the piece for the audience, while also pushing sound into the microphone. As a result, the microphone captured sound from two sources: sounds made by participants into the microphone, and sounds played back from the speakers. The sound entering the microphone was analyzed in real time by custom audio analysis software (FFT Control). Real-time control data derived from this analysis was sent to a particle system written in openFrameworks,[1] controlling various aspects of its motion, color, and composition (the mapping used was similar to that used in *Emulsion Juice* and *Ces Balons de Lumière*, although with a different particle system). The result of this process was that abstract, colorful moving images on the screen were driven by sounds entering the microphone–the relationship was direct, immediate, and obvious, yet subtle and varied. Different sounds triggered different color palettes based on their spectral content.

[1] http://openframeworks.cc/

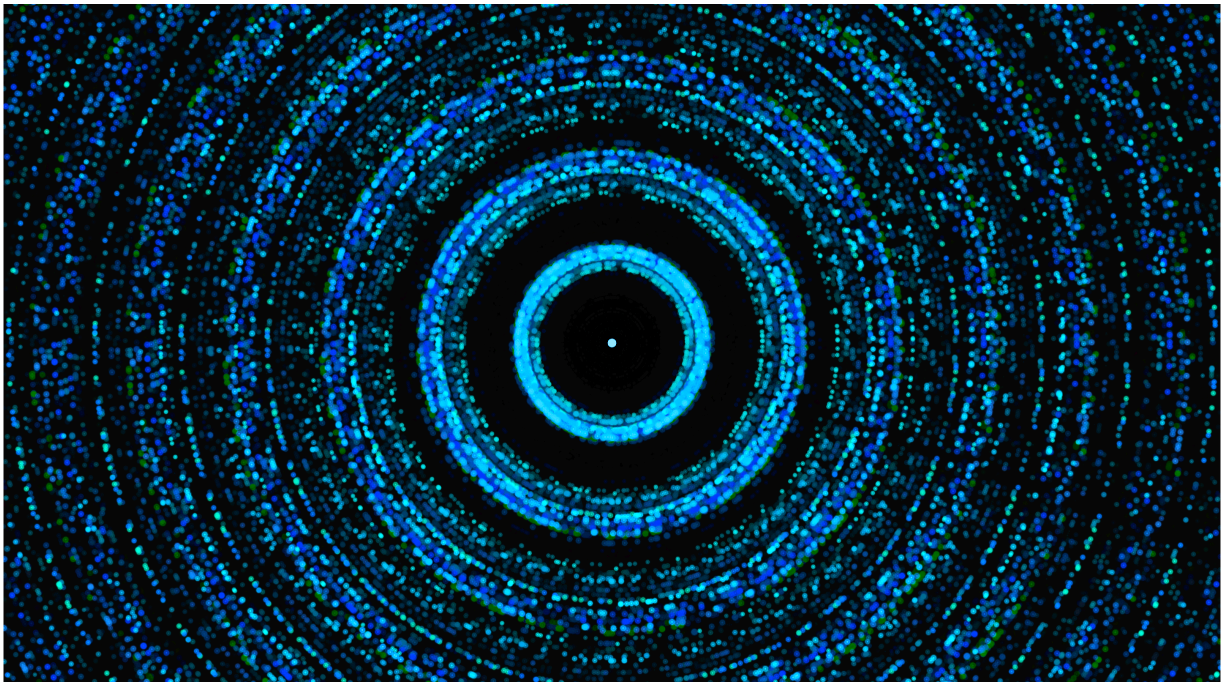

Figure 25. A still from *Audiovisual Ouroboros*

The particle system then sent real-time control data derived from its own generative motion–the position of particles on the screen, their speed, their color, collisions with the edge of the screen–to audio software, which triggered and affected sounds according to these messages. These sounds were played back from the speakers, where they entered the microphone.

The audiovisual feedback loop can be summarized as follows (starting, arbitrarily, with sound, as once the loop was completed and a feedback process begun, the loop really had no beginning or end): sound entered the microphone, triggering moving images; the parameters of these moving images triggered sounds; these sounds were played back through speakers and then captured by the microphone (completing the loop).

In relationship to the symbol of the Ouroboros, the microphone could be likened to the snake's mouth, the speakers to its tail, and the particle system and its accompanying sounds to the body of the snake (constantly nourished and devoured by itself).

The feedback loop was not at all pure, but rather was shaped, mediated, and excited by several factors:

- Sound from participants, and ambient sound. This constituted a signal from "outside" the feedback loop. Participants were invited to insert sound into the loop (feeding the snake), which generated a visual response, and, in turn, a sonic response.

- The nature of audio analysis and how it drove visualization. The characteristics of audio quantified by the analysis determined how the sound was "digested," and which of its characteristics affected the visual system. The parameters of the visual system affected by the control data from the audio analysis (and the scale at which they were affected) determined how the visual system responded.

- The nature of the sonification of the visual images. The way in which the parameters of the particle system triggered and shaped sound determined the audio re-entering the microphone.

- The distance between the microphone and the speakers, and the characteristics of the acoustic space (and to a lesser degree the sonic response of the speakers and microphone). While the intention here was to minimize the degree to which the audio signal was affected when traveling from speaker to microphone, the advantage to sending the sound through physical space was that it allowed participants to interrupt and interact with the feedback loop (by moving or covering the microphone, for example), which helped them intuitively understand the feedback loop.

The installation involves two forms of sound-image correlation (sound-to-image and image-to-sound), and combines them in order to explore the nature of audiovisual feedback in the context of synaesthetic musical experience.

Sound and image were in a constant process of constructing, shaping and controlling each other. This meant that, although sound was the domain in which participants could interact with the

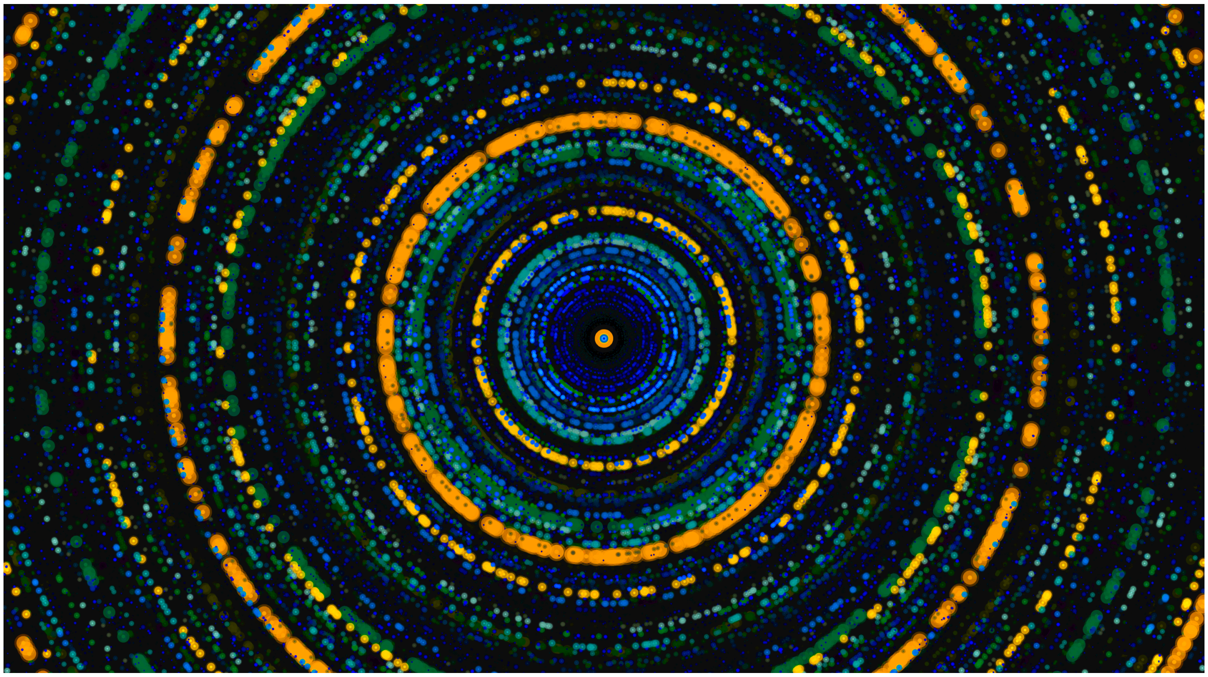

Figure 26. A still from *Audiovisual Ouroboros*

feedback loop, neither sound nor image felt primary in the perceptual experience created, but began to feel like two aspects of the same percept, interdependent and symbiotic.

