(BLANK)

Synaesthesia as a Model for Dynamic Media

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Arts in Design and approved by the MFA Design Review Board of the Massachusetts College of Art in Boston.

May 2009

Gunta Kaza, *Thesis Advisor*
Professor of Design
Dynamic Media Institute
Massachusetts College of Art and Design, Boston

Lisa Rosowsky, *Thesis Document Advisor*
Professor of Design
Chairperson, Communication Design Department
Massachusetts College of Art and Design, Boston

Jan Kubasiewicz
Coordinator of Graduate Program in Design
Professor of Design
Dynamic Media Institute
Massachusetts College of Art and Design, Boston

Special Thanks

To all of the people who have made this research possible:

Gunta Kaza, Jan Kubasiewicz, Brian Lucid, Joe Quackenbush & Lisa Rosowski.

Michael Golembewski, Ronald Bruce Smith, Ean White & Fred Wolflink.

Jaime, Dad, Mum, Theresa, Kris & my entire extended family.

The entire DMI community.

Rick McDermott & Jeff Keough.

In memory of Helen R. Owens & Kevin Fitzgerald.

Thesis Abstract

Synaesthesia & the Senses

14 Introduction
16 Your Senses
18 Animal Sense
20 Schools of Thought

History of Synaesthetic Media

24 Music and Color
28 Motion in Space
30 Film Music & Sound
36 Visual Music
40 Video
42 Music Video
44 Computer Sound & Image

Synaesthesia as a Model for Dynamic Media

48 Audio Mixing & Software
52 Hearing & Seeing
58 Hearing & Seeing as Complement
62 Effects: Visual & Auditory

Table of Contents

Case Studies

76 Overview
78 You Are Here Now
80 Panorama Scroll
82 Trevor Wishart Poster
84 Dynamic Volume Channel Studies
86 A.S. Mixer
88 Film Music Cues As a Way to Mark
88 Sound and Moving Image
90 cueTag
92 Sound Mandala
94 Track Color
96 A Size
98 Sound Shape Studies
100 ShapeMix

Bibliography

I much prefer the smell of real flowers to a can of artificial flower smells.

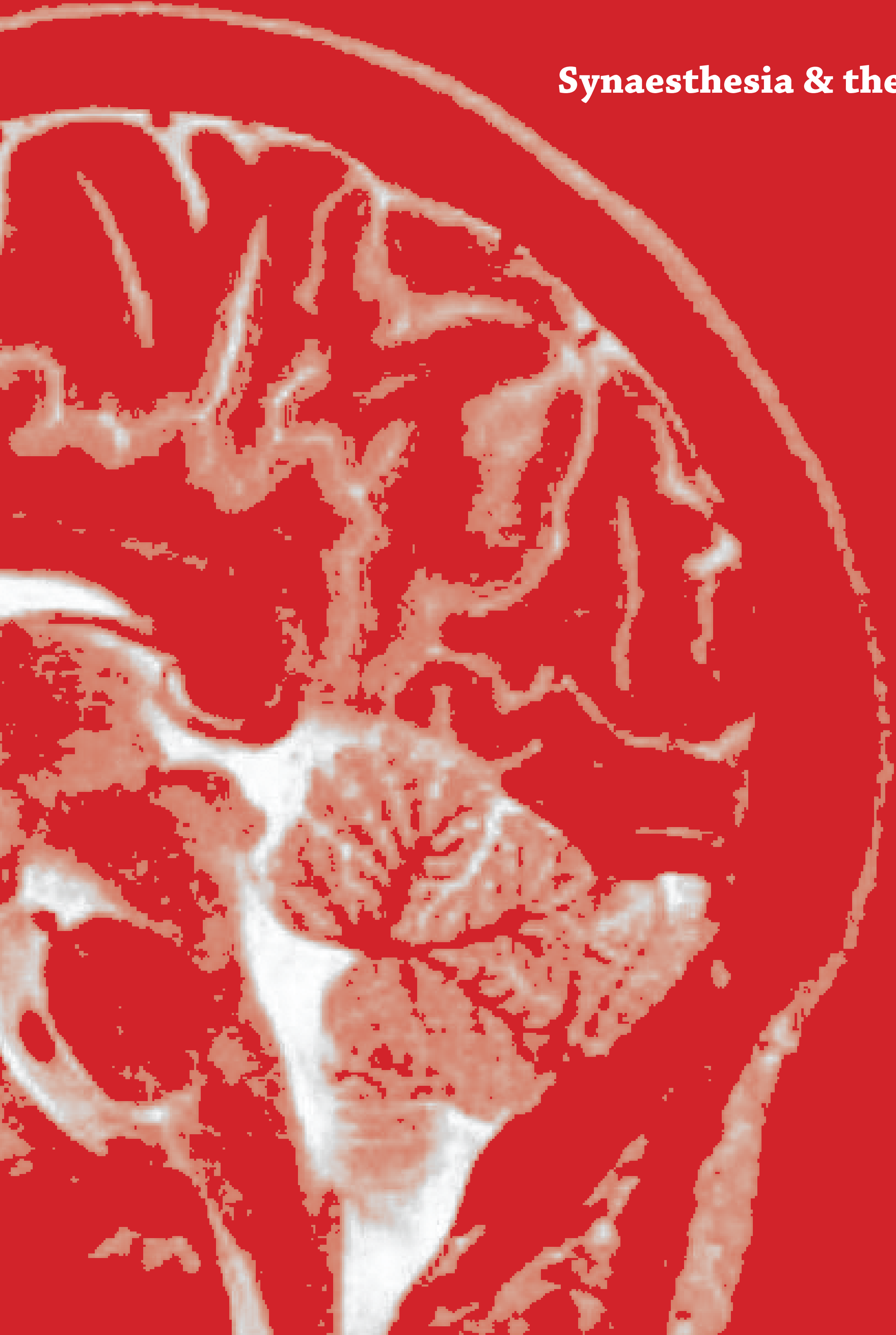
Thesis Abstract

Throughout human evolution our senses have evolved, and the tools we have invented have augmented our sensory exchanges with our surroundings. The extent to which we engage with the tools we invent, and how we use our senses to engage with them, has had a tremendous impact on our understanding of how these tools function.

This research examines the interrelationship between our senses as a means of more intuitive control of the computer-based tools we create and use. It challenges historical assumptions about audiovisual synaesthetic relationships, and proposes the adoption of perceptual relationships based on natural metaphor for building more useful experiences with these tools.

I have focused on the computer audio mixer interface in my research as an example of the type of human-computer interface that can be created by using natural motion and space cues, as well as other cues found in both the auditory and visual realm.

Synaesthesia & the Senses



Introduction

Reality has no particular form

–John A Waterworth, "Creativity and Sensation,
The Case for Synaesthetic Media."

What if you woke up one morning and you discovered you could hear colors? What if instead you thought certain numbers had personalities? If you could taste shapes? For a small population this is a permanent condition. It's called synaesthesia. Its Latin root words are "together" and "with the senses," meaning that in the person's mind, aspects of one sense are clearly associated with aspects of another sense.

My friend, Violet, has a form of synaesthesia that allows her to see whole numbers as colors. She said that learning multiplication tables was easier for some numbers than it was for others. The number 9 was a gorgeous color purple and that made all equations using the number 9 elegant. However, she didn't like the combination of the numbers 4 and 6, a leaf green and an indigo, because the resultant color was ugly to her.

We all have some form of synaesthesia or at least we did when we were very young and, if we held onto it, were able to develop strategies for memorizing and deciphering vast quantities of information. Cross-modal transference, a developmental form of synaesthesia, is linked to intelligence in early infancy. Intelligence is associated with the ability to identify a stimulus that had only been experienced through one sense by using another sense. "A baby who is able to recognize by sight a screwdriver that she has previously only touched, but not seen, is displaying cross-modal transference."[1] The level of this cross-modal transference by the time the infant reaches its first year correlates with intelligence scores later in life (Spelke, 1987; Rose, Feldman & Jankowski, 1999, 2003).

6 7 8 9 0

Our minds are capable of putting together powerful associations by the way we sense and how we perceive those senses. If infants who hold onto aspects of their synaesthesia are considered more intelligent later on, then synesthesia should be considered a useful model for how we work, the tools we use and the work we produce with those tools.

This thesis aims to use cross-modal sensory cues as a basis for creating those tools, specifically for computer software interface.

Your Senses

Our senses define the edge of consciousness, and because we are born explorers and questors after the unknown, we spend a lot of our lives pacing that windswept perimeter.

–Diane Ackerman, A Natural History of the Senses

Your world is a rich, sense-filled place. All around you, everywhere you go, from crowded subways to the dense jungles of the Amazon to the chilly Antarctic ice shelf, you depend on your senses to navigate, to sleep, to be self-aware. Without your senses, the world would be dark, soundless, tasteless and without odor. You would not know the comfort of a warm summer morning, the pain of a scalding tea kettle or the wet freeze of an ice cube. You would never itch. You would never know the difference between your toes and your pinky.

Imagine you are waiting for a subway train. With your eyes, you see the oncoming car approaching the platform. Your ears hear its click-clack, while the pressure from its exit from the tunnel manifests into a whooshing sound and a prickly feeling across the hairs on your exposed forearm. Your nose takes in the small chemical particles of the stale, moldy underground air and your taste buds confirm this. The train stops. The light next to the automatic door blinks in concert with the "ding dong" chime, alerting you that the doors are about to open. People exit. A blonde woman wearing a blue scarf exits, her perfume, Chanel No.5, fills the air.

In these five seconds you took in a lot of information. Your eyes received millions of images. The little hairs in your ears vibrated more than two hundred thousand times. The nerve endings on your skin were ready to feel pain in case the train doors accidentally closed on you. Your skin would also let you know if the air conditioning were a little too cold, if someone bumped into you or if your skin were a little dry. During this time, your senses worked in harmony, giving your body and mind cues on how to operate.

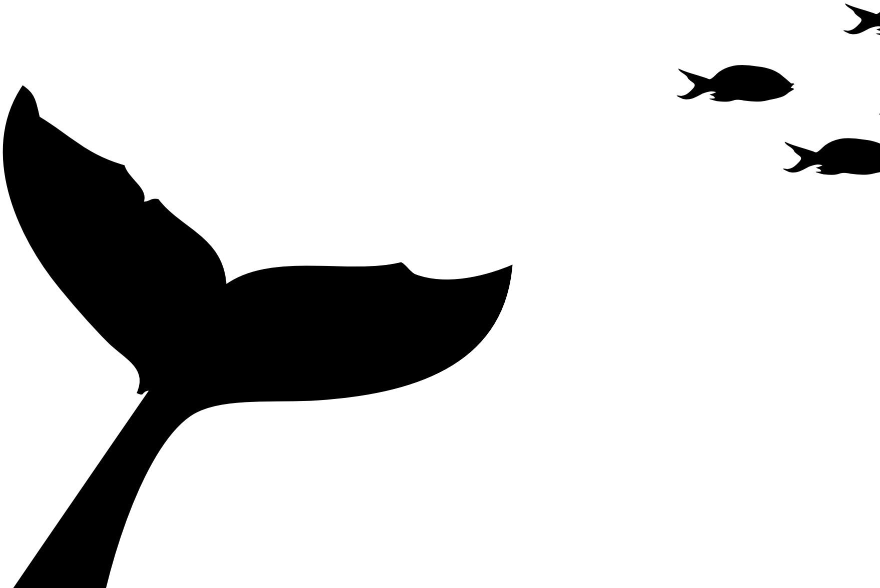
Many of our everyday phrases conflate the senses. When someone says "I see what you mean," do they mean they actually see what you have said, or it that also a product of what they have felt, heard, touched and interpreted? Likewise, the phrases: "something smells rotten in Denmark," "can't touch this," or "in your eyes?" are all a way of trying to breach the barriers between our senses?

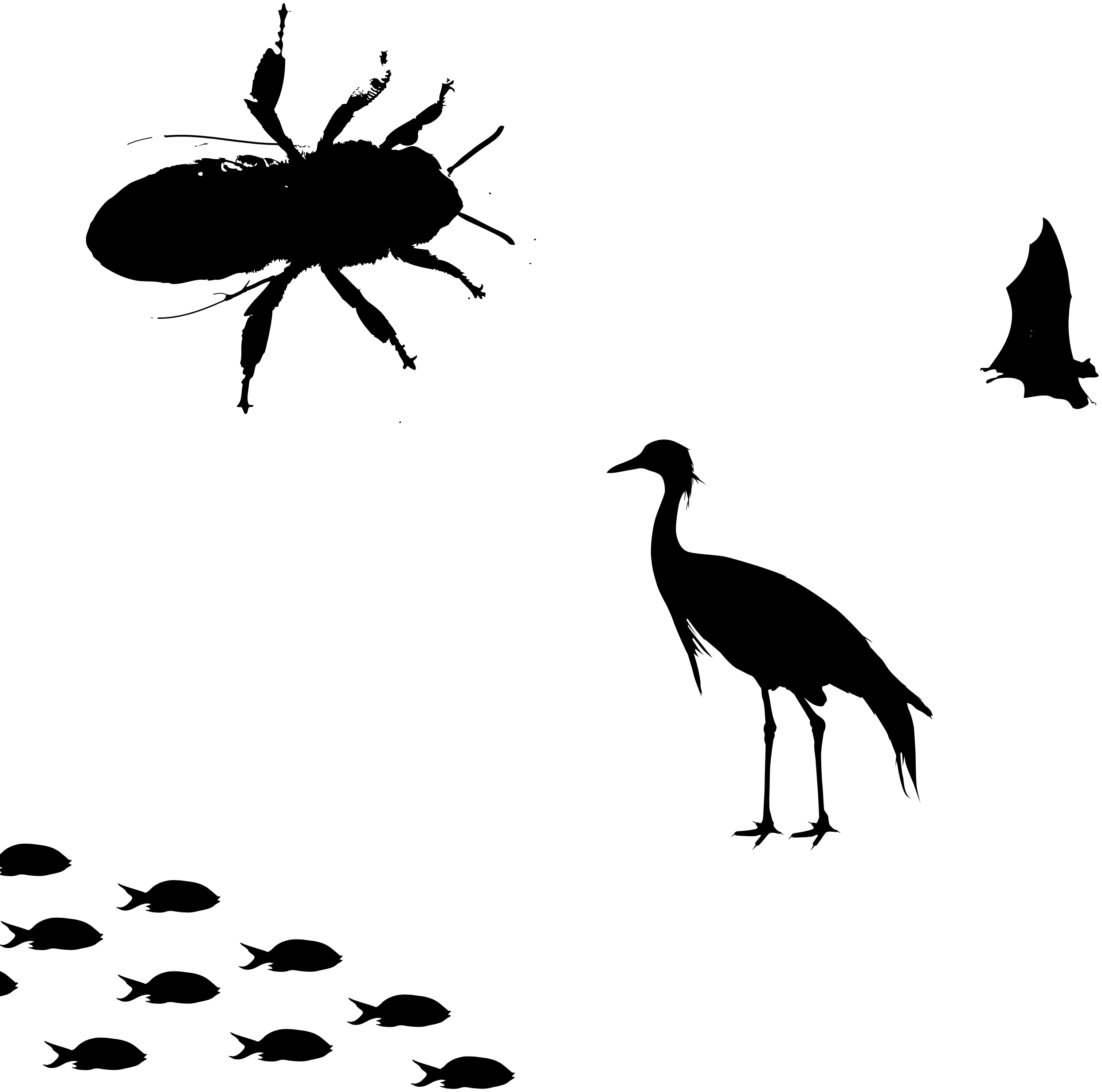
Animal Sense

We tend to take our senses for granted. Our animalian relatives rely on their senses for survival. Our human, mammalian body is just one model of many. Our hearing is good, but your next-door neighbor's Labrador Retriever can hear sounds over two miles away and track the blood of wounded prey one-quarter mile away. He can sniff out the scent of potential mates that have visited your back yard last week.

A common house cat can pinpoint the exact location of movement without turning her head. She can perch on precarious ledges and stay balanced with the help of her inner ear and her tail. She can jump up onto kitchen a table in near-perfect darkness. You, mere human, would probably bump into something and hurt yourself chasing her.

Whales can hear across oceans. Fish swim in schools and identify their depth with pressure and magnetic sensing. Sharks and bats use echolocation to orient themselves and seek out prey. The ancient jellyfish has a special form of orientation that keeps it upright. Insects pack a lot of sense in a small package. Bees can use polarized light to see in amazing detail. They can tell the difference between an immature, a ripe and a withering flower in less-than-ideal conditions. Flies have compound eyes which can detect motion lightening-fast, which is why you can't catch them.





Schools of Thought

But specialization is in fact only a fancy form of slavery wherein the "expert" is fooled into accepting his slavery by making him feel that in return he is in a socially and culturally preferred, ergo, highly secure, lifelong position. But only the king's son received the Kingdom-wide scope of training.

–Buckminster Fuller, Operating Manual for Spaceship Earth[3]

We have looked at how animal and human senses influence perception. How we have developed those senses into forms of expression gives us clues about how to put them back together.

We could examine any two senses, but hearing and vision are the two most talked about in the arts that it seems right to begin here.

We could devote our lives to the study of one sense and still not have enough time to skim the surface. Scientists and doctors often dedicate their entire careers to studying one sense organ, often building from their predecessors' knowledge. In the arts we have done the same; creating music schools for the study of sound and art schools for the study of the visual.

This specialization sets up a discourse between the schools of music and the schools of art. Specialization limits people to partial knowledge and practice and keeps them from understanding the broader picture.

Musicians and artists adopt strict rules that govern what parts of one another's domain they can use to enhance their own. Notation is a form of type used to govern the rules of a musical composition. Musicians use the visual cues of notation on sheet music to interpret music and play music. If they invent their own systems for writing their music other musicians can't read it. Artists interpret music in only the broadest of terms, as inspiration or in the context of gallery openings.

The performance of music is about delivering acoustic information to an audience through musical instrumentation and composition. The practice of music is detailed, centered around playing particular notes with particular instruments according a set of rules not unlike math equations.

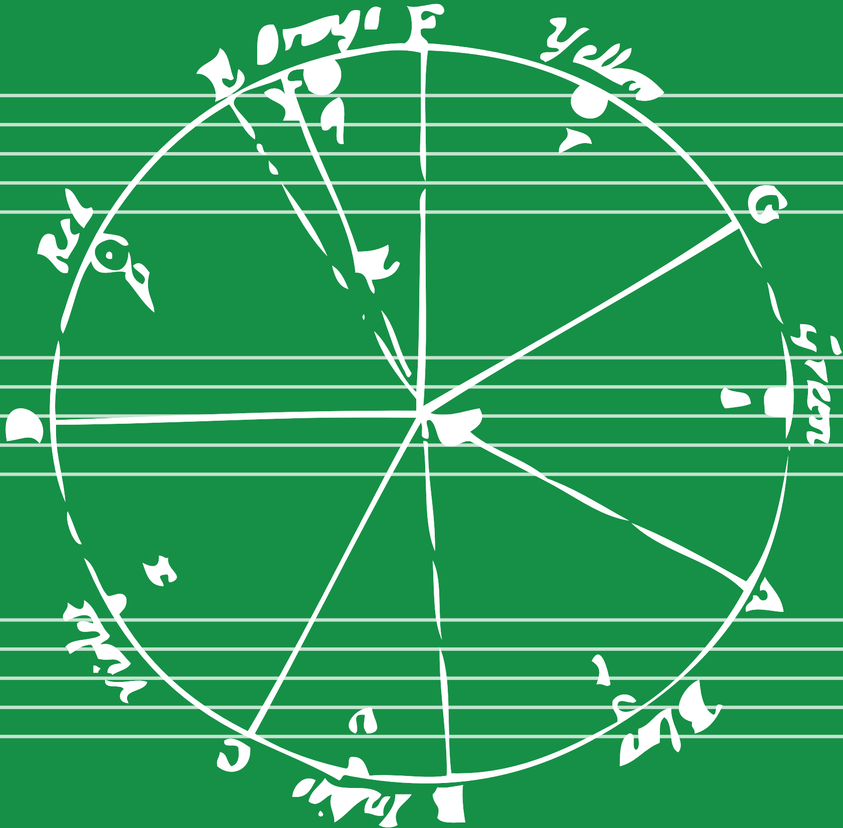
Time is central to music. When you go to a music concert or performance, you must arrive at a predestined time. The performance will then start at a specific time and will last for a predetermined amount of time. When the performance is over, if it was recorded, it can be played at a later time–however, it will never sound the same as it did when we heard it live.

Traditionally, in the visual arts, we engage in the study of two or three dimensions in static time. We create artifacts which are meant to be studied in the future in the same form as when they were made.

When you go and visit art, it can be on your own time, as long as the visited art is accessible. You can look at a piece of artwork from all its possible viewpoints and when you're done, you can do it again. Museums spend a lot of time and money to ensure that the Miro painting on the wall appears exactly as it did ten years ago.

