Representing Music with Visual Space and Color

Ryan McGee

MUS 274 | Winter 2010

Introduction

Visual space has been accepted as an analog to musical pitch [1]. The most obvious example of this relationship may be reading a musical score. As the positions of the notes move vertically up or down over time, the musician responds by playing pitches of ascending or descending frequency. Now, consider auditory space. In traditional music scores there may be placement of the performers, but there is no notation or consideration for the spatialization of sound. However, spatialization is a dimension used heavily by composers of computer and electronic music. If visual space relates to musical pitch, then what is a visual representation for auditory space? Do visual spatial movements relate to auditory space as well as these movements relate to pitch? Does visual color relate to pitch as strongly as visual space? Answers to these questions are not only relevant to music but also impact the fields of sonification and visual music.

Background

There exist two primary views concerning the mapping of visual sources to sound [2]. The first is known as indispensible attributes mapping. A dimension is defined as indispensible if and only if it is a prerequisite for perceptual numerosity (the ability to perceive more than one entity) [1]. Since different spatial locations are required to perceive separate visual objects, we call space an indispensible property of vision. Likewise, separate frequencies are required for us to perceive the presence of more than one simultaneous note, so pitch is an indispensible property of sound. Thus, indispensible attributes mapping agrees with the accepted relationship between visual space and musical pitch on the basis that both properties are indispensible to their respective mediums.

There is further support for indispensable attributes mapping. Our bodies are more sensitive to changes in visual space and pitch than to changes in the dispensable domains of visual color and auditory space [2]. Also supportive of indispensible attributes mapping is the observation of the Gestalt principle of rectification in both visual space and musical pitch. Gestalt rectification refers to the constructive aspect of perception in which our sensory systems have the ability to perceive more than what is actually present in the stimulus. Figure 1 shows an example of visual rectification. An example of rectification in the auditory domain is the ability to hear a fundamental frequency even though it is not present when playing only harmonics of that fundamental [4].

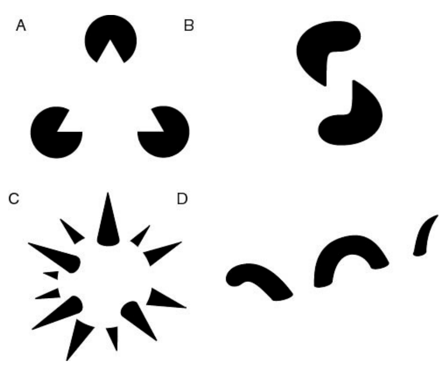


Figure 1. Visual Rectification [3]

The second view concerning the relationship between visual and sonic dimensions considers space and time relations rather than indispensability and is known as spatiotemporal mapping [2]. Using spatiotemporal mapping one would argue that auditory space, not pitch, is analogous to visual space. Spatiotemporal mapping also argues that visual color, not space, is analogous to pitch because both color and pitch are based on the frequency of waves.

There are previous experiments in support of spatiotemporal mapping. A study by Datteri and Howard concluded that an inverse relationship exists between the frequency of light and the frequency of sound [5]. For example, it was shown that relatively low frequency colors such as yellow mapped to relatively high audio frequencies greater than 1000 Hz. Image sonification techniques from Stanford's Center for Computer Research in Music and Acoustics have used spatiotemporal approaches for image sonification [6]. The Stanford group also introduces the concept of a soxel, the smallest discrete component of a given sound (analogous to a pixel in the visual domain) [7]. Mapping is achieved by translating the color and brightness of each pixel of an image to the pitch and loudness of a corresponding soxel [7]. Additional support for the spatiotemporal approach lies in a paper by Kubovy and Valkenburg that points out the influence the movement of visual objects can have on the perceived spatial location of auditory objects. When test subjects were presented with a sound and an off-center flash of light, the auditory location of the sound appeared to be shifted in the direction of the flash [2].

Though it is out of the scope of this particular experiment, timbral relationships should be mentioned when considering audio-visual relationships. Sound timbre is also known as sound color, so it is not surprising that many find it appropriate to map visual color to timbre rather than pitch. Yet, it has been argued that pitch is simply a dimension of timbre, implying that mapping visual color to pitch is actually a mapping of color to timbre [8]. Visual musicians often chose to map the shape and color of an image or animation to the timbre of a sound. Dewitt uses modulation techniques to produce a rose-like shape that maps to an amplitude-modulated sound in one of his works [4]. Dewitt also relates complex timbres to complex colors, while simple tones are black and white or colorless [4]. Abbado addresses the problem that there exists no visual notation for the timbre of complex, synthetic sounds and proposes mapping harmonic timbres to matte images and inharmonic timbres to shiny, reflective images [8]. When mapping to pitch, Abbado considers the corresponding frequency of visual shapes rather than the frequency of the wavelengths of visual colors [8]. For example, smooth visual shapes would have lower audio frequencies than edged shapes.