An audiovisual feedback loop could not have existed before technology made it possible to translate sound into image, and vice versa, in real-time. Analog video is the first technology that made this possible, and computers capable of generating and affecting digital video in real-time greatly expanded the possibilities.

While a feedback loop is one kind of structure that involves both sound-to-image and image-to-sound mapping, other structures with a bi-directional and multi-dimensional mapping between sound and image are possible, and should be explored in future work. Because such mapping puts sound and image on a more equal footing in terms of how each affects the synaesthetic relationship, it may be a more powerful technique for creating synaesthetic audiovisual relationships than mappings which only flow between sensory domains in one direction. It may also help to reveal the unique properties of audiovisual synaesthetic systems as a medium.

Chapter 5
ControlMaps

5.1 Motivation

History and contemporary musical and artistic practice are rich with attempts to create synaesthetic relationships between sound and image. However, as there are few software packages specifically designed for crafting such synaesthetic relationships, synaesthetic music is by necessity created using either a combination of available music technology and visual technology, or custom technology (and often all three). Many attempts (perhaps the majority of them) to correlate sound and image in a synaesthetic way rely mostly or entirely on manual composition in mostly separated visual and sonic production processes. While this can yield excellent results, it is one of many available approaches. By contrast, while modern computers have now opened up of a vast range of possibilities in audiovisual parameter mapping, the use of such techniques, and their average complexity, are limited.

Because parameter mapping (discussed in A.3) is an important technique for creating synaesthetic audiovisual relationships, protocols for sending real-time control data between devices and applications play a major role in synaesthetic music. The use of such control signals in a musical context has its roots in the use of control voltage (CV) in synthesizers. CV is also used in modular analog video synthesizers, and to connect and integrate sound and video synthesizers. While analog synthesizers using control voltage have seen a resurgence, MIDI took the place of CV as the predominant method in electronic music technology for sending real-time control messages. Open Sound Control is a newer, network-based standard (which, despite its name, has ironically seen greater adoption in visual software than in sound and music software).

DAWs and VJ software have built-in facilities for linking incoming control messages to their internal parameters. There is also a range of recent software designed specifically for translating, routing, scaling, and mapping real-time control data. Osculator[1] and Junxion[2] (both OS X only) are standalone applications that enable users to translate and route a wide range of different types of control data. They work with MIDI and OSC but also with messages from HID devices (including wireless game controllers like Nintendo's Wii Remote), serial sensor data, and other kinds of control events. They are widely used as a method for translating one kind of control data to another (for example, converting MIDI to OSC or vice versa), and routing it to the proper destination. They both have features for rescaling data in a linear fashion, and Junxion has basic functionality for remapping values of control data according to a user defined curve. Mapulator[3], which is a Max for Live[4] device, takes this functionality a bit further, with the ability to edit the curves graphically (as opposed to programming them numerically, as in Junxion), and the ability to output scaled data to multiple parameters in Ableton Live.[5] Its chief limitations are that it only works inside Live (and not with standards like MIDI and OSC), and that it is dependent on Max. Other notable software with similar functionality includes Bome's MIDI translator[6] (Windows and OS X), and MIDIPipe[7] (OS X).

Software dedicated to managing and shaping how control data flows between different applications can be very useful in the context of synaesthetic music. When parameter mapping is used to create correspondences between sound and image, the dynamics of the real-time control data that facilitate the mapping will dictate the nature of the relationship. Software that facilitates experimentation with, and fine-tuning of the dynamics of such control data would be very useful for synaesthetic music. Existing software dedicated to this purpose is not very flexible in how it can manipulate control data. Custom, ad-hoc solutions for specific situations can be created using the programming environments discussed in Appendix B (B.6), but this is potentially time-consuming, and requires programming to be a part of the creative process.

1 http://www.osculator.net/
2 http://steim.org/product/junxion/
3 http://www.djtechtools.com/2012/02/16/mapulator-advanced-midi-mapping-for-ableton/
4 https://www.ableton.com/en/live/max-for-live/
5 https://www.ableton.com/en/live/
6 http://www.bome.com/products/miditranslator
7 http://www.subtlesoft.square7.net/MidiPipe.html

ControlMaps seeks to help simplify and speed up the process of crafting relationships between sonic and visual parameters via parameter mapping, by providing a flexible yet intuitive interface for the scaling, nonlinear translation, and routing of real-time control data.

5.2 Functionality and implementation

ControlMaps is a flexible system for scaling and routing control data with a simple and intuitive interface (**Figure 27** shows a screenshot of the current prototype). It allows the user to take in a control signal as MIDI or OSC messages, scale the messages according to a graphically editable curve, and output the value as MIDI or OSC to a chosen destination. An arbitrary number of such mappings can be created, edited, and saved together as a set.

A breakpoint editor allows users to visually edit the curve using a visual vocabulary that will be familiar to users of music software. Straight line segments and Bézier curves can be created by adding control points using the mouse.

A vertical line superimposed on the curve visually indicates where the incoming value falls on the curve, making it easier to perceive the relationship between the graphed scaling curve and its effect, and helping users adjust the nonlinear scaling of a parameter intuitively based on the results achieved, in real-time (much in the same way a DAW user might edit the automation curve of an audio effect parameter while listening to the results of that automation, informed by the playback position indicator's intersection with the automation curve).

A standalone application built in C++ is currently in development using the cross-platform library Juce.[1] A prototype was created in Max to better understand considerations regarding the structure of the application and the design of its user interface. The Max prototype is fully functional and can be run in OS X. It uses the sa.function external by Toshiro Yamada[2] (also used by Mapulator) for interpolating values according to a graphically-editable curve.

[1] http://www.juce.com/
[2] http://cycling74.com/toolbox/sa-function/

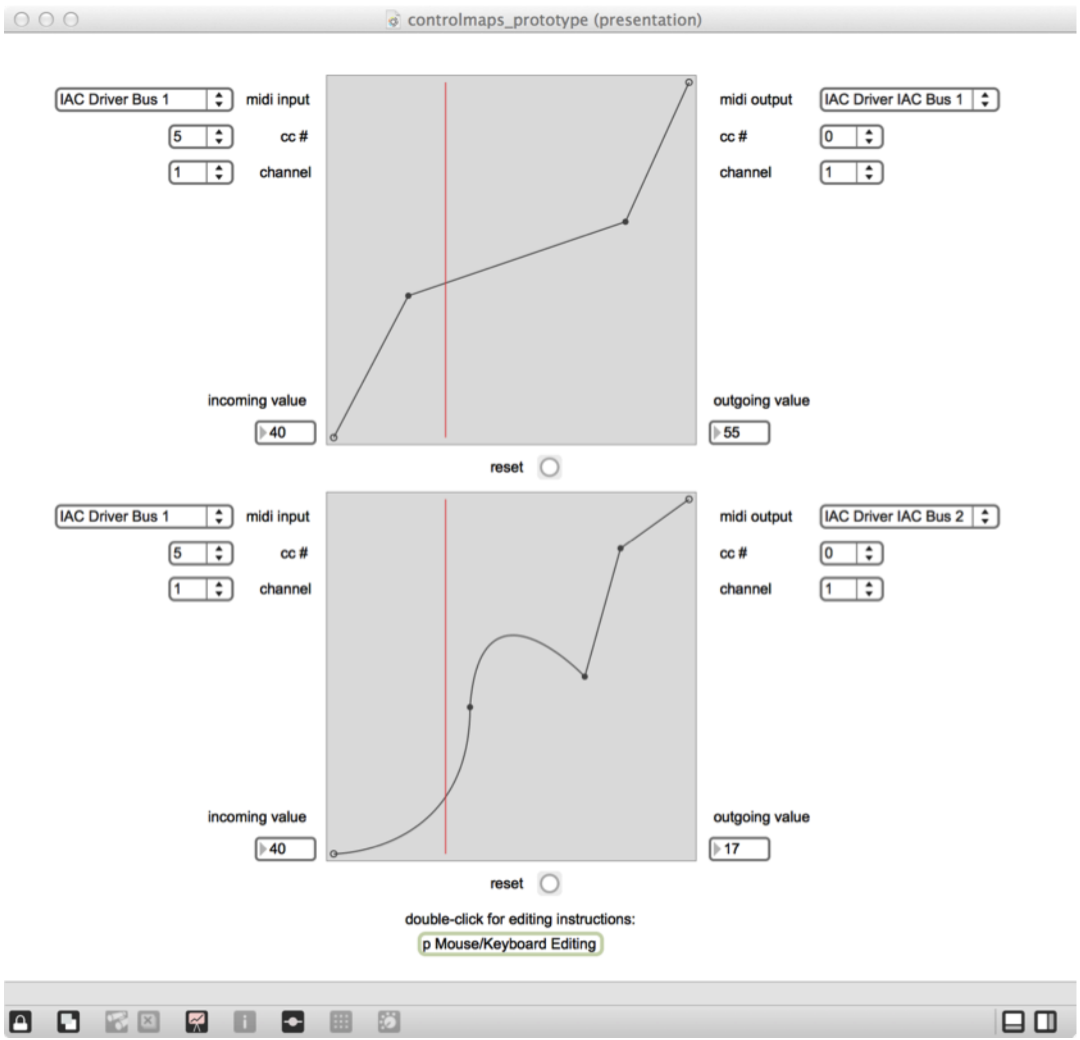

Figure 27. ControlMaps Max prototype

5.3 Use cases

The live performance of electronic music often involves the manipulation of parameters using MIDI control surfaces with knobs or faders; one type of audiovisual parameter mapping used in synaesthetic music extends this technique to the visual realm by linking a pair of corresponding sonic and visual parameters to the same live control input. Similarly, automation in a DAW (or in any other timeline-based environment that outputs control data) could control an aspect of sound and also be routed outside the DAW, to a visual production environment, as a MIDI or OSC control signal.