The temporal differences between the artifacts of art and the performance of music make for difficulty in creating relationships between the the two. If music lives in time and artifact is supposed to be timeless, how do we begin to merge the two effectively? In what forms have music and art worked together over the development of these two arts?

History of Synaesthetic Media



Music and Color

Associating music with image has roots in color-music theory, proposed in different forms by many different people over several centuries. Pythagorus suggested that planets and stars were in cosmic harmony. Plato theorized that heavenly bodies had color and harmonic relationships according to their position in the sky. Leonardo Da Vinci sketched and wrote about the relationship of music to color.[4] Sir Isaac Newton too suggested a divine relationship between light and sound wavelengths in his treatise "Opticks."

In the early 1720s, Louis Bertrand Castel took Newton's proposal suggesting that visible music could be produced by associating seven colors with seven keys on a harpsichord, corresponding to seven units of scale. He attached colored tape to each key and placed a candle behind the keyboard [5], so that when he pressed each note, colors would illuminate from within the instrument.

I procured an organ, and experimented by building an attachment to the keys, which would play with different-colored lights to correspond with the music of the instrument. *–Louis Bertrand Castel*

In 1893, Bainbridge Bishop published a paper that described his construction of at least several color organs:

His paper and the examples he uses from his organ describe not only colors, but tints and shades to describe harmonies and tone in music.

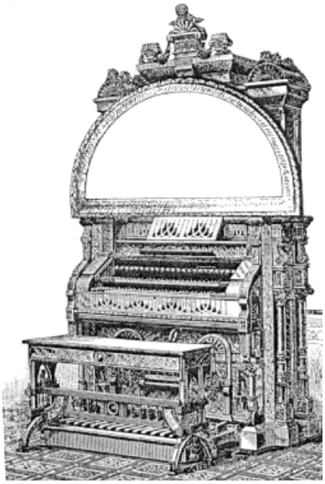
In 1895, Alexander Rimginton, a maker of color organs, published the paper *A New Art: Colour-Music* [6] noting his own scheme for associating color according to the rainbow. He described his new art (we're assuming he didn't know about Bishop) as a combination

Left: Bainbridge Bishop's color organ.
Middle: Alexander Rimington next to his color organ.
Right: Paul Klee's polyphony.

among “three novel elements into the use of colour-viz. time, rhythm, and instantaneous combination.” He continues: “If we measure the rate of vibration at the first visible point at the red end of the spectrum, we shall find it is approximately one-half what it is at the extreme violet end.” [7]

Early 20th century composer Alexander Scriabin associated color with tonality in his piece *Poem of Fire*. According to biographer Leonid Sabaneyev “he juxtaposed the ‘allied colors’ (arranged in spectrum) and the ‘allied tonalities’” arrange in the circle of fifths, a way of working with harmony. [8] Scriabin’s method brought up yet another question about what metrics to use between light and sound.

Painters have also tried to describe the relationship between color and music. In 1911, Wasilly Kandinsky painted his *Impression III* after attending his friend Schoenberg’s concert the previous night. He claimed that his series of music inspired paintings, including one name *Green Sound*, were meant to be “heard” by the viewer. In the same year in which Scriabin’s debut of *Poem of Fire* debuted, Paul Klee suggested the idea of picture polyphony. He noted: “Underlying such art there must be some sort of structured order...defined format, a system of articulation and rules to be both strictly observed and departed from.” [9]



There have been many experiments, but the relationship between color and note does not appear to be unanimous. There are too many versions of the note-to-color theory and no real consensus, principally because notes and color systems differ from culture to culture and system to system, and from synaesthete to synaesthete.

There are differences in the way music and color are measured. Mathematically, color could be calculated by additive, subtractive, RYB, Munsell's system, by temperature, saturation or any number of parameters. Music scale could be calculated against Western concert pitch, Indian 22 tone śruti scale, the gamalon 5 tone sléndro scale, or one of many others. Who is to say that any one system of note or color is right?

If the use of color is arbitrary, its use is practical to help differentiate between musical notes. Separation of color is more important than what color is chosen; almost any color system will do.

Musical associations do not necessarily need to be notes. They could be notes or collections of notes. So, as long as we can identify between sounds or groups of sounds, it does not matter and almost any musical scale will do.

C	C#	D	D#	E	F	F#	G	G#	A	A#	B		
🟥	⬜	🟧	⬜	🟨	🟩	⬜	🟦	⬜	🟪	⬜	🟪	*1704*	Issac Newton
🟦	🟩	🟩	🟩	🟨	🟨	🟧	🟥	🟥	🟪	🟪	🟪	*1734*	Louis Bertrand Castel
🟦	⬜	🟪	⬜	🟥	🟧	⬜	🟨	⬜	🟩	⬜	🟩	*1816*	George Field
🟥	🟧	🟧	🟨	🟨	🟩	🟩	🟦	🟪	🟪	🟪	🟪	*1844*	D.D.Jameson
🟫	🟥	🟧	🟨	🟨	🟩	🟩	🟦	🟪	🟪	🟫	⬛	*1881*	Theodor Seemann
🟥	🟥	🟧	🟧	🟨	🟩	🟩	🟩	🟩	🟪	🟦	🟪	*1893*	A. Wallace Rimginton
🟥	🟥	🟧	🟨	🟨	🟩	🟩	🟩	🟪	🟪	🟥	🟥	*1893*	Bainbridge Bishop
🟨	🟩	🟩	🟦	🟪	🟪	🟪	🟥	🟥	🟥	🟥	🟧	*1910*	H. von Helmholtz
🟥	🟪	🟨	🟦	🟦	🟥	🟦	🟧	🟪	🟩	🟦	🟦	*1911*	Alexander Scriabin
🟥	🟥	🟧	🟧	🟨	🟩	🟩	🟩	🟦	🟪	🟪	🟥	*1930*	Adrian Bernard Klein
🟥	⬜	🟧	⬜	🟨	⬜	🟩	🟩	⬜	🟦	🟪	🟪	*1940*	August Aeppli
🟥	🟧	🟧	🟨	🟨	🟩	🟩	🟩	🟦	🟪	🟪	🟥	*1944*	I.J. Belmont
🟩	🟩	🟩	🟦	🟪	🟪	🟪	🟥	🟥	🟧	🟨	🟨	*2004*	Steve Zieverink

Color scales according to famous synaesthetes and progenitors of synaesthesia.

Motion in Space

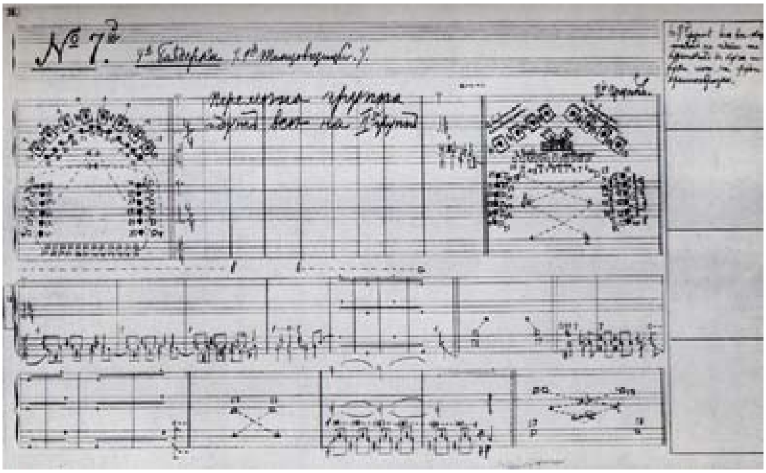
The use of space and motion could be used to combine image and sound in a synaesthetic way. Movement in natural space creates changes in sound and image in harmony. Dance, theatre, film, video, video games and the tools that enable their creation are examples of effective use of sound and image in motion.

There are notable (and beautiful) examples of music as integral part of the motion of dance. Indian Bharat Nayam is a two thousand year-old form of traditional dance, carefully choreographed to the movements, changes and rhythms of off-stage drummers and singers. Ballet, too, incorporates carefully choreographed music as part of theatrical performance.

There are too many variations in choreographed body movement and musical arrangement for dance performance to create repeatable, measurable associations between visual and auditory cues. Even

when the musical accompaniment remains the same from production to production the interpretation of motion can vary from dancer to dancer and from choreographer to choreographer. If there were a ballet, dance (or opera or theatrical) performance that were synaesthetic, the associations would have to be scripted in detail, note-for-note and movement for movement so that each movement would correspond to a note, its motion in a repeatable form. We haven't seen this yet.

The inventions of film, video and, most significantly, the computer made synaesthetic experimentation possible where stage acting and dance could not, and moved us closer to understanding how we can make meaningful (and lasting) associations between sound and image.



Left: Bharat Nayam dancer Dr. Minati Mishra.
Bottom left: Ballet Dancer Corinne Blum.
Bottom right: The Ballet Bayadere in Stepanov Choreographic Notation.

History of Synaesthetic Media

Film Music & Sound

Music in film may seem natural, but real life doesn't have a soundtrack.

The history of sound on film is important to our understanding of how sound and moving image can relate to one another in our modern context. Recorded or mechanical media can offer a repeatability that stage performance cannot.

The emergence of sound on film is important to the synaesthetic experience not only because of the obvious marriage between sound and image, but because sound gave cues about the film's narrative to the audience. The invention of sound film marks the first time sound and image could be recorded and repeatedly played back together.

The newly created sound-image medium allowed creators to experiment with the two senses in a way that had never been explored before. It could be experimented with, and quite remarkably, played back and repeated. Film introduced the time axis that was missing from earlier experiments in synaesthetic media. Through motion we could start to understand how the two senses could inform one another.

Peter Lorre in character for 'M,' arguably the first film where sound played an active role in the narrative of the cinematic story.

Kamillen

Silent Film

Silent film was a natural extension of the vaudeville acts which preceded them. To theatrical narrative they added the repeatability of film.

Early on silent film was accompanied by music. People would go see silent films as much for the music as for the film itself, especially if the film had low production value. From the large orchestras of the big cities to the single organ or piano players of the small towns, music played a role in conveying emotion where moving image could not alone. It also had a lesser known practical role: to drown out the projector's loud mechanical sound.

There were several basic kinds of music for silent film.[10] Purely improvisational, the first kind usually involved a solo player on a piano or organ, who followed the film and played music according to the mood of the scene. Prearranged scores worked well with larger arrangements of musicians. Most of the compositions came from stock pieces of music that were jury-rigged to each film. It was quite possible the same pieces of music could be used for a variety of films.

Another kind closely resembles what we now regard as film music. A continuous score of music was composed to match a specific film and its scenes by a composer who worked closely with the film director. The music would then be played as it was conceived. This kind of music was rare, as arrangements (and ability) differed from theatre to theatre.

Mechanical Sound on Film

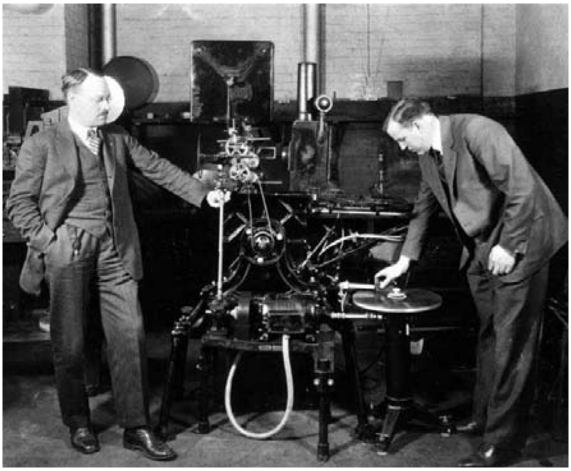
Sound-on-film is particularly important to our view of image and sound in motion because it put the two in the same place, reliably, at the same time. Music, not dialogue or sound effects, played a primary role in the history of recording and playback technology of film and greatly influenced the working mode of those trying to create associations between sound and image.

Technically there simply was no easy way to synchronize the playback of sound with film, never mind recording dialogue or natural sounds without the continuous noise of pre-war film cameras. Most early recording attempts were made by lip-synching the recorded voice with a filmed sequence.

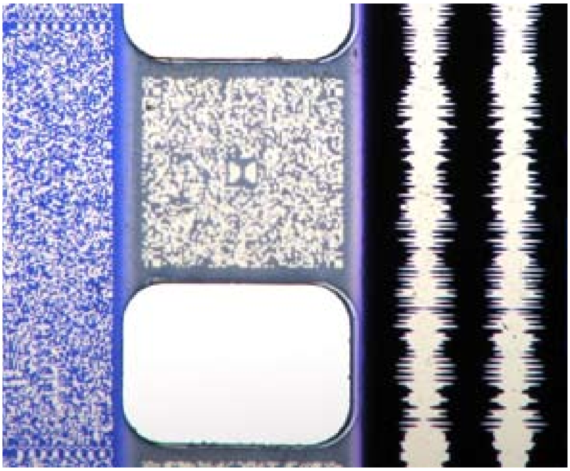
The Vitaphone, a direct decedent of Thomas Edison's rotating cylinder, made its appearance at precisely the time when live music at film theatres was at its artistic grandiosity. Good press and Warner's commitment to eradicating live music made the Vitaphone so commonplace that by the time a groundbreaking film would appear on its medium, there would be no going back.

That film was The Jazz Singer, in 1927, starring Al Jolson. The film contained only a marginal amount of dialogue. Its popular musical content and the words "Wait a minute, Wait a minute, you ain't head nothin' yet!" wooed the audiences. The success of this film prompted a lot of theaters to wire for sound and put a lot of unhappy theatre musicians out onto the streets. [11]

While Warner used the Vitaphone, a contraption that was prone to skipping and falling out of sync with film, Fox opted for "sound-on-film." This was the version all film studios would adopt eventually. It worked by placing an optically read track on the side of the film, side-by side with the image. For every frame of film, there would be an inch or so of mono optical track.



E. B. Craft demonstrates Vitaphone, from AT&T Archives.



The optical track, on the right of the sprocket, shows the amplitude of the signal.

Film Music

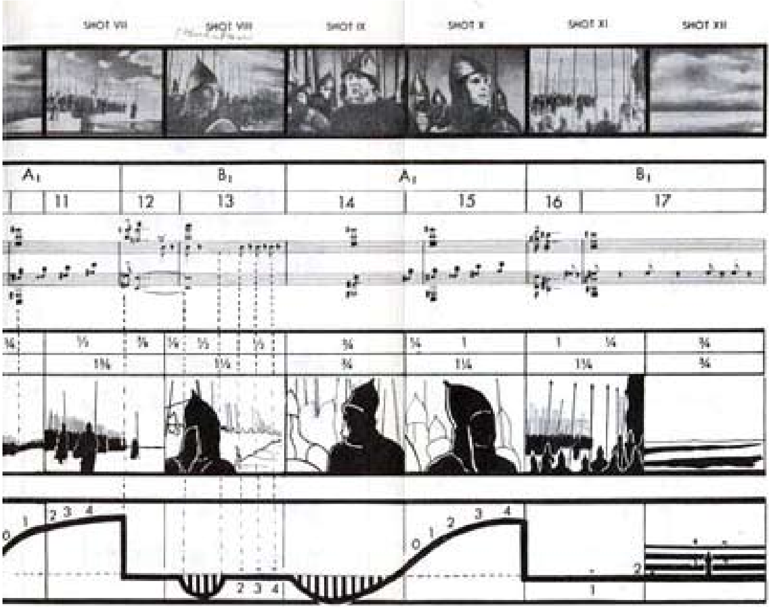
After the introduction of the "talkies" music and sound would develop different roles in film. Musicals continued to place music to the front. Animation and dramatic releases would use music to imitate speech, give ironic commentary, evoke the internal emotions of character, or contrast the scene. [12]

Film music and image work together in ways that vary from dominance over one another to a middle ground, or synthesis. In 1941, critic Paolo Milano suggested that music had relationships with film image that ranged from subordinate to dominant. On one end of the spectrum, the image would take the reins while music remained neutral. On the other end neutral images would subordinate to very powerful and emotional music. Somewhere in the middle, there existed what Milano calls a "counterpoint" where music and image would operate at aesthetic equality. [13] The image and music formed a medium synthesis giving credence to neither as dominant narrative. Films, such as Disney's 1940 film *Fantasia* fall into this category.

Fritz Lang's 1931 film *M* marked the first time sound (actually music) played the most crucial role in a film. During the course of the narrative, the killer whistles the *In the Hall of the Mountain King* from *The Peer Gynt Suite*. Later, he is identified by a blind balloon seller by that very same whistle. If this were a silent film, the cue would have been lost or the opportunity to use sound as a cue would have never happened.

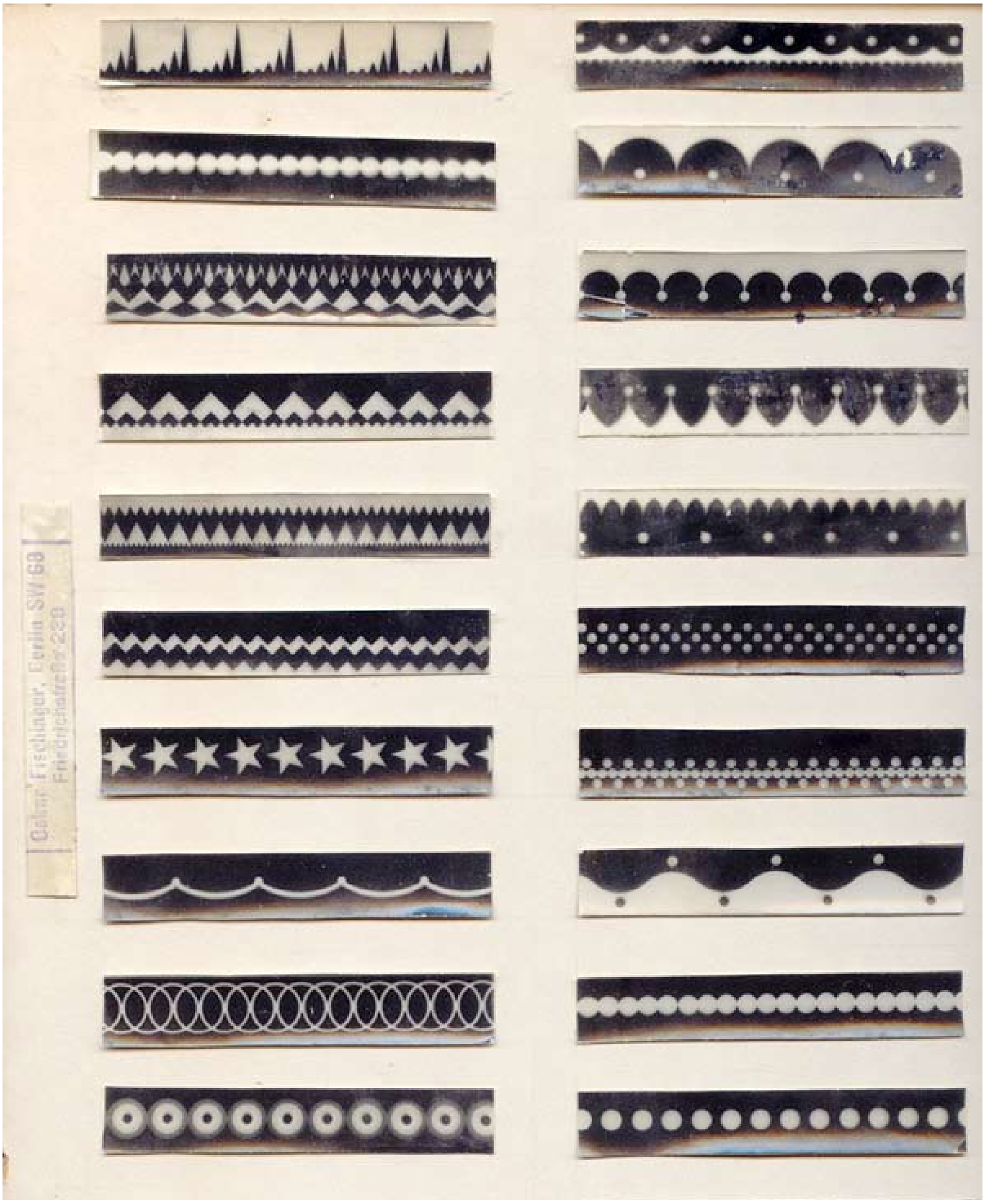
Sergei Eisenstein's editing style provided a vehicle for music to play a primary role in delivering narrative. He experimented with forms of film editing that he termed "montage," a way of cutting short sequences of images to show a passage of time. He would often use music to "express the movement of the rhythm of the motion picture."[14] He collaborated with composer Prokofiev in 1937 on the film *Alexander Nevsky* and through his montage style edited a battle sequence to fit the pre-composed piece of music. The score for the film was built collaboratively, with pictures of the storyboard alongside the notes of the staff. This was the first time an audience saw this kind of music and image collaboration.