The field of sonification attempts to achieve an accurate and useful mapping of data to sound. The data to be sonified may or may not be related to the visual domain, but mappings to auditory pitch, timbre, intensity, and location are of primary interest. Concerning the accuracy of data sonification, there are no conclusions as to the consistent superiority of indispensible attributes or spatiotemporal mapping approaches. Many sonifications explore multichannel auditory displays that employ several speakers. When completed, The University of California, Santa Barbara's Allosphere, a unique multimodal scientific and artistic instrument for sonification and visualization, will have 512 channels of sound. In such cases, auditory space becomes an important dimension for the mapping of data to sound.

Hypothesis

It has been established that visual space is one accepted analog to auditory pitch. This paper theorizes that visual space will relate equally well to both auditory pitch and auditory space. Furthermore, it is expected that visual color will relate to pitch as strongly as visual space relates to pitch. This experiment uses variables of fit, pleasantness, and preference to determine the strength of each audio-visual relationship. Expectations are not aimed at proving superiority for spatiotemporal mapping over indispensability attributes mapping, but seek to find relatively equal amounts of fit, pleasantness, and preference for visual space to pitch and visual space to auditory space relationships. Also expected are roughly equal amounts of fit, pleasantness, and preference between visual space to pitch and visual color to pitch relationships.

Methodology

To test the hypothesis this experiment compared the indispensible attributes based relationship of visual space to pitch to spatiotemporal relationships of visual color to pitch and visual space to auditory space. For each relationship both direct and inverse mapping methods were used. For visual color a direct mapping implied that colors of higher frequency (shorter wavelength of light) relate to higher pitches and colors of lower frequency related to lower pitches. For the visual space to pitch relationship a direct mapping implied ascending frequency with upwards vertical movement and descending frequency with downwards vertical movement. Lastly, for the visual space to auditory space relationship a direct mapping implied that the sound pans towards the right for right horizontal movement and towards the left for left horizontal movement.

The inverse mappings related lower frequency colors to higher frequency sounds, ascending pitch to downwards vertical movement, and right panning to left horizontal movement. While direct mappings may intuitively seem to be stronger relationships, the purpose of testing inverse mappings was twofold. First, this experiment was an opportunity for a subsequent confirmation of the inverse color to pitch relationships found by [5]. Secondly, the two other inverse mappings serve to prove through ratings of fit, pleasantness, and preference that their direct counterparts are in fact stronger relationships. Figure 2 provides a summary of the 6 stimuli required for this experiment.

Color to Pitch (Direct and Inverse Frequency Mapping)

1) Direct mapping => red = lowest pitch
2) Inverse mapping => red = highest pitch

Visual Space to Pitch

3) Direct mapping => up = higher pitch
4) Inverse mapping => down = higher pitch

Visual Space to Auditory Space

5) Direct mapping => left = left panning
6) Inverse mapping => left = right panning

Figure 2. Summary of Experimental Stimuli

Stimuli were presented via laptop screen and stereo headphones. 15-second animated audio-visual sequences numbered according to the 6 relationships in Figure 2 were presented to subjects. For 1 and 2 subjects were presented with a square that changed color according to pitches in a melody. The melody was produced using simple sine wave tones in the key of A major and had an equal amount of ascending and descending pitches. For 3 and 4, subjects were presented with a white square that moved in direct or inverse vertical directions according to the same melody as 1 and 2. For 5 and 6 the same vertical motion of the square was translated to amounts of left and right movements in the horizontal plane and the melody replaced with a single 220Hz A note that panned left or right according to the visual movements. For sequences 3-6 in which visual color was not a parameter a white cube was used so that color would not affect results. Likewise, for 5 and 6 a 220Hz A tone was used so that pitch did not influence results. For 1 and 2 the colored square remained in the center of the screen, and for 1-4 the panning of the sound remained centered. Figure 3 shows examples of experiment stimuli.