By default, these techniques usually create linear relationships between the sonic and visual parameters. ControlMaps allows for such relationships to be easily customized, scaling (either linearly or according to a curve) incoming control data differently for each destination parameter.

This allows users to experiment flexibly and quickly with the dimensions and proportions of the relationships created through parameter mapping. Such manipulation of a nonlinear mapping relationship could be the site of creative decision-making about the synaesthetic audiovisual relationship. For example, a visual object could be moved on a path across the screen in correlation with the panning of a corresponding sound; ControlMaps could be used to adjust how quickly the object moves across different parts of its range, perhaps slowing down towards the edges of its range and moving more quickly across the center part of the range (or vice-versa). The curve describing the position and proportions of these changes in speed could be edited graphically and intuitively while observing the resulting changes in real-time. Another creative technique might involve creating forms of visual motion that wobble or oscillate in response to linear change in an incoming control signal.

This kind of non-linear mapping can also help address and experiment with differences in how visual and auditory perception respond to energy levels and frequency relationships in stimuli (for example, in *Caruso's Dream*, subtle logarithmic scaling was used on the lighting control data that had been generated from an audio analysis, in order to help the relationship between brightness and volume feel natural and appropriate). In the case of perceived pitch and loudness in sound, changes that are perceived as linear actually reflect exponential changes in frequency and energy level. Similar non-linear relationships between energy levels in stimuli and the resulting perception exist in vision. (Nundy and Purves 2002) Adjusting these kinds of curves could be especially important in correlating different modes of sensory perception that respond differently to stimuli.

Scaling according to a user-defined curve can also be thought of as a way of manipulating the dynamics of a control signal, much in the way dynamics are manipulated by compressors, expanders, and limiters in audio production. For example, curves that act like a compressor (dividing incoming values above a defined threshold, according to a defined ratio) might be used to adjust how a visual parameter responds to a signal generated by an audio analysis, perhaps reducing the intensity of transients above a threshold while preserving the dynamics of the signal below that threshold. This kind of technique can help users adjust control signals into an appropriate range with great flexibility–different ranges of values in the incoming signal can be compressed, expanded or shifted linearly. It opens the door to sculpting audiovisual relationships based on parameter mapping in the way a skilled music producer sculpts the dynamics of sound.

In addition to its use as a tool for audiovisual parameter mapping (which was its main motivation), ControlMaps has applications in traditional (non-synaesthetic) musical practice. It can be used to create complex "macro" controls in which one control source–a knob, slider, or other source of control data–affects many sonic parameters in different, potentially nonlinear ways. Ableton Live has built-in functionality that allows linear one-to-many mappings within the software; Mapulator extends this functionality further within Live, introducing nonlinear mapping; ControlMaps is a platform for this kind of mapping that could be used in conjunction with a wide variety of software and hardware.

5.4 Future development

Features planned for future development of ControlMaps include:

- Features to assist in managing and editing large numbers of mappings: duplicating, copying, and pasting mappings; creating groups of mappings and visually expanding/collapsing these groups.

- Real-time control over control points, for the automation and live manipulation of changes to the curves. This can open up another dimension of complexity and real-time control in the creative manipulation of mapping relationships.

- Real-time switching between sets of mappings, muting/bypassing of channels, in response to control messages. This can be likened to switching between presets or scenes in music production software.

- The ability to chain multiple mappings together in series without routing control data outside the application; the ability to multiply together multiple curves, collapsing their cumulative mathematical effect into one curve.

- The creation of a library of useful, pre-defined curves.

ControlMaps has so far seen limited use in creative work. Future development will be informed by further creative experimentation with the current prototype.

Chapter 6
FFT Control

FFT Control is an application that builds upon real-time spectral audio analysis using the Fast Fourier Transform (FFT), providing low-level features, in both OSC and MIDI formats.

6.1 Motivation

While there are now vast possibilities for extracting musically meaningful information from sound in real-time, and applying this information to real-time visual content, the amount of experimentation with these techniques, and technical and aesthetic complexity of it, is relatively low.

FFT-based audio analysis and feature extraction provides a powerful way to extract musically meaningful information from sound, and modern computers make it possible to conduct such audio analysis in real-time. This is a powerful method for applying information about the sonic content of an audio signal to visual content, and it is the basis for a great deal of contemporary synaesthetic work. Many automated music visualizers use spectral analysis, and the technique has seen widespread use in musical and artistic contexts.

However, there is little software for using real-time spectral analysis in this way, and the capabilities of what does exist are limited. A lot of VJ software (discussed in B.1) has built-in FFT analysis capabilities that allow users to control various visual parameters with data extracted from an audio signal in real-time. This is a powerful feature, but often has limitations. Usually this functionality only works within a particular piece of VJ software, and can only be used to control internal parameters (lacking the ability to output a control signal). And although a wide range of documented methods for feature extraction in FFT analysis of audio signals now exist, usually only a few types of feature extraction are implemented (with most software only allowing the extraction of volume across a user-selectable or pre-defined frequency range).

While FFT analysis is used extensively in music software, it is usually used to visualize a spectrum of sound as part of an interface for sculpting sound, and rarely does it allow for data gained from the FFT to be output as real-time control data. Notable exceptions to this are Blue Cat Audio's FreqAnalyst Pro[1] and StereoScope Pro,[2] which allow some features from real-time audio analysis to be output as MIDI control data.

This chapter presents FFT Control, a VST plugin which attempts to facilitate experimentation and creation using spectral analysis, making it more accessible, faster, and more integrated into musical workflows and technological ecosystems.

6.2 Functionality and implementation

The first prototype of FFT Control (discussed in Chapter 3) was written in ChucK[3], with a GUI written in Processing that communicates with ChucK via OSC. It visualizes the spectrum by graphing the FFT bands, and allows the user to graphically select three frequency bands per channel which can each output a fluctuating control value that tracks the average magnitude across the selected frequency range (a range of FFT bins). The control signals generated are output as OSC. This prototype of FFT Control has been integral to a variety of projects, as documented in Chapter 3 and Chapter 4. Tracking overall energy levels across defined frequency ranges is a useful technique, but it is only one of many possible audio analysis techniques.