The previous year, Prokofiev had completed a children's story, *Peter and the Wolf.* It created relationships between instrumentation and character. The main character of Peter was depicted by a string quartet and his deputies the bird, the duck, the cat among others, were represented by the flute, the oboe and the clarinet, respectively. Throughout the piece each characters' instrument responds to spoken narration as a form of dialogue that informations the viewer of major narrative events in the story.



Score Eisenstein used to chart shots against musical score in 'Alexander Nevsky.'

Visual Music



Oskar Fischinger's optical designs.

At around the same time sound was introduced to film, animators started to work with music and sound.

One of the most visible of these animators was Oskar Fischinger, whose animation was created to respond to prerecorded musical compositions. Sequences of shapes in various colors flew, stuttered and disappeared according to movements of the music. Groups of triangles, for example, represented string sections, while major movements would change the scenes and usher in other colored shapes. [15]

Along with the works of Mary Ellen Bute and Norman Mclaren, these were the earliest animated visual representations of music. These works were painstakingly drawn and painted by hand and then shot onto film frame by frame.

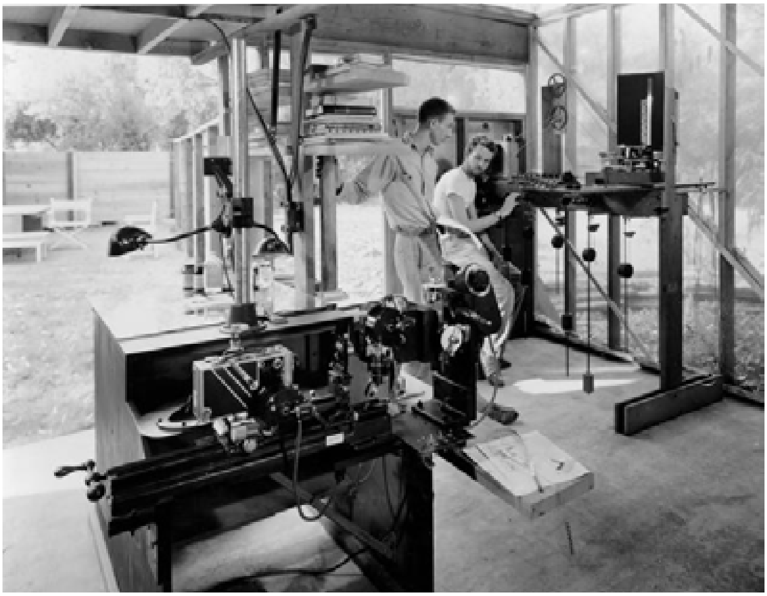
Although these early attempts were excellent and beautiful examples, the problems with synchronous representation and imperfection of hand artistry did not create accurate, consistent relationships between sound and the image.

The pattern of concentric wave-circles which was often used in cartoon and silent film iconography to represent the ringing of a door or alarm bell actually produced a buzzing clang sound when drawn in long rows and photographed onto the soundtrack area. *– Oscar Fischinger*

Around 1932 Rudolf Pfenninger and Fischinger began to experiment with drawing right on the optical track of the film to produce sounds. It was Pfenninger who first discovered that drawing wave-like forms onto the optical margin of the film would produce pure sine wave tones, similar (but arguably more electronic) to the sound produced by tone wheel organs. He did this because he wanted music but couldn't afford to record musicians. He, and later Fischinger, developed sets of drawings that read from the optical pick-up in a film projector, would produce simple sine tones. These drawings would form compositions. Depending on what note was needed, he would take the corresponding sine wave and optically print it. This was what he called "hand drawn sound. " [16]

Laslo Maholy-Nagy took this idea and photographed the hand drawn sound as the visual counter-track in his piece *Tönendes*, so that one could "see the same forms that one was also hearing." [17] Although this form of synaesthetic cinema could be called the most pure form of visualizing sound it was difficult to watch. Critics cited the form and sound combinations as "mechanical, almost soul-less." [18]

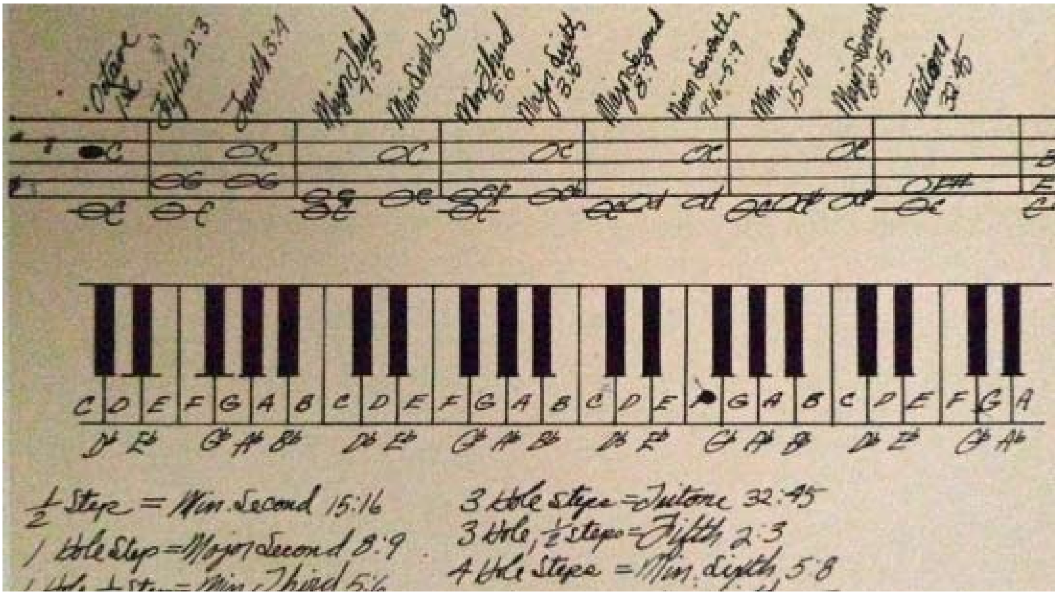
As students, John Whitney and his brother James created a series of music and image animations called *5 exercises* for the 1949 Experimental Film Competition in Belgium. They conceived of the project using a single score, a technique they developed to handle the synchronization of sound and image. In one part of their studio they had a 16 mm optical printer, and just behind it, an instrument constructed of pendulum bells. These two "instruments" were linked mechanically by an optical wedge[19], a valve that recorded the pendulum's shape as a sound, similar to Maholy-Nagy's technique, but more in the spirit of photography meeting analog computer.



John and James Whitney in their workshop.

At the heart of John and James’ work was the concept of visual harmonies, ways of creating shapes and lines with light that correspond to musical harmonies using the temporality of film. They searched for analogies in the natural world as well as current ideas in physics that gave them clues to how these relationships might manifest themselves.

This work, as well as the theories and work of others in this chapter has lived in in the works of video and computer artists. As the technology matured, artists like John and his brother James seized the opportunity to put their ideas of light and image relationship to the test.


Octave 1:2
Fifth 2:3
Fourth 3:4
Major Third 4:5
Min Sixth 5:8
Min Third 5:6
Major Sixth 3:5
Major Second 8:9
Minor Seventh 9:16 - 5:9
Min. Second 15:16
Major Seventh 8:15
Tritone 32:45
C D E F G A B C D E F G A B C D E F G A B C D E F G A
D♭ E♭ G♭ A♭ B♭ D♭ E♭ G♭ A♭ B♭ D♭ E♭ G♭ A♭ B♭ D♭ E♭ G♭ A♭
½ Step = Min Second 15:16
1 Hole Step = Major Second 8:9
1 Hole ½ Step = Min Third 5:6
3 Hole Steps = Tritone 32:45
3 Hole, ½ Steps = Fifth 2:3
4 Hole Steps = Min. Sixth 5:8

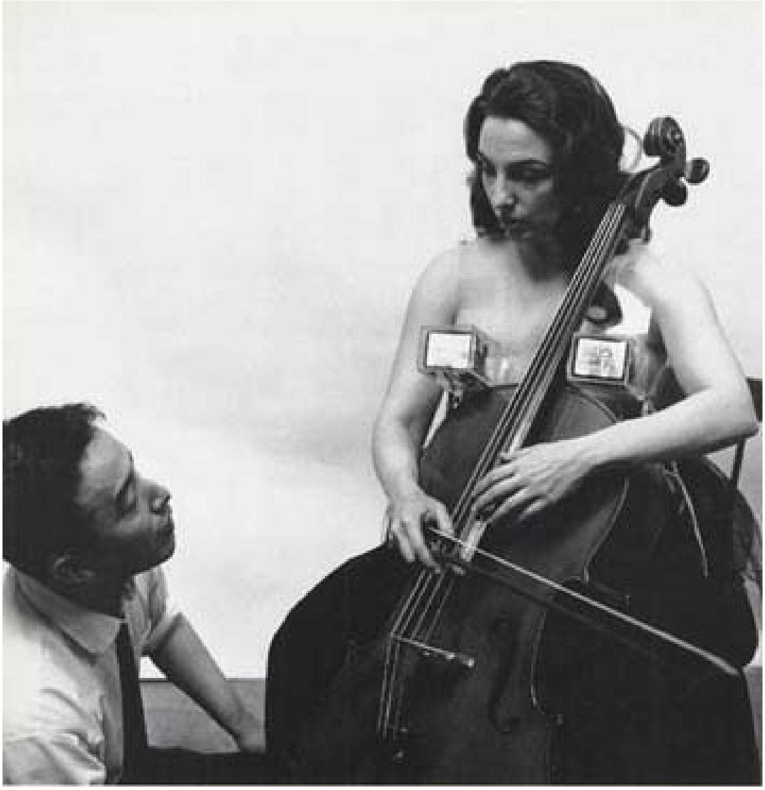
One of many examples of the Whitney’s synaesthetic scoring technique.

Video

Video, like film, combines image and sound on a single medium. Film editing was originally a laborious process that involved a recut of both optical and film tracks and then bringing them back together in a final "mix." Video could be cut and recut without remastering. Video had its drawbacks, though. Constant reediting resulted in video and audio degradation.

There have been a few notable video artists. One was Nam Jun Paik, who brought television into the gallery. Paik created multi-screen, multi-image video sculptures that ran loops on dilapidated TV monitors, sometimes with sound. His 1961 *TV Bra for Living Sculpture* combined two small television screens attached as a bra to violist Charlotte Moorman. As she played, the images modulated to the tones of the music, an example of an early dynamic piece.

Terry Riley's 1968 *Music With Balls* combined video, kinetic sculpture and music. Riley's music was prerecorded saxophone and organ tones played as beats. This was in turn played back on speakers embedded in two large spinning black spheres, with a silver pendulum in the middle.[20] The circular motion of the music and sculpture was filmed, recorded, transferred to video tape and edited. The result was a truly synaesthetic video tape, one that dealt with the Acoustics and vision of physical space combined with the motion of the kinetic structure.



Nam Jun Paik and Charlotte Moorman performing ‘TV Bra for Living Sculpture.’

Nam Jun Paik’s TV Cello.

Music Video

Michael Gondry's video for Daft Punk's 'Around the World.'

Music videos provided a chance to marry music and image in a way that had never been explored before. Unfortunately, most of the output was nothing more than a marketing vehicle for pop stars, resembling a television commercial than a form for self expression. Most music videos portrayed band members pantomiming song lyrics and instruments.

As music video matured, there were few creations that brought the motion of sound to the small screen. Peter Gabriel's 1986 video *Sledgehammer* used stop-motion animation to illustrate the text of the lyrics. The Art of Noise's video for *Close to the Edit* utilized Fischinger-style animation and live action to accompany the audio.

There were video directors and musical acts who extended the language of visual music beyond illustration. They broke the convention of focusing on the band or some aspect of the music and used the visuals as the vehicle for telling the musical story.

Film director Michael Gondry directed two excellent examples of music videos that extended the language of music into vision.

Gondry's 2002 *Star Guitar* was a video that appeared to be a continuous shot of video shot out of a passenger train in time with the music. Catenary poles marked the rhythm track while retaining walls signified the introduction or ending of a synth pad. The video was a meticulously edited recombination of footage based on a pre-composed visual score that the director developed.

Before *Star Guitar* Gondry produced a video for Daft Punk's 1997 song *Around the World*. It was an homage to Fischinger as interpretive dance. Members of the dance group descended a staircase, mimicking the melody line of the song. As the song progressed, dance group members moved to specific parts of the music. Mummies advanced across the screen to the beat. Robotic words "around the world" were complemented by four members dressed in robot suits moving stiffly.

Emergency Broadcast Network (EBN) were a group in the mid 90s that used sampled video footage from political propaganda to create lyrics and musical accompaniment. Each musical sample had a video complement as its source and their work was every bit as visual as it was auditory (and the other way around). Later, they developed a first-of-its-kind video sampler that, when played through a keyboard would display the companion video to the audio.

Video work, like its predecessor film, is a difficult medium with which to create repeatable relationships between moving image and sound. EBN's video sampler was the natural evolution to their video tape-edited methods and presented an opportunity to marry the two. Not only was this evolution important to the precise timing they needed to produce their audiovisual experience in real time, it utilized the computer as the vehicle to create repeatable experience where video could only approximate–at best.

Computer Sound & Image

The computer presents an excellent opportunity to combine sound and moving image in a repeatable, precise and reliable method. Media contained within the computer can be stored and retrieved, yes, but the most important aspect of this *New Media* is that it can be programmed to manipulate several outputs based on one or more inputs.

If watching a film or video is a passive viewing experience, nature is anything but passive. The computer is an extension of our curious nature. It's idle until we tell it to do something. It needs human input. It can do anything it's capable of carrying out, but it's useless if no one's on the receiving end. There are computers that don't need human interaction in order to carry out a task, however, a human will always be the audience of its outputs, and the outputs of that storage–that data, whether it's in the input or output stages–is all sensory.

There are many examples of sensory inputs and outputs using a computer. A computer keyboard is an example of a tactile interface. It's a direct descendent of a typewriter, the touch method for translating speech (audio) into text (written word). The computer translates this into characters that can be interpreted and stored into memory and displayed for immediate feedback to the inputer.

A car's computer is a kinesthetic emulation of our body: It tells you when its oil needs changing, when it's going too fast, or when one of its doors is open (ajar!). Small computers in smoke alarms take data in the form of CO_2 detection and when it reaches a certain threshold, a red light will blink and the speaker will produce a high, piercing noise until the threshold is diminished.

Perhaps some of the best examples of software programs that try and emulate our real world perception are video games. Sound has always been a part of commercially available "video" games.

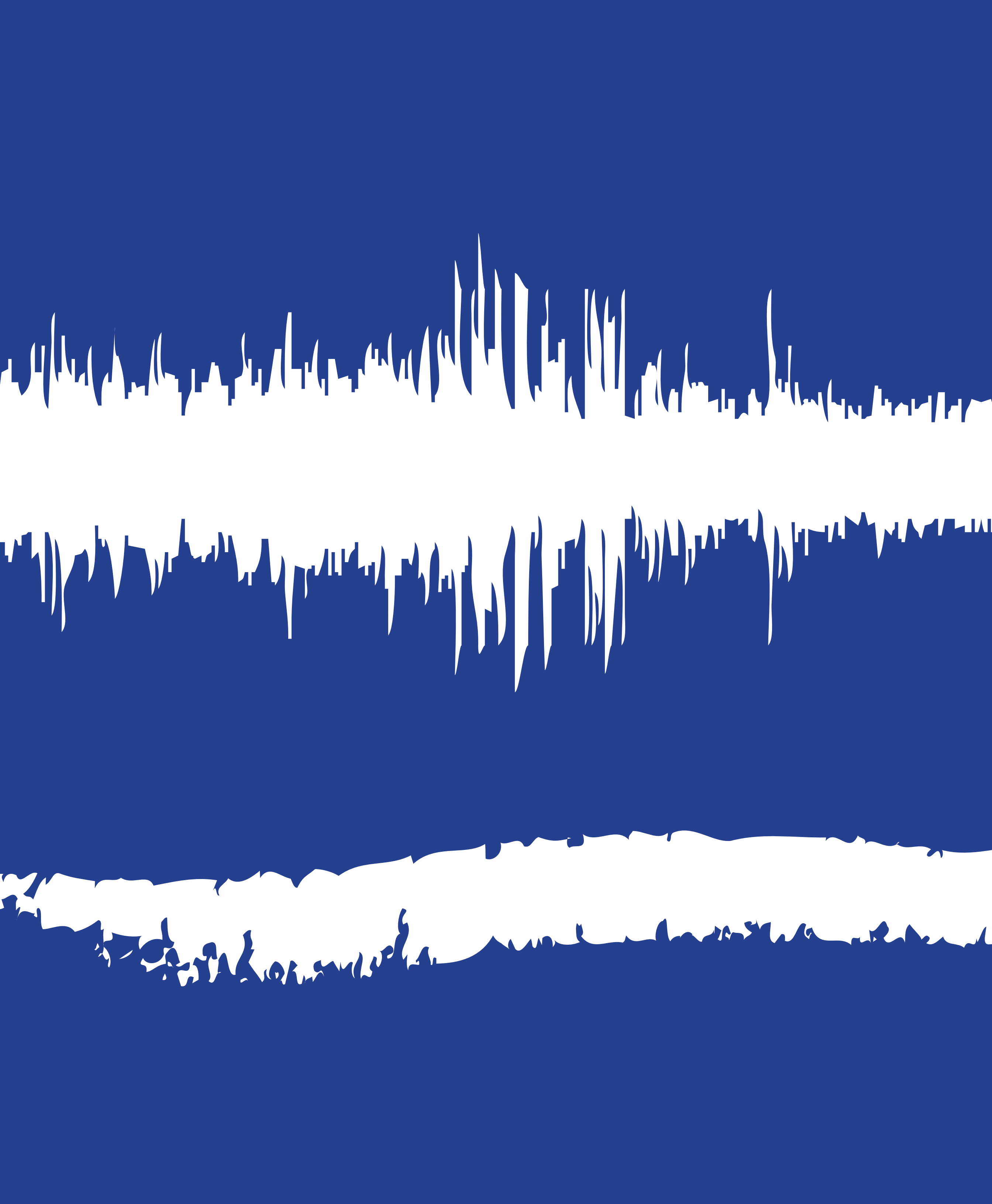
Pong, the first commercially available video game, used sound as a feedback mechanism. When a player successfully paddled the ball, a bell sound would play in a particular key. Hitting a wall would produce another sound. A missed pong would produce a higher bell.

Modern first-person-view games used our two dimensional space to emulate a three dimensional view (and sound) of the world. Early first person video games such as Doom brought us 3D space without 3d sound. As the technology became available, sound went into 3D space. Modern video games like Bungie's Halo 3 provide an opportunity to present noisemakers in natural 3D space relative to the gamer's current position.

3-73
THE NEWEST 2 PLAYER
VIDEO SKILL GAME
PONG
from ATARI CORPORATION
SYZYGY ENGINEERED
The Team That Pioneered Video Technology
PONG
FEATURES
• STRIKING Attract Mode
• Ball Serves Automatically
• Realistic Sounds of Ball Bouncing, Striking Paddle
• Simple to Operate Controls
• ALL SOLID STATE TV and Components for Long, Rugged Life
• ONE YEAR COMPUTER WARRANTY
• Proven HIGH PROFITS in Location After Location
• Low Key Cabinet, Suitable for Sophisticated Locations
• 25¢ per play
THIS GAME IS AVAILABLE FROM YOUR LOCAL DISTRIBUTOR
Manufactured by
ATARI, INC.
2962 SCOTT BLVD
SANTA CLARA, CA 95050
Maximum Dimensions:
WIDTH 26"
HEIGHT 56"
DEPTH 24
SHIPPING WEIGHT: 150 Lb.

Screen-based video games (and its passive cousin film) don't actually portray three dimensions. The computer monitor is a two dimensional surface that displays three dimensions with limited peripheral viewing. You can locate a noisemaker in the space you can see, but once the noisemaker has exited the screen you can only hear them.

As of the writing of this thesis we have not created an output device capable of giving us three dimensional viewing. *That will change.*



Synaesthesia as a Model for Dynamic Media

Audio Mixing & Software

Computers combine things to make new knowledge at such high speed that we cannot absorb it. They affect not just the things we buy or the things we know, but the things we do.

–Alvin Toffler

The tools we use to create our environments have a direct influence over how we treat image and sound together. As we have discovered, if all we have are candles and a harpsichord and some colored vellum, we're only going to be able to create so much. We must take a good look at the tools that create our environments to get a better sense of how to deal with the intersection of vision and sound.

For example, the way we mix audio in software and how we perceive sound-making for media has a powerful influence over the output. It's a bit of a chicken and egg problem. If we're trying to create an environment that fundamentally changes the way we perceive our media, we have to also change the tools that create that environment. Audio mixing is one of those fundamental tools.