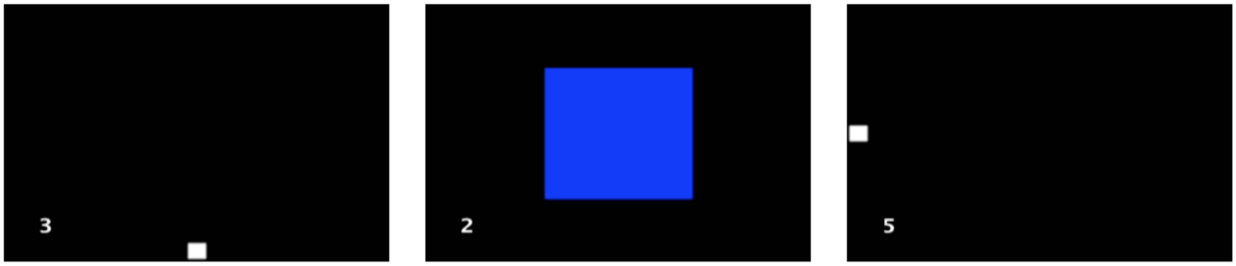


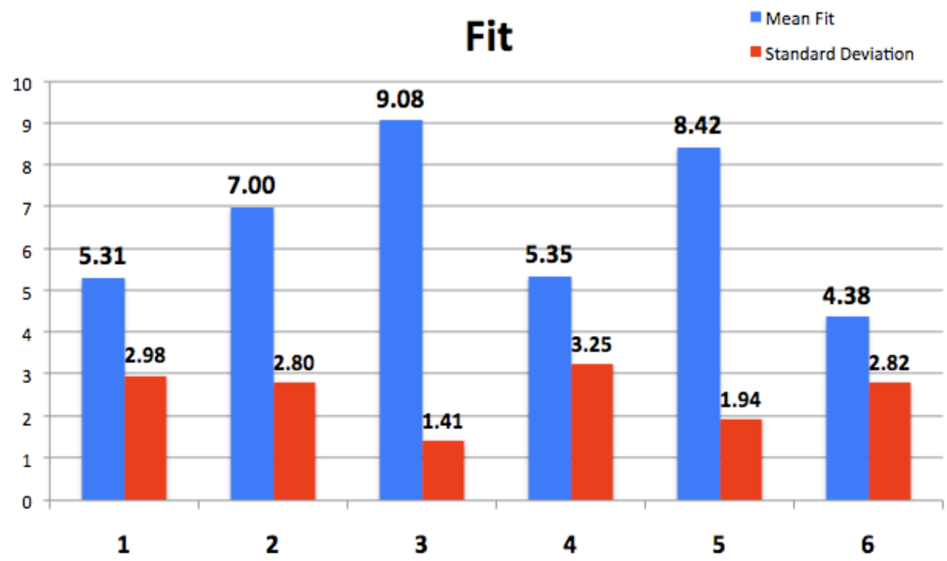
Figure 3. Examples of Experiment Stimuli

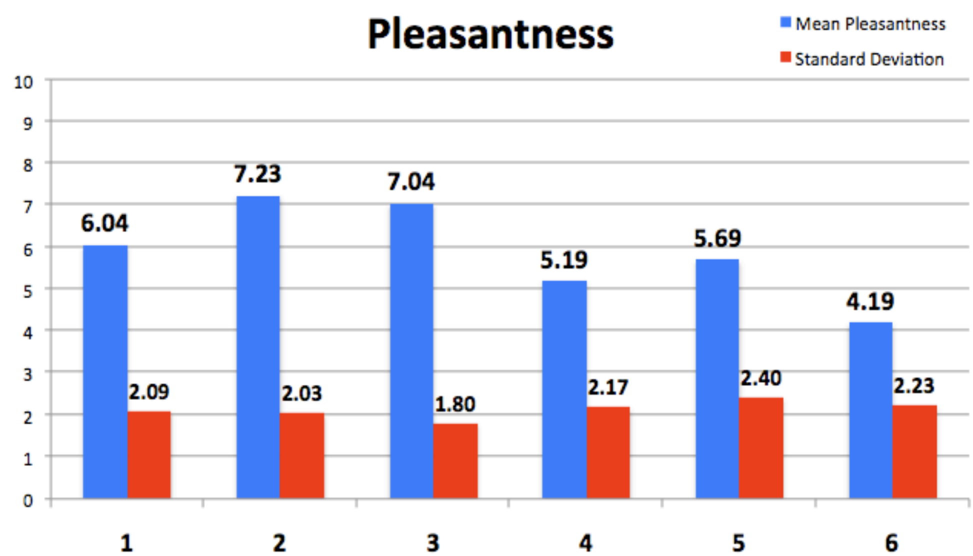
A total of 26 subjects were used in this experiment. All of the subjects were students in an undergraduate musical acoustics course and had experience reading music. Subjects viewed the stimuli in monitored groups of 2-3 students at separate computer stations with headphones. The stimuli were presented in random order for each subject. Subjects first viewed each of the stimuli without responding and then a second time when they had 30 seconds after each scenario to mark their responses. Data was collected using a response form on which subjects marked a response on ranging from 1 (low) to 10 (high) regarding degrees of fit, pleasantness, and preference for each of the sequences. Asking subjects about the degree of fit required them to think objectively about the best matching scenario, while asking for the degree of pleasantness and preference allowed them to rate subjectively.

