UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Visualizing Sound: Cross-Modal Mapping Between Music and Color

Permalink
https://escholarship.org/uc/item/7px9h0gg

Author
Griscom, William

Publication Date
2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Visualizing Sound: Cross-Modal Mapping Between Music and Color

By

William Griscom

A dissertation submitted in partial satisfaction of the

Requirements for the degree of

Doctor of Philosophy

In

Psychology

In the

Graduate Division

Of the

University of California, Berkeley

Committee in charge:

Professor Stephen Palmer, Chair
Professor David Whitney
Professor Robert Levenson
Professor Edmund Campion

Fall 2014

Abstract

Visualizing Sound: Cross-Modal Mapping Between Music and Color

By

William Griscom

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Stephen Palmer, Chair

Within the realm of perception, the phrase “cross-modal correspondences” refers to consistent associations between features in two different sensory domains, such as the fact that people commonly associate low-pitched tones with larger visual images and high-pitched tones with smaller visual images. These cross-modal correspondences play an important role in helping our brains solve the multisensory binding problem, that is, deciding what information from one domain to combine with information from another domain. It is less obvious what kind of role cross-modal correspondences might play in music perception and music visualization. In this thesis, I will describe three different research projects that explore questions about cross-modal correspondences within the context of music: 1) Do people have consistent associations between different colors and musical intervals and chords? 2) Do consistent visual associations exist for musical timbre? 3) What is the relationship between normal cross-modal correspondences and the neurological condition known as synesthesia, in which individuals actually experience visual sensations when they are listening to music. Across all three projects, I found that participants had a surprising level of consistency in their associations between visual and musical features, as well as intriguing evidence for a role of semantic and emotional features in guiding these cross-modal correspondences.

Chapter 1: Introduction

Part I: Multisensory Perception and Synesthesia

Our full experience of the world relies on all of our sensory capabilities: we know an object not only by its visual appearance, but also by it's distinctive auditory, tactile, and other sensory features. Our sensory systems can maximize the usefulness of these various streams of information for perception by combining input from multiple modalities into a single, unified representation, but doing so requires that we know which information in one sensory stream comes from the same external source as the information we are receiving in another sensory stream (Gibson, 1966; Spence & Driver, 2004). Importantly, we seem to use a wide variety of correspondences to guide this process, ones that go far beyond mere spatial and temporal coincidence (Driver & Spence, 1998). For example Kohler (1969) and later Ramachandran and Hubbard (2001) showed that people consistently associate the sound of the nonsense word "bouba" with a curved shape and the sound of the word "kiki" with a more angular shape. This appears to be true even across vastly different cultures, linguistic backgrounds, and age groups (Maurer, et al., 2006; Bremner et al., 2013).

This type of correspondence may be learned through experience with objects and behaviors in the world that follow consistent statistical patterns. In the bouba/kiki example, it is thought that we learn the association between a rounded mouth shape and the sound of open vowels, and generalize this to the "roundness" of written words and line drawings (Ramachandran & Hubbard, 2001). Another example would be that, due to the association between object size and resonant frequency, larger human beings tend to have deeper voices and bigger instruments tend to play lower notes, so a perceptual bias towards binding low-pitched sounds with large-sized objects would be useful for solving the multi-sensory binding problem. These associations among co-occurring low-level stimulus features such as size of pitch are sometimes called *statistical* correspondences (Spence, 2011), a term that stands in contrast to the idea of *semantic* correspondences based on higher-level features such as a shared linguistic term (e.g., "bright" sounds and "bright" colors) or shared symbolic meaning.

At this point there has been a great deal of basic psychophysical research done on mappings between low-level features of sound and images. Some consistent findings include associations between increasing auditory loudness and visual brightness (Bond & Stevens, 1969; Stevens & Marks, 1965) and between loudness and size (Lewkowicz & Turkewitz, 1980). Several studies have also found that higher pitches are associated with stimuli that are lighter (Marks, 1974; Wicker, 1968; Hubbard, 1996), smaller (Marks, 1987), and higher in the picture plane (Evans & Treisman, 2010). These associations have been shown to affect performance in a Stroop-like manner, such that response speed increases and accuracy decreases on recognition tasks when a simultaneous auditory stimulus matches a visual stimulus poorly along an analogous dimension (Bernstein and Edelstein, 1971; Evans & Treisman, 2010).

In addition to these studies of statistical correspondences between audio and visual features, there have been a smaller number of studies specifically investigating semantic correspondences in multisensory perception. For example, Karwoski et al. (1942) found that participants produced similar drawings in response to twelve different pieces of music played on a clarinet, and that these similarities were well-explained by categorization of the pieces and drawings along on semantic dimensions, such as *strong*, *happy*, and *exciting*. More recently, Palmer, Schloss, Xu, and Prado-Leon (2013) found even stronger evidence for regularities in semantic mapping between samples of classical music and color patches. When asked to choose the three colors that "went best with" and "went worst with" 18 selections of orchestral music, fully 94% of variance (r=.97) in the *happy/sad* ratings of the colors chosen to go with a given musical selection could be predicted from the *happy/sad* ratings of the musical selection (see Figure 1.2). This finding supports what Palmer et al. call the Emotional Mediation Hypothesis: the idea that cross-modal mappings between music and color are actually mediated by shared emotional meaning of the two sets of stimuli.

Some researchers (e.g., Martino & Marks, 2001; Hubbard, 1996; Osgood, 1981) also theorize that these patterns of cross-modal association may be related to the rare neurological condition called synesthesia, in which a person experiences a percept in one modality when they are presented with a stimulus in another modality or domain (for example, seeing colors when listening to particular pitches or chords). Estimates vary based on methodology, but somewhere between 1% and 5% of the general population are thought to experience some type of synesthesia (Hubbard & Ramachandran, 2001), with the most common forms being grapheme→color (in which uncolored letters produce a color experience in the same modality but in radically different dimensions) and time→space (in which units of time take on distinct spatial characteristics or locations). Only a very small percent of people experience truly multi-sensory synesthesias such as music→color, with some estimates putting it at about 20% of all synesthetes (Day, 2005).

Nonetheless, such synesthetes provide an interesting comparison group in studies of multi-sensory perception. Some researchers, such as Martino (2001) go so far as to argue that cross-modal associations in non-synesthetes may represent a form of "weak synesthesia" and that the distinction between normative perception and synesthesia should be thought of as quantitative in nature rather than qualitative. In this framework, synesthetes would be expected to share the same type of associations as non-synesthetes, but to experience these associations as direct conscious experiences rather than as abstract conceptual associations. Although controversial, there is currently some evidence to support the idea that synesthesia is related to non-synesthetic cross-modal associations. For example, Ward et al. (2006) found that although sound→color synesthetes have more precise and stable color associations than non-synesthetes, both associate lighter colors with higher pitches (see Figure 1.3) and more saturated colors with certain complex timbres. The alternative hypothesis would be that synesthetic associations do not obey the kind of relationships found in normal perception, but rather follow patterns that are either totally idiosyncratic or unique to synesthetes as a group**.**

Part II: The Present Research

Prior to the present research, studies of both sound-color synesthesia and cross-modal mapping from music have primarily focused on low level features such as loudness or pitch. However, comparatively little is known about what kind of visual associations listeners make with emergent features of musical sounds such as interval harmony, chords, timbres, melodies, and tempos, despite the fact that many sound→color synesthetes report having distinct responses to such musical features (Cytowic & Eagleman, 2009). The study of such emergent features presents an interesting test of the relationship between synesthesia and standard multi-sensory perception by allowing us to directly compare patterns of association between the two groups.

In this thesis, I describe three lines of study that shed new light on the question of whether synesthetic associations might be structurally similar to common cross-modal associations, as well as revealing information about cross-modal mappings more generally.

1) Do untrained, non-synesthetic individuals make consistent cross-modal mappings from basic musical units of harmony (e.g., intervals and chords) to colors? Previous research (Maher, 1980; Costa et al., 2000) has demonstrated that intervals of different sizes have consistent psychological or emotional qualities across listeners (e.g., a perfect fifth is rated as being more steady, friendly, and sonorous than a minor second), and thus it seems likely that semantic/emotional cross-modal mappings are possible for this class of stimuli, similar to those found by Palmer et al. (2013) for classical music. In Chapter 2 of this dissertation I describe how, using all possible intervals in an octave and a set of three-note chords built from these intervals, we explored the ways in which simple and increasingly complex harmonic structures map onto dimensions of color experience such as lightness, saturation, and hue. By using color pairs in addition to single colors, we also looked at mappings for relational features, such as color harmony, that are not necessarily present in single colors, but that may still be relevant for mapping to relational auditory stimuli such as intervals and triads. By collecting rating data on relevant emotional and semantic variables for both musical stimuli and color stimuli, we were also able to test the relative strength of emotion and low-level sensory features as explanations for mappings between the two domains.

2) Do non-synesthetic individuals show consistent cross-modal mappings from instrumental timbres to colors? Timbre is particularly interesting as a musical feature because it is inherently multidimensional, consisting of aspects that are temporal, spectral, and spectro-temporal in nature. For this reason, I describe in Chapter 3 how we looked at single and paired color mappings for instrumental timbres as well as spatial and temporal image features such as onset/offset timing and edge contrast. Using timbres of single instruments and more complex stimuli composed of multi-instrument timbres, we investigated whether mappings that are relevant for single-instrument timbres generalize to more complex timbral mixtures.

3) What is the relationship between these various cross-modal mappings and the visual experiences evoked in synesthesia? If cross-modal mappings are truly an example of "weak synesthesia" as Martino (2001) proposed, we should find the same pattern of responses for synesthetes and non-synesthetes when averaged across individuals, despite the differing sensory experiences of the two groups. Additionally, if color→music synesthesia is a result of direct communication between sensory areas, as proposed by researchers including Hubbard and Ramachandran (2001), we should expect to see mappings between music and color that are based on low-level sensory features rather than high-level semantics. Chapter 4 of this dissertation describes a study we conducted in which, by collecting data from 15 music→color synesthetes as well as an equal number of matched non-synesthetes, we were able to examine the extent to which these groups show similar structural and/or semantic mappings between musical features and colors. The results have important implications for developmental theories of synesthesia, as well as how we think about the relationship between synesthesia and cross-modal perception more generally.

Figure 1.1: Kiki and Bouba, from Ramachandran & Hubbard (2001)

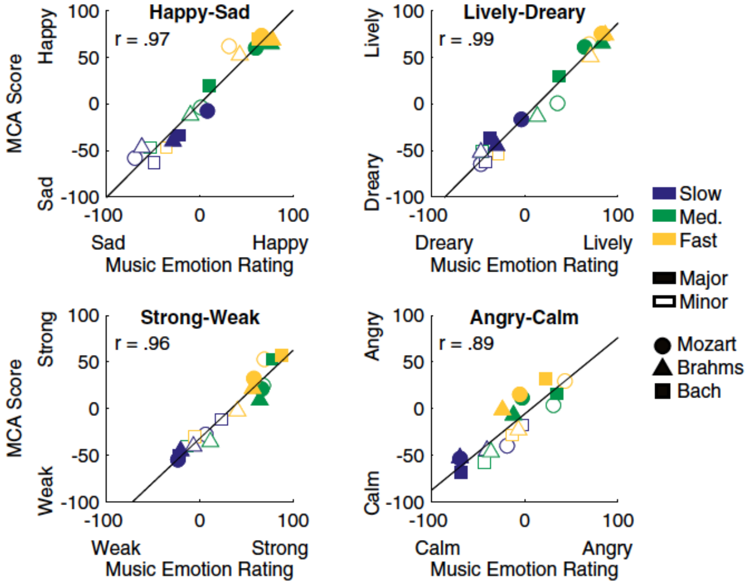

Figure 1.2: These data from Palmer et al. (2013) show the extremely high correlations between emotion ratings for a given piece of music (x-axis) and the average emotion rating of colors chosen as going well with that piece of music (y-axis) for four different emotion dimensions.

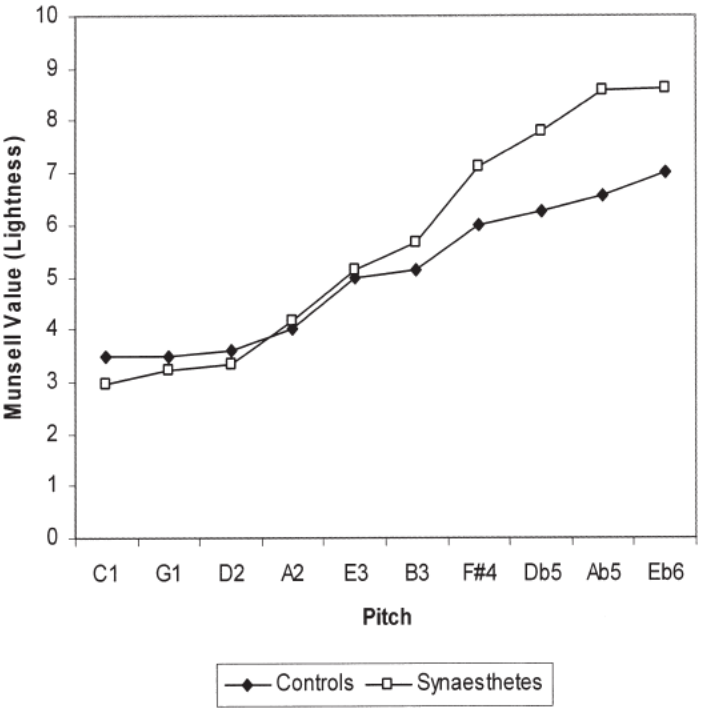

Figure 1.3: Ward et al. (2006) found that both synesthetes and non-synesthetes associate higher pitches with lighter colors.

Chapter 2: Cross-Modal Correspondences from Musical Intervals and Chords to Colors in Non-Synesthetes

Although there have been some preliminary studies of sound-color relationships in synesthetes (Ward, Huckstep and Tasnikos, 2006; Zamm, Schlaug, Eagleman, and Loui, 2013), until now there has been no systematic study of associations between colors and musical intervals or chords in non-synesthetes. This means that it is actually an open question whether untrained, non-synesthetic individuals even make consistent cross-modal mappings for basic units of harmony into the visual domain. If there are consistent associations across individuals, it also remains to be seen whether those mappings are based on low-level sensory correspondences, on emotional/semantic[1] associations, or something else completely.