There are several steps to digital multitrack recording and mixing. The computer is both a recorder and a mixer, emulating the tape recorder and mixing desk of days past–in one place. A sound passes into the computer through analog to digital convertors from a microphone or directly from an instrument. Sounds are recorded onto different tracks to keep them separated until mix-down. Sounds are then edited and then mixed using different types of electronic or psychoacoustic effects. Most programs even have music composition sections that play virtual instruments or trigger events to external synthesizers via MIDI (Musical Instrument Digital Interface).

The software interface to the recording, composition and playback section of programs such as MOTU's Digital Performer and Avid's Pro Tools utilize the x and y axes to show signal amplitude along time. MIDI programs allow you to compose music with pitch in time. These paradigms work very well for music creation, so why is it that the mix interface is merely a copy of the its analog counterpart?

The paradigms for the software tools I just mentioned grew up at the same time as the software itself, while the mixing window did not. MIDI programs do borrow from piano roll notation quite heavily, but the specification for MIDI came out of the computer control of physical keyboards, not an emulation of the keyboards themselves. The same goes for the recording interface. The recording window emulates the tape machine while using the computer as a tool to show what the tape machine cannot: a history of amplitude along a time axis. If both of these tools used the design methodology of the mixing window they would not be as useful.

The layout of a digital multitrack recorder's mixing environment closely resembles its analog counterpart. The long rectangular channel module harkens back to the days when recording console channels could be removed, repaired and placed either back into the same channel or into a different channel altogether. Engineers would oftentimes keep spare channels to plug into the main board of the console, just in case there was a problem with an installed board's circuit–hence, the need to isolate each channel and keep it modular.

Within the computer environment, the long rectangle makes it easy for us to see that each channel is different than the next. All of the channel's functions are contained in this one space in the same relative form as its analog counterpart.

Each channel in both analog and digital settings serves as a mental placeholder for a particular track, whether it's on tape or a file in the computer. However, in the digital world this relationship is not physical.

A mixer designation in the computer is the same as the multitrack recorder channel. There is no separation per se. Each track has a corresponding mixing channel and tracks can be assigned and reassigned to any output without reconnecting any physical connections.

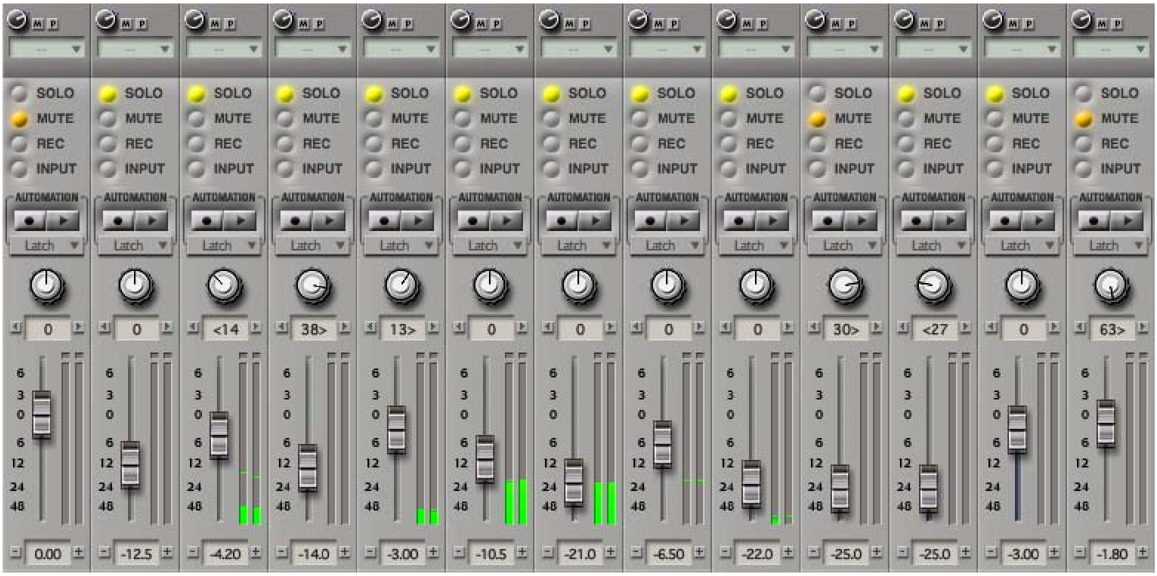
Channel position has no relevance to the position of the mix. There is no reason a channel needs to be in any particular order. Track one from a multitrack recorder could be playing back on track five of a recording console. The same goes for the software paradigm. Track order does not correspond to any kind of hierarchy or logical ordering other than the arbitrary designation of tracks by the engineer.

Reasoning the Interface

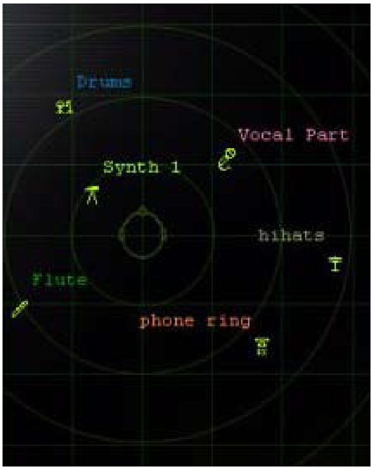
Tools evolved on the computer apply much more natural solutions than tools emulating their analog counterparts. MIDI composition programs apply a version of notation that is a suitable answer to long form notation without compromising the intent of the tool. Sound mixing applies a carbon copy of its ancestor in an attempt to give familiarity to the mixing engineer, but compromise the usefulness of the computer.

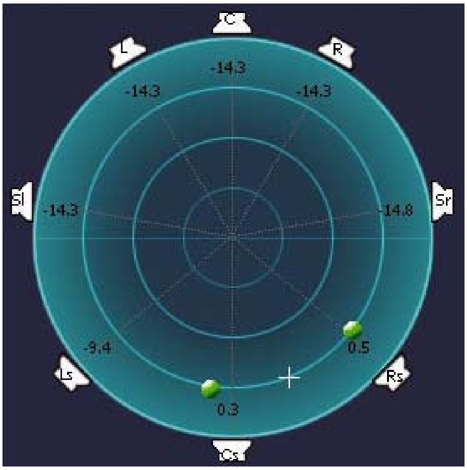
Latter day 3D Software panning tools get closer to describing how software can emulate real space. A 3D panner is a tool that allows the mixer to put a sound in space by placing the sounds in relationship to the listener. The view is from above. The listener is represented in the middle, with sounds represented as points around them. The more to the left the sound is, the more that sound is heard in the left ear. The further the sound is from the listener, the quieter that sound will be.

Software tools that use the eyes and ears as complement are much more useful than tools that don't. If we want to re-envision the mixing interface in the natural world, we need to take a long look at what our ears and eyes are good for and then recombine their best attributes in the most useful way possible.



The mix window for Digital Performer.





3D Panner Studio.

Sonar Surround Pan.

Hearing & Seeing

In order to think about how to represent audiovisual objects physically in space within a digital mixing environment, it's important to break down what our eyes and ears are good for. The evolution or each sense and how they influence and complement one another gives us clues on how to use them in our software mixing environment.

The Ear

On the side of your face, right behind your jawbone, you can locate a distinctive looking flap of skin. This is your outer ear, or pinna, the part that protects your delicate ear canal from unwanted debris. Its unique shape of folds acts as an amplifier for a range of sounds around 4000 Hz, or the piercing top-end of a human baby's cry.

Your ear flaps are designed to determine some sound direction, although poorly so. Your pinna are stuck to the sides of your head. Locating anything making noise behind your ears is difficult unless you turn yourself to face the noisemaker. A house cat, on the other hand, only needs to turn her ears to find the location of a noisemaker.

The Auditory Canal

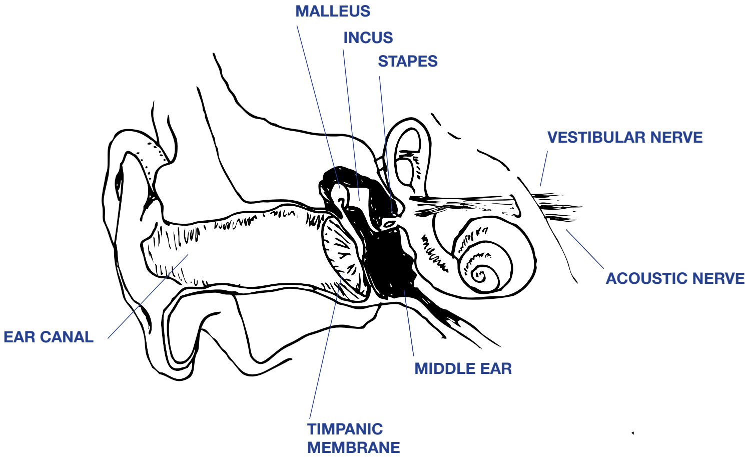
If you had the ability to peer into your ear right now. you might be able to see down the auditory canal right to your ear drum. Your ear canal is just about the size of a number 2 pencil and about as long as half of the first section of your pinky (I wouldn't recommend sticking either in your ear). That's not very long at all, but it's designed to resonate a specific frequency range.

Everything physical has a resonant frequency. Listen to a car's trunk rumbling as it rolls down the main street blaring hip hop music. That's the result of the speaker pushing a loud frequency at the resonance of the car's metal, shaking it immensely.

Your auditory canal's resonant frequency is around 3000 Hz[21], or the lower end of a baby's cry.[22] This is no mere coincidence. Your outer ear is specifically adapted to respond to a baby crying out above all other noisemaking.

Inside the Ear

On the other end of your auditory canal is your eardrum, the gatekeeper to the mysterious universe of hearing. Every sound you hear, whether it's distant or near, makes your eardrum vibrate. This thin, taut piece of skin responds to the pressure of these sounds, which are transformed the type of mechanical motion of bones.



On the other side of your eardrum is a pocket of air containing three tiny bones, measuring only 18 mm in total length. Like a telegraph, the purpose of these three bones is to conduct airborne sound accurately to your inner ear. If any of these three bones were missing, your fine-tuned sense of hearing would be more like hearing underwater, which is why fish don't need middle ears.

The malleus, called this because of its hammer-like shape, moves in synchronous action with your eardrum, which is vibrating as you read this. It connects to the anvil, which in turn connects to the stapes, which is firmly planted, like a foot, against the oval window.

On the other side of the oval window is a nautilus-shaped-liquid-filled tube called the cochlea (Latin for snail). Inside of the cochlear tube are two canals: the vestibular and the tympanic. The tympanic helps regulate the amount of liquid the oval window needs in order to respond to incoming vibration.

Without the vestibular canal, you would have no sense of orientation. Imagine a small sack of liquid with a ball inside. Hair lines the entire interior surface of the sack. When the ball moves around, the hair connects with nerve endings that send messages to the nervous system on your whereabouts in gravity.

The hair-lined cochlear duct sits literally in the middle of the two canals, as in the vestibular canal, hair lines the inside of your cochlear duct, but instead of a ball, there's fluid. The action of the stapes moves the cochlear fluid from the reverse-side of the oval window. The hairs inside of the duct conduct a broad range of frequency signals to the central nervous system on a separate path from your orientation sense.

Why We Hear

Orientation was one of our first major evolutionary senses.[23] Without the hearing sense, your aquatic ancestor could still detect low frequency sounds through this vestibular organ. Enough sound pressure at a low frequency have given you input that something nearby is happening, but not until it was too late and you had become someone's early morning snack.

As evolutionary time progressed, hearing developed as a way of locating other animals, hungry predators mostly, into three-dimensional space. As we flapped our way out of water and onto land, the middle ear's bone-relay adaptation allowed us to hear higher frequencies so that we could determine the distinct auditory qualities of predators, prey and landscape.

Landscape is different from waterscape. In waterscape it is possible for a water-borne animal to detect its position in relation to others in proximity by an evolved pressure sense. Water compresses progressively higher as depth increases. Water can can create pockets of pressure around geologic features as well.

This sense, however, is not available nor useful to us air breathers. Hearing has similar purpose by detecting air pressure at different frequencies. Since we cannot determine our depth, our brain filters the frequencies into distinct sound patterns, which we identify as discrete noisemakers.

If you need a broader range of frequencies to determine what you are hearing, you need two ears to determine where the sound originates. Your ears are placed at opposite sides of your head, facing away from one another. If a sound is predominant in one ear, your mind will take the difference heard in the other ear and determine the sound's relative location.

The Eye

Your eyes are on the front of your head; they are binocular, they perceive space quite well. You are a predator. If your eyes were on the sides of your head, like a bunny, you would be a vegetarian whose eyes are looking for flora, not fauna. You would only have 10 degrees of binocular vision and your perception of depth would be relatively nonexistent.

Safely tucked inside of the front of your skull are two globes that are just under 1 inch in size each. Most of your brain's sense functions are dedicated to interpreting neural signals from the eye. The protective outer layer of your eyeball is called the sclera. You can normally only see the corners, called the "whites." Your sclera gives way to your cornea, a translucent membrane that does most of the

work of refracting light through to the inside of your eyeball. Sandwiched between the cornea and the lens is your iris, whose principle job is to filter the amount of light entering into your eyeballs through the pupil. The wider the iris contracts the pupil, the more light will be collected. If it's bright, your pupil will become smaller, letting in less light. Together with the cornea, the lens projects the image to the back of the eyeball, up-side down, like a camera obscura. It's up to your brain to decode this information and put it right-side up.
This area is known as the retina.

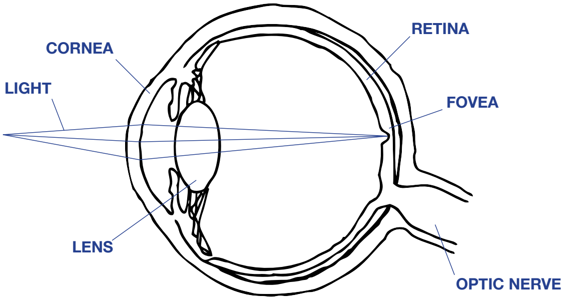
Inside the Eye

Contained in the retina are about 120 million rods and 6 million cones, both photoreceptor neurons. These receptors contrast with auditory receptor nerves, which are mere translations from vibrating cilia. Rods help give general shape to objects, but without cones, objects would appear fuzzy and without color. Cones help to give definition and depth. Attached to the sides of your eyeballs are several muscles that form around the optic nerve leading to your brain. They control the lateral (sideways) and medial (up and down) movements, in addition to keeping your eyeball firmly planted inside of your skull.

To suppose that the eye with all its inimitable contrivances for adjusting the focus to different distances for admitting different amounts of light and for the correction of spherical and chromatic aberration could have been formed by natural selection seems I freely confess absurd in the highest degree.
–Charles Darwin, *The Origin of Species* [24]

Why We See

Scientists at the European Molecular Biology Laboratory [25] found that our cones and rods originated as light sensors in our brain. They studied the brain of a marine worm, Platynereis dumerilii, a "living fossil" and found that the cells in its brain resembled those found in the rods and cones of our eyes. Some of the cells, they argue, still remain in our brain millions of years later and regulate our circadian rhythms.



There are many simpler forms of photo-receptivity in animalia. Light sensitive fluid exists inside of bacteria that serves only to tell light from dark [26]. It is presumed they contain these receptors to tell up from down. Jellyfish contain the most rudimentary light-sensing organs that are essentially cups containing a light-sensitive surface.

Your more evolved, complicated eye can detect more than just simple light. If you cover one eye, you can still see all of the objects in the room relative to one another. You can still detect shading, texture and perspective. You can tell the light post across the street is farther away than the pen on your desk.

Hearing & Seeing as Complement

A further consideration is, that owing to the singularly extensive development of mechanical physics a kind of higher reality is ascribed to the spatial and to the temporal than to colors, sounds, and odors; agreeably to which, the temporal and spatial links of colors, sounds,and odors appear to be more real than the colors, sounds and odors themselves. The physiology of the senses,however, demonstrates, that spaces and times may just as appropriately be called sensations as colors and sounds.

–Ernst Mach, Analysis of Perception [27]

Space & Time

It appears that both the eyes and ears started as competing senses for orientation. The eyes developed in single-celled organisms as a way of detecting sunlight in oceans. The ears developed from the orientation sense we still retain in our inner ear. Our ears and eyes developed differently over time, but both give our brain cues for how we perceive the world.

The eyes and ears share similar characteristics. You have two of each, for instance, one on each side of your head. This gives both sense organs the ability to detect your position in space. Both sense organs can also detect the passage of time. Your eyes and ears can detect change. However, both sense organs are adapted to different dimensional conditions.

Our eyes' perception is better for space. Our ears' perception of acoustic vibration is better suited to time. As far back as 1886 Mach reasoned that spatial symmetry is directly perceptible to the eye whereas temporal symmetry is not directly perceptible to the ear. Unlike vision, the human ability to parse musical rhythms inherently involves the measurement of time intervals.

Differences: Space

Both your eyes and ears can collect spatial information and locate objects in the landscape, but your eyes are better at it. Your ears can locate objects in a general sense, but lack the ability to re-orient themselves physically the way other animals with directional ears can.

We address more acoustic attention to what we can see because our hearing augments our visual sense of space.

In bright contrast, your eyes can, with accuracy, extract visual information precisely. Since we have two levels of seeing with rods and cones, we can perceive a high level of detail. We have the ability to turn our eyes in parallel. We can also detect where things are in space relative to us using the space between our eyes, or parallax. In hearing, we can do none of these.

Differences: Time

Both your eyes and ears can detect the change of information over time, but your ears are better at it. Your eyes can detect change, but that change is measured with receptors which are more suited to detecting space. They can pinpoint objects, but are not as well suited to detect and recall patterns in time. Our recall of musical melodies exceeds our recall of visual patterns.

Your ears, however, extract features as a function of time. We perceive sound as linear–in time. J.J. Johnston notes that a glitch in time to the ear is much more perceptible than a glitch in time to the eye.[28] If you watch a video on YouTube and a frame drops out, your eyes will compensate with a memory of the last perceived “good” frame. However, if there is an audio glitch, the sudden difference in perception in loudness will be much more noticeable–and very jarring.

Sense Illusions: Influence

The eyes and ears can play tricks on one another. Influence from one sense can completely alter the functions of the other.

For example, you could view a YouTube video with high quality audio, yet the perceptible quality of the video will not be improved. Regardless of the fidelity of the audio signal, it will never improve video's perceived quality. However, the inverse is true. According to an AES paper, "When subjects are asked to judge the audio quality of an audiovisual stimulus, the video quality will contribute significantly to the subjectively perceived audio quality." [29] A high quality video will bump our perceived quality of the audio, even if the quality of the audio is sub-par.

The McGurk Effect (McGurk and MacDonald,1976) study shows that our perception of phonemes is influenced by what what is actually spoken versus what is seen or heard. Psychologists created an experiment using video playback in which the visual signals were in direct conflict with the auditory spoken syllables. Subjects were presented with video of a woman forming the syllables for "ga-ga" while the synchronized audio track gave a different sound "ba-ba." when people closed their eyes, they clearly heard "ba-ba." With the sound turned off, they could reasonably identify the woman as saying "ga-ga." However, with both turned on, they perceived the woman as saying a whole different thing: "da-da." [30]

How you hear influences your seeing and how you see influences your hearing. If you're going to create a useful software interface that involves both sound and moving image, it's important to realize that the associations you make have to be carefully thought out and meaningful to the tools you're creating.

Effects: Visual & Auditory

There can be no completely intimate visible and audible music until audiovisual unison is achieved.

–Ralph K. Potter

Using Space

We've described effects that alter our perception of one sense based on the other's input. If we want to create a synaesthetic audio mixing experience, we need to find auditory and visual harmony. Are there effects that are equivalent between sight and sound? How do we begin to categorize them?

There are two parts to the theory of audiovisual effects as a synaesthetic mixing experience. The first part of the theory is based on what we can see with our eyes and hear with our ears in physical space without specialized instruments such as microscopes of oscilloscopes to augment our perception

The second part is based not on what we see or hear as distinctive elements, but rather on the attributes of light and sound waves in space.

Landscape & Soundscape

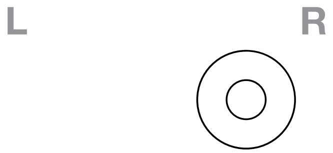
Each visual noisemaker is a distinct player in the landscape. Physiologically, we are separate from one another. We can distinguish between different noisemakers visually because they are separate entities.

Space & Localization

Localization allows us to locate objects in space using our eyes and ears together as one sense organ. Objects in our left field of audiovisual sense will appear more perceptively “left.” The same goes for the right.

You can use the X axis to place each noisemaker in the stereo field. If we place objects in space according to this positioning using stereo space (or even panoramic space) we can hear and see these objects in relative position to both our eyes and ears. This relative placement on this horizontal axis represents our pan position.

