# Real-Time Sound/Image Manipulation and Mapping in a Performance Setting

THOMAS CIUFO
Music/Special Studies
Brown University
1 Young Orchard Ave
Providence, RI 02912
USA
Thomas_Ciufo@Brown.edu

Abstract

Although the idea of combining sound and image is not new, recent developments in computing have made this idea more tangible. Real-time image and sound processing systems enable artists to affect sound and image generation, playback, and processing based on expressive performance attributes. One unique potential of these new performance works is that of sound-to-image and image-to-sound mapping, making sound and image structurally integrated. To achieve this integration, specific aspects of the audio can be analyzed and used to directly control the manipulation of specific aspects of the image, and visa versa. The real efficacy of this medium comes from combining the strength of interaction, real-time processing and sound/image linking and mapping into a singular work that exploits intermodal forms of expression. In this paper, I will present a brief history of the field, as well as my own research on real-time sound/image composition.

Key-words: Image, Mapping, Analysis, Performance

## 1 Sound/Image Background: Early Foundations

"So strongly do these color phenomena appeal to me that I venture to predict that in the not very distant future there may be a color art analogous to the art of music--a color music." Albert Michelson, *Light Waves and Their Uses*. 1899 [1]

The thought that there may be some relationship between sound and image can be traced back to ancient Greek philosophers, like Pythagoras and Aristotle who speculated regarding the correlation between musical scale and the rainbow spectrum of hues [2]. In *De Sensu*, Aristotle comments on this relationship: "Colors may mutually relate like music concords for their pleasantest arrangement; like those concords mutually proportionate" [3]. It would be many years before these ideas and speculations would materialize in the form of color/sound instruments, such as Father Louis Bertrand Castel's ocular harpsichord, built around 1730, or the various color organs of the late 1800's, such as those by Rimington and Bishop.

Two notable color music performances were given around the turn of the 20th century and apparently neither one was very well received [3]. Alexander Rimington, Professor of Fine Arts at Queen's College performed on the color organ in St. James's Hall in London in 1895. His one-to-one mapping of a single note to a specific color proved problematic and the London critics were not kind, commenting on the "restless flicker" on the screen [3]. In 1915, a color organ was used in the performance of Alexander Scriabine's symphonic poem, *Prometheus: A Poem of Fire*, but apparently the performance (at least the visual aspect) was not too successful. A review in the *New York Times* (March 21, 1915) pointed out what may have been the performance’s most prominent shortcoming: "So far as the lights were concerned, it could not be discovered how they added to or intensified the meaning of the 'music.' They were continually shifting and melting, but without visible relation to the sounds. In the midst of what seemed to be one phrase the lights would change half a dozen times. There was no variation in intensity as the music grew more emphatic: at the height of its proclamation there was the same pleasing variety of yellows, oranges, violets, purples, and emeralds as there was in the beginning."

Color music research and experimentation continued through the 20$^{th}$ century and included various incarnations of the color organ, including the Clavilux by Thomas Wilfred and the Sarabet, by inventor and musician Mary Greenewalt. In 1930 Wilfred founded the Art Institute of Light in New York, and by 1933 it included a lumina theatre with studios and laboratories, as well as a program series with lectures and recitals [3].

Other artists were also interested in the intersection between sound and image, especially abstract filmmakers such as Viking Eggeling, Harry Smith, Walter Ruttmann, James and John Whitney, and Oskar Fischinger. Fischinger was so interested in this field that in the late 1940's he invented a color instrument called the Lumigraph which was played by directly manipulating a simple latex screen with the hands. Painters and sculptors also contributed to this field, especially Bauhaus artists Kandinsky and Klee, who speculated on similarities between music and painting and Moholy-Nagy who investigated the physical properties of color and light [4]. Another contribution came from Adrien Bernard Klein's book "Color-Music: The Art of Light."