There are good *a priori* reasons to believe that consistent sound-to-color maps for intervals and chords might exist and might even be consistent across individuals. Palmer et al. (2013) recently found that non-synesthetic participants report color associations with short excerpts of classical music that show interesting consistencies across individuals. A linear model using participant-generated ratings of emotional content for both colors and music was able to explain the vast majority of the variance (as high as r=.99 for some dimensions) in participants' choices for which colors went best with the music. For example, colors that were rated as high on the "happy" dimension (which were on average lighter, warmer, and more saturated) were more frequently associated with music that was similarly rated as high on the "happy" dimension (which tended on average to be faster in tempo and major in mode), and colors that were rated as high on the "sad" dimension (which were on average darker, cooler, and less saturated) were associated with music rated as high on the "sad" dimension (which tended on average to be slower in tempo and minor in mode). The same pattern was found for four different emotion dimensions: *happy/sad*, *calm/angry*, *lively/dreary*, and *strong/weak* (see Figure 1.2 for scatterplots).[2]

One of the strongest predictors of the emotion ratings for a given piece of music in Palmer et al.'s data was major vs. minor mode (see Figure 2.1). Given that the mode of a piece is largely determined by the type of intervals and chords it contains, it is not unreasonable to expect similar effects to be evident at the level of individual intervals and chords. It has long been theorized that intervals have specific emotional associations determined by the ratio of frequencies between their component notes (see for example Meyer, 1956), but it is only relatively recently that this has been tested empirically. For example, Costa et al. (2000) used the semantic differential technique developed by Osgood et al. (1957) to establish that intervals do have distinct emotional characteristics, both for trained and untrained listeners, and that these differences can be well-described using a low-level emotion space with the three dimensions of valence (*good/bad*), activity (*active/passive*) and strength (*strong/weak*). More recently, Oelmann and Lang (2009) had listeners rate four different intervals (major second, major third,

---

[1] We use the term "emotional/semantic" associations because it is often difficult to locate the dividing line between emotional and non-emotional semantic dimensions. This way we do not have to justify which dimensions are explicitly emotional and which ones are not: e.g. is *strong/weak* or *harmonious/disharmonious* an emotional dimension?

[2] One detail which may be worth mentioning here is that out of the four dimensions tested in Palmer et al.'s 2013 paper, three were confirmed to be strongly bipolar (happy/sad, r=-.94; strong/weak, r=-.87; lively/dreary, r=-.99), but the last dimension was much less so (calm/angry, r=-.51). For this reason in some of the subsequent experiments described below, we use two different dimensions in place of calm/angry (not angry/angry and calm/agitated) which better represent the underlying factors.

perfect fourth, and perfect fifth) using a large number of adjective rating scales, and found that there were significant differences between the intervals on many of the dimensions (for example, a perfect fifth is more *robust, vigorous,* and *hopeful*, whereas a major second is more *unhappy, frustrated,* and *hazy*). These differences were more pronounced when the raters were trained musicians, but were still statistically significant even when using responses from untrained naïve listeners.

Based on these results, it seems plausible that individual intervals carry sufficient emotional associations for participants to make color matches via emotional mediation. On the other hand, it is also entirely possible that emotional mediation is present in visual matches for complex musical passages (as demonstrated in Palmer et al.'s experiments) but does not hold for smaller isolated elements of music. This could happen either because individuals do not have systematic visual mappings for isolated intervals and chords or because those mappings are better explained by non-emotional factors such as low-level sensory correspondences. The clearest way to compare these hypotheses is to ask volunteers to make color matches for different intervals and then determine whether we can predict a significant portion of the variance in their choices using models based on shared emotional or semantic meaning. Those results can then be compared with models based on low-level sensory features.

Experiment 1: Intervals and Color Pairs

Methods

Participants. We recruited 21 undergraduate participants from the Berkeley Research Participant Pool. None had any color-vision vision defects, as tested using the Dvorine Pseudo-Isochromatic Plates. All gave informed consent and were naïve to the purpose of the study, and the Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

We did not formally screen participants for music-color synesthesia in this study (or any of the studies described in this dissertation, other than those in Chapter 4), but given the extremely low prevalence of any form of synesthesia in the general population (generally thought to be around 1%; for a detailed comparison of different estimates see Ramachandran and Hubbard, 2001) and the even lower prevalence of music-color synesthesia (estimated at around 0.2%), we felt confident that any significant results of this study would arise primarily or entirely from the behavior of non-synesthetes.

Stimuli. We generated 24 audio stimuli using multi-sampled recordings of a grand piano taken from the Kontakt Orchestral Library.[3] The stimuli consisted of pairs of notes forming different intervals across a range of two octaves. In each stimulus, the notes of the interval were played first sequentially (one at a time, with the lowest note first) and then simultaneously, with the duration of each note being one second. Twelve stimuli were created by holding the first note constant at middle C, and varying the second note through all possible positions in the chromatic scale up to an octave. Another twelve stimuli were created by transposing those intervals down an octave. During the experiment, participants listened to all of the audio stimuli at a fixed volume over Sennheiser HD-280 headphones.

The color stimuli we used were 18 color pairs selected from the Berkeley Color Project (BCP) set of 37 colors. These colors were originally used by Palmer and Schloss, (2010). Figure

---

[3] Thanks go to Dan Levitin of McGill University for advice on the audio software and sample libraries used in this and other experiments.

2.2 shows examples of what the BCP colors look like, and Table 2.1 shows the coordinates of the colors in CIE xyY color space. The 18 color pairs were chosen because they represented extremes on one of nine different colorimetric and emotional dimensions and were also close to average on the other dimensions (see Figure 2.3 for examples of colors used for each dimension), based on data from Schloss and Palmer (2011). The full list of color pairs used for the stimuli in this experiment is reported in Table 2.2.

Each color pair was displayed as a small square (100 x100 pixels) centered within a partly occluded larger square (300 x 300 pixels) against a neutral gray background (CIE x = 0.312, y = 0.318, Y = 19.26). All of the stimuli were presented on a 20" iMac (2007) computer monitor (1680 × 1050 pixels; 60-Hz refresh rate) in a darkened room from a distance of approximately 70 cm. The chromaticity and luminance functions of the red, green, and blue guns were measured using a Minolta CS100 Chroma Meter, and the measurements were then used to calculate the appropriate RGB values to ensure accurate presentation of the CIE xyY values for the chosen colors. All displays were generated and presented using Neurobehavioral Systems Presentation software.

Tasks. Each participant listened to the intervals played over headphones. For each interval two color pairs were presented, one on the left and one on the right side of the screen. The participant was asked to indicate which color pair "went best" with the audio stimulus by pressing the left or right key on the keyboard. Rather than view all possible combinations of color pairs for each audio stimulus, participants were only asked to choose between two color pairs that were strongly opposing for one of the relevant dimensions (e.g., *happy/sad*). Thus for each interval, every participant made 18 color choices. Trials were not blocked and were randomized across both visual and audio stimuli.

Subsequent to completing this task, participants were asked to rate all 37 visual stimuli and 24 audio stimuli on five emotional/semantic line-rating scales (*happy* vs. *sad*, *calm* vs. *angry*, *strong* vs. *weak*, *active* vs. *passive*, and *harmonious* vs. *disharmonious*). Participants rated these dimensions on a 400-pixel continuous rating scale with the dimensions labeled at the right and left endpoints of the scale. These trials were blocked by rating dimension, such that all stimuli were rated on one dimension before participants moved on to the next dimension. Prior to any given block of ratings, participants listened to all of the stimuli and were asked to choose two of the stimuli which were most extreme for that dimension (for example, the most *happy* and most *sad*), and instructed to use these as anchoring points for the ends of the scale. Explicitly anchoring the rating scales in this way ensured that all participants were using the scales in the same way and making full use of the range of possible ratings.

Results

Participant ratings of harmony for the interval stimuli are plotted in Figure 2.4. The overall pattern of harmony ratings is not dissimilar to the results of tonal hierarchy experiments conducted by Krumhansl (1979), in which participants listened to either an ascending or descending scale and were then asked how well a final note fit as the end point to that scale. Notably, Krumhansl found no evidence for the tonal hierarchy in listeners without musical training, who instead seemed to use a strategy based solely on interval size. These results indicate that a consistent tonal hierarchy can in fact be elicited from untrained participants simply by using a different task structure.

The average emotional/semantic rating scores for all of the interval stimuli in this experiment are reported in Figure 2.5. Average ratings on all five of the emotional/semantic dimensions differed significantly across the 24 different interval stimuli: *happy/sad* (*F*(11,220)=16.74, *p*<.001), *calm/angry* (*F*(11,220)=13.91, *p*<.001), *active/passive* (*F*(11,220)=5.26, *p*<.001), *strong/weak* (*F*(11,220)=4.07, *p*<.001), and *harmonious/disharmonious* (*F*(11,220)=17.52, *p*<.001). Additionally, we found that the main effect of octave was significant for all of the dimensions we tested: *happy/sad* (*F*(1,20)=23.76, *p*<.001), *calm/angry* (*F*(1,20)=2.32, *p*<.05), *active/passive* (*F*(1,20)=7.76, *p*<.001), *strong/weak*, (*F*(1,20)=20.58, *p*<.001), and *harmonious/disharmonious* (*F*(1,20)=2.12, *p*<.05). On average, the higher octave intervals were perceived to be happier (ΔM=75.74, SE=6.29), calmer (ΔM=18.29, SE=7.88), more active (ΔM=46.29, SE=8.56), weaker (ΔM=-69.83, SE=7.87), and more harmonious (ΔM=18.53, SE=8.76) than lower octave intervals. There were no significant interaction effects between octave and interval for any of the dimensions we tested.

To explore the dimensional structure of the emotional/semantic rating data, we computed the pairwise correlation matrix (see Table 2.3). It is immediately apparent that the dimensions are not independent of each other. For example, ratings of happy/sad were significantly correlated with all other tested dimensions. An exploratory factor analysis using maximum likelihood estimation confirmed that the emotion rating dimensions are highly redundant for these stimuli, with 90% of the variance in the ratings explained by a two-factor solution heavily weighted towards the dimensions *happy/sad* and *strong/weak*. The resulting factor weightings are reported in Table 2.4.

To investigate the possibility of emotional mediation in participant color choices, we looked more closely at the relationships between various different perceptual and emotional/semantic dimensions of the intervals and those of the colors chosen as "going well" with them. We calculated a score for each interval and for each rated dimension of the colors that we will henceforth refer to as the Sound-Color Association (*SCA*). The *SCA* represents a weighted average of the ratings of colors that were chosen as going well with a given interval for a particular dimension of the colors. A separate *SCA* score was calculated for each rating dimension according to the formula,

$$\mathrm{SCA_D(i)} = \Sigma\ ((\mathrm{p(i_L)}-.5) * (\mathrm{D(i_L)} - \mathrm{D(i_R)}))$$

where $D(i_L)$ and $D(i_R)$ are the ratings of color pairs on dimension *D* for the left (*L*) and right (*R*) color pairs when presented with interval *i* on a given trial, and $p(i_L)$ is the probability of choosing the left (*L*) color pair over the right (*R*) color pair for interval *i* on that trial.

To give an example, to calculate the SCA for the *happy/sad* dimension for the interval of a perfect fifth, we would first look at each of the 18 color comparisons for that interval and calculate the difference in average *happy/sad* ratings for the two color stimulus options in each comparison. For example, participants would have to choose between the color pairs LY-SY vs. DB-A1, which show a strong difference in *happy/sad* ratings (mean difference of 297.76), whereas DG-SG and DR-SR does not (mean difference of 40.1). We then multiply this lightness difference by the percentage of participants that chose one of the pairs over the other in the 2AFC task. If participants consistently chose LY-SY (light-yellow on saturated-yellow) over DB-A1 (dark-blue on dark-gray) when listening to a perfect fifth, then this would increase the lightness SCA for that interval, and if they consistently chose DB-A1 over LY-SY then this would decrease the SCA. Whether they chose DG-SG (dark-green on saturated-green) or DR-SR (dark-red on saturated-red) would have much less of an impact, because the difference score between those pairs is nearly an order of magnitude smaller. We averaged these weighted values across all of the different 2AFC comparisons that participants made. The end result is that if participants consistently chose the colors they subsequently rated as *happy* when they heard an interval of a perfect fifth, then $SCA_{happy}(perfect\ fifth)$ would be a highly positive number, and if they consistently chose colors they subsequently rated as being *sad* then it would be a highly negative number.

We calculated SCAs in this way for the five emotion dimensions and also for the four color appearance dimensions (red/green, blue/yellow, light/dark, and saturated/unsaturated). Average emotion ratings for the visual stimuli are reported in Table 2.5, and color values for the stimuli are reported in Table 2.6. By correlating harmony ratings for all of the intervals with the SCA values for those intervals along each rating dimension, we found that the colors chosen to go with more harmonious intervals were, on average, lighter (r=.84, p<.001), greener (r=.82, p<.001), bluer (r=.60, p<.05), and more saturated (r=.74, p<.01) than the colors chosen to go with less harmonious intervals, which were correspondingly darker, redder, yellower, and less saturated (grayer).

Why might harmonious (vs. disharmonious) intervals be chosen as "going better" with lighter, greener, bluer, and more saturated colors? One possibility is a statistical association: i.e., perhaps harmonious musical intervals occur more frequently when lighter, greener, bluer, and more saturated colors are seen, and this real-world statistical association has been learned by our participants. One problem with this hypothesis is that it does not actually explain the purported statistic association. Why should lighter, greener, bluer, and more saturated colors be more likely to be seen when more harmonious music is being heard? This is a difficult hypothesis to test directly without having access to the prior sensory experiences of our participants. Nevertheless, it is possible that such correlations exist in movies and other audio-visual media to which our participants are frequently exposed. If so, that fact would require an answer to the further question of why the media were created with these particular audio-visual correlations in the first place. Moreover, there is an alternative hypothesis that we can more easily evaluate given the data we collected in this study: perhaps these sound-to-color associations are mediated by common emotional/semantic associations.

To explore this possibility, we correlated participants' emotional ratings of the 12 musical intervals with the corresponding SCA values for those emotion dimensions (i.e., the weighted average emotion rating of the colors chosen as going well with the intervals). The results showed strong evidence for emotional mediation: more harmonious colors were chosen to go with more harmonious intervals (r=.97, p<.001), happier colors were chosen to go with happier intervals (r=.99, p<.001), and calmer colors were chosen to go with calmer intervals (r=.97, p<.001). There were also less extreme trends for stronger intervals to be chosen to go with stronger colors (r=.69, p<.05) and more active intervals to go with more active colors (r=.54, p=.07).