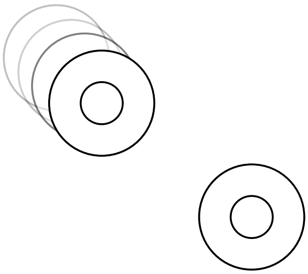
It is theorized that blind people who experience brief periods of sight before blindness have a better sense of space than people who have congenital blindness. If this is true, we can assume that the two senses create spatial recognition better than just one alone, or at least that sight enhances hearing.

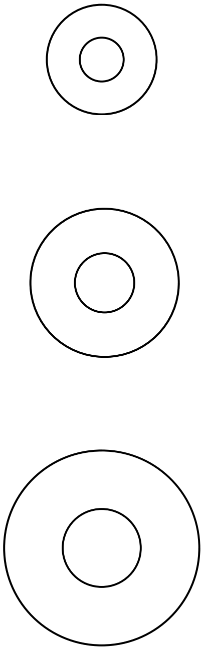


Practical Physics: Collisions

Position in space is physical. It is impossible to place two noisemakers in the same position at the same time. To our ears, it may appear that two sounds could come from the same space, but to our eyes this is never true. In an alternate, quantum universe we could place one physical object inside of the other, but in our universe this is never true; it is physically impossible.

Representations of objects in software space should never occupy the visual or auditory space at the same time. We should always place objects next to one another at their respective boundaries. That means that we should never represent convergence of visual elements as a representation in space with audio and expect it to be synaesthetic.





Size and Absolute Volume

Volume is relative in space. Something farther away from you could emit sound and then move closer to you and emit the same sound at the same volume as before and be perceptibly louder and visually larger. In theory this could work as a synaesthetic representation in space. However, something physically farther away could also emit a louder sound than something closer to you.

This starts to create a problem. In real space, you can use acoustic and visual cues to determine relative distance, but in software, at least in the near term, this is made difficult by two-dimensional displays and stereo speakers.

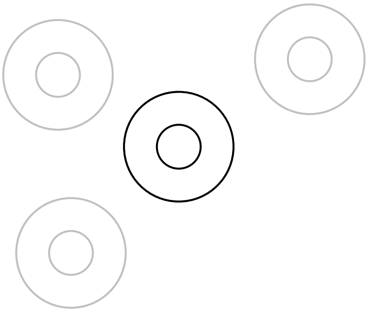
In 3D space, the determination of an object or sound's height is difficult to detect. We have two ears on either side of our head that have difficulty locating objects in space without our eyes. Elevation is even more difficult for our ears to detect. A cat meowing at you from the floor will have the same pitch and loudness as a cat meowing at you the top of a bookcase (although a cat on top of the bookshelf has an advantage over you).

3D space becomes a very messy space in which to represent a synaesthetic experiment if we want to maintain image-to-audio continuity.

If we constrain our field of vision to two dimensions, we can control our parameters by placing every object at the same relative distance from us, the same way we can scatter a bunch of papers on a desk. Every visual element will then be the same relative size.

We can control the relative gain of each object by using the y axis, which at this point has become arbitrary in our field of view. If we move objects on the y axis, we can see the relative volumes of each object in space.

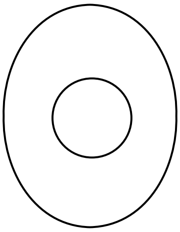
The question becomes: What direction do we use to represent higher volume? We could use the position closest to the “floor” or the bottom of the Y axis as the representation of maximum volume, but this does not have an analog in our world. Instead, we could use the maximum height of the Y axis to determine maximum gain. Mixing “up” equates to maximum volume and “down in the mix” equates to less volume and they are synonymous terms in the audiovisual mixing world. In the visual world it this analogy be like watching a balloon rise to its peak.



Audiovisual Muting

A visual mute is when an object is at its full volume (or brightness) at one moment and then loses its visual gain the next moment. This effect is similar in audio, only that muting results in the sudden and total loss of signal.

We could viually represent audio muting by increasing a shape’s transparency over a dark background and by reducing the volume at the same time. We would show muting on one or more objects or conversely create a “solo” effect, where only only one object in space is illuminated as if by some kind of spotlight stage effect.

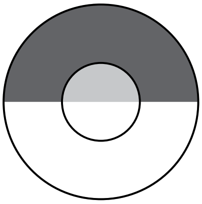


Limiting & Compression

Compression is an audio effect that changes dynamic range by raising low-intensity signals while lowering high-intensity signals[31]. This increases the overall perceptible intensity of an audio recording while lowering its overall dynamic range across frequencies.

The visual analog to this is to “squeeze” an object. A beach ball that is squeezed contains the same amount of air as a normal beach ball, only it’s visibly compressed. In the audio realm, a visual element could still retain most of its characteristics, but be visually perceptible as more dense or loud.

A real world equivalent is to hold you hand over your throat and squeeze while speaking. Both your throat and your speech will be perceptibly compressed.



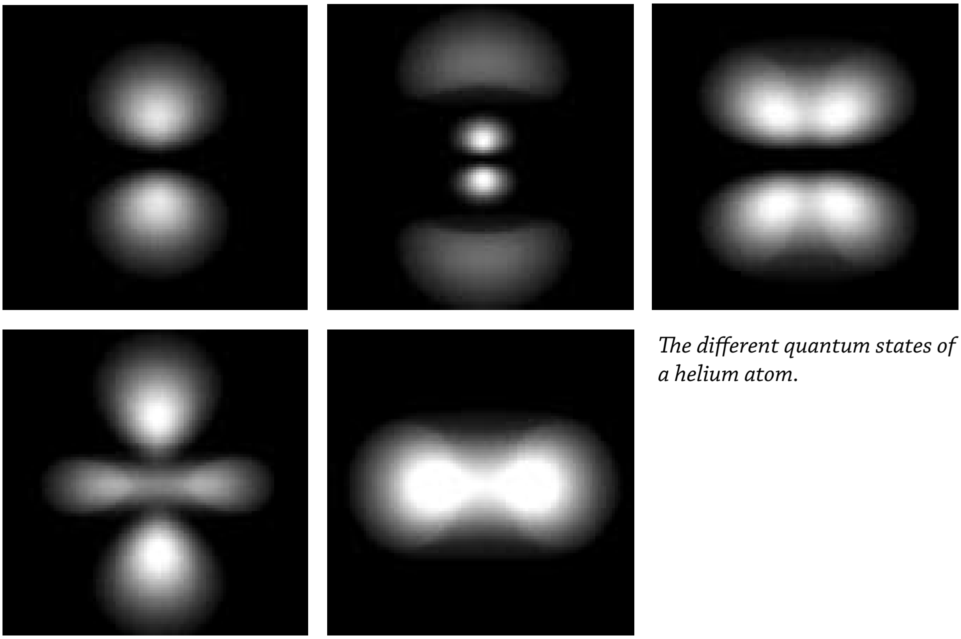
Filter

In audio, a filter is an elimination of a frequency range through equalization. A very simple high or low pass filter allows its target frequency range to pass while eliminating its opposite frequency range. For instance, a lowpass filter will eliminate the high frequency range, revealing only the low tones of a sound.

In the visual realm, a filter prohibits light waves from passing. The visual counterpart to a lowpass filter could show a darker shade on the top end of the shape and brighter on the lower end. The darker part of the shape would appear similar to the muting effect, with its transparency revealing more of the dark background.

Edge Effects

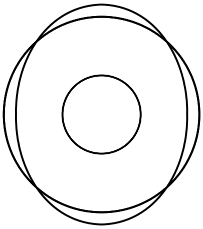
Edge effects are those that deal with synesthetic sound and image on an objects' wave length. These effects take each original audiovisual source and somehow "blur" or manipulate its appearance. Phase, delay, echo and reverb are examples of edge effects.



The different quantum states of a helium atom.

Phase & Cancellation

When two of the same audio frequencies arrive at the ear simultaneously they cancel each other out and you don't hear them. That's because the wave forms of the frequency literally overlap exactly and are not audible to the ear. It's called phase cancellation. Noise cancellation headphones work this way. There are tiny microphones on the outside of the headphones that feed the signal from the outside into the headphones, so that only the source recording comes through the headphone speakers.

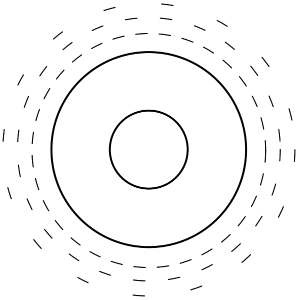


Audio phasing is when the wave forms don't quite overlap. You will hear a rhythmic pop or beat when the wave forms don't overlap and when they do, phase cancellation will kick in.

Similarly, when a light source is fed through two pinholes (or slits) and projected onto a screen (Young, 1881) [32] they will appear to pulsate when the waves cross over. Since both projections of light originate from the same source, they will both be in phase with one another and will pulsate in the same frequency. This is called coherence. [33] It looks a lot like what audio phase sounds like.

Delay

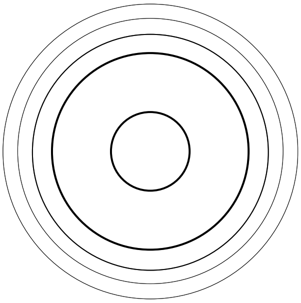
The delay effect in audio is literally a delay between the emission of a sound and its repetition. [34] Try clapping your hands in a medium sized space, like a lecture room. You'll hear a sound of the clap followed by a near version of itself (you might even hear some phasing). What you'll hear is your original sound and then your delayed sound at a diminished volume within milliseconds.



A similar thing happens with sight, only the visual delay is purely perceptive, rather than spatial. Imagine a pendulum swinging from left to right in front of you (Pulfrich, 1922). Now imagine that a piece of frosted glass is placed in front of your right eye. The pendulum will appear closer to your left side. The disparity of illumination causes your right eye to perceive the pendulum as moving in an elliptical orbit because the retina is receiving the image of the pendulum bob lagging somewhat behind the uncovered eye. What you're actually seeing in your right eye is a version of the pendulum bob after it's already passed, much the same way you hear a sound that is less loud, right after you heard the real sound. It would appear that loudness and brightness in the afterimage and after sound have a relationship.

" The oscillating bob is never seen where it actually is, but appears at a place in its path a little farther behind its true position."– Alfred Lit [35]

Echo

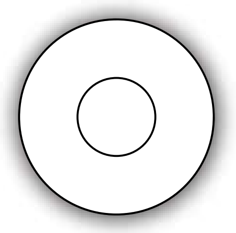


Imagine you are standing on the top of a ridge overlooking a desert canyon. Above you is the top lip of the canyon. Now yell "hello" very loudly. Wait. It will come back to you. Each time you hear your own voice, it will get softer and softer as it bounces off the walls and loses its energy as it makes its way back to you.

This is the audio echo effect. It's a natural extension of the delay effect, only the walls are much longer and the ceiling higher. The delay of the canyon is much longer. It is an effect that can be perceived in one ear or both ears, although the effect is much more pronounced in space if heard in both ears.

Light can also arrive to us in echoes. In 1987 a star went supernova in the Tarantula Nebula. This explosion was significant because we could see it on earth with the naked eye. More significant was what happened ten years later. Instead of continually fading out the supernova appeared to get much brighter[36] through x-ray and radio telescopes, which detected bands of light we cannot see with our naked eyes. The explanation for this was that the light bounced off of neighboring space dust. Light and sound can arrive at a location more than once, lessening in intensity with each turn.

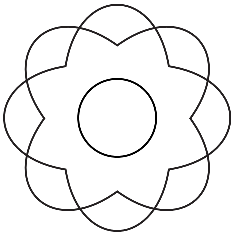
Soft Edges: Reverberation



Instead of a sound that simply hits the wall and reflects back to you, a reverberation effect is made up of many scattered reflections that return to you. They're actually diffusions of diffusions hitting the crevices of the interior space, causing a softening of the sound. The reflections last longer or shorter, depending on the intensity of the sound and the size of the space. Imagine our valley, only covered with a dome or a cathedral space.

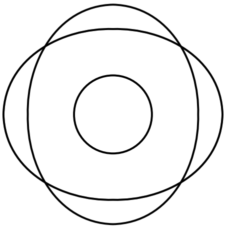
Light diffusion is the same effect. In film, as well as still photography, a light is placed inside of a box and projected through a translucent material. This is the classic light diffuser, although frosted or stained glass in a church produces a similar result. When a model is photographed using a soft box, he or she appears ten years younger. That's because the light is diffused, causing the light to scatter in many different directions as it approaches the model, softening wrinkles and harsh shadows.

Further Studies



This body of work is only the surface of what we could achieve by using space as a metaphor. There are many further studies that could extend the language of mixing audio using visual cues in natural space. I have used examples of physics, biology and psychology in our perception of space to show what kinds of associations are possible between light and sound using this method.

There are effects that haven't been addressed in this argument. The act of chorusing, a delay effect with a low frequency modulation on the delayed signal, could be shown by wobbling the delay ring.



A low frequency modulation on a signal without the delay could be shown as a series of overlaid waves.

Beyond this 2D simplistic environment, we could look at more complicated psychoacoustic effects. We could try and produce the Doppler effect by showing the modeled effects of compression on the front of the audiovisual object and release behind it as it moves in space.

The use of physical metaphors in audio software is already being taken seriously. Izotope's iDrum and Intua's BeatMaker iPhone touch interface based applications use the Pan/ Volume methodology proposed in this thesis to put single parts into space. A French company, JazzMutant, is offering a touch sensitive screen that allows an artist to send a part's effect into Pan space, similar, but different, to the demonstration I've offered here. Golan Levin and Zack Lieberman's *Manual Input Stations* proves that gesture and gravity in space can be used to create sound and image in motion. The ReacTable project at Universitat Pompeu Fabra proves that tangible objects can be used to represent sound objects in real space and manipulated by touch.

What this thesis proposes is that with careful vetting, a natural, physical environment can not only pave the way for more useful use of a computer, it can be used a more creative use of a computer to unlock those synaesthetic associations we all have lurking in our psyches. While adopting the old modernist mechanical metaphors in software might work in the short term, following how our eyes, ears, hands and possibly our noses and tongues use space works much better in the long term in augmenting our environment.

PS

WE
NEEDED
HELP

TOOLS

STICK

BOWL

POTTERY

SPEAR/
LONG STICK

CROUCH &
POKE

SIT

RUN

ALL INVOLVE
HANDS — THOSE
OPPOSABLE THUMBS
OF OURS

QUITE USEFUL

EVOL

WANT
SUPER
TO

WE DIDN'T
EVOLVE BY
GENERATING NEW
SENSE ORGANS —
WE EVOLVED
BY MAKING TOOLS
to HELP US — BY EXTENDING
OUR MINDS

DU
AR

WAR

SYNAESTHESIA
SYN SENSE
WITH SENSATION

N PERCEPTION

PERCEPTION - VISION - CATS CAN SEE IN THE DARK

SOUND

TOUCH

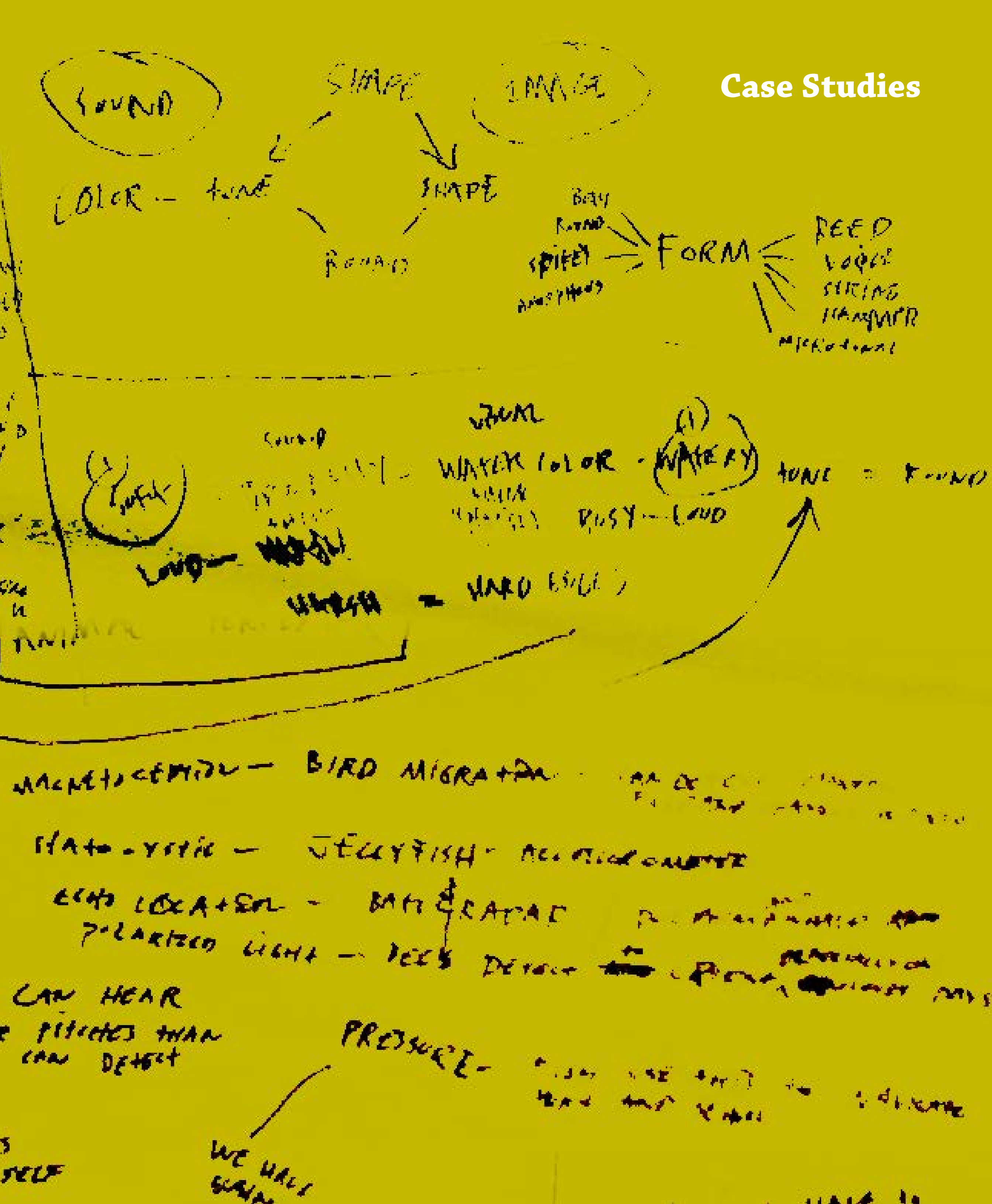
SMELL — DOGS CAN SMELL MORE THAN US

BOTH
HIGHER
WE

BALANCE

+ KINAESTHETIC - GIVES US RELATIVE POSITION OF PARTS

Case Studies



Case Studies

Overview

The following projects gave me tremendous insight on the possibilities of combining image and sound into a single experience. The projects here are presented in chronological order.

You Are Here Now

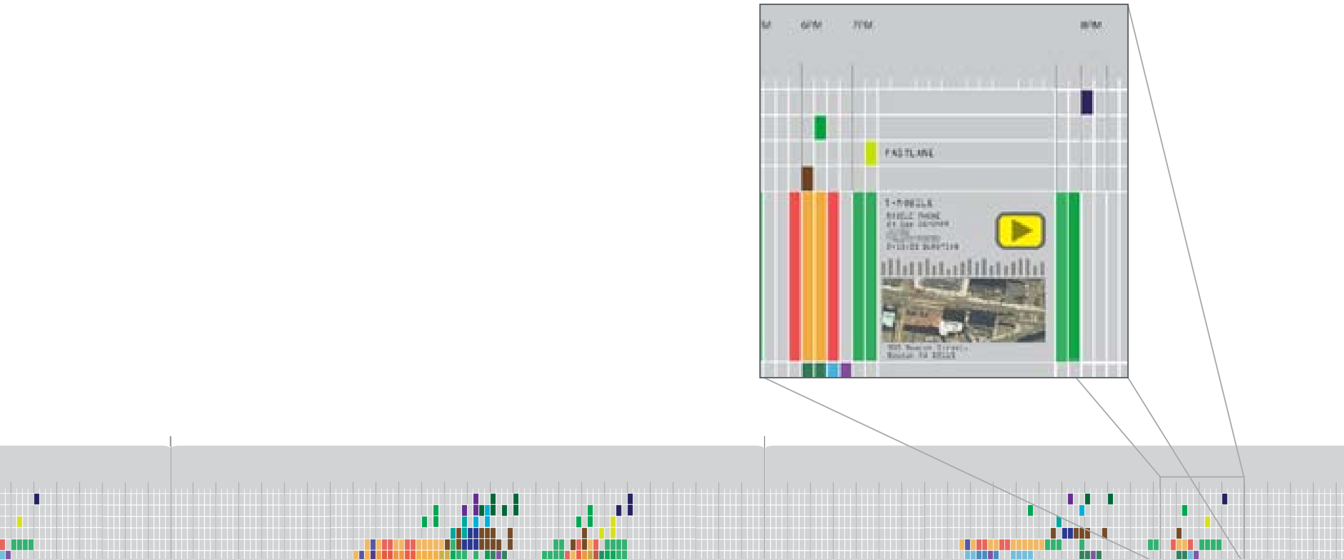
The purpose of this project was to map personal data over an entire lifespan.