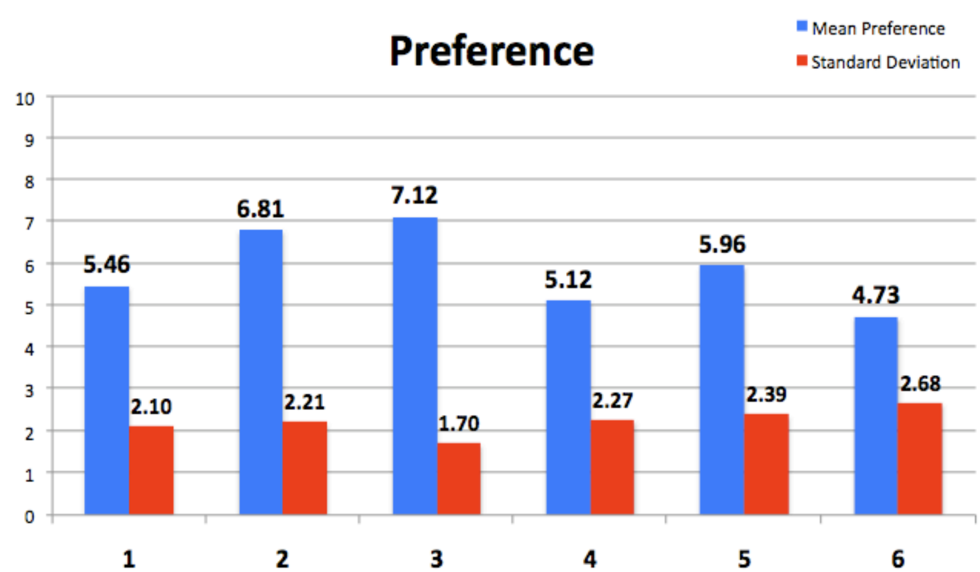
Results

Charts showing the means and standard deviations for degree of fit, pleasantness, and preference for each of the stimuli are below. A key for each sequence number is also provided.

Sequence Number Key
1) Direct frequency relationship of visual color to auditory pitch
2) Inverse frequency relationship of visual color to auditory pitch
3) Direct relationship of visual space to auditory pitch
4) Inverse relationship of visual space to auditory pitch
5) Direct relationship of visual space to auditory space
6) Inverse relationship of visual space to auditory space


Fit
Mean Fit
Standard Deviation
10
9
8
7
6
5
4
3
2
1
0
5.31
2.98
7.00
2.80
9.08
1.41
5.35
3.25
8.42
1.94
4.38
2.82
1
2
3
4
5
6


Pleasantness
Mean Pleasantness
Standard Deviation
10
9
8
7
6
5
4
3
2
1
0
6.04
2.09
7.23
2.03
7.04
1.80
5.19
2.17
5.69
2.40
4.19
2.23
1
2
3
4
5
6


Preference
Mean Preference
Standard Deviation
10
9
8
7
6
5
4
3
2
1
0
5.46
2.10
6.81
2.21
7.12
1.70
5.12
2.27
5.96
2.39
4.73
2.68
1
2
3
4
5
6

In terms of fit the direct mappings of both visual space to auditory pitch and visual space to auditory space stand out as the highest rated with the lowest standard deviations. For pleasantness the inverse mapping of visual color to auditory pitch and direct mapping of visual space to auditory pitch are the highest rated. In terms of personal preference the direct mapping of visual space to auditory pitch and the inverse mapping of visual color to auditory pitch are the highest rated. Note that only visual color has higher responses for its inverse mapping.