## 2 The Computer Age

Manipulating images with analog machines is cumbersome and limiting, due in part to the lack of flexibility inherent in the physical nature of these machines. Once constructed, they are very difficult to reconfigure and therefore produce a fairly narrow range of images. The emergence of computing technology would soon make available a radically different approach to the ideas (and problems) of color music. Golan Levan comments on this in his MIT masters thesis: "Computer technology has made it possible for visual music designers to transcend the limitations of physics, mechanics and optics, and overcome the specific/general conflict inherent in electromechanical and optomechanical visual instruments" [5].

Pioneering work was done in this new area by animation/film artist John Whitney, including an influential text on the subject called *Digital Harmony*, published in 1980 [6]. With the expansion of computing technology came a broad range of new applications, including digital sound synthesis, motion graphics, and 3D animation. This fundamental interest in relating sound and image continues today and is being explored by a range of artists, musicians, scientists, programmers and VJ's.

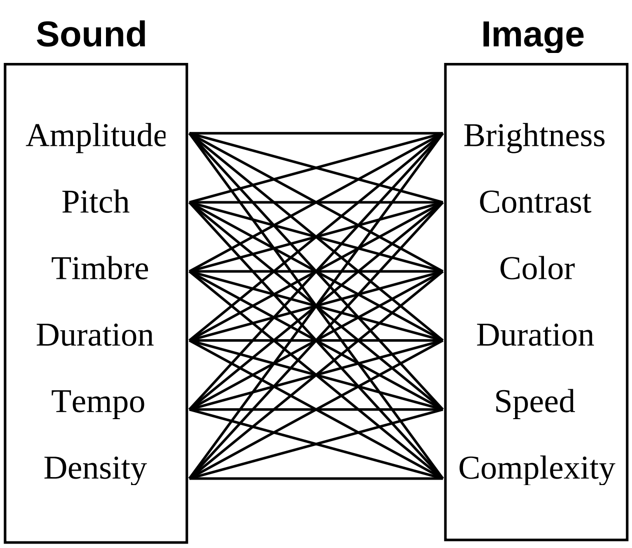
But as I look at the development of visual music, one important aspect seems to be neglected. Most of the work done in this field does not utilize an actual integration of sound and image on a structural level. By this I mean analyzing the sound or image and using that analysis to manipulate aspects of the other media. Two notable exceptions are the interactive video and music pieces by Don Ritter [7] and the graphic/sound environments of Golan Levin [5]. In most performance works, if some type of structural integration is attempted it is often only temporal simultaneity, in an attempt to imply integration by the presentation of images and sound happening at the same time.

## 3 Concepts of Linking and Mapping

Linking can be thought of as a general perceptive function and describes the intersection or interaction of sound and image. Mapping deals specifically with how these perceptual aspects connect and influence each other. The core issue regarding mapping, (whether or not there is some absolute correspondence) is well defined by Levin: "One more theme which has cut across more than four centuries of color-music research, from Castel's to that of the present day, is the question as to whether there are any 'absolute' correspondences between sound and vision. It is one of the deepest issues in the field; some have felt that it is best answered through empirical studies in psychology, while others have denied the possibility of any essential mappings, and instead held that the matter is simply an aesthetic one, best handled by design. The truth is almost certainly somewhere in between" [5].

Kapuscinski uses the term linking to refer to the combining of sound and image on equal terms and states that " Even an unlikely collision of sound and image can cause both of them to be evaluated with equal attention. It may even combat the usual dominance of sight over hearing" [8]. The important issue is not the equality between the media forms, (which is very dependent on individual perception) but how the different media interacts. According to Kapuscinski, linking in the simplest sense has to do primarily with temporal coincidence and correspondence. These correspondences or relationships can deal with tempo, density, level of activity, texture, dynamics and other similar perceptual qualities.

Mapping addresses how these perceptual aspects can be connected and used to influence each other. These mappings will vary greatly from piece to piece, depending on the specific concept or intent of the work. To facilitate this flexibility, most software attempts to address mapping in a modular way that allows for easy reconfigurability. Also of great importance is how these mappings might change or evolve over the course of the piece. The following illustration shows a possible mapping matrix with the ability to connect any element to any other element (or elements) in either direction.