If the government could collect data about you across your whole life and consolidate it into one repository, what would it look like? My answer was to create a a proof-of-concept video that would demonstrate how CCTV footage, mobile phone logs, utility bills, ID access logs, email transcripts and biometric data could be organized in one big scary place.

Time & Space

I created a continuous scrolling device that started at the beginning of a "citizen's" life. The proposed interface was a multi-touch screen embedded in a public kiosk (accessed by finger print identification and iris scan, of course!). The "citizen" could scroll with their finger from birth to now by swiping from right to left.

The scrolling device first divided into years. When the "citizen" wanted to dig deeper into a year, they could "pull" apart the year in question using two fingers across the horizontal axis. They could continue to do this until they arrived at a particular hour of a day. They could then "pull" at vertical axis to show the transaction detail (or in this case an entire phone conversation!).



Color & Shape

In the initial draft of the interface, each group of data points had a particular shape. For instance, ID log data was represented by a round rectangle icon, while email was characterized by pill-shaped icon. Each data point within each group had its own color.

Conclusions & Hindsight

This project made me aware that space is an oft neglected primary viewpoint for software interface. The software interfaces I encounter use windows as the primary viewpoint, not the whole landscape itself. This interface broke away from popular modernist notions that transfer existing physical interfaces like knobs and switches onto the screen for no purpose but to provide a familiar looking metaphor.

Case Studies

Panorama Scroll

Interface & Space

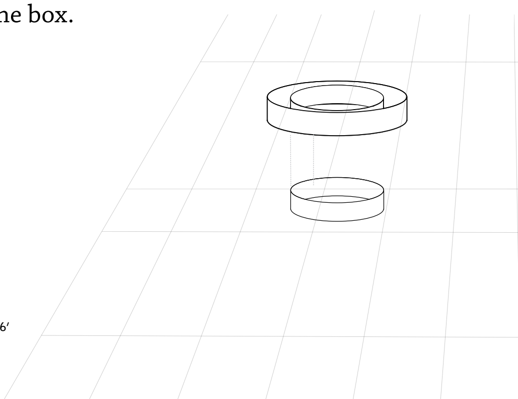
I wanted to create an way of navigating through space that didn't involve touching an interface. In real space we use our feet as our walking interface and our hands as our tactile interface. What if we could emulate that in so-called "virtual" space?

Space

The world was constructed out of individual photographs, stitched together to create "panoramas." Panoramas in this world allowed for 360 degrees of seamless rotation in space, similar to our real world. People weren't actually in 3D space; they were staring at a projected image. They could only see about 1/3 of the space at any given point.

Pan & Interface

People could navigate by placing a hand between two parallel planes (let's call it a box). By moving their hand in left, the panorama image on the screen would scroll to the left. By moving their hand to the right, it would scroll to the right. The interface would scroll faster as the hand moved closer to the coincident side of the box.



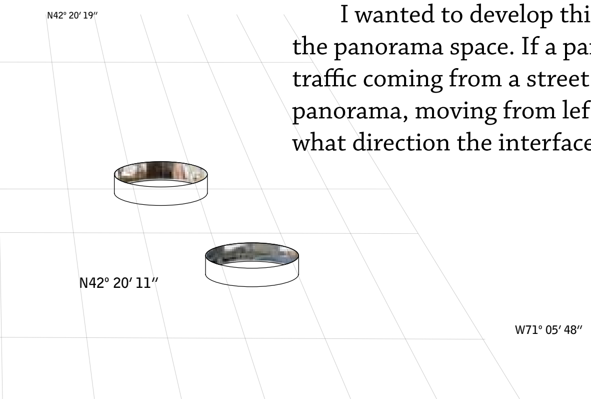
The second version of the interface allowed the second hand as the interface. A person could place both hands in the box and pull them apart to zoom in on an area of interest or navigate to an entirely different panorama space.

Conclusions & Hindsight

The interface required a lot of physical effort to work. A person had to crouch down and place their hand in the box or sit down and hold their arm for a length of time, which could put a lot of strain on their shoulder. In the end the interface itself just wasn't a natural way to work. Perhaps the same language could be developed using the whole body instead of just a single part of the body.

The panorama idea was a direct extension of the continuous scroll theme from the "You Are Here Now" project, but instead of looking at space through the lens of time, it looked at time through the lens of space. Using space as a primary lens allows people to explore data on their own terms, as it is in real life.

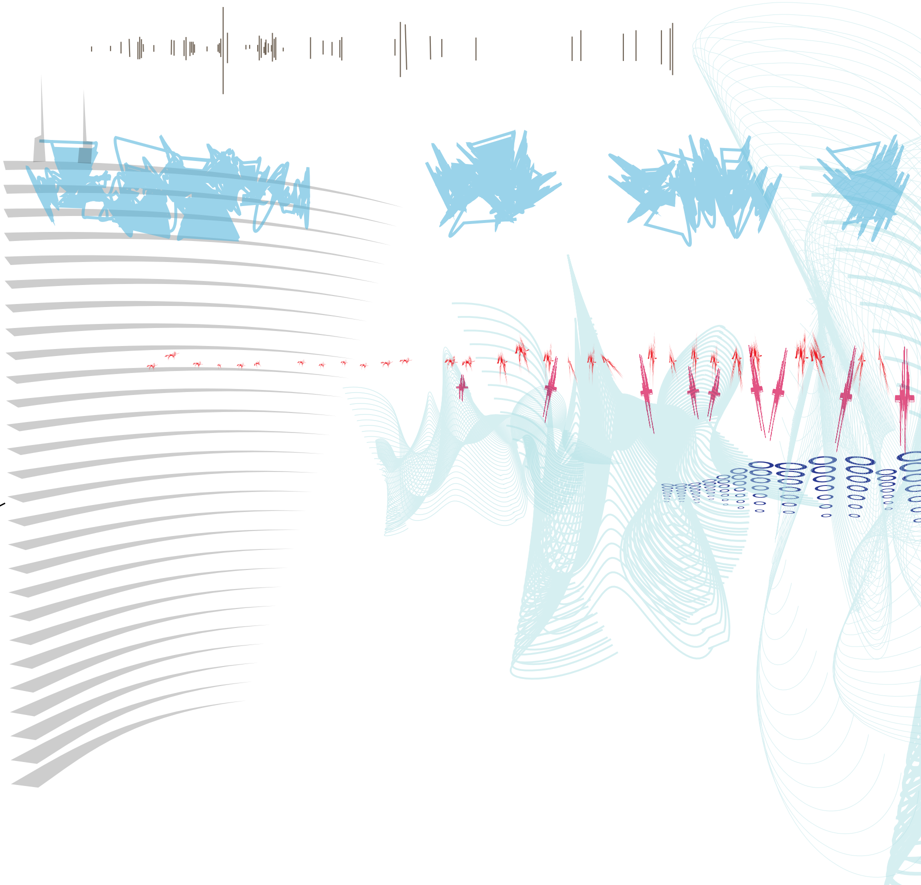
I wanted to develop this idea further, incorporating sound in the panorama space. If a particular part of the space had sound, like traffic coming from a street, the sound would rotate along with the panorama, moving from left to right or right to left, depending on what direction the interface was scrolling.



Trevor Wishart Poster

Time, Color, Part, Note & Shape
I created a visual score to a very difficult piece of music in time as an exercise to represent sound. The music was more of a sound art piece with some parts literally bleeding into one another than a properly charted piece of music with defined notes. This made representation difficult, but not impossible.

Time, Frequency, Volume, Color, Note/ Part, Shape
I charted the music as time, using the X axis to show the number of seconds. Each part was placed on the timeline according to its frequency and became larger as its volume grew or shrank as it became softer. Although each part was placed according to its



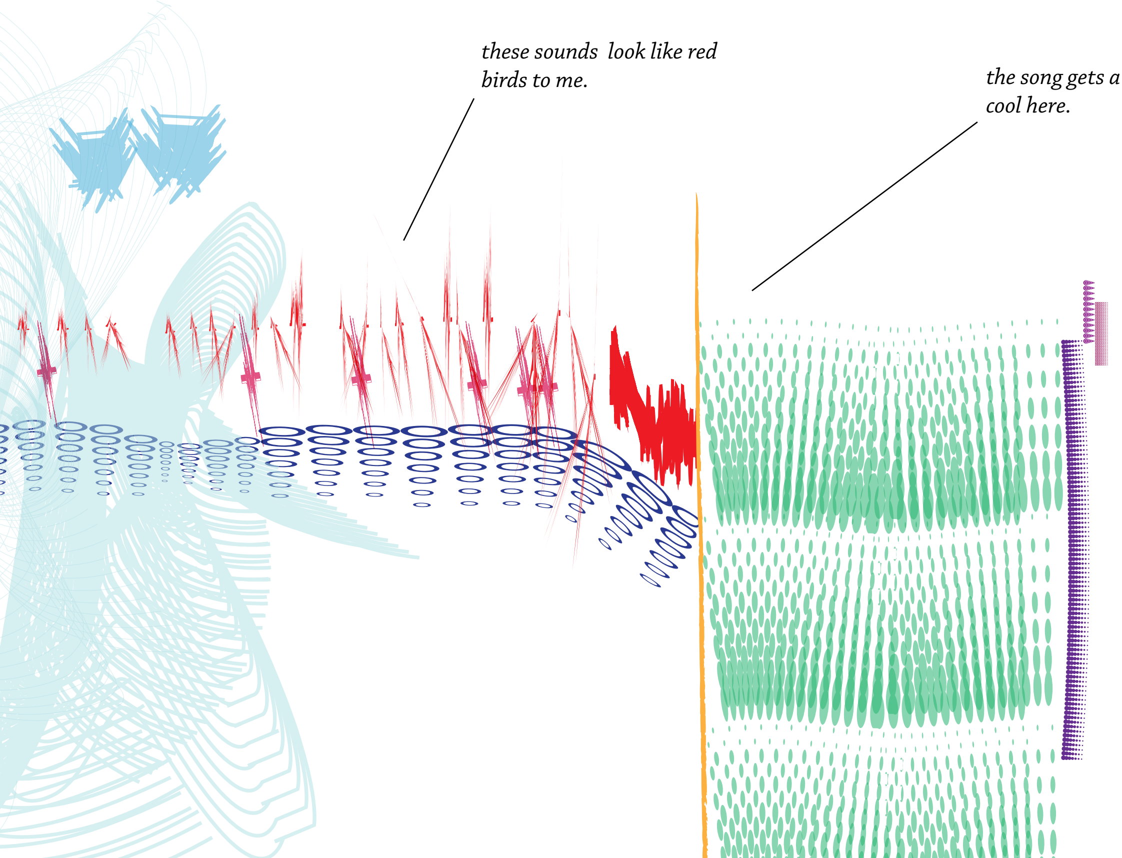
I measured the frequency range of each sound using Fourier frequency analysis.

the songs starts with a series of throaty, gutteral sounds that span the frequency range.

frequency, I chose to take artistic liberty to represent each part in color and shape the way I had heard it. There were crow-like sounds that I thought looked like winged, warm birds in fiery red, while whooshing noises sounded like they should be cool and amorphous clouds.

Conclusions & Hindsight

As difficult as it was to identify individual sounds in the mix, I found I could at least group families of sounds into identifiable shape patterns. This helped in my own understanding of the musical piece. I then started to think about the identity of sounds as shapes and how I could use this in sound representation.



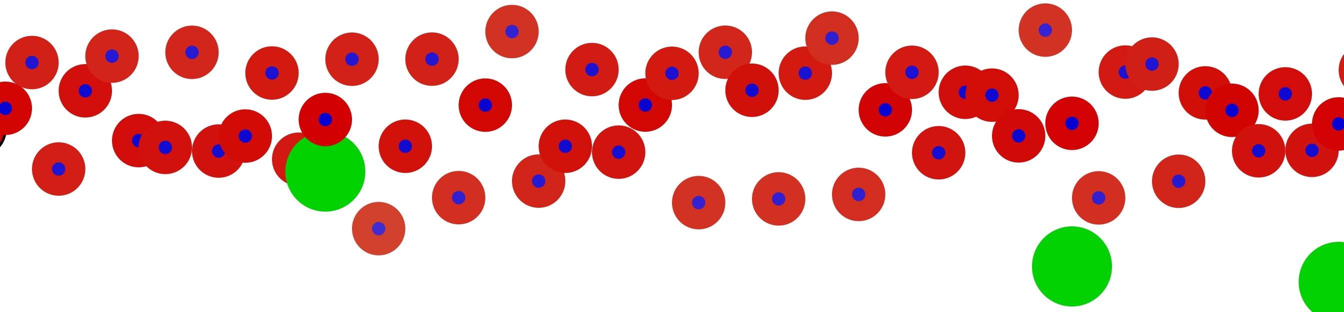
these sounds look like red birds to me.

the song gets a bit cool here.

Dynamic Volume Channel Studies

Time, Volume & Part
Over the course of the program I completed several studies on the representation of volume over time. In all examples, the volume of the incoming audio was represented by a number of dots dividing a millisecond across the Y axis.

The first two experiments used single channels sources as the input, with controls to change the quantity, the size and color of the dots across the X axis. The volume dots would jump up and down according to the volume of the incoming audio. The first example used a sound file as an input. The second example used live microphone input.

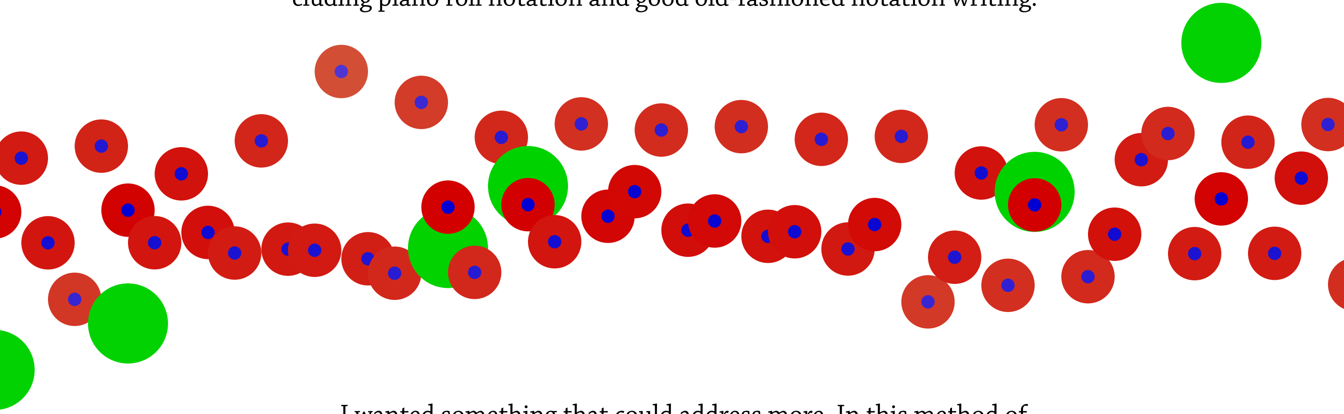


I introduced multiple channels in the third version. The multiple channels were represented by different colors according to relative frequency range. For example: bass sounds were shown further down the Y axis than cymbal sounds, which were placed at the top. In cases where two sounds clearly occupied the same frequency range, I arbitrarily placed one above the other for visual clarity.

Conclusions & Hindsight

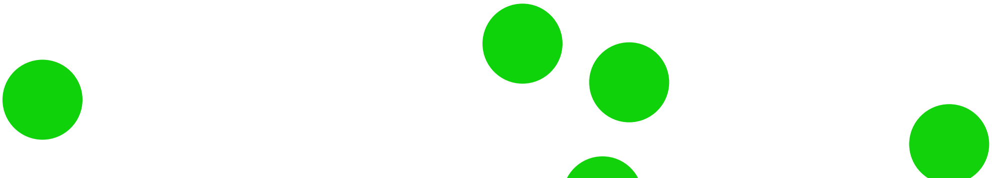
This method is excellent for visualizing sound in time. The third multipart version gives people active cues for when specific musical parts come in and how loud they are when they play. Some visual people said it gave them a chance to look at music in a way they never understood before.

However, I became weary of using time as the primary axis. It's not very useful for manipulation, unless you can change frequency or time dimensions of the parts themselves. There are already many good software models for direct manipulation out there already, including piano roll notation and good old-fashioned notation writing.



I wanted something that could address more. In this method of direct manipulation it's very difficult to manipulate the pan position of a sound or its volume. Even if you could show volume as a graph on the bottom of the screen (most music sequencing programs do) or use iconography, it would be a lot of codification and a lot of windows to follow.

I hadn't realized yet at this time that I was looking at redefining the mixer interface, not the musical composition interface.

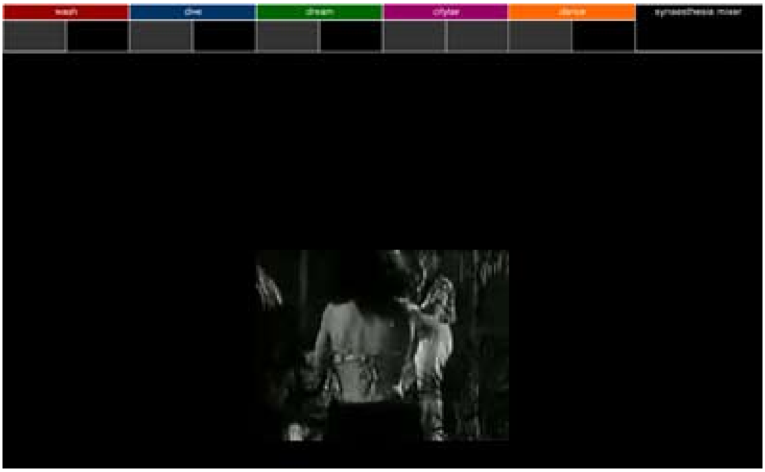


A.S. Mixer

Color, Time and Dynamic Media

The A.S. Mixer* was a software proof-of-concept project that created arbitrary associations between moving image, sound and color. On the top of the screen were a row of buttons with a label. When one of the buttons was clicked, a corresponding silent movie file would pop up in its corresponding place on the screen and a music part would play. All of the musical tracks were synchronized so that when parts were turned on and off, they would form a single musical piece.

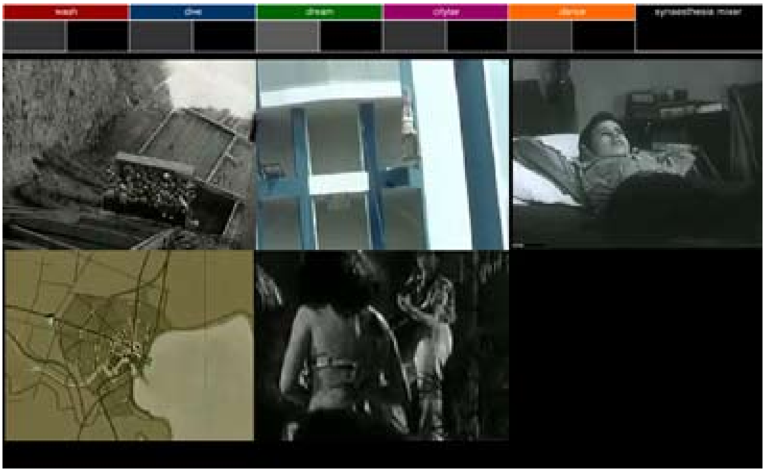
The labels were meant to show the general feeling of both the musical and video parts, not referring to either the instrumentation or the video content. When the orange colored “dance” section was turned on, the sound channel played a rhythmic percussive instrument synchronized with a video channel displaying a hula girl dancing. When the green colored dream section was turned on the sound channel played a soft organ chord synchronized with a video channel showing a boy lying on his bed in the act of daydreaming.



Conclusions & Hindsight

The A.S. Mixer was my introduction to playing with color labels, image and musical content as one synchronous piece. Sounds are generally given the role of short-burst feedback in software applications, such as when the Mac OS system gives you an alert or a warning. I hadn't seen it played out in a long, musical format like this before.

It would be an interesting alternative to either musical sequencing or video DJing to cram both media into one place in a fully-functioning piece of software based on this idea. If I had more time, I would like to add the capability of replacing both the videos and the audio on the fly and introduce the video's own audio track to the mix as well. A playback sequencer using video keyframes and piano roll notation could also benefit this application.