[1] http://www.bluecataudio.com/Products/Product_FreqAnalystPro/
[2] http://www.bluecataudio.com/Products/Product_StereoScopePro/
[3] http://chuck.cs.princeton.edu/

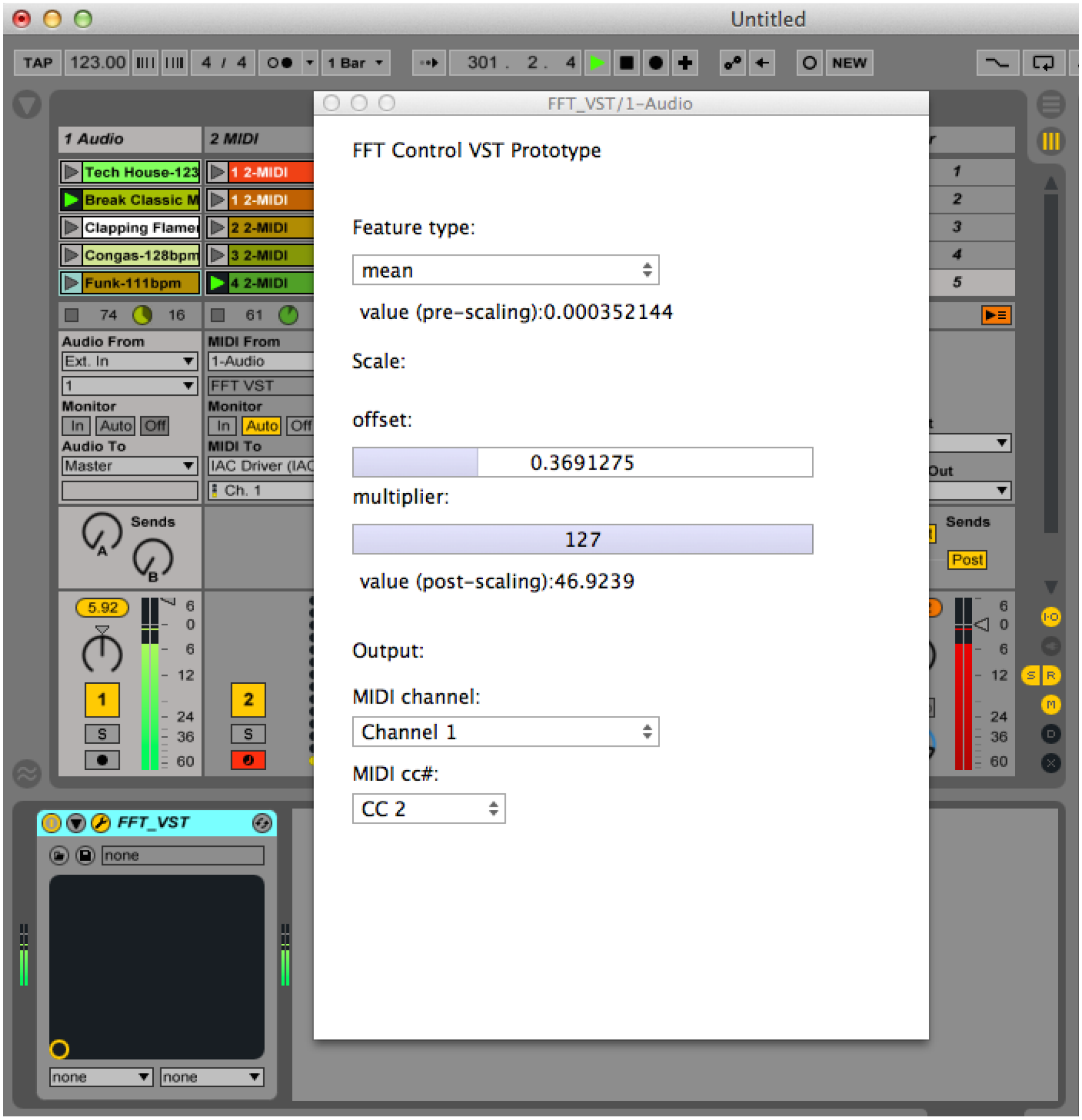

Figure 28. FFT Control on a track in Ableton Live

The version of FFT Control in current development is a cross-platform VST/Audio Unit plugin. It is developed using the Juce library, and uses the LibXtract library by Jamie Bullock[1] to extract features from audio, sending the result as a MIDI control signal to the host application, which can then route it to other tracks, to parameters of other devices, or outside the host to

[1] http://libxtract.sourceforge.net/

other applications or hardware MIDI devices. **Figure 28** shows the current working prototype of FFT Control running on a track in Ableton Live, analyzing audio from a drum loop and sending MIDI CC messages back to Live, where a MIDI track is receiving those messages.

As depicted in **Figure 28**, the application's interface allows users to select a feature type from a dropdown list, displays the initial result of that calculation (a fluctuating numeric value), scales it according to user-defined parameters, and outputs it to a selected MIDI channel and CC number.

Audio entering the plugin is passed on unaltered. This allows FFT Control to be inserted at any point in a chain of effects, analyzing the audio at that point without altering the signal. It also means that multiple instances of FFT Control can be chained together for the extraction of multiple features from the same source. Future versions of the software will allow multiple features to be extracted in the same instance of the plugin, and from the same FFT calculation, improving efficiency.

The first version of FFT Control focuses on the extraction of "instantaneous features." From *An Introduction to Audio Content Analysis Applications in Signal Processing and Music Informatics*:

> The term *instantaneous feature*, *short-term feature*, or *descriptor* is generally used for measures that generate one value per (short) block of audio samples. An instantaneous feature is not necessarily musically, musicologically, or perceptually meaningful all by itself, and it is frequently referred to as a *low-level feature*. A low-level feature can serve as a building block for the construction of higher-level features describing more semantically meaningful properties of the (music) signal (such as tempo, key, melodic properties, etc.). (Lerch 2012)

Such features were chosen because of their immediate application as continuous control signals that could be mapped to visual parameters (creating constant motion in that parameter). Higher-level features are of interest for future development. Feature extraction types implemented in the current prototype of FFT Control include mean, spectral centroid, rolloff, fundamental frequency, smoothness, spread, noisiness, and sharpness.

6.3 Use cases

As a real-time audio plugin, FFT Control integrates modularly into a vast ecosystem of existing audio software. This includes DAWs, live performance software, VJ software, and visual programming environments. Multichannel audio analysis can be easily integrated into established musical platforms and workflows. Control signals can be generated which represent some aspect (or multiple aspects) of each instrument and musical element in a piece of music created, recorded, or performed in a DAW, and the complexity and size of this process is limited only by processing power.

The control signals generated can be routed by the host to other applications in order to control visual content in real-time, opening up a wide range of possibilities for experimentation with the mapping of audio analysis features to visual parameters, and making such experimentation faster and more accessible. FFT Control could be used in conjunction with ControlMaps in order to shape the control signals generated.

Such techniques can be used in composition, live performance, and installations. The power of these techniques in synaesthetic music is that they allow the use of perceptually and musically meaningful data about sound to be applied in real-time to visual systems, opening up a vast range of possibilities for crafting rich and meaningful synaesthetic audiovisual relationships. The goal of FFT Control is to make the use of real-time audio analysis more accessible for intuitive creative experimentation that does not necessarily require a technical understanding of the underlying mathematical processes, and to allow this experimentation to take place within the context of established music technology.

While it is not the focus of this thesis, FFT Control's spectral audio analysis can also be used to control parameters of sound synthesis and processing. Sidechain compression and vocoding are examples of similar processes in which the dynamics of one signal are used as a source of control over processing executed on another signal. There is a range of software that enables such audio processing shaped and driven by spectral analysis; FFT Control could also be used for such processing by combining it modularly with other audio effects, allowing custom spectrally informed processing. This kind of spectral audio processing could also become the basis for audiovisual effects, by routing the same spectrally-derived control data to visual parameters.

6.4 Future development

Not all of the feature extraction methods provided by LibXtract have yet been implemented in FFT Control; taking full advantage of LibXtract's functionality is a high priority for future development.

LibXtract was designed in such a way that feature extraction methods can be cascaded efficiently, extracting multiple features from one block of incoming audio, without duplicating the FFT calculation. While multiple instances of FFT Control can be chained together in a host (allowing for the extraction of multiple features from the same source audio), it would be more efficient to use LibXtract's built-in capability for extracting multiple features, within one instance of the plugin (this would also make it easier to save a grouping of feature extraction functions).

FFT Control's built-in scaling functions provide a basic ability to manipulate the range of control data generated by the application. Of interest for future development is more intelligent scaling functionality that takes into account the range of data generated by particular feature extraction methods, as well as the desired range of values for a particular output format (for example, 0-127 for MIDI CCs, 0-16383 for MIDI pitch bend, 0.-1. for OSC messages).

Chapter 7
Conclusion

7.1 Summary

This thesis has investigated the relationship between technology and synaesthetic music–the history of that relationship, its current nature, and where it can lead. We are in a period of explosive growth in the capabilities afforded by computers for generating and processing sound and image in real-time, and for correlating these processes. Ever-increasing processor speeds and widespread interest in the musical correlation of sound and image (synaesthetic or otherwise) have facilitated a growing ecosystem of software that is capable of the creation of synaesthetic music.

A great deal of software for shaping sounds and images is firmly rooted in one modality or the other. Therefore, most software that deals with the production and manipulation of images–obviously an important part of creating audiovisual synaesthetic content–is generally not rooted in musical concepts and practices, but in concepts and workflows from established visual arts. As a result, software that facilitates the production and manipulation of images using musical concepts and workflows is rare and underdeveloped. This limits the ease and immediacy with which musicians and artists can explore integrated audiovisual production as a musical process, taking advantage of music's established techniques and processes of experimentation, jamming, and composition.