Discussion

These data show that non-synesthetic observers do, in fact, produce consistent and systematic color associations as going better with different musical intervals. Although interval harmony was correlated with color appearance features, the best predictors of participant color choices were the common underlying emotional/semantic dimensions, indicating that the emotional mediation hypothesis seems to hold even at the level of individual musical intervals. The fact that only three of the five emotional dimensions showed a level of significance comparable to that seen in Palmer et al. (2013) is actually interesting and potentially important. It could indicate that interval harmony is responsible for conveying emotional meaning along the *happy/sad* and *calm/angry* dimensions, whereas other musical features that are present in a full recording of an extended musical piece, such as timbre, loudness, and/or rhythm, are responsible for conveying emotional meaning along the *strong/weak* and *active/passive* dimensions.

Experiment 2: Intervals and Single Colors

In Experiment 1, we used color pairs as visual stimuli in order to capture the inherently relational aspects of color harmony, which we thought might be important for matching colors to harmonious and disharmonious intervals. In the present experiment we wanted to see whether participants could make the same kind of association judgments using choices among single colors. This has the tangential benefit of helping us understand how (if at all) participants understand the idea of "harmony" existing within a single color. It also allows us to use a task design similar to that used by Palmer et al. (2013) which gives the participant significantly more freedom, as they are able to choose for themselves which sets of colors which they feel "go best" (and "go worst") with an audio stimulus rather than making a forced choice between two alternatives.

Methods

Participants. We recruited 17 participants from the Berkeley Research Participant Pool. None had color-vision deficiency as tested using the Dvorine Pseudo-Isochromatic Plates. All gave informed consent and were naïve to the purpose of the study. The Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

Stimuli. The auditory stimuli in this experiment were the same as those described in Experiment 1. However, rather than having participants choose between color pairs, we displayed the entire set of 37 BCP colors on the screen simultaneously (see Figure 2.7 for an image of the screen layout). As can be seen in that figure, the colors were organized into four hue circles (actually displayed as squares in the four quadrants) defined by "cut" (light, dark, muted, and saturated in quadrants going clockwise from 12 o'clock) with a neutral (achromatic) color of the average lightness for that cut at the center of each hue circle. This meant that medium gray appeared at two locations (at the center of the muted and saturated cuts) and black and white appeared at the top, outside any of the hue circles. Each colored patch was presented as a square 100 x 100 pixels in size.

Design. Participants listened to the entire set of intervals in a random order. For each interval they were asked first to choose the best three colors to go with the interval (in order) from the set of 37. The interval was then repeated, and they were asked to choose the worst three colors to go with the interval (in order). Colors were selected by clicking on them in the display, and each time a participant chose a color from the display that color disappeared from the screen to prevent the participant from choosing the same color multiple times.

Subsequent to the color choice task, participants were asked to rate all 37 colors and all 24 audio stimuli on five emotional/semantic dimensions using a 400-pixel line rating scale, blocked by rating dimension and stimulus category. The five dimensions were the same as in the previous experiment (*happy/sad*, *calm/angry*, *active/passive*, *strong/weak*, and *harmonious/disharmonious*).

Results

The results of this experiment are plotted in Figure 2.8. Once again, we found that interval size significantly affected participant ratings on all emotional/semantic dimensions: *happy/sad* ($F(11,187)=12.92$, $p<.001$), *calm/angry* ($F(11,187)=6.93$, $p<.001$), *active/passive* ($F(11,187)=3.94$, $p<.001$), *strong/weak* ($F(11,187)=1.77$, $p<.05$), and *harmonious/disharmonious* ($F(11,187)=11.66$, $p<.001$).

We also saw significant main effects in all of the ratings dimensions for the high vs. low octave of the interval: *happy/sad* ($F(1,17)=26.07$, $p<.001$), *calm/angry* ($F(1,17)=13.15$, $p<.001$), *active/passive* ($F(1,17)=23.74$, $p<.001$), *strong/weak* ($F(1,17)=5.02$, $p<.001$), and *harmonious/disharmonious* ($F(1,17)=33.5$, $p<.001$).

For this experiment, Sound-Color Association scores were calculated differently from those in Experiment 1 because of the difference in participant response method (choosing the three best/worst colors rather than two-alternative forced choice). Here we used a formula analogous to that used by Palmer et al. (2013) in their study on music-to-color associations for classical orchestral music:

$$SCA_D(i)= (3 * D(i_{C1}) + 2 * D(i_{C2}) + D(i_{C3}) - 3* D(i_{I1}) - 2* D(i_{I2}) - D(i_{I3}))$$

where D is the relevant dimension, i is a given interval, $C1$ through $C3$ are the colors chosen as being most “consistent” with that interval (in order), and I1 through I3 are the colors chosen as being most “inconsistent” with that interval (in order). The choices are weighted such that the first color chosen has greater influence on the resulting SCA score than subsequent color choices, and the second color chosen has greater influence than the last. The average emotion ratings and color dimensions for all of the stimuli used to calculate these scores can be found in Tables 2.7 and 2.8.

The correlations between interval ratings and SCA scores were highly significant for *harmonious/disharmonious* ($r=.86$, $p<.001$), *happy/sad* ($r=.97$, $p<.001$), and *calm/angry* ($r=.92$, $p<.001$), somewhat lower for *active/passive* ($r=.79$, $p<.01$) and lowest of all for *strong/weak* ($r=.44$, $p=.15$). These correlations are comparable to the corresponding correlations obtained in Experiment 1.

The mean color appearance ratings of the colors chosen as going with each of the interval stimuli can be seen in Figure 2.9. Colors chosen to go with more harmonious intervals were on average significantly greener ($r=.71$, $p<.001$) and bluer ($r=.64$, $p<.001$) but not significantly lighter ($r=.11$, $p=.73$) or more saturated ($r=.05$, $p=.87$) than the colors chosen to go with less harmonious intervals. These results are somewhat different from those obtained in Experiment 1, where all of the color dimensions were significantly related to interval harmony. The discrepancies may indicate that when given free choice to select which colors they believe best match an interval, participants do not consistently relate lightness and saturation to the degree of interval harmony. They may also indicate that the lightness and saturation effects evident in Experiment 1 can only be properly evaluated when relational color associations are available.

Discussion

We found that participants were in fact able to complete this task using single color matches and that they gave largely comparable results to the 2AFC design using color pair stimuli. The same three emotion rating dimensions again showed effects consistent with emotional mediation, although on average the correlation values were slightly lower than those in Experiment 1. We also saw that interval harmony significantly predicted participant color choices, although in this case the effect was limited to the two hue dimensions and did not extend to lightness and saturation.

Experiment 3: Simultaneous and Sequential Intervals

The audio stimuli used for Experiments 1 and 2 included both sequential and simultaneous harmony: first the notes were played one after the other, and then together. In the present experiment we investigated whether the same color relationships would be present if listeners heard only the simultaneous or only the sequential parts of the stimulus, as these presentations may be perceived differently. For example, there is good reason to believe that the amount of destructive interference in simultaneously presented tones largely determines the amount of displeasure that some people experience when listening to dissonant intervals (see McDermott et al., 2010), and this is a physical feature that is entirely absent in the sequential stimulus.

Methods

Participants. We recruited 36 participants from the Berkeley Research Participant Pool. None had color-vision deficiency as tested using the Dvorine Pseudo-Isochromatic Plates. All gave informed consent and were naïve to the purpose of the study. The Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

Stimuli. The audio stimuli consisted of the same set of intervals used in Experiments 1 and 2, except that they were presented either as sequential pairs or as simultaneous intervals. In the simultaneous condition, the two notes of the interval were played together for one second, followed by a second of silence, and then played again for one second. In the sequential condition the notes were played in order from low to high, at a rate of one note per second, and then played again after a one second pause, for a total of two repetitions. The visual stimulus was the same array of the 37 BCP colors described in Experiment 2.

Design. Alternating participants were assigned to the two conditions, one of which heard the simultaneous intervals and one of which heard the sequential intervals for all stimuli. Participants listened to the entire set of intervals in a random order. For each interval they were asked to first choose the three colors that went best with the interval (in order) from the set of 37. The interval was then repeated, and they were asked to choose the three colors that went worst with the interval (in order).

Subsequent to the color choice task, participants were asked to rate all of the visual and audio stimuli on five semantic dimensions using a 400-pixel line rating scale, blocked by rating dimension and stimulus modality. The five rated dimensions were the same as in Experiments 1 and 2.

Results and Discussion

Figure 2.10 shows the mean ratings for each emotional/semantic dimension in both the sequential (left column) and simultaneous presentation conditions (right column). Main effects of interval were evident for all dimensions except *active/passive*, which was not significant in either condition (see Table 2.9).

There were significant effects of octave on every dimension for both groups, with the sole exception of *calm/agitated* in the sequential condition, which was marginally significant: *happy/sad* (F(1,17)= 13.95, 2.69, p<.001, .01), *not-angry/angry* (F(1,17)= 3.56, 4.06, p<.001, .001), *calm/agitated* (F(1,17)= 3.12, 1.90, p<.001, .06), *active/passive* (F(1,17)= 12.14, 12.86, p<.001, .001), *strong/weak* (F(1,17)= 8.13, 6.10, p<.001, .001), *harmonious/disharmonious* (F(1,17)= 6.49, 7.10, p<.001, .001) for sequential and simultaneous presentation, respectively. In both the sequential and simultaneous presentation groups, intervals in the higher octave were rated as being happier, calmer, more active, weaker, and more harmonious.

We also calculated SCA scores using the formula described for Experiment 2. With the sole exception of *strong/weak* in the simultaneous interval condition, SCA scores were strongly correlated with emotional/semantic ratings for all dimensions and for both groups. In the simultaneous condition, interval ratings of emotion were significantly correlated with SCA for *happy/sad* (r=.92, =.90, p<.001, .001), *not-angry/angry* (r=.89, 89, p<.001, .001), *calm/agitated* (r=.49, .50, p<.05, .05), *weak/strong* (r=.23, .76, p=.27, <.001), *active/passive* (r=.85, .86, p<.001, .001) and *harmonious/disharmonious* (r=.92, .87, p<.001, 001) for the simultaneous and sequential conditions, respectively. For a visual comparison of the correlations for the two presentation types, see Figure 2.12. Notice that the correlations are almost identical except for *weak/strong*, where the ratings of the simultaneous presentations are much weaker than of the sequential presentations.

Overall, the emotion ratings for intervals in the two tasks were highly correlated (*happy/sad* r=.85, p<.001, *not-angry/angry* r=.74, p<.001, *calm/agitated* r=.58, p<.01, *weak/strong* r=.67, p<.001, *active/passive* r=.77, p<.001, and *harmonious/disharmonious* r=.84, p<.001). Somewhat surprisingly, the colors chosen as going with the intervals, as indexed by the average SCA values on the four color appearance dimensions, were significantly related only for yellow-blue, (r=.79, p<.001) and light-dark (r=.95, p<.001), with saturation (r=.13, p=.54) and red-green (r=-.01, p=.96) producing essentially no relation.

We also ran a series of difference score t-tests to determine whether the emotional ratings differed between the simultaneous/sequential conditions. The results were not significant for the dimensions *happy/sad* (t(11)=.76, p=.46), *not-angry/angry* (t(11)=-.58, p=.57), *calm/agitated* (t(11)=-.28, p=.79), and *strong/weak* (t(11)=-1.31, p=.22) but were weakly significant for the dimensions of *active/passive* (t(11)=-2.69, p<.05) and *harmonious/disharmonious* (t(11)=2.83, p<.05), with the sequential intervals being rated as overall more harmonious and less active (t(11)=-1.45, p=.175) and red vs. green (t(11)=1.48, p=.167) were all non-significant.

The fact that both groups showed strong evidence for emotional mediation indicates that simultaneous vs. sequential harmony may not be a highly relevant factor for color associations with two-note intervals. Both types of interval presentations convey emotional distinctions to listeners, and the type of emotions conveyed by different intervals were strongly correlated across conditions. In terms of the colors chosen, these emotional similarities were conveyed primarily by the dimensions of lightness and saturation, with less consistency in hue dimensions. There are some small differences between the conditions, however. For example, the sequential intervals were perceived as being more harmonious overall, which aligns well with the idea that destructive interference is an important part of our perception of harmony and dissonance. Nonetheless, destructive interference is clearly not the only important feature of intervals, as participants made systematic matches based on perceived harmony in both conditions. This provides further evidence that interval-to-color matching may be based on high-level emotional meaning rather than low-level acoustic features.

Experiment 4: Chords and Color Pairs

Having demonstrated that participants have consistent color associations for different intervals, we decided to investigate whether similar results could be obtained for the more complex harmonies present in three-note chords. Do chords also show effects of emotional mediation? And, if so, to what extent can we predict participant responses to chords based on their responses to its various two-note component intervals?

Methods

Participants. We recruited another 20 participants from the Berkeley Research Participant Pool. None had color-vision deficiency as tested using the Dvorine Pseudo-Isochromatic Plates. All gave informed consent and were naïve to the purpose of the study. The Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

Stimuli. We generated recordings of 21 three-note chords consisting of middle-C and all possible combinations of two additional notes in the major scale an octave above middle-C. We also generated recordings of all of the component two-note intervals of these chords (see Figure 2.13 for an illustration of the three component intervals in a three-note chord). In a similar fashion to the presentation of two note intervals in Experiment 1, the notes in each stimulus were played first sequentially and then simultaneously, using a multi-sampled piano timbre with a duration of one second per note. Visual stimuli were the same set of color pairs used in Experiment 1.

Design. Each participant listened to the chords and component intervals presented as a single block in a random order. At the same time that each stimulus was presented, the participants viewed a display of two possible color-pair matches. For each chord or interval they were asked to choose which of the two color pairs they felt went better with the audio stimulus. Subsequent to the color choice task, participants were asked to rate all of the visual and audio stimuli on five semantic/emotional dimensions using a 400-pixel line rating scale, blocked by rating dimension and stimulus modality. The five dimensions were *happy/sad*, *calm/angry*, *active/passive*, *strong/weak*, and *harmonious/disharmonious*.

Results and Discussion

The average rating scores for the chord stimuli can be found in Table 2.10. We also ran a series of ANOVA tests to determine whether the chords differed significantly on their rating scores. As can be seen in that table, the chords differed significantly on only three out of the five emotion dimensions used (*happy/sad*, *calm/angry*, and *harmonious/disharmonious*).