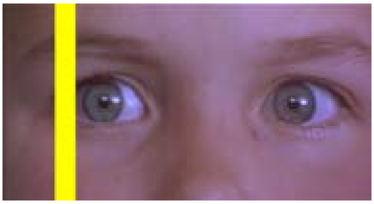


Film Music Cues As a Way to Mark Sound and Moving Image

Color, Time and Dynamic Media

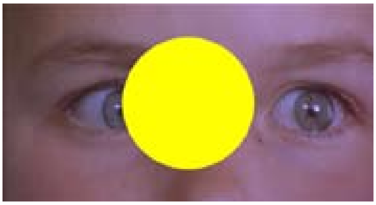
On a standard computer-based media player, you have access to the transport controls (play, pause, stop and maybe fast forward and rewind). On DVDs you have access to the forward-to-next-chapter and its sister back-to-previous-chapter, that allow you to transport the physical mechanism to a particular cue point. This cue point is fixed at the time the DVD is encoded and can't be moved.

Composers think of the film timeline as a set of movable cues. They use these cues to show them moments of significance or when a scene is about to end. It is possible to score a film without these cue markers, but it's a lot more precise to use them and it only requires the composer to play to the film, not worry about what's coming next. In example, composers use the "sync points" at the end of dialogue to cut from scene to scene or for "hits" in fight scenes and car chases[37]



A Streamer is an anticipatory visual cue before a punch.

This is a tall, thick line that moves from left to right on the screen between 2-3.33 seconds



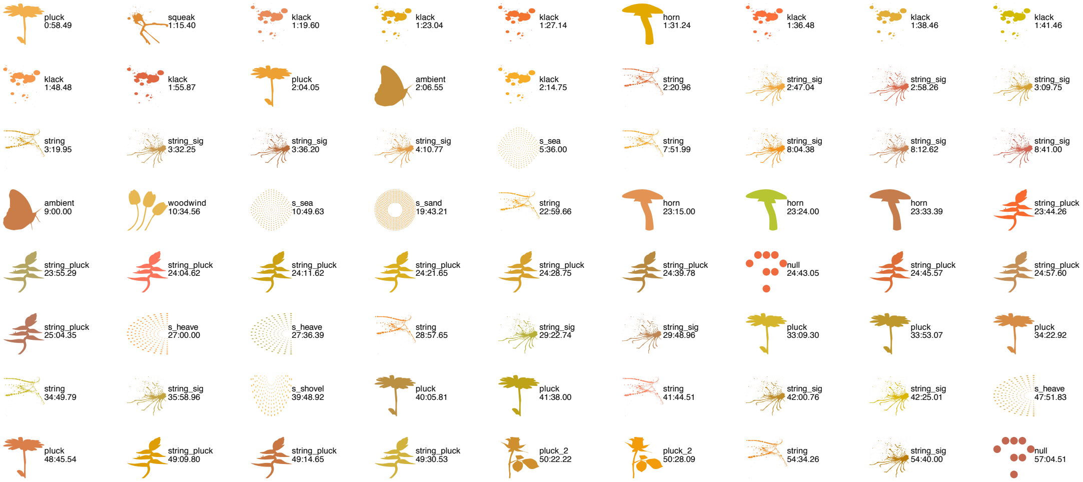
A punch delivered at intervals creates a visual tempo for the composer to write corresponding music.

A large dot in the middle of the screen for a second or less.

At the direction of Ronald Smith, I spent a considerable amount of time analyzing the music of the 1964 film Woman in the Dunes by director Hiroshi Teshigahara and composer Tōru Takemitsu. The music in the film forms a synthesis with the image and at times plays a dominant role.

After I watched the film, I placed cue points in the form of punches, streamers and flutter punches into every place music appeared and gave it a corresponding label. If the same musical hit happened more than once in the film I gave it the same name, so I could look for audiovisual patterns.

I found that I understood the film a lot better when I placed the cue points than when I had watched the film without them. I was able to determine, for instance, that music in a particular scene was timed not to the film editing, but deliberately timed a few seconds slower. This explained my uneasiness with the scene (and indeed my fascination with the film).



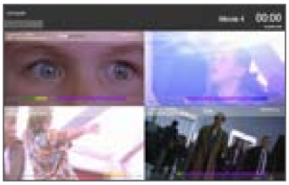
The first 72 musical cues of ‘Woman in the Dunes’ visualized as icons.

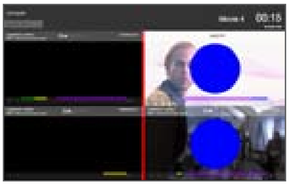
cueTag

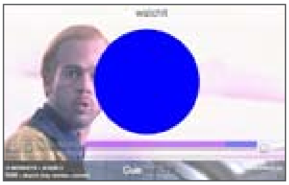
Color, Time & Dynamic Media

After I spent some time with Dunes, I reflected on an earlier film comparison project. The brief for that project was to create a tool that students could use to compare two films: *La Jetée* and *Twelve Monkeys*. *Twelve Monkeys* was based largely on *La Jetée*. Both films dealt with the cycle of life and death and they both detail the story of a boy at an airport who saw himself as a man shot down by a higher power.

If visual cues can help musicians create music for film, they can also be used by non-composers, students of film and film critics to interpret the relationship between the music and moving image. I saw this as an opportunity to revisit the original interface and to apply color and iconographic shape as meaningful cues in film study.





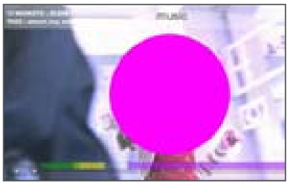


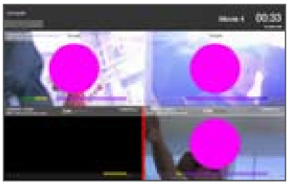
Films are broken down into scenes. This is an ideal place to compare films or between places in a single film. There are four separate scenes that deal with the death scene in Twelve Monkeys, so I started by putting all four scenes in the four quadrants of the computer screen. I then placed a typical playback mechanism with a timeline below each of the films and assigned each of the scenes titles. I could then review each of the films separately with their own playback devices or by using a master playback section found in the upper left hand corner.

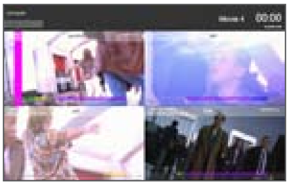
On the timeline itself, I created a system in which I could click on the timeline and create a cue tag. A cue tag is a duration of time on the timeline with an associated color and name. A cue tag could have been placed just about anywhere and for any purpose, but I chose to use music and sound as my guide.

If I placed the same named cue tag on other scene, the same color would appear in that scene's timeline. When I placed all of the tags, I could play each film separately and watch as the corresponding-colored streamer came across, or click on the colored cue tag on the timeline. When that happened, each film that contained that cue tag would snap to the beginning of the same section. All other scenes would disappear.

When I clicked on the master play button with all four airport scenes of Twelve Monkeys on the "gunshot" cue, I was surprised to hear (and see) that both the beginning and the end scene music were almost the same arrangement with the same tempo. I also noticed that in all four scenes, the death sequences were played out of time with one another (intentionally, I gather, by the director Terry Gilliam).



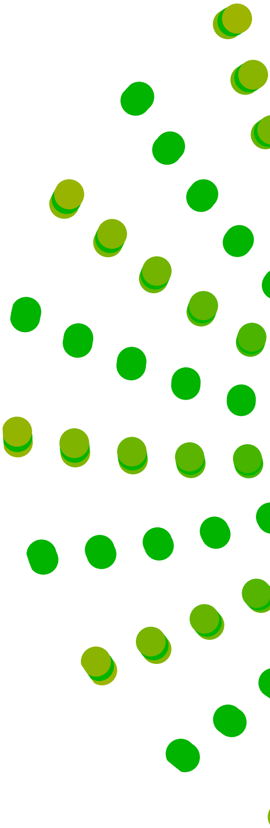




Conclusions & Hindsight

I learned about using color and moving shape to show associations by making the tool and by studying the films' content through the interface. However, I didn't get at how to use the tool as a way of understanding the direct relationships between sound and image. I wanted to explore more with the possibilities of using space as a more natural environment with image and sound, the way I had intended to with the panorama scroll interface, but with more user control, the way I had done with the A.S. Mixer.

Sound Mandala

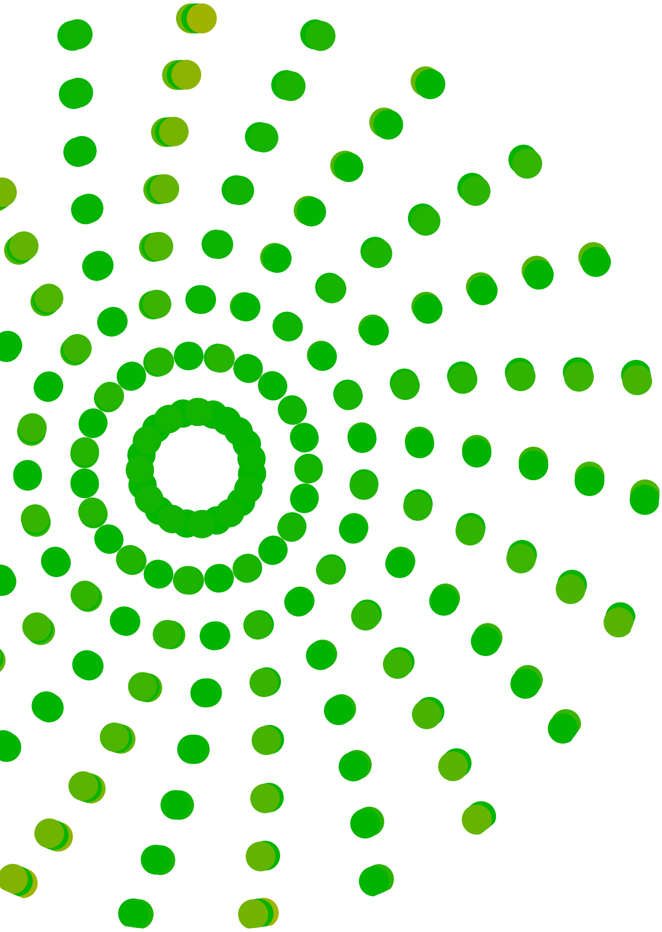


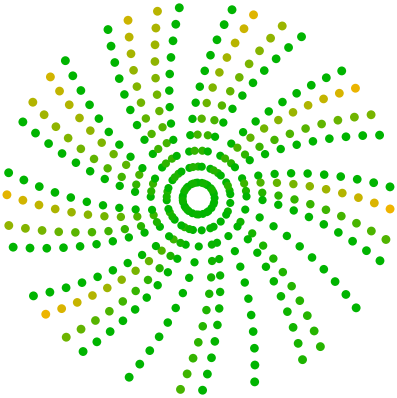
Math: Frequency and Visuals Based on Algorithm
After I explored the work of John Whitney, I wanted to try and use computational means to marry image and sound. The sound mandala was a series of spinning dots around a center. When the dots reached the meridian, they would play their corresponding sine wave notes.

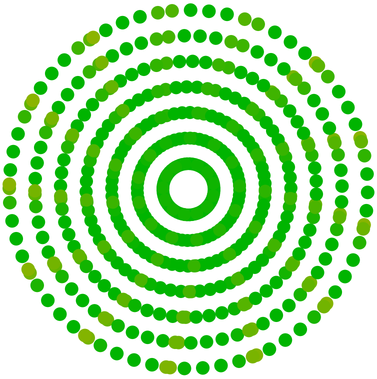
The first dot towards the center of the mandala represented the note G3, with each note after that in 12 tone scale intervals to the outer edge of the shape. Depending on how many dots were displayed, the mandala would play that many notes in the pattern. If the visual pattern changed, the notes would change their note patterns to match.

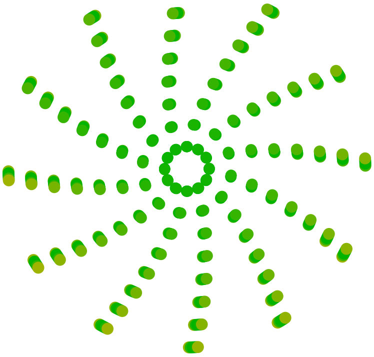
Conclusions & Hindsight
Although the mandala produced compelling visuals, it didn't do music any justice. After I completed my first draft, I researched who else might be working with John Whitney's theory and discovered Jim Bumgardner, a gifted flash programmer who had already done the same thing, only better, called the Whitney Music Box. [38] His mandala produced compelling visuals and notes based on better scale intervals and better sounds.

Finding out that someone else had already done this wasn't my only concern. The project didn't answer my questions about using space in a more natural or synaesthetic way. It did start to answer questions on how to build patterns from music notes, but I wasn't as interested in creating music as I was in placing music into visual space.









Track Color

Math: Frequency and Visuals Based on Algorithm

During the time I created the sound Mandala, I wanted to introduce more user-control through video color tracking. I took a piece of example code from the programming language Processing that tracked color, and made the height of the screen represent notes on the scale. I could train the program to track a particular color, like bright yellow, and move the swatch of color up and down in front of the camera to produce a music piece.

Conclusions & Hindsight

The results were weird and interesting, but I didn't get a sense that I was using space for anything more than a virtual Theremin. I could have trained the camera to use the X axis for something meaningful, but I had already decided that the interface had no meaning beyond playing with math.

I had the camera track my skin tone to produce sounds. It was fun for a while, but it got annoying.

A Size

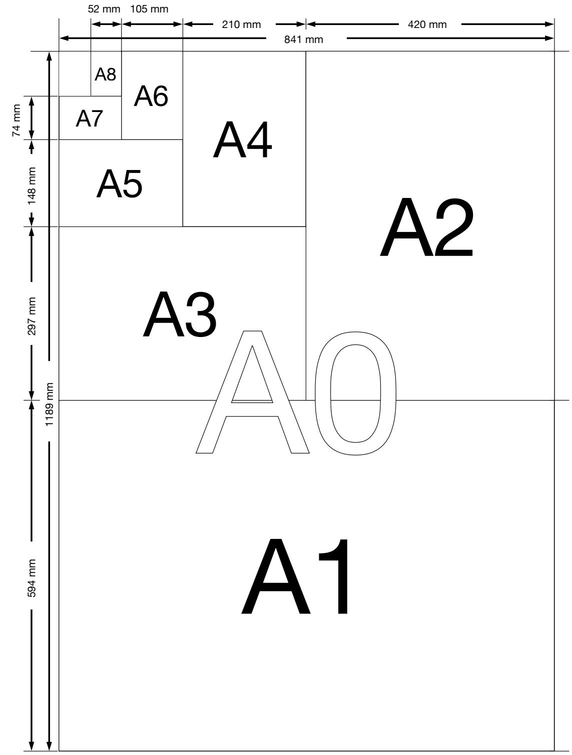
Math, Space, Size & Volume

While I was looking at using the vertical space, I discovered that there were some similarities between image and sound. Musical scale and paper can be defined in a series of ratios, particularly with the notion of the key of A versus the international standard A size of paper. With paper, each progressively smaller A size is a folded half size of the number before it. In example, A1 is half the folded size of A0. In music, there is a similar convention. The frequency of the key of A in cycles per second for each progressive scale is doubled. For instance, the note Middle A is 440 while A2 is half that at 220, and A3 doubled at 880.

Conclusions & Hindsight

There seemed to be some kind of divine connection between the two, but then I remembered that both 'A' conventions were man-made, not a property of mathematical physics. I had to start looking elsewhere if I wanted to find a connection between sound and image.

I then studied the A sizing for a completely different reason and discovered that I could use space the same way I intended to use it in the panorama project. It's not how big or how tall that matters in space: It's where that matters most. After this realization I decided to abandon plotting frequency against an axis and dove right into what I had been previously avoiding: using the notion of software space as real space.



A5	880.00
A4	440.00
A3	220.00
A2	110.00
A1	55.00
A0	27.50

The international paper standard ISO 216 (in mm).
next to the international pitch standard ISO 16 (in Hz).

Sound Shape Studies

Shape, Size & Part

I collaborated with Gunta Kaza on a project to investigate a connection between sound and shape.

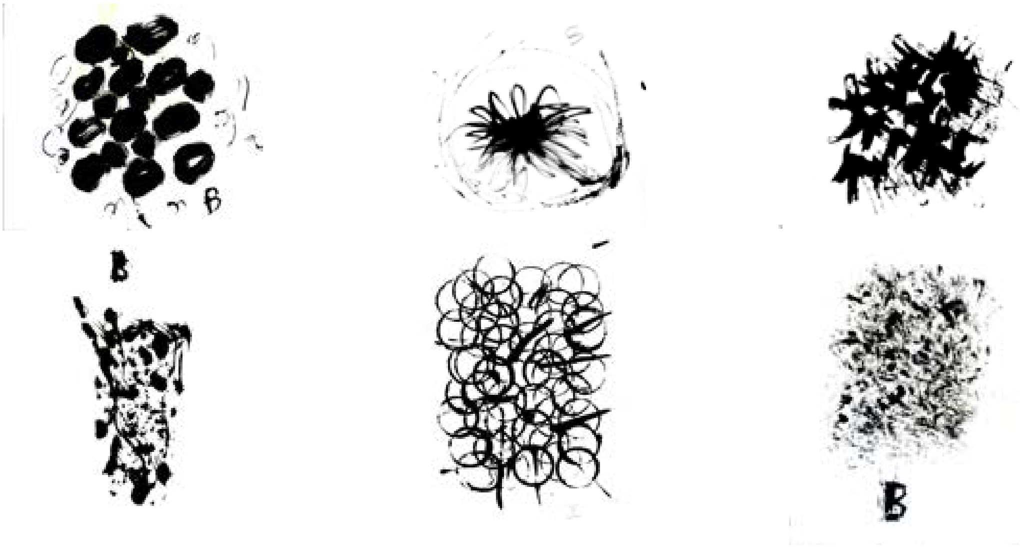
To construct the experiment I took 7 empty coffee cans and placed unique letters on each of them. I then placed various objects in each of the cans like rubber bands, bolts, dry chick peas, orzo, and more. We then brought the cans to her classroom and asked the students to draw what they heard–Without revealing the object inside. They were allowed to draw with whatever they wanted to using tempera paint. They weren't isolated from one another, so they were allowed to see what others were drawing (although they tended not to).

We had no expectations, but the drawings of objects that had distinctive sounds tended to look more alike than sounds that sounded more similar. For instance, orzo looked and sounded more soft than the rumbling of bolts.

Conclusions & Hindsight

I want to do much more with this work. There are a lot of possibilities: Isolating the participants, giving them all the same material to draw with, or involving people of various ages.

I realized right after the experiment that the cans must have played a large role in the sounds produced. The tinniness of the cans definitely influenced the output of the sound.

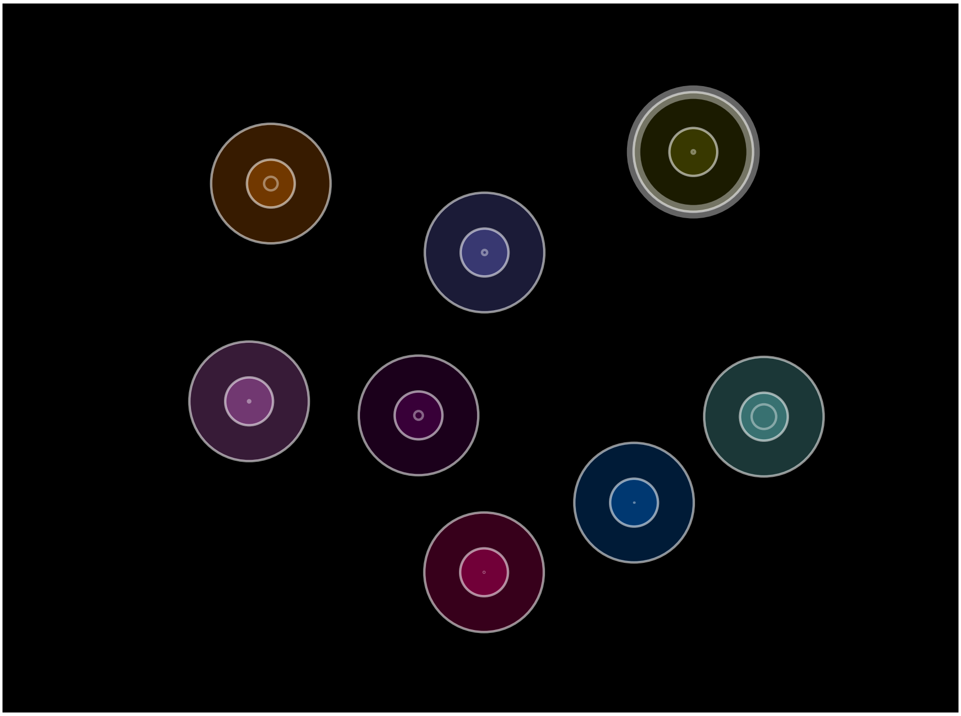


Top: Drawings from one of the sound cans.
Bottom: Students drawing the sounds.