The synaesthetic relationship–the nature of the connection between sound and image–is a central, defining aspect of synaesthetic music, yet there is relatively little technology that focuses directly on facilitating experimentation and creativity with types of connections between sound and image. Much of the existing software that enables this kind of interdisciplinary work requires of its users some skill in programming. Artistic processes where programming is central to the creative act can be beautiful, unique, and perfectly suited to some kinds of work. But the lack of accessible and standardized (yet powerful and flexible) tools in this realm limits the range of artists who will participate in creating synaesthetic music. For artists who do have programming knowledge, the need to engage in technical problem solving can sometimes interrupt the creative process. Limiting such interruptions enables artists to stay in a state of creative flow for longer periods of time, and can have a profound impact on an artist's ability to realize their vision.

The historical survey presented in Chapter 2 shows that the practice of synaesthetic music has advanced when new technology–whether developed specifically for synaesthetic music or for other disciplines–affords new kinds of technical connections between the production and manipulation of sounds and images. The survey of contemporary technology in Appendix B examines the current state of software enabling such connections and musical audiovisual workflows, identifying areas where further development could expand creative possibilities. The musical and artistic work documented in Chapter 3 and Chapter 4 further motivated and informed the planning and development of new software for synaesthetic music. The software prototypes and concepts presented in Chapter 5 and Chapter 6 attempt to help advance accessibility, interconnection, integration, and musicality in technical approaches to synaesthetic music, with the goal of expanding the possibilities for creative expression.

7.2 Primary Contributions

This section summarizes the primary contributions to the art and technology of synaesthetic music presented in this thesis.

1. Light Compositions for *Caruso's Dream*

The light compositions for *Caruso's Dream* synaesthetically embody and represent the music of Enrico Caruso in light moving through sculptural forms. While the light compositions, given their basis in existing pieces of music, were not the result of an audiovisual creative process in which sound and image are created and grow together, they were the result of synaesthetic techniques which correlate sound and image in a musical process. Caruso's music existed first (and exists independently of *Caruso's Dream*) but the experience imparted by *Caruso's Dream* can be considered a form of synaesthetic music. The aesthetic of the light compositions required the use of unique technical processes involving custom software, described in sections 3.3.2-3.3.5. The processes chosen, and the software created to help execute these processes, situated the creation of the lighting sequences in a musical context. The development of this software as part of a synaesthetic, musical creative process yielded insights relevant to the design of technology for synaesthetic music, and informed the development of software presented in Chapter 5 and Chapter 6.

2. *Emulsion Juice #1*

Emulsion Juice #1 is an audiovisual installation in which participants were invited to explore a synaesthetic interaction between sound and image using musical instruments. It explored the application of real-time spectral analysis (using an early prototype of FFT Control) to extract spectral information and use it to drive live generative visuals in a particle system. Via custom software, the piece successfully created a rich relationship between spectral content in sound and resulting colors in generated images.

3. *Ces Balons de Lumière*

Ces Balons de Lumière explored some of the techniques used in *Caruso's Dream* and *Emulsion Juice #1*, using audio analysis to create a relationship between spectral content and color, in this case using a modular synthesizer as a sound source, and exploring the creative possibilities in the mapping and performance of an audiovisual relationship between a synthesizer and reactive colored lights. By using a simple and direct yet total mapping from sound to image, the resulting temporal and qualitative synaesthetic relationships between sound and image were revealed for the audience with clarity and focus.

4. *Audiovisual Ouroboros #1*

This installation builds on many of the techniques used in *Caruso's Dream*, *Emulsion Juice #1*, and *Ces Balons de Lumière* (real-time audio analysis, spectrally-driven color, and a particle system), but explores the possibilities of audiovisual feedback in a synaesthetic music system involving both sonification and visualization. Like *Emulsion Juice #1*, it is an interactive installation, inviting participants to play with the audiovisual feedback system. The installation reveals some of the possible aesthetic properties of audiovisual feedback. A goal of synaesthetic music should be for sound and image to feel like they embody and create each other in a unified manifestation of one idea, rather than existing as separate streams that react and respond to each other. In *Audiovisual Ouroboros*, because sound and image constantly co-create each other, neither feels primary in the experience. This suggests that future work could explore the power of bi-directional mapping between sound and image in crafting synaesthetic relationships. Although such mappings can create feedback, it may also be desirable to use bi-directional mappings without feedback as a way of putting sound and image on more equal footing (in terms of their position in a hierarchy of creative process and/or perceptual experience).

5. ControlMaps

ControlMaps is new software that allows for quick and intuitive adjustment of audiovisual relationships based on parameter mapping. Its goal is to allow for easier and faster experimentation with audiovisual relationships in musical contexts. It is one of the first applications to allow for the creation and manipulation of non-linear mappings between sonic and visual parameters, using a graphical interface. It interfaces with a wide range of real-time musical and visual software and hardware via MIDI and OSC, with the goal of facilitating more complex and refined synaesthetic audiovisual parameter mapping, and making parameter mapping a faster and more accessible creative practice.

6. FFT Control

FFT Control is new software that enables the use of spectral audio analysis to generate real-time control signals, with several types of feature extraction available to the user. Early prototypes accepted audio input via software like Soundflower[1] and Jack[2], and output OSC messages. The current version runs as a VST plugin and outputs MIDI to the host. The goal of the software is

[1] https://github.com/RogueAmoeba/Soundflower
[2] http://jackaudio.org/

to allow for easier and faster experimentation with real-time audio analysis and feature extraction as a basis for generating and shaping real-time visual content, and to integrate this functionality into widely used music software and workflows. FFT Control is the first VST plugin of its kind with the capability of providing a wide range of spectral audio features as real-time control signals for flexible use in audiovisual mapping or spectrally-informed audio processing. By providing a range of perceptually and musically meaningful real-time data from audio analysis, and situating itself within the context of established musical workflows, FFT Control helps expand the possibilities for crafting synaesthetic audiovisual relationships in music.

7.3 Final Thoughts

There are a wealth of theories, inventions, scientific studies, experiments, and artistic work from throughout history that explore correlations between various aspects of sound and form, hearing and vision, music and the visual arts. Artistic manifestations and explorations of such correlations have usually required meticulous manual labor, custom technology, and a lot of time, but we are now moving quickly towards a future where this is no longer the case. The capabilities of modern personal computers can enable a vast range of real-time audiovisual expression, and in this context history inspires and challenges us to experiment, innovate, and create more quickly, deeply, and expressively with ideas that have fascinated human beings for thousands of years.

It is easier than ever to create synaesthetic music, yet there is much work to be done to give this medium a mature toolset. There is an increasing amount of software that can be used to create synaesthetic music. But much of the available software, while capable of creating such relationships, is not specifically focused on synaesthetic music, and is either firmly rooted in the workflows and vocabulary of more established mediums, or is open-ended and interdisciplinary but requires a significant amount of technical knowledge and programming ability. If we want to see synaesthetic music become an art form that is widespread and organic, with many people contributing to it, we need to push the integration of music and visual technology even further. Tools that integrate audiovisual production into established music-technological workflows can help achieve this.

As processor speeds continue to increase, and audio analysis and music information retrieval continue to develop and become more powerful, it will become increasingly possible for automated visualization driven by creative programming to achieve very satisfying, synaesthetically musical and meaningful results. However, at any point in technological development, a skilled artist using this kind of technology as a creative tool will do even better (at least until the point when artificial intelligence becomes as generally intelligent as humans). This kind of sound and music analysis technology can be used by humans as a creative tool, just as instruments and recording technology are used in music.

Modern technology should be used to explore the synaesthetic formulas and theories of the past and present. In the same way that music theory can help us write good music, by focusing our efforts and guiding our work within an organizing framework that allows us to explore different kinds of aesthetics and creative ideas, we should look to theories which correlate color, sound, light, pitch, texture, and timbre similarly. These systems can be explored, rules tested, broken, re-written. However, we should not search for the "correct" or "universal" formula for mapping sound and image, but rather for systems that help us express ourselves artistically with synaesthetic sound-image relationships. And as is the case in music, there will always be a place for intuitive experimentation that has no need for these theories or rules.