We decided to explore the dimensional structure of the rating data (as in Experiment 1), to see whether the relationships between dimensions were the same when using chord stimuli. The correlation matrix for emotion ratings given to three-note chord stimuli can be seen in Table 2.11. Much as before, the dimensions were found to be highly correlated, and an exploratory factor analysis using maximum likelihood estimation again arrived at a two dimensional solution for the data, with the two dimensions weighting heavily on *harmonious-calm-happy/disharmonious-agitated-sad* and *strong-active/weak-passive* (see Table 2.12). These factor loadings are similar to those reported in Table 2.4 from Experiment 1 ($r = .75$, $t(8) = 3.21$, $p < .01$).

Using the same formula as in Experiment 1, we calculated SCA scores for each chord based on the ratings of colors that were chosen as going well with that chord. We found that the average emotional rating for a particular chord was highly predictive of the weighted average of the emotional ratings of colors chosen to go with that chord for three of the five dimensions tested. The SCA correlations for *harmonious/disharmonious* (r=.90, p<.001), *happy/sad* (r=.86, p<.001), and *calm/angry* (r=.81, p<.001) were all significant, whereas the correlations for the dimensions *active/passive* (r=.29, p=.20) and *strong/weak* (r=.05, p=.83) were again substantially lower. Colors chosen to go with more harmonious chords were on average greener (r=.68, p<.001), bluer (r=.62, p<.01), lighter (r=.74, p<.001), and more saturated (r=.84, p<.001) than the colors chosen to go with less harmonious chords. These chord-to-color associations were thus similar to the interval-to-color associations reported in Experiment 1 (and replicated in the present experiment).

We were also interested in investigating the extent to which participants' emotional ratings of a given chord could be predicted from the emotional ratings of the component intervals of that chord. For each emotional dimension, we ran a forward stepwise regression with the ratings of the three component intervals on that dimension as predictors. These models were able to predict a significant portion of the variance in all five emotion rating dimensions, although with a much higher total for *happy/sad* (88%), *harmonious/disharmonious* (93%), and *agitated/calm* (91%) than for *active/passive* (33%), and *strong/weak* (39%) (see Figure 2.14 for a comparative illustration). This pattern of results mirrors those of the SCA correlations just reported and those obtained in Experiment 1, supporting the interpretation that *happy/sad*, *agitated/calm*, and *harmonious/disharmonious* are more consistent and salient dimensions for making sound-to-color matches for both chords and intervals than are *active/passive* and *strong/weak* .

It is worth noting that the single best predictor of chord harmony ratings was the outermost interval in the chord (i.e., the lowest and the highest notes, indicated as 1-3 in Figure 2.13), which predicted over 50% of variance in harmony ratings. This may seem surprising given the importance of the inner intervals in music based around major and minor triads, which have the same outer interval but distinctly different emotional connotations. However, this result is, in fact, consistent with theories of chord perception such as Parncutt's (1988), which predict that the outer notes will be more easily perceived than inner notes in a chord due to reduced effects of masking from nearby frequencies for the outer notes. The inner notes are perceptually masked to a greater degree than the outer notes because the middle note, which they both contain, is close to the fundamental frequencies of the notes both higher and lower in the chord.

The results of these experiments provide converging evidence that emotional mediation is relevant in color choices for basic units of musical harmony including both intervals and chords, particularly for the emotional/semantic dimensions of *happy/sad*, *angry/calm*, and *harmonious/disharmonious*. *Active/passive* and *strong/weak* appear to be less relevant to emotional mediation of sound-to-color matches for intervals than they are for full-scale classical music. Furthermore, the relationship between lower-level two-note units of harmony and more complex three-note harmonies seems to be linear and predictable, suggesting that an interval-based model of harmony might explain much of the emotional content (and corresponding color choices) for even more complex harmonic relationships.

Experiment 5: Chords and Colors with the Ascending Minor Scale

In Experiment 4, we generated chord stimuli using all possible combinations of notes in a major scale. This creates a wide variety of different chord types, some very harmonious and some very dissonant, but it is lacking in a few distinct chord shapes that are particularly important in classical Western music, such as chords with a minor third in the lower notes. The present experiment was designed to redress this lacuna. We also wanted to generalize the findings of Experiment 4 by using the single-color choice paradigm from Experiment 2 with chord stimuli, so we chose to address both of these issues in a single experiment.

Methods

Participants. The participants were 26 volunteers from the Berkeley Research Participant Pool. None had color-vision deficiency as tested using the Dvorine Pseudo-Isochromatic Plates. All gave informed consent and were naïve to the purpose of the study. The Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

Stimuli. For the audio stimuli, we used the union of the sets of notes from a C-major scale (C-D-E-F-G-A-B-C) and an ascending C-minor scale (C-D-E♭-F-G-A-B-C). Effectively, the only difference between these two is the substitution of a flatted third above the root note (E♭) in the ascending minor scale. From this set of notes we generated recordings of 28 three-note chords consisting of middle-C and all possible combinations of two additional notes, plus the 36 two-note intervals that make up the components of those chords. In any given stimulus the notes of the chord were played sequentially as an arpeggio and then simultaneously, as in Experiments 1, 2, and 4 using a multi-sampled piano timbre, with a duration of one second per note. The visual stimulus was an array of the 37 BCP colors, as described in Experiment 3.

Design. Participants listened to the entire set of chords and intervals in a random order. For each chord or interval they were asked first to choose the three colors that went best with the interval (in order) from the set of 37. The chord or interval was then repeated, and they were asked to choose the three colors that went worst with the interval (in order).

Subsequent to the color choice task, participants were asked to rate all of the colors and audio stimuli on five semantic dimensions using a 400-pixel line rating scale, blocked by rating dimension and stimulus category. The five dimensions were the same as in the previous experiment (*happy/sad*, *calm/angry*, *active/passive*, *strong/weak*, and *harmonious/disharmonious*).

Results and Discussion

The addition of a flatted third to the set of intervals in this experiment allows us to directly measure and assess the effect of minor vs. major thirds in chords. To examine this, we also conducted a two-way within subjects ANOVA, with the two factors being (1) the presence of a major/minor third in the lower part of the chord (two levels) and (2) the outer interval in the chord (five levels). The results showed significant main effects and interactions for both factors for the dimensions of *happy/sad* (factor 1 F(4,100)=12.182, p<.001; factor 2 F(1,25)=109.5, p<.001; interaction F(4,100)=16.07, p<.001) as well as *calm/angry* (factor 1 F(4,100)=17.40, p<.001; factor 2 F(1,25)=43.40, p<.001; interaction F(4,100)=29.52, p<.001) and *harmonious/disharmonious* (factor 1 F(4,100)=21.10, p<.001; factor 2 F(1,25)=32.86, p<.001; interaction F(4,100)=34.30, p<.001), but no significant effects for either *active/passive* (factor 1

F(4,100)=1.56, p=.19; factor 2 F(1,25)=.25, p=.62; interaction F(4,100)=1.48, p=.22) or *strong/weak* (factor 1 F(4,100)=1.05, p=.11; factor 2 F(1,25)=.38, p=.54; interaction F(4,100)=2.04, p=.10). Comparisons of the mean scores for these and the other chords included in this experiment are given Table 2.13**.**

We also calculated Sound-Color Association (SCA) scores for these data, using the method described above in Experiment 2, to get a measure of the average emotion of colors chosen as going with the chords. We then compared the SCA scores to the emotional/semantic ratings given by participants for each of the chords. The SCA correlations for *harmonious/disharmonious* (*r*=.86, *p*<.001), *happy/sad* (*r*=.85, *p*<.001), and *calm/angry* (*r*=.65, *p*<.001) were all significant, whereas the correlations for the dimensions *active/passive* (*r*=-.08, *p*=.71) and *strong/weak* (*r*=.19, *p*=.37) were again substantially lower and not significant. Colors chosen to go with more harmonious chords were on average significantly bluer (*r*=.68, *p*<.001) lighter (*r*=.86, *p*=.001) and more saturated (*r*=.48, *p*=<.01) than the colors chosen to go with less harmonious chords, but not significantly greener or redder (*r*=.11, *p*=.57). This pattern also conforms to the results of Experiment 4.

We also attempted to predict the emotional qualities of a chord from ratings of its component intervals using stepwise linear regression, as in Experiment 4. We found that we were still able to predict a significant percentage of variance for all dimensions, and the models fit best for the dimensions of *harmonious/disharmonious* (89%), *happy/sad* (72%), and *calm/angry* (79%) (see Figure 2.15 for more details).

General Discussion

The series of experiments described in this section provides strong evidence that the emotional mediation hypothesis for audio-visual matching, originally found using recordings of orchestral classical music (Palmer at al., 2013), also extends to the more basic units of musical harmony: two-note intervals and three-note chords. For the two-note intervals, the results generalize both to their simultaneous and sequential forms. Unlike the results of the previous experiments using classical orchestral music, not every type of emotional/semantic distinction we tested was relevant for these basic units of harmony. In particular, *happy/sad*, *harmonious/disharmonious*, and *calm/angry* ratings for the intervals and chords were very highly correlated with the corresponding ratings for the colors chosen as going best with the intervals and chords (see Figure 2.16 for a visual summary of the results across experiments). There were generally weaker or mixed correlations for *strong/weak* and *active/passive*, but we would expect these dimensions to be carried mainly by other musical features such as rhythm, loudness dynamics, and/or timbre.

We also showed that this emotional mediation holds whether the color associates are presented as relational pairs of colors in a 2AFC paradigm or as individual colors in a free choice, best/worst paradigm. Further, we demonstrated the approximate linearity of predicting semantic judgments of harmony of three-note chords from their component two-note intervals. This last finding has far-reaching implications, as it indicates that the color and emotional associates of a complex piece of music (possibly even a full orchestral recording) could be reasonably well predicted from a harmonic analysis of the color and emotional associates of the component notes, intervals, and chords of the score. Given sheet music or MIDI notation for a piece, it may even be possible to automate the analysis and visualization of harmonic features in terms of color in a way that is consistent, evidence-based, and psychologically meaningful.

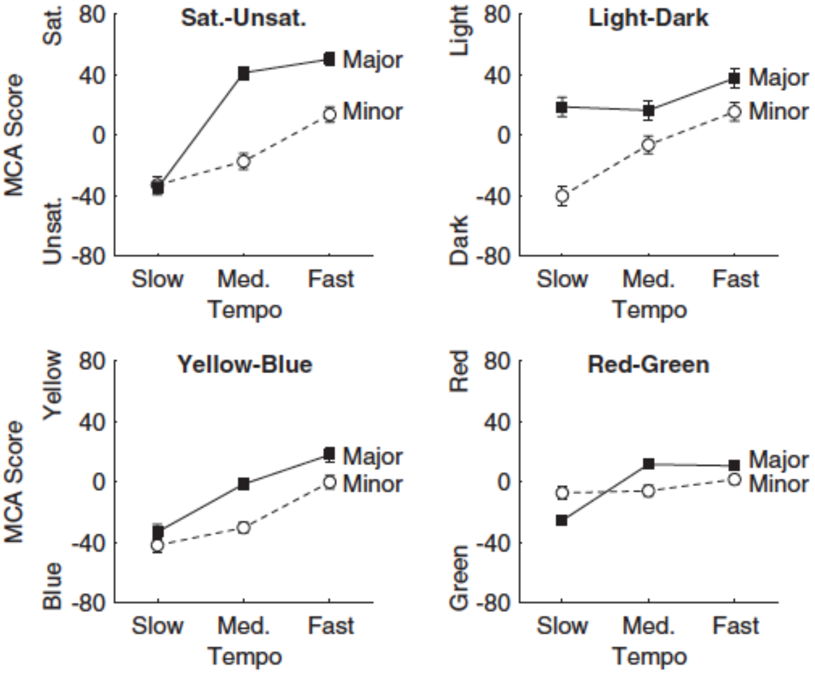

Figure 2.1: Data from Palmer et al. (2013) illustrating the effect of mode and tempo on participant color choices. The four graphs show effects on color saturation, color lightness, and two hue dimensions (yellow vs. blue and red vs. green).

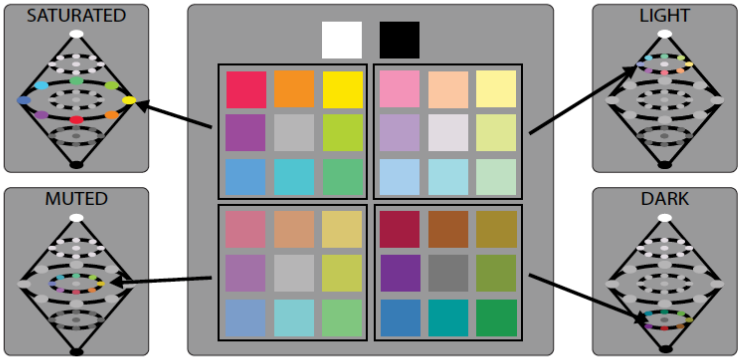

Figure 2.2: The 37 Berkeley Color Project colors, which consist of eight hues (red, yellow, green, blue, orange, chartreuse, cyan, purple) at four different lightness and saturation levels (saturated, light, muted, dark) plus five achromatic colors (black, white, and three levels of gray).

Colors	**x**	**y**	**Y**

SR	0.549	0.313	22.93
LR	0.407	0.326	49.95
MR	0.441	0.324	22.93
DR	0.506	0.311	7.60
SO	0.513	0.412	49.95
LO	0.399	0.366	68.56
MO	0.423	0.375	34.86
DO	0.481	0.388	10.76
SY	0.446	0.472	91.25
LY	0.391	0.413	91.25
MY	0.407	0.426	49.95
DY	0.437	0.450	18.43
SH	0.387	0.504	68.56
LH	0.357	0.420	79.90
MH	0.360	0.436	42.40
DH	0.369	0.473	18.43
SG	0.254	0.449	42.40
LG	0.288	0.381	63.90
MG	0.281	0.392	34.86
DG	0.261	0.419	12.34
SC	0.226	0.335	49.95
LC	0.267	0.330	68.56
MC	0.254	0.328	34.86
DC	0.233	0.324	13.92
SB	0.200	0.230	34.86
LB	0.255	0.278	59.25
MB	0.241	0.265	28.90
DB	0.212	0.236	10.76
SP	0.272	0.156	18.43
LP	0.290	0.242	49.95
MP	0.287	0.222	22.93
DP	0.280	0.181	7.60
A1	0.310	0.316	0.00
A2	0.310	0.316	12.34
A3	0.310	0.316	31.88
WH	0.310	0.316	63.90
BL	0.310	0.316	116.00

Table 2.1: Coordinates for each of the 37 BCP colors in CIE xyY color space.

Figure 2.3: Examples of color pairs chosen for the interval experiment because they represented extreme values on emotional and color appearance dimensions.