ShapeMix

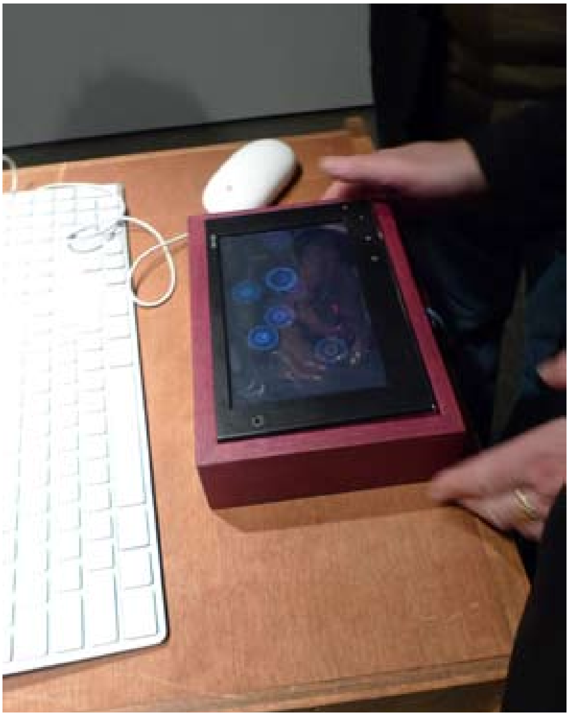
ShapeMix is the project I created as a testbed for several theories of audiovisual exchange. I wanted to create a place where one or more audiovisual objects could live in one space and be manipulated.

At the heart of the project is the notion of a confined space that represents left and right/ pan and up and down volume. each of the shapes (circles in this case) represents a sound. Inversely, each sound represents a shape. The two are synonymous. Each of the circles can be manipulated with sonic and visual effects that give audiovisual cues.

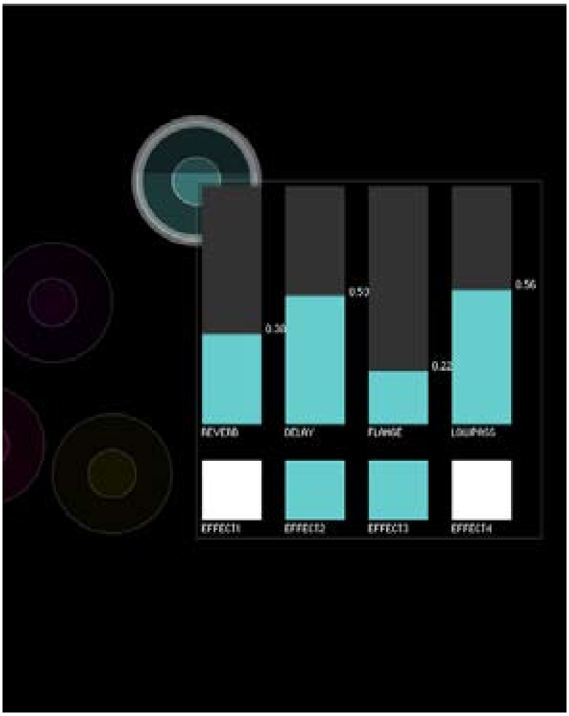


Eight sound shapes in various states of play within Volume/ Pan space in the laptop-based demo. The olive green shape with the soft edge in the upper-right-hand corner has a reverb applied.

I wanted to create a flexible software interface with an eye towards portability and ease of use. For the gallery show, I created a small touch interface on a screen inside of a box with an accelerometer inside. Because of the small size, the user could pick the box up and shake the shapes around. This probably could be done with a larger table, but I doubt it would have been easy to pick up and move around!



A tabletop version of shapeMix. The form factor of the small box allows a user to tip the box to move the shapes with an accelerometer.



Double tapping on a sound shape gives direct Manipulation of each channel's effects.

Several early experiemnts led me to the creation of the fully working prototype.

Beat Detection

I started working with a library based on Frédéric Patin's algorithms for detecting sound energy peaks. I took several recorded tracks from the same musical piece and assigned them to a circle. I then had the program enlarge each of the circles according to the beat detected.

The beat detection worked very well with percussive sounds (as you would expect). On the downside, it randomly assigned beats to sounds that didn't have as much definition. I dropped the idea of using beat detection in favor of volume data.

Pan & Space

I had circles placed in space, but it bothered me that they appeared in different places–yet sounded like they were coming from the same place. I experimented with audio pan by mapping the horizontal space of the circles in relationship to the left and right ears. One of the sounds to the far left was more apparent to the left ear. The two sounds near the middle sounded more in the center, with the one on the left leaning toward the left and the right one leaning toward the right.

Movement & Space

I had a single circle move by following the cursor. The movement of sound and shape from left to right worked well for panning, but nothing happened when I moved the circle up and down. I thought to use height as volume, making the circle bigger and the sound louder as it approached to the bottom of the screen and smaller (becoming almost nothing) when it reached the top.

Multiple Shapes & Volume

When I introduced multiple shapes, the size of the circles became a problem. Circles closest to the top would become difficult to find when I tried to click on them. Circles closest to the bottom would become difficult to locate and maneuver because they overlapped so much.

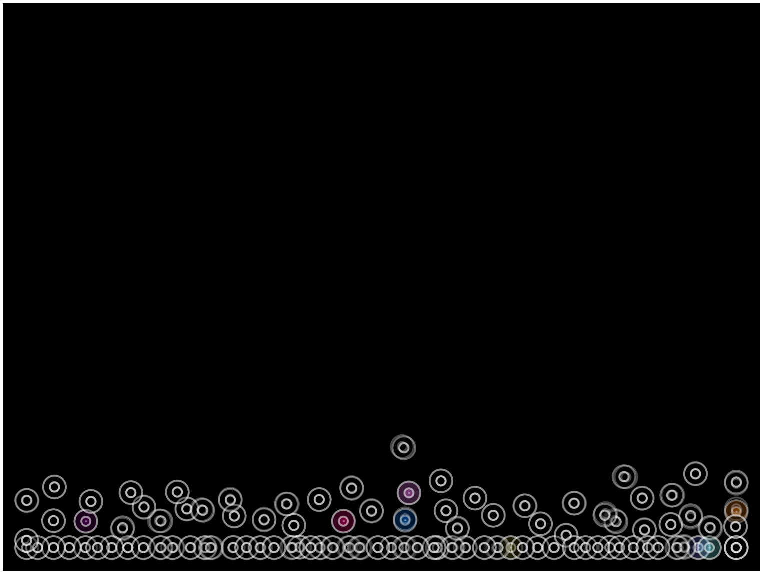
I settled on keeping the shapes the same size as one another so I could identify them as different objects of the same type. This would also make it easier later, when I would use a touch screen to move the objects around. If objects of different sizes were difficult for the mouse to locate, a touch screen with a lot larger clicking area would drive someone crazy trying to locate and move individual circles.

Since i didn't want the base shape of the circle to change, I had to represent volume in another way. There were several possibilities: represent volume as a non-filled second circle emanating from the middle; use the edge of the circle as volume; or use the opacity of the circle. I didn't want to use opacity because if I had any more than a few circles of similar base color it might get confusing. Using the circle's circumference for volume was mathematically difficult for the computer; the presentation method too limiting and having to read around the circumference would be odd.

I chose to represent the volume as a non-filled circle emanating from the center. It was far less distracting than the other two proposed methods and it made the interface easier to use.

Physics
I wanted to find a way to keep objects from overlapping too much on the screen. It didn't make sense to have two objects in the same place at the same time, since our version of physics doesn't support two things being in the same place at the same time. In audio mixing theory it's also prudent not to place two things in the same exact space (there are exceptions to this rule).

I adopted some simple physics principles of collision detection and separation. In addition to keeping the circles apart, I opted from them to "spring" or repel away from one another, like when you put two magnets with the same polarity near one another. I turned the



Gravity send the balls downward, bouncing off the bottom (and one another) until they settle, diminishing the sound.

whole space into a "box metaphor" so that the circles wouldn't pop out of the side and end up on the other side abruptly. I applied a simple gravity that, when turned on, would make all of the circles fall to the bottom and bounce like super balls.

The direction of volume had to change at this point. Up means volume-up and down means volume-down in audio mixing. In physics, rest means silence. When the volume was changed so that the bottom meant silence and the top meant full volume, it made sense because objects at the bottom didn't have any movement attached to them.

Later, I added an accelerometer that, when place on the bottom of an LCD screen laid flat on the table, would bend gravity in different directions, depending on which way it was turned. It reminded me of lifting a pool table and watching the balls fall to the bumper. This effect helped in instances where every sound needed to be nudged up or to the left a bit.

Adding gravity, spring and collisions to the mix had some fun unintended consequences. The act of randomly bouncing balls changed the volume and pan and created its own set of effects. Using a single ball to move the one next to it helped to widen the distance between two mixed objects, creating better audio separation. This reinforced the idea that natural metaphors can extend the language of motion using sound and image in physical space.

Effects

Effects seemed to have correlations between image and sound on a wave length level. Every time I came across an audio effect, I found a corresponding visual effect whether it was experienced by naked eye or on a smaller, molecular-atomic level. Panning is a psychoacoustic audio effect and the visual artifact of seeing an object in its relation to our heads.

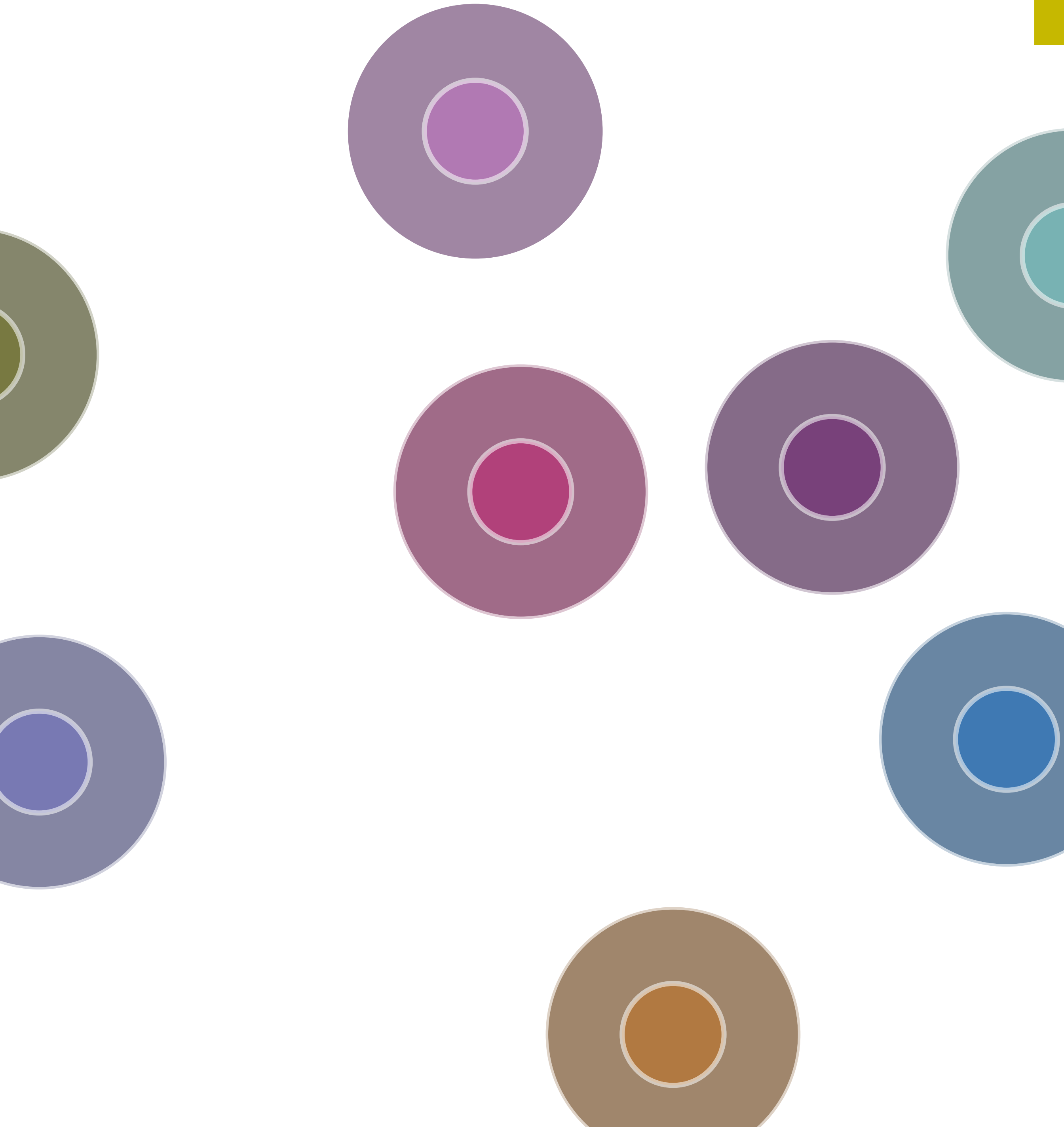
Delay was a little more difficult to comprehend, since the corresponding audio effect was based in astronomical physics. However, looking at the theory of audio delay, it too can be measured in wavelengths. This formed the basis of the edge effect, or effect based in levels much smaller than we can see translated to our naked eye as audiovisual correspondence.

Conclusions & Hindsight

If I want to extend this language, I will have to spend some considerable time analyzing and testing each effect type image and light analogy. I realize that both the research and the software environment are just in their infancies. I could create a set motion studies to look at the effects of a Doppler effect on the pitch or arc motion of a shape. I could look at the elasticity or rigidity of shape as a way of giving cues about sound. There are many others.

There are other metaphors beyond 2D space. Video could be used in place or along with shape to give a sense of space to both the sound and video. The ReacTable Environment could help give tactile feedback using physical markers on the screen that act as distinct objects. This could be useful in low light situations or with people with vision impairment. A true 3D space could give way to a totally immersive environment, however, until we invent a true 3d projection system, this is not possible.

This work was made possible with the computer, but its usefulness as a tool is limited. The mind is a much more powerful tool. The computer is merely a vehicle to enable the theory that there are connections between light and sound waves and that we behave in particular ways in space.



[1] Feldman, Robert S. *Development Across the Lifespan* (New Jersey, US: Pearson Prentice Hall), 165.

[2] Ackerman, Diane. *A Natural History of the Senses* (New York, US: Vintage Books, 1990), xv.

[3] Fuller, Buckminster. *Operating Manual for Spaceship Earth* (New York, US: E.P. Dutton & Co, 1971), Chapter 3.

[4] Pocock-Williams, Lynn (1992). *“Toward the Automatic Generation of Visual Music, ”* Leonardo, Vol. 25, No. 1 (1992), pp. 29-36.

[5] Potter, Ralph K. (Autumn, 1947). *“Audivisual Music,”* Hollywood Quarterly, Vol. 3, No. 1, pp. 66-78.

[6] Bishop, Bainbridge. *A Souvenir of the Color Organ, With Some Suggestions in Regard to The Soul of the Rainbow and The Harmony of Light*. Bainbridge Bishop, New York: New Russia, Essex County. 1893.

[7] Rimington, Alexander. *A New Art: Colour-Music*. Messrs. Spottiswoode & Co., New St. Square. June 13, 1895.

[8] B. M. Galeyev and I. L. Vanechkina (The MIT Press, August 2001). *“Was Scriabin a Synesthete?,” Leonardo, Vol. 34, Issue 4, pp.357-362.*

[9] Clauser, Henry R (MIT Press, 1988). *“Towards a Dynamic, Generative Computer Art,” Leonardo, Vol. 21, No. 2, pp. 115-122.*

[10] Miller, Patrick (Autumn, 1982 - Summer, 1983). *“Music and the Silent Film,”* Perspectives of New Music, Vol. 21, No. 1/2 , pp. 582-584.

[11] Hubbard, Preston J. (University of Illinois Press, Winter, 1985). *“Synchronized Sound and Movie-House Musicians,”* American Music, Vol. 3, No. 4, pp. 429-441.

[12] Gallez, Douglas W. (University of Texas Press on behalf of the Society for Cinema & Media Studies, Spring, 1970). *“Theories of Film Music,”* Cinema Journal, Vol. 9, No. 2, pp. 40-47.

[13] Milano, Paolo. (Blackwell Publishing on behalf of The American Society for Aesthetics, Spring, 1941) *“Music in the Film: Notes for a Morphology,”* The Journal of Aesthetics and Art Criticism, Vol. 1, No. 1, pp. 89-II.

[14] Gallez, Douglas W. (Blackwell Publishing on behalf of The American Society for Aesthetics, Spring, 1941). *“Music in the Film: Notes for a Morphology,”* The Journal of Aesthetics and Art Criticism, Vol. 1, No. 1, pp. 89-II.

[15] Fischinger, Oscar *(black-and-white Study No. 7, Optical poem 19).*

[16] Levin, Thomas Y. (The MIT Press , Summer, 2003). *“‘Tones from out of Nowhere’: Rudolph Pfenninger and the Archaeology of Synthetic Sound,”* Grey Room, No. 12, pp. 32-79.

[17] Levin, Thomas Y. (The MIT Press , Summer, 2003). *“‘Tones from out of Nowhere’: Rudolph Pfenninger and the Archaeology of Synthetic Sound,”* Grey Room, No. 12, pp. 32-79.

[18] Levin, Thomas Y. (The MIT Press , Summer, 2003). *“‘Tones from out of Nowhere’: Rudolph Pfenninger and the Archaeology of Synthetic Sound,”* Grey Room, No. 12, pp. 32-79.

[19] *Whitney, John. Digital Harmony: On the Complementarity of Music and Visual Art (New York, McGraw-Hill, 1980), 152.*

[20] Youngblood, Gene. *Expanded Cinema* (New York, E.P. Dutton, 1970),293.

Bibliography

21 *Schiffman, Harvey Richard. Sensation and Perception (New York, US: John Wiley & Sons, 2000), 326.*

22 Hallett, Ross. *Introductory Biophysics* (Toronto, CA: Methuen, 1976), 9.

23 Manley, Geoffrey A., *Evolution of the Vertebrate Auditory System* (New York: Springer, 2004), 97-98.

24 Darwin, Charles. *The Origin of Species* (London, UK: Oxford University Press, 1996), 190.

25 Heidelberg, "*Darwin's greatest challenge tackled, The mystery of eye evolution,"* European Molecular Biology Laboratory, http://www.embl.de/aboutus/news/press/press04/press28oct04/index.html

26 Parker, Andrew. New York, USA. *In the Blink of an Eye* (Basic Books, 2003), 189.

27 Mach, Ernst. *Analysis of Sensation* (Chicago, US, Open Court Publishing Company, 1914), 8.

28 Johnston, J.D. *"Audio Versus Video"* Audio Engineering Society of the Pacific Northwest.

29 Beerends, John G.; De Caluwe, Frank E. (Audio Engineering Society, May 1999). "*The Influence of Video Quality on Perceived Audio Quality and Vice Versa,"* JAES Volume 47 Issue 5 pp. 355-362.

30 Schiffman, Harvey Richard. *Sensation and Perception* (New York, US: John Wiley & Sons, 2000), 390.

31 Augoyard, Jean-François, Henry Torgue. *Sonic Experience* (Montreal, McGill-Queens's University Press, 2005), 28.

32 Young, Thomas. *The Library of Original Sources* (University Research Extension, New York), 442.

33 Born, Max. *The Principles of Optics* (University of Rochester, New York, 1999), 290.

34 Augoyard, Jean-François*, Henry Torgue. Sonic Experience* (Montreal, McGill-Queens's University Press, 2005), 37.

35 Lit, Alfred, *"The Magnitude of the Pulfrich Stereo Phenomenon as a Function of Binocular Differences of Intensity at Various Levels of Illumination,"* The American Journal of Psychology (April, 1949): 161.

36 Cowen, Ron. (February 22, 1997.). *"A Supernova Turns 10: Birthday of an Explosion,"* Science News, Volume 151, Number 8, pp. 120-121.

37 Richard Davis, *Complete guide to Film Scoring* (Berklee Press, Boston, 2000), page 154.

38 *http://www.coverpop.com/whitney/*

Design
This book is set in Adobe Chaparral Pro and Helvetica Neue.

The form factor of this book is based on the Major 6th 3:5 chromatic musical proportion proposed by Robert Bringhhurst in his book *The Elements of Typographic Style*.

The major chapters are color coded to RGBy (red, green, blue and yellow) based on an LED device RGBy by Makoto Hirahara and Shinya Matsuyama that changes color according to its surroundings.

Printed by ACME bookbinding, Boston, MA.

Software/ Hardware
The projects created for this thesis were coded/ programmed in a combination of Processing, Flash and Max/Msp with the help of snippets and open source code libraries by Daniel Shiffman, Ronald Smith, Mike Golembewski, Keith Peters, Jesse Kris, Ben Fry and Casey Reas, just to name a few.