7.4 Future Work

The creative work documented in Chapter 3 and Chapter 4 used early software experiments and prototypes, and motivated the refinement and development of FFT Control and ControlMaps. This software has only recently reached its current stage, and still needs more development, but it will be interesting to work with it creatively to see how it expands creative processes. Live performance of synaesthetic music is an area I have explored in the past, and would like to revisit using FFT Control and ControlMaps. Undoubtedly, further creative work with the software will inform further development of the software. For an artist/technologist, this can be an ongoing, iterative, and never-ending process of projects and techniques informing the refinement of technology. But ultimately, the software discussed here is a means to an end, not the end in itself, so its real value will be demonstrated not by enumerating its features but by using it to create more effective artistic work (on the other hand, there is not necessarily a problem with a never-ending cycle of technological development for creative purposes–in fact, it is often the dance between the progression of available technology and artistic intention which manifests new and interesting forms of music and art).

My goals for future work in the realm of synaesthetic music include:

- Developing more ways of connecting visual and musical systems, closing the distance between music and visual production and situating visual production in a musical context.

- Carrying on the tradition of color organs with contemporary technology: audiovisual instruments and effects that allow for the real-time musical performance of synaesthetic musical content.

- Experiments in audiovisual relationships with these tools.

- Installations that are synaesthetic & musical.

- Synaesthetic musical works in which the visual component is not confined to a screen, but becomes more immersive. Technologies and techniques like projection mapping, sculpture, virtual/augmented reality, 3D video, and holographic projection can help achieve this.

- Performances of synaesthetic music.

With the speed, flexibility, and ubiquity of modern computers, and the growth of software that enables audiovisual integration, audiovisual mapping is coming of age as a creative technique. The dream of synaesthetic music, in which sound and image are an integrated musical material that can be composed, performed, and manipulated in musical ways, is technically within closer reach than ever. With unprecedented capabilities at our disposal, the challenge now is for the practitioners of synaesthetic music–and technologists working to expand what is possible with this art–to deeply investigate the creative possibilities afforded when technology facilitates the synaesthetic integration of sound and image under the direction and control of an artistic and musical hand.

Appendix A
Approaches for linking sound and image

This appendix describes several categories of technical and artistic approaches for creating synaesthetic relationships between sound and image in musical contexts: composition, parameter mapping, harmony, and visualization/sonification.

A.1 Synaesthetic Composition

Perhaps the oldest way in which people have attempted to create audiovisual, synaesthetic music is simply through composition: crafting the visual and sonic material to complement each other, with the intention of integrating sound and image into a synaesthetic experience, but without the aid of any direct or automated technical linkage between the two. This kind of synaesthetic production usually involves painstaking manual correlation between sound and image, which can sometimes involve some sort of systematic mapping, but which is executed manually (with varying degrees of systematic purity). Without the aid of a technical process that can automate the execution of such mappings, however, the compositional intention of the artist is always primary in creating the audiovisual relationship, and mapping has not yet really become a material or tool with which to experiment. There is not a clear distinction between a system that exists in the mind of an artist and one worked out on paper, but there is a clear distinction between a manually executed system and one that is automated.

All audiovisual musical techniques, (including those discussed in subsequent sections of this appendix) can be placed on a spectrum between manual composition and completely automated visualization or sonification. All techniques relevant to the creation of synaesthetic music involve some kind of composition, in a sense: decisions are made by a human being about what to create in both the auditory and visual realm, and about how specifically to correlate sound and image (even when such correlations are largely programmed, human decision-making and creativity has gone into the programming).

Much abstract film art that has explored synaesthetic music is based primarily on composition, involving a meticulous sequencing of visual images in sync with music. Some examples of this are "Swinging the Lambeth Walk"[1] by Len Lye and "An Optical Poem"[2] by Oskar Fischinger (though Fischinger's use of "animated sound" in other pieces is an example of sonification).

Contemporary examples of synaesthetic audiovisual correlation via manual composition include "IV.7"[3] and "IV.10"[4] by Beeple, "_grau"[5] by Robert Seidel, "Fragment"[6] by Emptyset, and "Omicron"[7] by AntiVJ.

A.2 Harmony

John Whitney extensively explored the concept of visual harmony, and correlations between musical structures and visual structures born out of the same harmonic principles. He documented his work in this area in his book *Digital Harmony: On the Complementarity of Music and Visual Art*.

Lissajous curves are an example of visual harmony that arises directly from two sources of sound (they are also an example of sonification). Contemporary artists who have explored using harmony in this way include Memo Akten (in his "Simple Harmonic Motion"[8] series), Alexander Dupuis (in "That Which Pulls"[9]), and the duo Cyclo[10] (Carsten Nicolai and Ryoji Ikeda).

[1] https://vimeo.com/92722063
[2] https://www.youtube.com/watch?v=they7m6YePo
[3] https://vimeo.com/4277069
[4] https://vimeo.com/39157873
[5] https://vimeo.com/2669327
[6] https://vimeo.com/77125797
[7] https://vimeo.com/41486619
[8] http://www.memo.tv/simple-harmonic-motion/
[9] https://vimeo.com/78398891
[10] http://www.ryojiikeda.com/project/cyclo/

A.3 Parameter Mapping

Parameter mapping involves an automated linking of sonic and visual parameters. It facilitates the automatic temporal linkage of sound and image (sounds and images, or changes to them, can be triggered simultaneously, or trigger each other) as well as mathematical links between qualities of sound and image (parameters controlling aspects of the generation or processing of sounds and images can be placed in mathematical relationships with each other).

Color organs were perhaps the first form of parameter mapping, and usually involved a very simple mapping of discrete pitches to discrete colors. More complex real-time audiovisual parameter mapping became possible with video and sound synthesizers, and with computers. These technologies have facilitated the possibility of experimenting with and "playing" audiovisual correlations in a live, musical fashion. Artists can try out various relationships between sound and image without having to commit as much time to manually composing these correlations, and can iterate and refine these correlations quickly. Parameter mapping is perhaps one of the most flexible ways of linking sound and image, and it becomes steadily more powerful as increasing processor speeds facilitate more complex real-time generation and processing of moving images. Real-time audiovisual parameter mapping makes it theoretically possible to link any kind of sonic quality to any kind of visual quality, and therefore has and will continue to have a central role in synaesthetic music and art.

Parameter mapping is used extensively in contemporary synaesthetic music and art. Examples include Ryoichi Kurokawa's "syn_" (Kurokawa 2013a) and "oscillating continuum" (Kurokawa 2013b) and "Lumière" by Robert Henke. (Henke 2014)

A.4 Visualization of sound, and sonification of visual information

Sound visualization is the process of creating an image from information about sound: it involves either a direct translation and transformation of sound data (digital or analog) into the visual realm, or some form of audio analysis that drives or shapes a visual system. Sonification of visual images is the converse of sound visualization: it involves the translation or transformation of visual information into the sonic realm, or an analysis of visual information driving sound synthesis.

Sonification/visualization could be described as a form of parameter mapping; these types of techniques do not have clear boundaries between them, but rather exist on a spectrum. One distinction that can be made, however, is that visualization and sonification always involve the use of existing sonic or visual data either in analysis or in translation/transformation, whereas parameter mapping usually describes a linkage between processes that create or act on sound and image media simultaneously (often procedurally *before* the sound and image exist either as media or physically).

Although there are forms of sound visualization where the result is a static image of some kind (such as spectrograms), the term more often refers to the real-time generation of moving images in response to and synchronous with sound (and "static" sonification is not a possibility, as sound is defined by time).

Cyclo's "id#00" (cyclo 2013) (and other work by cyclo, which is a collaboration between Ryoji Ikeda and Carsten Nicolai) exemplifies visualization (and also demonstrates principles of visual harmony, via sound visualization that is essentially a digital version of the Lissajous experiment). Ryoji Ikeda's "test pattern" (Ikeda 2014) is another example. "No-Input Pixels" by Alexander Dupuis (Dupuis 2013) is an example of sonification.

Appendix B
Survey of contemporary technology for synaesthetic music

This appendix surveys contemporary technology (primarily software) that can be used in the creation of synaesthetic music. This survey is not intended to be comprehensive, but rather focuses on specific features well suited to synaesthetic musical work.

There is a wide range of highly developed technology for the production of music and audio, and for the production of moving images. Music technology and visual production technology often play fundamental roles in the creation of the auditory and visual components of synaesthetic music. This survey is not concerned with all aspects of such software, but focuses on areas which impact the relationship between sonic and visual components, and enable (to varying degrees) the integration of sonic and visual creative practices more often treated as belonging to distinct disciplines.

This survey approaches the intersection of audio and visual production technology from a musical perspective, seeking primarily to document the ways technology can help situate the real-time production and manipulation of moving images in the context of musical processes and practices. Given music's history and roots as a sonic art form, such a perspective tends to privilege processes that begin with sound and attempt to represent the sonic dimension of music in visual form. On the other hand, the concept of synaesthetic music broadens the scope of music, and suggests that sound and image can be integrated to such a degree where neither is primary, and each is an essential aspect of the music. This survey seeks technology that leads us in the direction of such integration.