## 4 Analysis

One fundamental approach to mapping is to start by examining the important perceptual features of the media elements used in a composition or an improvisation. As the sound or musical structure is being formed, what expressive attributes are available? Some obvious gestural elements in the sonic realm are amplitude (volume), pitch, timbre (tone quality), duration, tempo, rhythm, and density. A similar analysis process can be applied to the visual materials or generative system, including the analysis of brightness, color, contrast, duration, speed and complexity. The visual analysis will be highly dependent on the image content used or generated in the composition and may be severely limited due to the primitive image analysis options in most affordable real-time software.

The image types can be loosely arranged into three broad categories.

- Graphic base images (including three-dimensional environments)
- Prerecorded filmic images and/or still images
- Live camera images viewed or captured during the performance

Obviously these areas overlap and can often be combined in a single work. The analysis structure can be very challenging as the different elements (sound and image) may be progressing and developing simultaneously and will influence and shape each other and the nature of the analysis. One simplification is to drive the analysis primarily from one direction, preferencing either image or sound.

## 5 Specific Mapping and Manipulation Issues

Sometimes, very simple direct mappings are most effective (audio amplitude mapped to image brightness for example) while in other situations, more complex interrelated mappings may be desirable. In discussing his audiovisual piece *Dynamics*, Adriano Abbado outlines the three primary mapping modalities utilized [9].

The first is between timber in sound and shape-surface in the images. The second mapping is established between the locations of the sounds and images in space, and the third mapping creates a correspondence between perceived intensities of the aural and visual elements (loudness and brightness).

Aural/visual complexity is a central issue to be considered when dealing with mapping. The complexity can vary widely from a simple mapping between complex images and sounds, to a very complex mapping between simple images and sounds. Complexity can also be built up by combining several relatively simple interactions. Another aspect of the complexity equation is the level of bi-directionality utilized in the map. By utilizing bi-directional, multimodal mapping, highly responsive and reactive environments can be created. One way to describe complexity in this context is in reference to the perceptibility of the image/sound or sound/image interaction. In my work, I often reside somewhere in the middle of this continuum, allowing the sound/image relationship to be perceptible and expressive, but not simply resolved.

The field of real-time audio manipulation is well covered in the literature [10] so I will only briefly mention some of the readily available options. In a flexible real-time environment such as Max/Msp [11] generative structures, such as synthesis and resynthesis, prove especially malleable and effective in sound/image composition. Additionally, sampling, variable rate playback, granulation, filtering, and spacialization can all be interactively controlled and manipulated. Computer based real-time image manipulation is somewhat less mature, but is undergoing rapid development and expansion. A range of options exists, and some broad categories can be outlined. All of these categories can be mixed in various combinations. It is interesting to note that all of these processes have a similar counterpart in the audio realm.

- Filmic elements: variable rate playback, collage and montage, both spatial and temporal
- Recording/buffering: recorded materials (motion or still) can then receive filmic treatment or general processing
- Live input: images from a live camera or Internet stream
- Generative elements: drawing (lines, shapes, points, and objects, including three-dimensional) and algorithmic processes (fractals and lissajous)
- Processing: effects (scale, warp, solarize, emboss) color manipulation (hue, saturation, contrast, filtering), keying, blurring, and deformation

## 6 Current Concerns/Future Directions

My deepest concern and greatest challenge regarding sound/image composition is the issue of form and content. While these issues are not adequately addressed in this paper, they must be placed in the foreground if this field is going to mature. It is to easy to string together an unrelated series of images and sounds, processed through a range of unrelated effects. To progress beyond this superficial level of image and sound simultaneity, I am always seeking a compelling conceptual, content based approach that will steer or inform the sound/image integration. Only through this deeper engagement with content and form will we be able to move beyond the most obvious aspects of this medium.

Current tools offer reasonably sophisticated options for sound manipulation and analysis, but real-time video manipulation (and especially image analysis) is lagging behind. (For a discussion of various design approaches, please see Levin 2000).

Another concern revolves around current computing speeds and software optimization. In an effort to include high feature counts, most current software is too demanding for current low cost hardware, resulting in unacceptably low frame rates and poor image resolution. While these limitations can be exploited in certain contexts, they impose a certain aesthetic that is not always desired.