Dimension	Inner	Outer Color
Happy	Light Yellow	Saturated Yellow
Happy	Saturated Yellow	Saturated Blue
Angry	Saturated Red	Dark Gray
Angry	Saturated Red	Dark Red
Active	Saturated Red	Saturated Blue
Active	Saturated Yellow	Saturated Blue
Strong	Saturated Red	Dark Blue
Strong	Saturated Blue	Dark Red
Harmonious	Light Blue	Saturated Blue
Harmonious	Saturated Blue	Dark Blue
Red	Light Red	Saturated Red
Red	Dark Red	Saturated Red
Blue	Light Blue	Saturated Blue
Blue	Dark Blue	Saturated Blue
Light	Light Gray	Light Blue
Light	Light Yellow	Light Gray
Saturated	Saturated Yellow	Saturated Red
Saturated	Saturated Red	Saturated Blue
Sad	Dark Blue	Dark Gray
Sad	Dark Yellow	Dark Gray
Calm	Light Blue	Muted Blue
Calm	Dark Blue	Saturated Blue
Passive	Medium Gray	Light Gray
Passive	Medium Gray	Dark Gray
Weak	Light Red	Yellow Red
Weak	Light Yellow	Medium Gray
Disharmonious	Light Green	Dark Yellow
Disharmonious	Saturated Blue	Dark Yellow
Green	Light Green	Saturated Green
Green	Dark Green	Saturated Green
Yellow	Light Yellow	Saturated Yellow
Yellow	Dark Yellow	Saturated Yellow
Dark	Dark Gray	Dark Blue
Dark	Dark Yellow	Dark Gray
Desaturated	Dark Gray	Medium Gray
Desaturated	Light Gray	Medium Gray

Table 2.2: Two color pairs were chosen to represent extreme ends of each emotion and color appearance dimension.

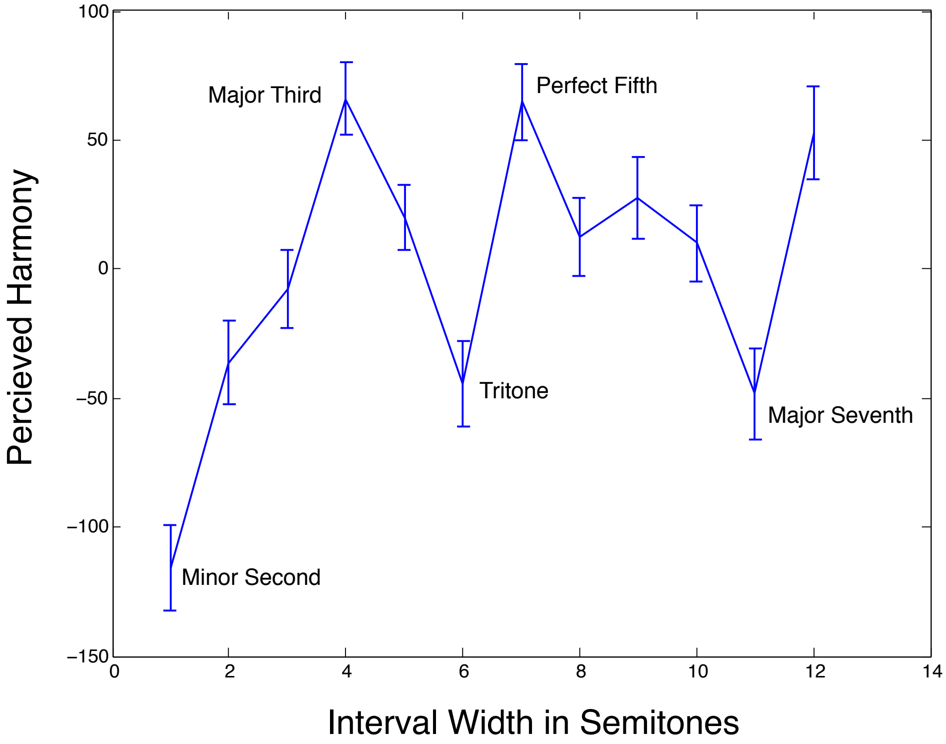

Figure 2.4: Average harmony rating for the set of interval stimuli in Experiment 1, averaged across octaves.

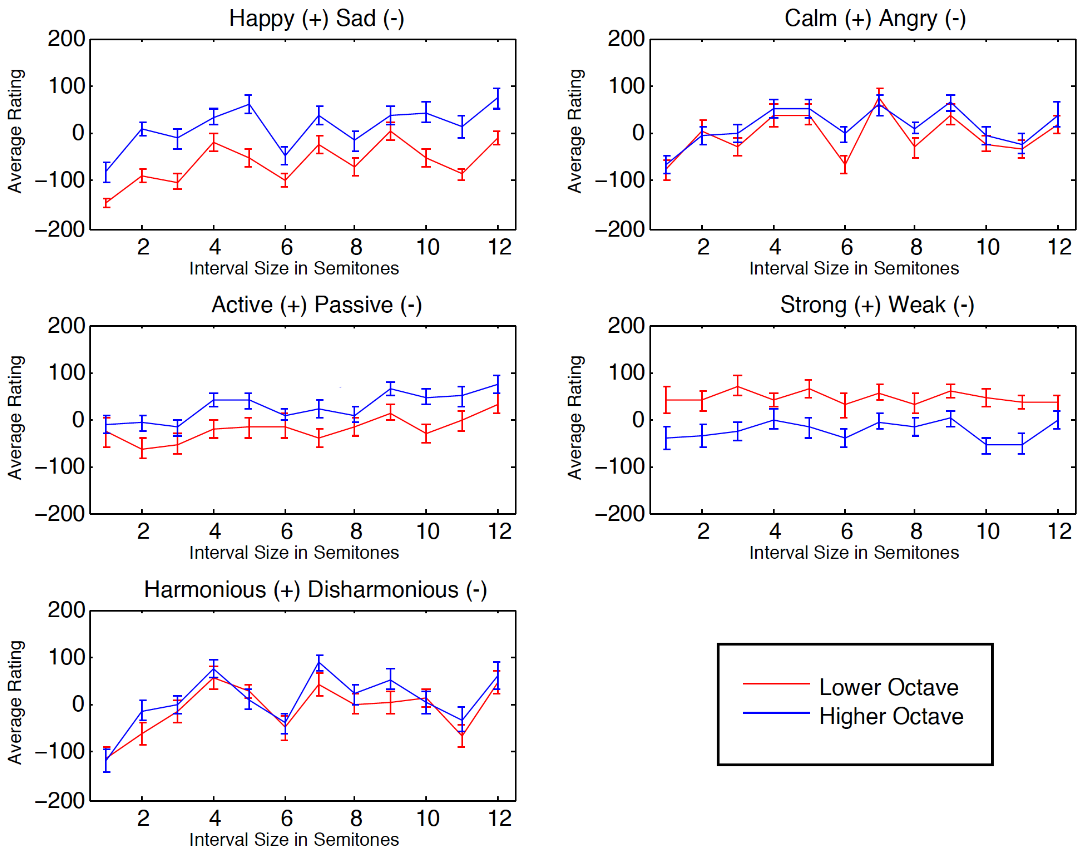

Figure 2.5: Average emotion ratings for the interval stimuli from Experiment 1, plotted with SEM as error bars.

	Happy (+) Sad (-)	Calm (+) Angry (-)	Active (+) Passive (-)	Strong (+) Weak (-)	Harmonious (+) Disharmonious (-)
Happy (+) vs. Sad (-)	1.00	0.73	0.79	-0.48	0.72
Calm (+) vs. Angry (-)	0.73	1.00	0.28	0.08	0.84
Active (+) vs. Passive (-)	0.79	0.28	1.00	-0.55	0.42
Strong (+) vs. Weak (-)	-0.48	0.08	-0.55	1.00	0.09
Harmonious (+) vs. Disharmonious (-)	0.72	0.84	0.42	0.09	1.00

Table 2.3: Correlations between emotion rating dimensions for interval stimuli in Experiment 1. Positive scores were coded for the first-mentioned pole of the bipolar scales and negative scores for the second-mentioned pole.

	Weightings	
Dimension	**Factor 1**	**Factor 2**
Happy vs. Sad	1.00	0.00
Calm vs. Angry	0.73	0.57
Active vs. Passive	0.79	-0.28
Strong vs. Weak	-0.48	0.75
Harmonious vs. Disharmonious	0.72	0.56

Table 2.4: Factor weightings for two-dimensional solution to factor analysis of emotion dimensions in Experiment 1. Positive scores were coded for the first-mentioned pole of the bipolar scales and negative scores for the second-mentioned pole.

Mean Emotion Ratings for Color Pairs

	Happy (+) Sad (-)	Calm (+) Angry (-)	Active (+) Passive (-)	Strong (+) Weak (-)	Harmonious (+) Disharmonious (-)
A1-DB	-97.14	-12.24	-125.76	-35.24	1.19
A2-A1	-160.81	32.9	-131.33	-105	56.57
A3-A1	-118.43	20.71	-120.81	-41.48	-3.38
A3-A2	-61.38	78.1	-107.81	-84.62	53.14
A3-LB	31.33	130.95	-6.24	-65.33	60.76
DB-A1	-133.95	11.1	-121.86	-53.33	-2.43
DB-SB	19.38	106.62	-7.95	11.57	61.95
DG-SG	42.1	95.19	71.9	65.48	66.43
DR-SR	2	-117.1	129.29	145.48	32
DY-A1	-141.48	9.1	-108.71	-67.05	-35.95
DY-SY	49.1	-13.86	103.14	63.05	7.14
LB-SB	78.33	158.43	26.43	3.29	107.67
LG-SG	110.33	119.24	74.76	46.14	106.14
LR-A2	6	94.38	-32.67	-81.48	10.71
LR-DY	-56.95	0.05	-38.24	-21.14	-63.14
LR-SR	71.57	-69.9	141.24	113.57	55.81
LY-A2	-0.1	65.52	-40.9	-24.29	26.52
LY-A3	18.62	88.48	-15.76	-28.29	56.71
LY-SY	163.81	18.67	170.19	101.71	85.19
SB-DB	-43.05	67.76	-50.67	25.76	76.43
SB-DR	-43.14	-80.67	31.67	79.43	-66.57
SB-DY	-58.1	3	-54.43	-18.05	-91.48
SR-A1	-71.76	-42.29	-12.9	11	-43.62
SR-DB	-21.19	-18.38	21.57	59.14	-76.29
SR-DR	-26.24	-120.76	42.81	82.95	56.71
SR-SB	65.19	18.76	76	71.86	-39.29
SY-SB	117.48	95.71	102.29	73.52	19.95
SY-SR	103.05	-70.81	175.9	140.62	17.14
ANOVA F(35,700)	F=29.90 p<.001	F=12.86 p<.001	F=32.46 p<.001	F=12.9 p<.001	F=6.2 p<.001

Table 2.5: Average emotion/semantic ratings for each of the color pairs used in Experiment 1. Each color pair is labeled using a four character code; the first two characters refer to the inner color and the last two colors refer to the outer color. Each chromatic color is described by its hue (R=red, O=orange, Y=yellow, H=chartreuse, G=green, C=cyan, B=blue, P=purple) and cut (L=light, M=muted, D=dark, S=saturated); see Figure 2.2 for a visual representation of the relationship between the various hues and cuts. Achromatic colors are indicated by the codes A1=dark gray, A2=middle gray, and A3=light gray.

Color Pair	**Mean Color Appearance Rating**			
	Saturated (+) Unsaturated (-)	Yellow (+) Blue (-)	Red (+) Green (-)	Light (+) Dark (-)
A1-DB	-46.05	-91.83	-8.93	-82.06
A2-A1	-149.08	-0.86	-0.69	99.73
A3-A1	-144.87	-4.17	0.52	40.24
A3-A2	-149.08	-0.86	-0.69	99.73
A3-LB	-87.63	-52.97	-8.77	138.07
DB-A1	-46.05	-91.83	-8.93	-82.06
DB-SB	82.40	-170.16	-20.44	-21.05
DG-SG	81.18	-11.75	-173.71	-15.57
DR-SR	141.26	-1.51	174.65	-45.54
DY-A1	-74.91	35.92	-8.68	-65.58
DY-SY	79.41	137.60	-12.86	19.74
LB-SB	38.15	-137.30	-18.73	86.10
LG-SG	39.02	-6.79	-126.52	88.46
LR-A2	-71.77	11.27	37.64	66.28
LR-DY	5.07	54.04	28.12	13.95
LR-SR	95.66	16.20	121.03	52.26
LY-A2	-74.92	70.79	-3.83	93.38
LY-A3	-68.03	73.48	-4.16	146.88
LY-SY	83.61	169.17	-6.80	119.21
SB-DB	82.40	-170.16	-20.44	-21.05
SB-DR	106.49	-88.07	80.96	-30.00
SB-DY	53.54	-42.41	-20.18	-4.56
SR-A1	15.67	-1.92	84.23	-27.27
SR-DB	117.17	-83.59	73.25	-36.59
SR-DR	141.26	-1.51	174.65	-45.54
SR-SB	144.13	-80.25	72.73	33.74
SY-SB	135.23	13.19	-13.63	73.59
SY-SR	169.99	99.76	80.05	58.05

Table 2.6: Average color appearance ratings for color pairs used in Experiment 1, based on data collected by Schloss and Palmer (2010). See Table 2.5 for an explanation of the Color Pair coding scheme.