For the purposes of this survey, the technology discussed is grouped into several categories (which overlap to some degree): VJ software, audio and video routing software, real-time visual effects, sound visualization software, video synthesizers, programming environments (both visual/node-based and text-based), audio analysis software (for controlling visual systems and generating real-time control data) and software for routing, scaling and mapping real-time control data.

B.1 VJ software

If creating a visual analog to music can be achieved through analogous technical processes, the role of VJ software in the visual production process could be likened to the role played by sequencers and digital audio workstations (DAWs) in music (especially electronic music). VJ software is intended as the overall environment for visual production and manipulation–it generally integrates video playback and sequencing, video effects, generative visuals (in some cases), video compositing, and mixing. These processes can be likened, respectively, to audio playback and sequencing, audio effects, MIDI sequencing and virtual instruments, and audio mixing (with the analogy here somewhat strained, as the process of summing together multiple video streams is more complex and involves many more potential creative decisions than summing together multiple audio streams).

Both DAWs and VJ software allow for live manipulation of parameters, although in this regard, VJ software, which is by definition designed for live performance, is more specifically similar to performance-oriented DAWs like Ableton Live and Bitwig Studio.[1]

VJ software connects to the world of music technology via MIDI (and often also supports Open Sound Control). It is generally able to synchronize itself to music software or hardware via MIDI time code or MIDI clock, and accepts MIDI messages as a way of sequencing and controlling various parameters, allowing it to be controlled by hardware MIDI controllers designed for musical purposes, by a DAW, or any other music software or hardware that sends MIDI.

VJ software usually has some audio functionality of its own. It can usually play back video files with audio, manipulating and sequencing corresponding video and audio simultaneously. It can sometimes also load audio effects plugins, which opens up the possibility of shaping sound and

[1] https://www.bitwig.com/en/bitwig-studio.html

visual parameters in the same integrated real-time environment. VJ software also often incorporates basic forms of real-time spectral audio analysis that can be used to control visual parameters.

Like support for third party audio effects plugins in DAWs, VJ software often has support for third party visual effects plugins (see B.3). It also increasingly has support for the routing of streams of video between applications (see B.2).

B.2 Audio and video routing software

Soundflower and Jack are flexible systems for routing audio between applications and across networks, and can be useful for routing audio between music software and visual software. Syphon[1] (OS X), Bigfug Video Stream System[2] (cross platform), and Wyphon[3] (Windows, currently in development) are systems that offer similar functionality for video. These video routing systems can function as FFGL plugins (see B.3), which makes them compatible with a wide range of real-time visual software.

These tools allow various pieces of audio and visual software to be combined and integrated flexibly, in a modular fashion.

B.3 Real-time visual effects: plugins and open standards

Real-time visual effects plugins parallel audio effects plugins in terms of their role in the real-time production process and its technological ecosystem. They make it possible to encapsulate relatively small-scale applications with a particular generative or processing function, and integrate these applications into larger systems easily and flexibly. In the case of audio plugins, there are a few industry-standard plugin formats that are widely adopted (VSTs and Audio Units

[1] http://syphon.v002.info/
[2] http://www.bigfug.com/software/video-streaming/
[3] https://wyphon.wordpress.com/

are the most common), which means that the most basic elements of sound production–which replace the synthesizers and effects units of the hardware studio–can be combined and configured modularly in various host environments. Several open formats for real-time video effects have started to emerge as standards in recent years: Freeframe GL[1] plugins, GLSL[2] shaders, and most recently, the ISF[3] format introduced by Vidvox (which is a container for GLSL shaders).

Freeframe was created to fulfill the need for an open, cross-platform, standardized, and technically independent plugin format for live visual software. Version 1.5 of the standard "includes extensions to FreeFrame 1.0 that enable real time rendering of stunning graphics and video effects with OpenGL compatible graphics cards," and is known as Freeframe GL ("FreeFrame Open Realtime Video Effects" 2014).

GLSL is a cross-platform, high level shading language that can be used on any system with a graphics card that supports OpenGL. GLSL shaders are sets of strings that are sent to the driver for the graphics card, which compiles them into instructions for that particular hardware at runtime. This makes it easy for GLSL shaders to be edited and modified on the fly; there are websites such as Shadertoy[4] and GLSL Sandbox[5] that allow users to create, share, and edit GLSL scripts, which are executed in the browser via WebGL. Shadertoy encourages the development of audio-reactive GLSL shaders with a built in option to play back and conduct real-time FFT analysis on an audio file, exposing the FFT spectrum data to the shader.

Developer Vidvox (maker of VJ software VDMX), which supported the development of Freeframe as an open standard, recently announced that they have created a new format to make GLSL shaders usable as modular visual effects plugins:

> ISF stands for "Interactive Shader Format", and is a file format that describes a GLSL fragment shader, as well as how to execute and interact with it. The goal of this file

[1] http://freeframe.sourceforge.net/
[2] https://www.opengl.org/documentation/glsl/
[3] http://vdmx.vidvox.net/blog/isf
[4] https://www.shadertoy.com/
[5] http://glslsandbox.com/

> format is to provide a simple and minimal interface for image filters and generative video sources that allows them to be interacted with and reused in a generic and modular fashion. ISF is nothing more than a [slightly modified] GLSL fragment shader with a JSON blob at the beginning that describes how to interact with the shader (how many inputs/uniform variables it has, what their names are, what kind of inputs/variables they are, that sort of thing). ("Introducing ISF Video Generators and FX" 2014)

Vidvox plans to make the code open source after gathering feedback.

B.4 Visualization software

Sound visualization software is often associated with entirely automated visualizers built into mp3 players, designed for passive and casual use (i.e. Winamp visualization plugins, and the iTunes visualizer). These visualizers have likely been, for many people, an introduction to the idea of music with a visual component. While they have grown more musically intelligent in recent years (the most recent version of the iTunes visualizer, created by artist Robert Hodgin (Hodgin 2014), is an example of this) in general there is a relatively fixed visual palette and audiovisual relationship (or small range of these things). Although they can be works of art in themselves, they are not designed as creative or interactive tools. These visualizers can sometimes produce a high degree of correlation between sound and image, but the effect is usually temporary and limited in its depth. The complexity of the audio analysis involved (and musical analysis, if any) is usually fairly basic. Without artistic intention involved in crafting the audiovisual relationships, without the sound and image informing one another through the lens of a human mind, it is difficult for automated visualizers to feel like they do justice to the sonic variation, musical structures, and creative choices in the music.

iTunes music visualizers can be created in Quartz Composer, a visual programming environment (discussed in B.6). VSXu[1] and Magic[2] are contemporary software environments specifically

[1] http://www.vsxu.com/
[2] https://magicmusicvisuals.com/

designed for creating custom sound visualizers. These applications allow users to build visualizers by connecting audio analysis modules to visual generators and effects modules. They are notable and applicable to the creation of synaesthetic music because of their ability to customize visual content and its relationship to real-time control data generated by an audio analysis.

B.5 Video synthesizers

Analog video synthesizers have seen something of a renaissance in recent years, and there is now a range of contemporary analog video synthesizers ranging from inexpensive and relatively simple to professional and complex.

On the simpler end of the spectrum, boutique electronic instrument maker Critter & Guitari makes several video synthesizers that respond to audio according to several different fixed modes. ("Black & White Video Scope" 2014; "Rhythm Scope" 2014; "Video Scope" 2014) The Pixelmusic 3000 is an open source video synthesizer by Tarikh Korula, inspired by the Atari Video Music (an early-1980s device for visualizing music), which similarly has a limited number of modes.

LZX Industries[1] is a company that makes a high-end modular video synthesizer system that can be linked to Eurorack audio synthesizer modules. The system has a variety of different modules, including a dedicated module for translating control voltages between the range used in the LZX system and that used by Eurorack format sound synthesizers. Dave Jones (who built video synthesizers in the 70s) is now developing a Eurorack format modular video synthesizer system.[2]

[1] http://www.lzxindustries.com/
[2] http://www.jonesvideo.com/

B.6 Programming languages

Programming environments oriented towards music and the arts have been central to contemporary synaesthetic music and art. Many of these environments are nodal patching environments in which programs are constructed visually, while others are text-based programming languages (and some environments contain both of these types of programming). Some of these environments integrate functionality for both sonic and visual work, while others focus on either sound or images but are extensible and connect to other programs and environments.