Conclusions

It was hypothesized that visual color would relate to pitch as strongly as visual space relates to pitch. Results for pleasantness and preference of the inverse mapping of visual color to pitch confirm this part of the hypothesis. Though the mean values show that visual color was slightly more pleasant but less preferred than visual space, the standard deviations are large enough to inhibit decisiveness by mean values alone. Thus, for this experiment it is valid to conclude that the pleasantness and preference for both visual color and visual space were equal when relating to pitch. Moreover, higher ratings across all variables for the inverse mapping of color to pitch over the direct mapping of color to pitch confirm Darreri and Howard's proposed inverse color relationship [5]. On the other hand, it was not shown that subjects believed the visual color to pitch relationship to fit as well as the visual space to pitch relationship. However, the high fit ratings for the visual space to pitch relationship could be attributed to the fact that all subjects had experience reading music and have thus been further accustomed than non-readers when using vertical movements to indicate pitch.

The second part of the hypothesis predicted that visual space would relate equally well to auditory pitch and auditory space. This part of the hypothesis was confirmed by similarly high ratings of fit for both direct mappings of visual space to pitch and visual space to auditory space. While the mean fit rating was slightly higher for the visual space to pitch relationship, standard deviations are again large enough to prevent a distinct difference in results. However, ratings for pleasantness and preference argue our hypothesis and show that visual space relates more strongly to pitch than to auditory space. Higher ratings for pleasantness and preference in regards to the visual space to pitch relationship could again be attributed to the subjects' music reading experience.

Overall, it is observed that visual color is as pleasant and preferred as visual space when mapping to pitch. Also, visual space is observed to fit equally well to both pitch and auditory space. Though visual color does not fit as well as visual space for pitch and visual space is not as pleasant or preferred for auditory space as for auditory pitch, it is believed that both hypotheses were adequately confirmed and that discrepancies are due to the lack of variety in the musical backgrounds of the experiment subjects.

In future experiments it will first be imperative to include subjects that have no formal musical training or experience reading music. Though vertical space to pitch relationships are seen outside musical notation (in cartoons for example), those with experience reading music have additional exposure to the visual space to pitch mappings in scores and may tend to rate those relationships higher than color mappings. Also interesting would be to include a number of subjects with a background in electronic music composition. Electronic and computer musicians may have more exposure and appreciation for auditory space and rate those relationships accordingly.

Timbre must also be addressed in future experiments when visual color is mapped to sound. This experiment used simple sine tones to avoid timbral influence, but there should be a comparison of visual color mapped separately to timbre and pitch. Lastly, the experimental stimuli should be expanded into 3-dimensional visual and auditory space. It proved to be difficult to generate sounds coming from above or below the listener using only headphones. Future experiments should map to 3-dimentional visuals to 3-dimentional auditory space by utilizing horizontally and vertically surrounding speaker arrays like those seen in many new media venues.

References

[1] Roger Shepard, "Pitch Perception and Measurement," in Music, Cognition, and Computerized Sound, Perry R. Cook, Ed., MIT Press, 1999, pp. 149-165.

[2] Michael Kubovy, David Van Valkenburg, "Auditory and Visual Objects," Cognition, Vol. 80, pp. 97-126, 2001.

[3] "Gestalt Psychology," [Online document], Wikipedia, Available at http://en.wikipedia.org/wiki/Gestalt_psychology

[4] Tom DeWitt, "Visual Music: Searching for an Aesthetic," Leonardo, Vol. 20, No. 2, pp.115-122.

[5] Darcee L. Datteri, Jeffrey N. Howard, "The Sound of Color," in Proceedings of the 8th International Conference on Music Perception & Cognition, Evanston, IL, 2004, pp. 767-771.

[6] Woon Seung Yeo, Jonathan Berger, "Application of Image Sonification Methods to Music," [Online document], (2005), Available at https://ccrma.stanford.edu/~woony/publications/Yeo_Berger-ICMC05.pdf

[7] Zune Lee, Jonathan Berger, Woon Seung Yeo, "Mapping Sound to Image in Interactive Multimedia Art," [Online document], (2004), Available at https://ccrma.stanford.edu/~zune/sources/papers/papers.files/ccrma2004.pdf

[8] Adriano Abbado, "Perceptual Correspondences of Abstract Animation and Synthetic Sound," Leonardo, Vol. 1, Supplemental Issue, pp. 3-5, 1988.