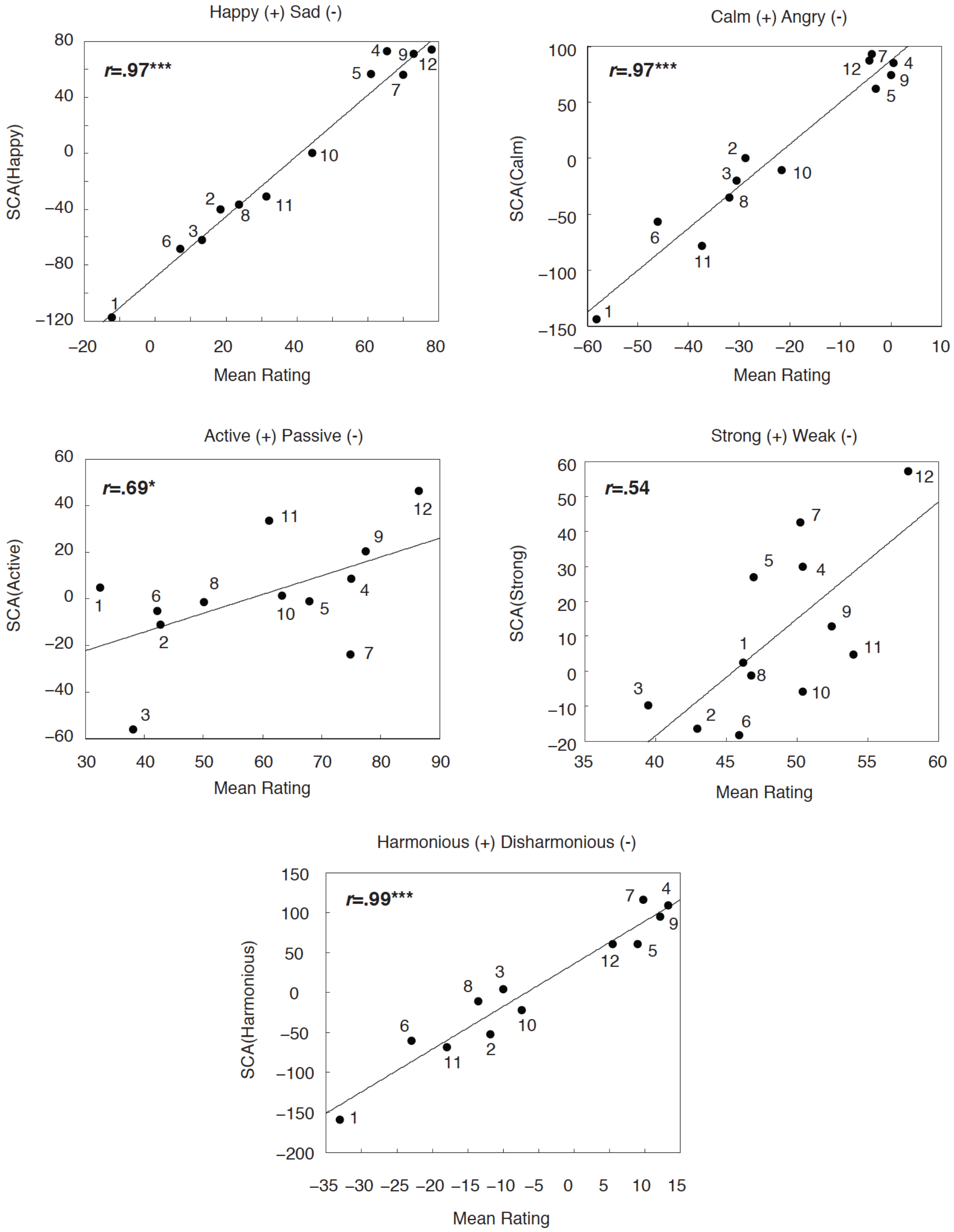

Figure 2.6: Scatterplots illustrating the correlations between SCA scores and mean participant ratings for the five semantic/emotion dimensions. Pearson's *r* is noted in the top left corner of each plot (**p*<.05; *** *p*<.001).

Figure 2.7: Display used for single color choice tasks. Participants were asked to choose which three colors "went best" and "went worst" with each sound from this set of 37 colors.

Emotion Ratings for Individual Colors

	Happy (+) Sad (-)	Calm (+) Angry (-)	Active (+) Passive (-)	Strong (+) Weak (-)	Harmonious (+) Disharmonious (-)
SR	60.47	48.88	78.06	97.71	-10.00
LR	53.76	-7.29	26.41	14.06	49.47
MR	35.18	-7.47	7.29	14.82	-4.71
DR	-10.71	75.82	23.76	8.35	-32.12
SO	80.59	22.35	86.47	97.94	34.18
LO	47.06	-32.76	15.88	27.71	57.94
MO	0.65	-32.47	-4.35	3.18	4.65
DO	-64.29	44.82	4.76	-35.12	-65.29
SY	105.59	-1.24	73.71	109.29	38.94
LY	66.71	-32.53	14.65	40.71	55.53
MY	39.53	-15.35	3.29	32.29	0.94
DY	-37.59	17.76	-10.29	-36.35	-52.29
SH	69.12	7.59	48.06	90.71	5.29
LH	44.71	-40.82	8.18	52.24	44.06
MH	-13.71	-39.76	9.35	9.47	-4.94
DH	-40.00	14.94	-11.76	-17.00	-62.00
SG	52.82	-25.24	46.06	57.94	35.94
LG	38.82	-30.59	2.12	17.29	54.65
MG	-14.47	-47.18	-7.53	0.59	18.65
DG	-21.82	-16.41	-1.12	2.47	-10.06
SC	60.82	-16.71	57.47	63.12	26.12
LC	30.53	-40.29	-5.00	37.24	82.65
MC	20.12	-55.76	-16.00	-25.76	7.24
DC	-18.06	-16.12	-8.53	-21.00	2.53
SB	18.88	-16.65	40.41	35.88	29.88
LB	28.71	-54.53	-19.47	-3.65	76.29
MB	-14.71	-52.18	-15.65	-16.47	11.18
DB	-34.65	-30.71	0.47	-39.35	3.00
SP	27.18	0.06	28.71	51.47	-8.06
LP	46.53	-64.29	10.82	25.59	54.06
MP	3.76	-17.88	-2.59	-14.94	-25.18
DP	-22.35	3.94	26.18	-18.18	-21.06
BK	-81.35	75.18	11.71	-58.82	-103.53
A1	-90.82	4.29	-23.59	-68.65	-65.82
A2	-89.29	-38.88	-41.47	-77.82	-40.82
A3	-2.53	-63.06	-1.29	-22.94	39.59
WH	58.24	-23.29	11.82	18.35	80.18
ANOVA F(36,592)	F=9.25 p<.001	F=3.92 p<.001	F=1.99 p<.001	F=5.11 p<.001	F=5.53 p<.001

Table 2.7: The average emotion ratings given by participants for individual colors in Experiment 2.

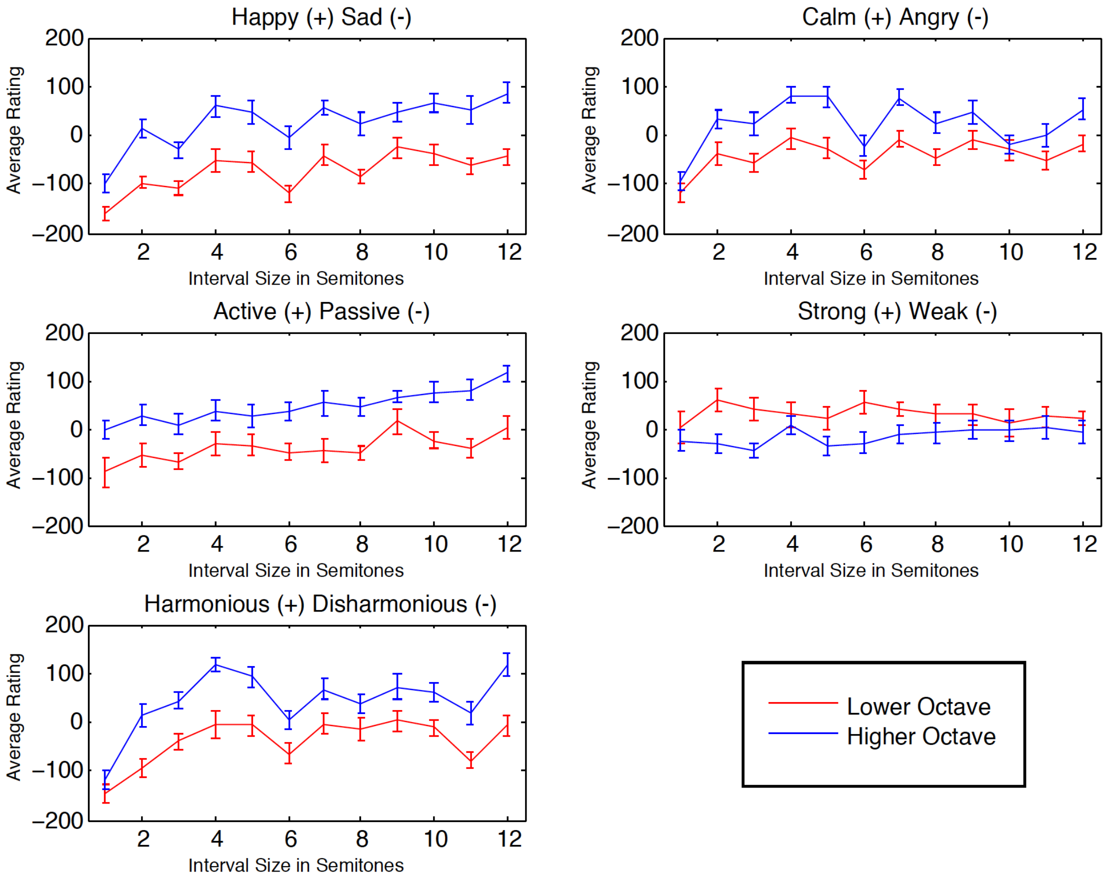

Figure 2.8: Average emotional/semantic ratings for the interval stimuli from Experiment 2.

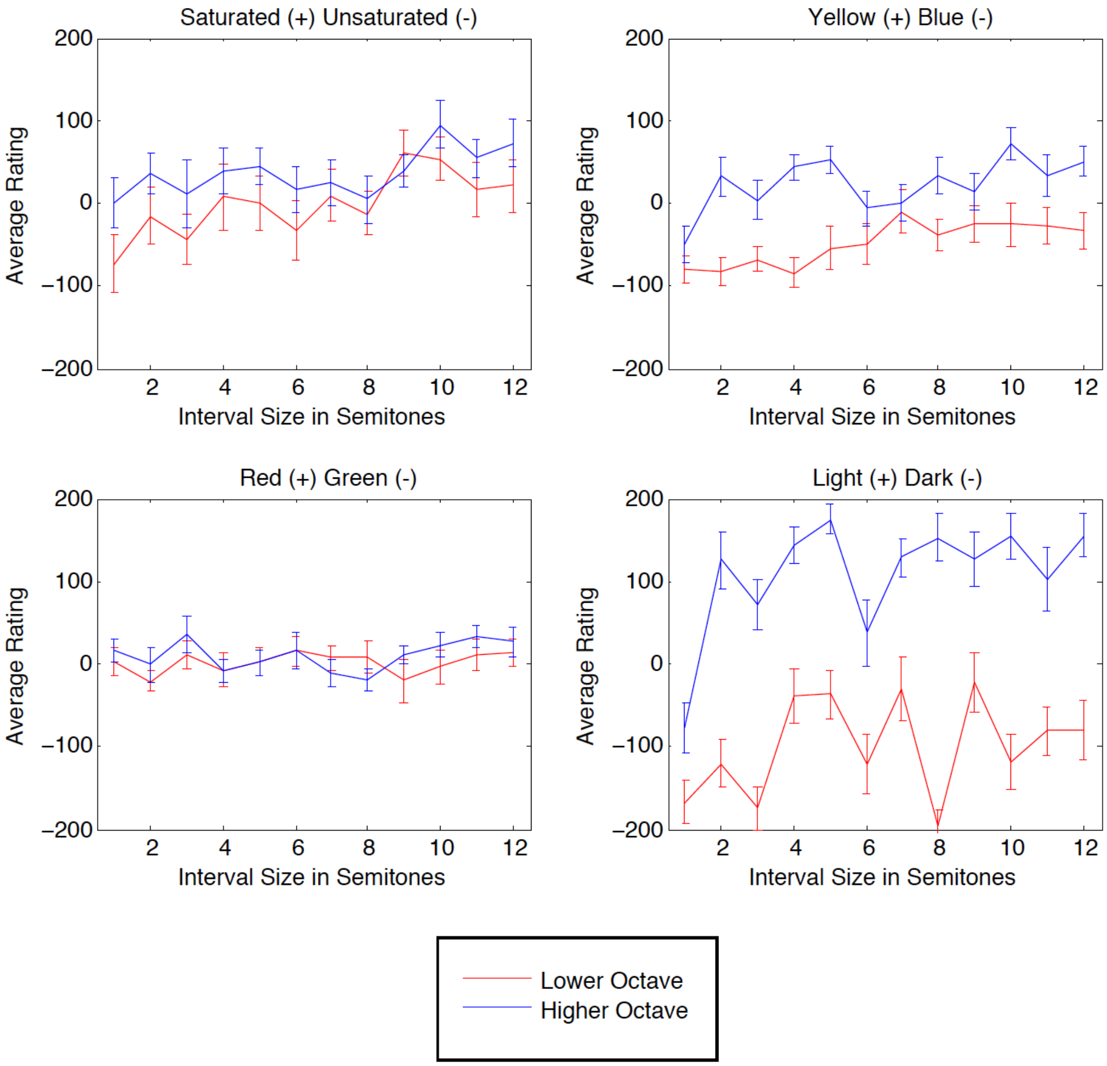

Figure 2.9: Average SCA ratings (based on color appearance ratings) for the six colors chosen as going best/worst with the interval stimuli in Experiment 2.

Color Appearance Ratings for Individual Colors

	Saturated/Unsaturated	Yellow/Blue	Red/Green	Light/Dark
SR	178.89	6.32	166.41	18.2
LR	12.43	26.07	75.65	86.32
MR	7.32	12.84	98.29	-12.2
DR	103.62	-9.33	182.88	-109.28
SO	153.91	96.14	46.06	39.94
LO	-25.32	52.28	35.37	124.2
MO	-18.6	48.61	45.9	44
DO	45.34	29.84	56.27	-108.55
SY	161.09	193.2	-6.31	97.89
LY	6.13	145.13	-7.3	140.53
MY	-4.45	136.72	-8.86	64.56
DY	-2.28	82	-19.4	-58.42
SH	135.26	113.36	-103.22	96.35
LH	-27.04	112.09	-50.48	123.13
MH	-43.53	60.79	-96.83	57.07
DH	-14.64	18.02	-138.87	-28.5
SG	102.98	-0.89	-159.91	61.01
LG	-24.94	-12.68	-93.14	115.9
MG	-31.94	-2.43	-120.82	36.02
DG	59.38	-22.61	-187.51	-92.16
SC	103.13	-115.48	-53.98	69.14
LC	-41.15	-84.76	-33.77	125.33
MC	-33.77	-100.72	-49.7	38.94
DC	35.43	-88.83	-94.31	-74.39
SB	109.36	-166.82	-20.96	49.29
LB	-33.06	-107.77	-16.51	122.91
MB	-41.34	-144.26	-20.95	27.49
DB	55.45	-173.49	-19.92	-91.38
SP	112.06	-51.56	53.52	-8.22
LP	-15.17	-39.52	36.84	121.84
MP	6.06	-54.44	39.43	19.88
DP	84.45	-62.7	53.04	-130.84
BK	-90.77	-19.05	0.67	-199.38
A1	-147.55	-10.16	2.05	-72.75
A2	-155.96	-3.54	-0.36	46.23
A3	-142.19	1.83	-1.02	153.22
WH	-125.85	4.1	-0.51	191.16

Table 2.8: Color dimensions for individual color stimuli used in Experiment 2, based on data for Schloss and Palmer (2010).