Max (formerly known as Max/MSP/Jitter), Puredata (Pd),[1] vvvv,[2] and Touch Designer[3] are examples of node-based visual programming environments that have a high degree of functionality with both sound and images. Each of these environments can be augmented using code-based languages, and some integrate nodes that allow for text-based programming in the context of the visual programming environment.

Examples of node-based environments primarily geared towards generating and processing visual images include Quartz Composer, Vuo,[4] Isadora,[5] and Nodebox.[6] Node-based environments oriented towards generating and processing audio include Audiomulch,[7] Plogue Bidule,[8] and Audulus.[9]

While node-based visual environments have likely been the most widely used programming environments in artistic technical experimentation with synaesthetic audiovisual relationships, several text-based programming environments should also be noted. Many artist-oriented text-based coding environments focus primarily on visual production. Processing is the most widely

[1] https://puredata.info/
[2] http://vvvv.org/
[3] https://www.derivative.ca/
[4] http://vuo.org/
[5] http://troikatronix.com/isadora/about/
[6] https://www.nodebox.net/
[7] http://www.audiomulch.com/
[8] http://www.plogue.com/products/bidule/
[9] http://audulus.com/

used example of this. Cinder[1] and openFrameworks, while featuring some audio functionality, are also primarily visually oriented (though their basis in C++ makes it more possible to integrate them with other libraries). As noted in section B.3, GLSL has also emerged as a platform for visual and audiovisual artistic experimentation (although it has broader roots and uses). While all these environments are oriented primarily towards visual production, they all can receive real-time input, making it possible to integrate them into systems involving parameter mapping, visualization, or sonification.

ChucK, C Sound,[2] and Supercollider[3] are text-based programming languages for sound and music that, while oriented towards audio, can similarly be integrated into audiovisual systems via real-time control messages.

Shadertone[4] is a very new language that integrates both GLSL and Supercollider, with the goal of providing an environment for audiovisual experimentation. Pixilang,[5] a language used primarily (perhaps exclusively) by the developer who created it, is also audiovisually oriented in a unique way, and can compile native code across many platforms.

By allowing the custom interconnection and integration of sonic and visual processes, artist-oriented programming libraries and packages have been instrumental in facilitating an explosion of contemporary experimentation with synaesthetic sound-image relationships.

B.7 Software for routing, mapping, and scaling control data

Existing software for routing, mapping, and scaling real-time control data is discussed in Chapter 5, as context and motivation for the development of new software for that purpose.

[1] http://libcinder.org/
[2] https://csound.github.io/
[3] http://supercollider.github.io/
[4] https://github.com/overtone/shadertone
[5] http://warmplace.ru/soft/pixilang/

B.8 Real-time audio analysis software

Existing software for generating control data from real-time audio analysis is discussed in Chapter 6, as context and motivation for the development of new software with the working title FFT Control.

Bibliography

"Black & White Video Scope." 2014. *Critter & Guitari*. Accessed April 7. http://www.critterandguitari.com/collections/video-synthesizers/products/black-white-video-scope.

Campen, Cretien van. 2009. "The Hidden Sense: On Becoming Aware of Synesthesia." *TECCOGS* 1: 1–13.

Carterette, E.C., and M.P. Friedman. 1978. *Perceptual Coding*. Handbook of Perception, v. 8; v. 236. Academic Press. https://books.google.com/books?id=JFIgAQAAIAAJ.

Castel, Louis Bertrand. 1757. *La Musique En Couleur*.

cyclo. 2013. *id#00*. https://vimeo.com/73860675.

Cymascope.com. 2014. "Cymatics History." Accessed April 6. https://cymascope.com/cyma_research/history.html.

Daniels, Dieter, and Sandra Naumann. 2010. *Audiovisuology Compendium: An Interdisciplinary Survey of Audiovisual Culture See This Sound*. Walther König, Köln.

Dolscheid, Sarah, Shakila Shayan, Asifa Majid, and Daniel Casasanto. 2013. "The Thickness of Musical Pitch Psychophysical Evidence for Linguistic Relativity." *Psychological Science* 24 (5): 613–21. doi:10.1177/0956797612457374.

Dupuis, Alexander. 2013. *No-Input Pixels*. https://vimeo.com/77643568.

"FreeFrame Open Realtime Video Effects." 2014. Accessed April 7. http://freeframe.sourceforge.net/about.html.

Gadalla, Moustafa. 2000. *Egyptian Harmony: The Visual Music*. Greensboro, NC, U.S.A.: Tehuti Research Foundation.

Gage, John. 1993. *Color and Culture: Practice and Meaning from Antiquity to Abstraction*. Berkeley: University of California Press.

Henke, Robert. 2014. "Lumière Technical Stuff." http://roberthenke.com//technology/lumiere.html.

Hochenbaum, Jordan Natan. 2013. "L'arte Di Interazione Musicale: New Musical Possibilities through Multimodal Techniques: A Dissertation Submitted to the Victoria University of Wellington and Massey University in Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Sonic Arts, New Zealand School of Music." http://muir.massey.ac.nz/handle/10179/4525.

Hodgin, Robert. 2014. "Magnetosphere iTunes Visualizer." Accessed April 7. http://roberthodgin.com/magnetosphere-itunes-visualizer/.

Ikeda, Ryoji. 2014. "Test Pattern." Accessed April 8. http://www.ryojiikeda.com/project/testpattern/.

"Introducing ISF Video Generators and FX." 2014. *VDMX*. Accessed April 7. http://vdmx.vidvox.net/blog/isf.

Kurokawa, Ryoichi. 2013a. *Oscillating Continuum*. https://vimeo.com/61348050.

———. 2013b. *Syn_*. https://vimeo.com/61347519.

Lerch, Alexander. 2012. *An Introduction to Audio Content Analysis Applications in Signal Processing and Music Informatics*. Hoboken, N.J.: Wiley.

Maurer, D., and C. Mondloch. 2004. "Neonatal Synesthesia: A Re-Evaluation." In *Attention on Synesthesia: Cognition, Development and Neuroscience*. Oxford University Press. http://psych.mcmaster.ca/maurerlab/Publications/Maurer_NeonatalSynesthesia.pdf.

McLaren, Norman. 1940. *Dots*. National Film Board of Canada. https://vimeo.com/32645760.

McLaren, Norman, and William Jordan. 1953. "Notes on Animated Sound." *The Quarterly of Film Radio and Television* 7 (3): 223–29. doi:10.2307/1209819.

Moritz, William. 1997. "The Dream of Color Music, And Machines That Made It Possible." *Animation World Magazine*, April. http://www.awn.com/mag/issue2.1/articles/moritz2.1.html.

Nitai Das, Jahnava. 2014. "Vedic Conception of Sound in Four Features." *Tattva Prakasha, Volume 1, Issue 7*. Accessed April 5. http://veda.krishna.com/encyclopedia/vedicsound.htm.

Nundy, Surajit, and Dale Purves. 2002. "A Probabilistic Explanation of Brightness Scaling." *Proceedings of the National Academy of Sciences* 99 (22): 14482–87. doi:10.1073/pnas.172520399.

Rees, Torben. 2009. "Chladni Plates: The First Step towards Visualizing Sound." In *Explore Whipple Collections*, edited by Whipple Museum of the History of Science, University of Cambridge. http://www.hps.cam.ac.uk/whipple/explore/acoustics/ernstchladni/chladniplates/.

"Rhythm Scope." 2014. *Critter & Guitari*. Accessed April 7. http://www.critterandguitari.com/collections/video-synthesizers/products/rhythm-scope.

Schwierin, Marcel, and Sandra Naumann. n.d. "The Musicality of Abstract Film." In *Audiovisuology Compendium*.

Shaw-miller, Simon. 2013. "Synaesthesia." In *The Routledge Companion to Music and Visual Culture*, by Tim Shephard and Anne Leonard. Routledge.

Spielmann, Yvonne. n.d. "Video, an Audiovisual Medium." In *Audiovisuology Compendium*, edited by Dieter Daniels and Sandra Naumann.

Thaler, Lore, Stephen R. Arnott, and Melvyn A. Goodale. 2011. "Neural Correlates of Natural Human Echolocation in Early and Late Blind Echolocation Experts." *PLoS ONE* 6 (5). doi:10.1371/journal.pone.0020162.

"Video Scope." 2014. *Critter & Guitari*. Accessed April 7. http://www.critterandguitari.com/collections/video-synthesizers/products/video-scope.

Walker, Peter, J. Gavin Bremner, Uschi Mason, Jo Spring, Karen Mattock, Alan Slater, and Scott P. Johnson. 2010. "Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences." *Psychological Science* 21 (1): 21–25. doi:10.1177/0956797609354734.