## 7 In Closing

The area of interactive image/sound composition has broad potential and could be utilized in many different types of creative works. The idea of color music is not new and a great deal can be learned by examining its history. The level of image/sound integration currently available, especially through analysis and mapping, offers new opportunities and challenges. Sound/image analysis and mapping are unique and expressive aspects of this intermodal medium. For this reason, they deserve original research and creative application.

References


[1] Michelson, Albert, *Light Waves and Their Uses*, The University of Chicago Press, 1903.

[2] Moritz, William, The Dream of Color Music, and Machines that made it Possible, *Animation World,* Issue 2.1, April, 1997.

[3] Wilfred, Thomas, Light and the Artist, *Journal of Aesthetics and Art Criticism*, June, 1947.

[4] Chipp, Herschel, *Theories of Modern Art*, University of California Press, 1968.

[5] Levin, Golan, Painterly Interfaces for Audiovisual Performance, MIT Master Thesis, 2000.

[6] Whitney, John, Digital *Harmony, On the Complementarity of Music and Visual Art*, McGraw-Hill, 1980.

[7] Ritter, Don, Interactive Video as a way of Life, *Music Works*, vol. 56, pp. 48-54, 1993.

[8] Kapuscinski, Jaroslaw, Compositions with Sounds and Images, http://www-crca.ucsd.edu/~jarek/ideas_top.html.

[9] Abbado, Adriano, Perceptual correspondences of Abstract Animation and Synthetic Sound, MIT Masters Thesis, 1988.

[10] Roads, Curtis, The *Computer Music Tutorial*, MIT Press, 1996.

[11] Max/Msp, http://www.cycling74.com/index.html.

Collopy, Fred, Visual Music in a Visual Programming Language, IEEE *Symposium on Visual Languages*, pp.111-118, 1999.

DeWitt, Tom, Visual Music: Searching for an Aesthetic, *Leonardo*, Vol 20, No. 2, pp.115-122, 1987.

Ox, Jack, The 21st Century Virtual Reality Color Organ, http://www.bway.net/~jackox/colororgansite/index.html.

Peacock, Kenneth, Synesthetic Perception: Alexander Scriabin's Color Hearing, *Music Perception*, Summer Vol. 2, No. 4, pp. 483-506, 1985.

We installed the sets on a stage for live performance, and rehearsed particular scenes of a musical. In live performance, using projection-based augmented reality technology enhances technical and theatrical aspects which were not possible with existing video projection techniques. The projected images on the surfaces of actorâ€™s costume could not only express the particular scene of a performance more effectively, but also lead the audience to an extraordinary visual experience.Â  We selected a theater, and installed stage set, stage objects, and the real-time projection mapping system, as illustrated in Figure 3 . Figure 3. Installation of stage sets, stage objects, and the real-time projection mapping system. Then, particular musical scenes applying our system were rehearsed. It's a real-time effect which can be combined with other effects in the mastering rack or the effects rack. In the Multitrack View, you can also vary pitch over time by using automation lanes. Choose Effects > Time and Pitch > Pitch Shifter effect, and set the following options: Pitch Transpose. Contains options that adjust pitch: Semi-Tones Transposes pitch in semi-tone increments, which equalmusical half-notes (for example, the note C# is one semi-tone higher than C). A setting of 0 reflects the original pitch; +12 semi-tones is an octave higher; -12 semi-tones is an octave lower.Â  To quickly determine which Precision setting to use, process a small selected range at each setting until you find the best balance of quality and processing time. Pitch Settings. Control how audio is processed See 5 steps to real-time process your instrument in the DAW. And What is (audio) direct monitoring? If you are sure you have to write your own, you can use libraries for real-time audio processing (as far as I know, C++ is better for this than C#). These libraries really works. They are specially designed for realtime.Â  apply your audio effect to the PCM (if the effect is a volume change over time, this could be done by progressively altering a volume factor between 0 to 1, multiplying the factor against the PCM). convert back to audio bytes. write to the SourceDataLine. All these steps have been covered in StackOverflow posts. The link tutorial does some simplification in how file locations, threads, and the stopping and starting are handled.