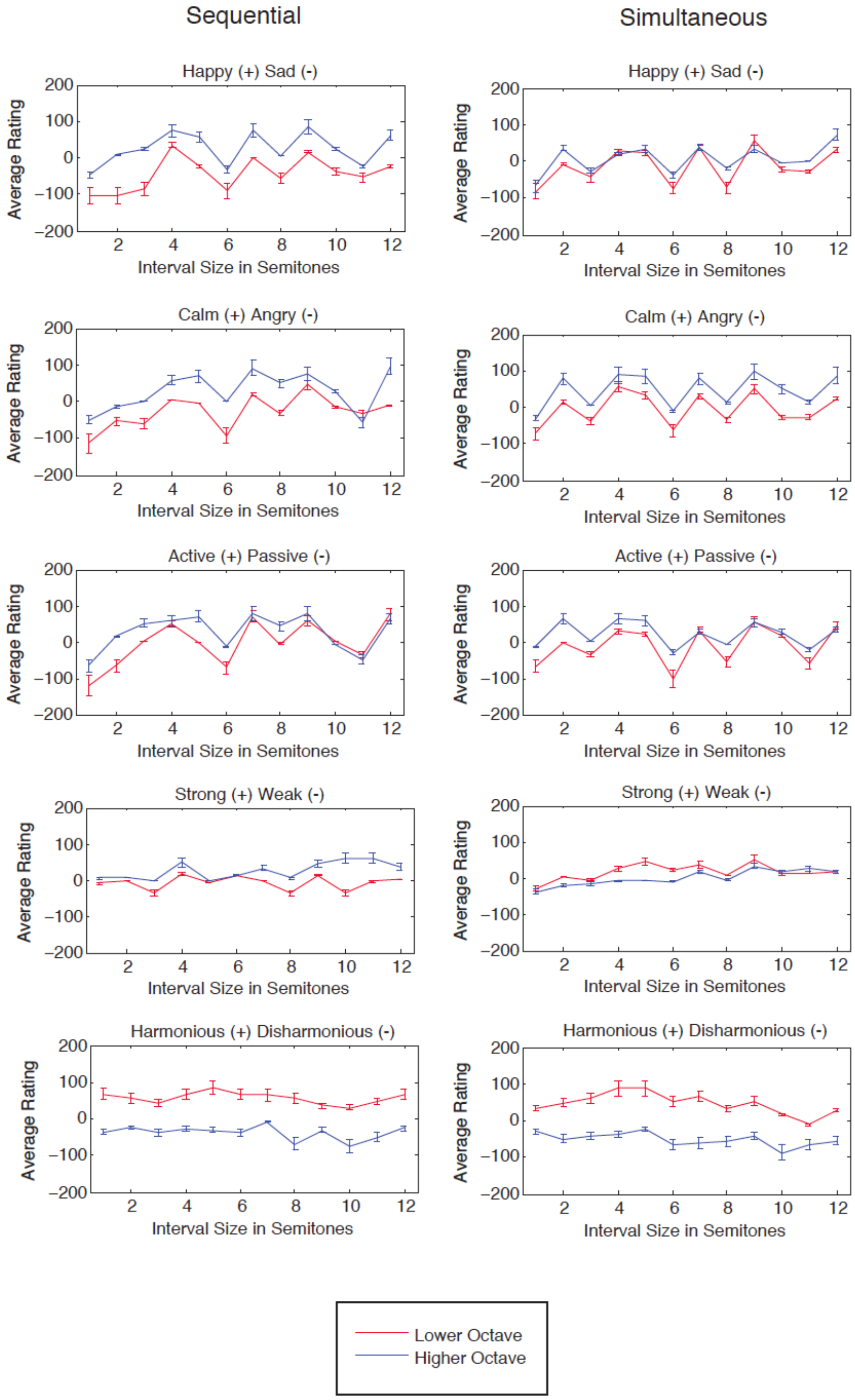

Figure 2.10: Average emotion ratings for interval stimuli presented sequentially and simultaneously in Experiment 3.

ANOVA Results for Emotion Dimensions of Simultaneous and Sequential Intervals						
F(11,198)	**Happy (+) Sad (-)**	**Angry (+) Not Angry (-)**	**Calm (+) Agitated (-)**	**Active (+) Passive (-)**	**Strong (+) Weak (-)**	**Harmonious (+) Disharmonious (-)**
Sequential	F=5.00 p<.001	F=7.13 p<.001	F=4.73 p<.001	F=1.15 p=.29	F=7.99 p<.001	F=7.06 p<.001
Simultaneous	F=8.93 p<.001	F=7.04 p<.001	F=7.89 p<.001	F=1.35 p=.13	F=6.72 p<.001	F=12.27 p<.001

Table 2.9: Statistical tests of interval effects on emotional/semantic ratings for intervals presented sequentially and simultaneously in Experiment 3.

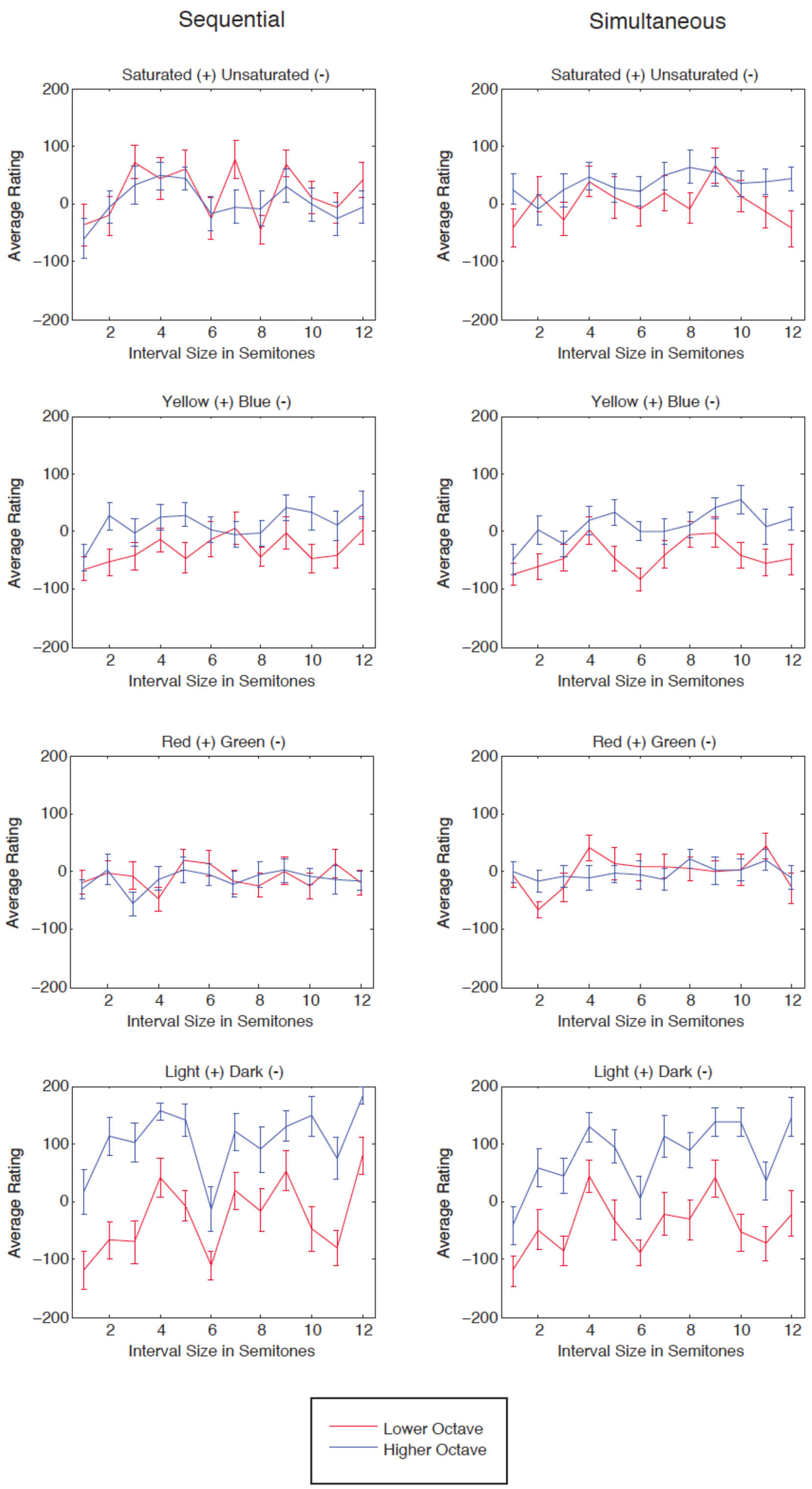

Figure 2.11: Average SCA scores for color appearance ratings of the colors chosen as going best/worst with intervals presented simultaneously and sequentially in Experiment 3.

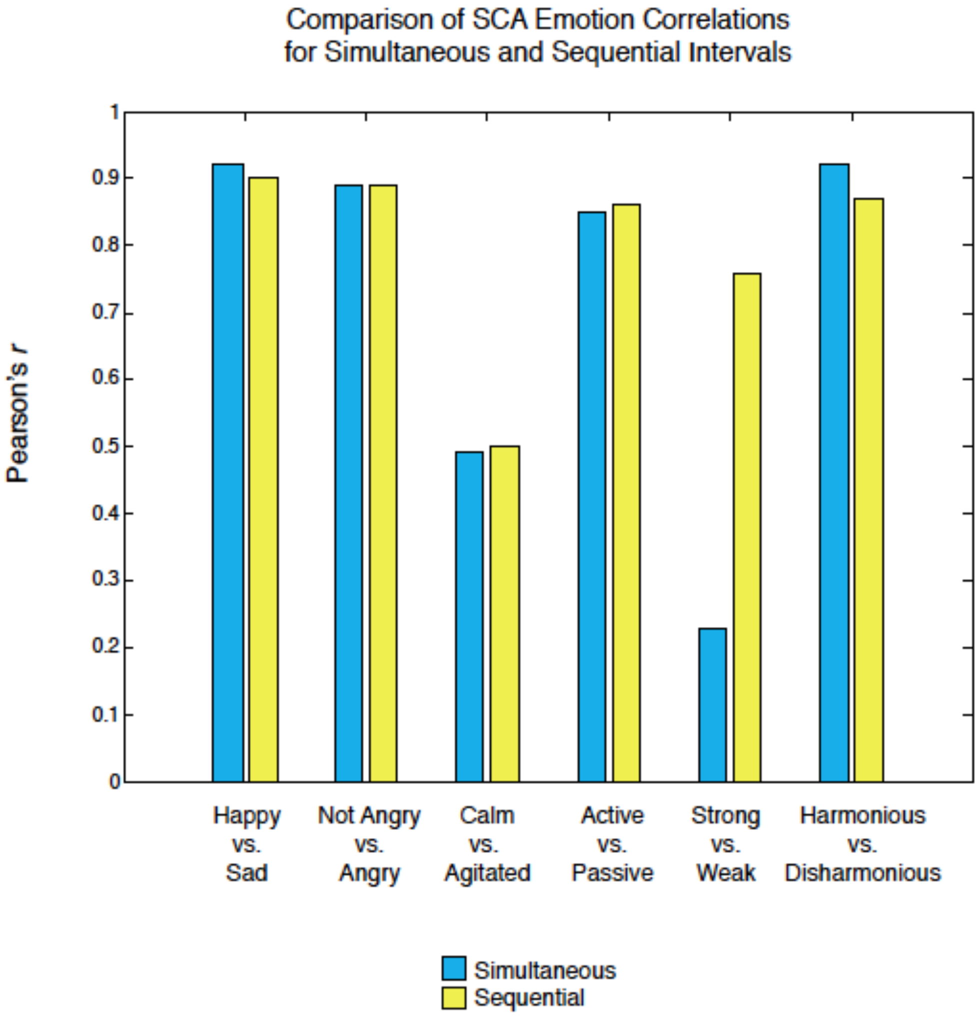

Figure 2.12: Correlations between emotional/semantic ratings and SCA scores for simultaneous and sequential presentations of the interval stimuli.

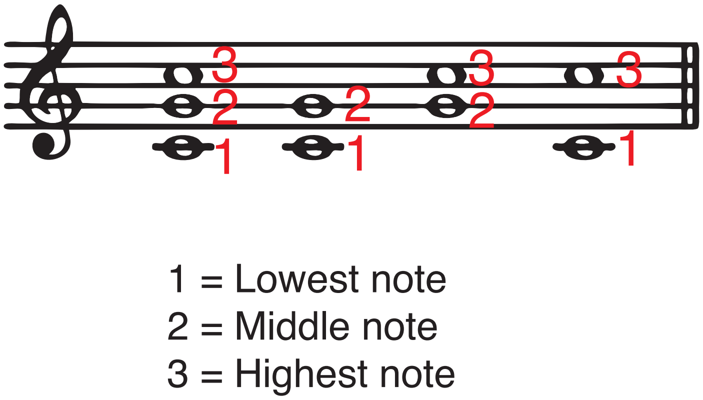

Figure 2.13: Illustration of the three component two-note intervals in a three-note chord.

Mean Emotion Ratings for Chords

Chord	Happy (+) Sad (-)	Calm (+) Angry (-)	Active (+) Passive (-)	Strong (+) Weak (-)	Harmonious (+) Disharmonious (-)
CDE	-53.43	-106.07	3.07	33.57	-57.86
CDF	-62.43	-91.86	4.64	14.79	-39.36
CDG	-79.14	-59.43	-29.64	51.07	-9.93
CDA	-50.36	-46.00	-29.86	39.71	-16.93
CDB	-73.79	-52.14	-6.07	20.07	-31.50
CDC	5.00	-26.29	43.21	48.21	-19.43
CEF	-75.86	-118.00	-40.79	17.21	-101.43
CEG	-11.64	25.86	33.50	46.14	25.86
CEA	-38.43	-46.50	-15.79	15.86	-31.43
CEB	-42.79	-3.86	9.71	-15.93	6.57
CEC	43.29	53.29	38.57	32.14	87.50
CFG	-67.00	-77.21	-5.36	12.36	-59.79
CFA	46.71	-18.93	20.29	62.14	44.07
CFB	-66.07	-70.43	-23.86	45.00	-59.29
CFC	14.64	15.50	42.71	65.86	63.64
CGA	-65.79	-37.71	18.07	33.86	-36.64
CGB	-14.86	-1.93	-1.14	43.43	5.64
CGC	50.00	18.93	20.50	55.07	67.86
CAB	-47.93	-48.50	-11.43	9.79	-51.36
CAC	54.43	36.21	57.93	58.36	37.21
CBC	-11.07	-15.14	18.21	-0.50	-22.93
ANOVA F(20,380)	F=4.2 p<.001	F=3.97 p<.001	F=1 p=.47	F=0.65 p=.87	F=3.83 p<.001

Table 2.10: Average emotional/semantic ratings for all of the chord stimuli. Each chord is labeled according to its three component notes.

	Happy vs. Sad	Calm vs. Angry	Active vs. Passive	Strong vs. Weak	Harmonious vs. Disharmonious
Happy vs. Sad	1.00	0.92	0.43	0.11	0.90
Calm vs. Angry	0.92	1.00	0.57	0.02	0.94
Active vs. Passive	0.43	0.57	1.00	-0.35	0.57
Strong vs. Weak	0.11	-0.02	-0.35	1.00	0.01
Harmonious vs. Disharmonious	0.90	0.94	0.57	0.01	1.00

Table 2.11: Correlations between the five emotional/semantic rating dimensions for the audio stimuli used in Experiment 4.

	Weightings	
Dimension	**Factor 1**	**Factor 2**
Happy vs. Sad	0.95	0.18
Calm vs. Angry	0.975	-0.03
Active vs. Passive	0.57	-0.62
Strong vs. Weak	0.01	0.57
Harmonious vs. Disharmonious	0.96	-0.03

Table 2.12: Factor weightings for the two-dimensional solution to a factor analysis of the correlation matrix (see Table 2.11) from the data in Experiment 4.

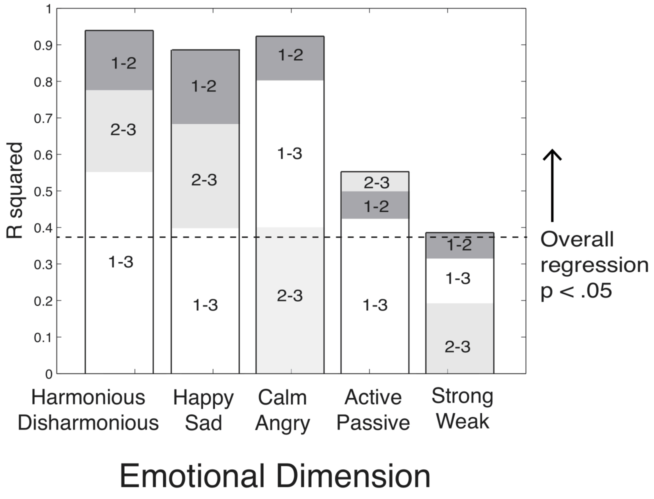

Figure 2.14: Results of regression modeling for five emotion dimensions using ratings of component intervals as predictors. Bar height represents total *r*-squared, with subcomponents indicating partial *r*-squared accounted for by each interval predictor. The label 1-3 indicates variance predicted by the ratings of the interval between the highest and lowest notes in a chord, 1-2 indicates variance accounted for by the lower two notes, and 3-2 indicates variance accounted for by the higher two notes.

	Mean Emotion Ratings of Chords Derived from Major and Minor Scales				
Chord	**Happy (+) Sad (-)**	**Calm (+) Angry (-)**	**Active (+) Passive (-)**	**Strong (+) Weak (-)**	**Harmonious (+) Disharmonious (-)**
C D E♭	-151.00	-157.65	44.65	7.35	-145.35
C D E	-44.65	-44.71	46.59	39.24	-45.35
C D F	-5.00	7.71	-13.12	-35.41	50.12
C D G	9.41	24.82	-23.18	-21.18	7.12
C D A	-46.65	0.47	5.00	7.35	-33.53
C D B	-105.06	-46.76	44.12	24.53	-60.12
C D C	-71.12	-66.59	11.12	33.71	-36.41
C E♭ E	-114.24	-126.47	25.29	-25.35	-166.76
C E♭ F	-110.06	-57.71	7.65	-54.00	-59.35
C E♭ G	-111.00	-99.18	52.12	29.00	-78.53
C E♭ A	-74.24	-83.59	-11.94	-47.06	-80.71
C E♭ B	-127.00	-122.71	23.18	-4.24	-130.71
C E♭ C	-117.29	-99.53	4.24	-3.53	-101.59
C E F	-124.24	-119.76	24.76	-14.12	-102.12
C E G	97.71	107.12	12.06	-1.00	100.59
C E A	-35.35	8.35	-19.24	-0.41	-26.82
C E B	-41.29	-75.29	28.76	-24.41	-80.12
C E C	77.00	99.76	9.29	68.76	134.24
C F G	-20.18	17.53	-17.35	-41.53	40.12
C F A	67.24	107.82	-53.41	-12.29	124.00
C F B	-64.94	11.53	-40.29	-16.76	-59.41
C F C	32.29	21.53	-32.18	-7.29	65.59
C G A	-36.41	-1.82	-28.94	-20.53	4.06
C G B	-7.76	-5.06	7.59	-30.41	-7.18
C G C	46.59	12.12	33.12	60.88	62.76
C A B	-59.06	-48.24	-12.65	15.71	-59.88
C A C	49.12	40.88	55.88	66.41	32.53
C B C	-1.76	-15.53	13.53	9.18	-25.53
ANOVA F(27,432)	F=17.60 p<.001	F=16.26 p<.001	F=1.74 p<.05	F=2.18 p<.01	F=16.42 p<.001

Table 2.13: Average emotion ratings of chord derived from both the C major and ascending minor scales in Experiment 5.

Mean Color Dimensions for Chords Derived from Major and Minor Scales

Chord	Saturated (+) Unsaturated (-)	Yellow (+) Blue (-)	Red (+) Green (-)	Light (+) Dark (-)
C D E♭	-16.37	-32.67	-15.02	-18.85
C D E	-35.61	-49.54	-40.41	-48.83
C D F	42.09	12.76	16.69	86.44
C D G	-0.21	-0.58	-7.90	44.88
C D A	-6.92	-0.69	-31.81	107.56
C D B	45.56	4.53	-21.34	106.94
C D C	3.39	5.25	-16.33	21.88
C E♭ E	28.46	-2.02	-17.26	92.58
C E♭ F	-68.50	-39.64	-11.31	-68.13
C E♭ G	-13.72	-25.35	-20.46	-35.68
C E♭ A	-13.80	-25.71	-5.19	45.18
C E♭ B	35.17	2.48	-22.71	117.16
C E♭ C	-57.49	-14.82	-43.06	52.08
C E F	-16.62	33.97	-36.99	108.24
C E G	-6.75	-6.67	-23.43	61.62
C E A	-19.82	-36.65	-1.00	-98.62
C E B	-18.19	-36.30	-27.66	-58.67
C E C	-24.62	-29.32	-21.43	20.62
C F G	-18.05	-26.43	3.80	-13.71
C F A	-31.30	-31.77	-28.69	-19.74
C F B	104.85	40.53	52.54	78.62
C F C	-37.09	-54.29	-30.47	-96.30
C G A	23.75	45.10	16.42	134.00
C G B	-19.94	32.79	-3.85	103.52
C G C	-52.35	7.56	-12.24	66.64
C A B	65.72	46.86	0.57	167.87
C A C	84.59	41.91	3.80	147.28
C B C	33.57	59.53	-9.56	154.18
ANOVA F(27,432)	F=2.54 p<.001	F=2.37 p<.001	F=1.15 p=.27	F=9.47 p<.001

Table 2.14: Average color dimensions for the SCA scores of chords derived from both the C major and ascending minor scales in Experiment 5.

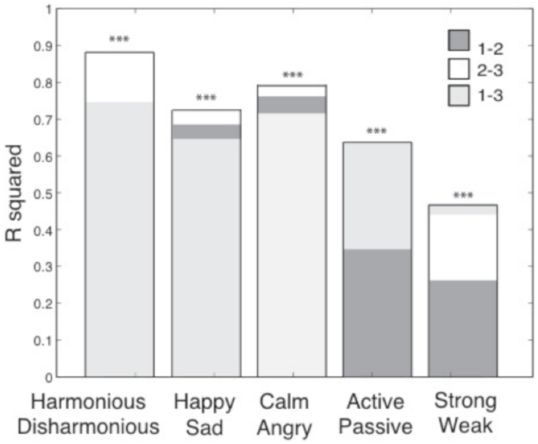

Figure 2.15: Results of a forward multiple regression analysis predicting ratings of chords from ratings of their component intervals using data from Experiment 5. Bar height represents total *r*-squared, with subcomponents indicating partial *r*-squared accounted for by each interval predictor. The label 1-3 indicates variance predicted by the ratings of the interval between the highest and lowest notes in a chord, 1-2 indicates variance accounted for by the lower two notes, and 3-2 indicates variance accounted for by the higher two notes.

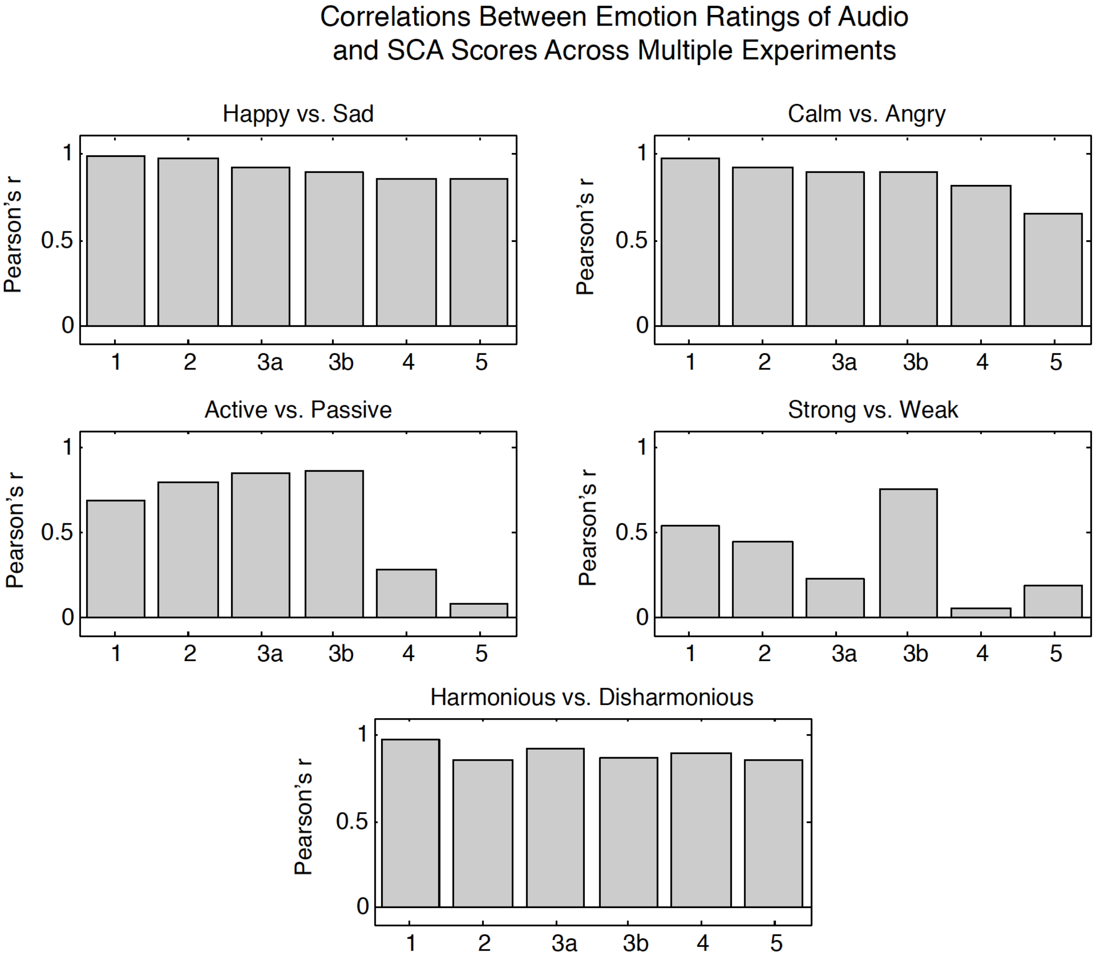

Figure 2.16: Visual comparison of the results of SCA correlations across the five experiments described in this chapter. Experiments 3a and 3b refer to the simultaneous and sequential conditions from Experiment 3, respectively. A high SCA correlation means that emotion ratings of the audio stimulus were highly predictive of the emotion ratings of colors chosen to go with that stimulus.

Chapter 3: Visual Associations with Instrumental Timbres

Introduction

Music consists of more than just harmonic and melodic relationships among pitches, and there are many other dimensions, such as overall pitch range, loudness, rhythm and timbre, that can be equally important for a full appreciation of musical structure and emotion. Of particular interest for this chapter is one of these other dimensions: namely, instrumental timbre.

A common definition of timbre is essentially a definition by negation: timbre is “the way in which musical sounds differ once they have been equated for pitch, loudness, and duration” **(**Krumhansl, 1989**)**. Timbre is also sometimes called 'tone color' (from the German *klangfarbe*, introduced by Helmholtz, 1885), and common English language descriptions of timbre often use color terms as a metaphor (e.g., a 'bright' tone vs. a 'dark' tone). However, there has been very little research prior to the present work that has systematically investigated cross-modal mappings between between timbre and color, or the extent to which various dimensions of timbres might map onto dimensions of color appearance**.**

In general, timbre is a poorly understood aspect of auditory perception. Because it is not defined by a set of features, but by what is left over after other features have been accounted for, timbre is an inherently multi-dimensional attribute of musical sound rather than a single thing (Grey, 1977). The lack of a clear definition or set of a priori features for timbre has led many researchers to use techniques such as multi-dimensional scaling (MDS) to attempt to analyze it **(**e.g., Elliot et al. 2013; McAdams et al., 1995; Iverson and Krumhansl, 1993; Wessel, 1979; Grey, 1977**)**. MDS is set of exploratory data analysis techniques based on ideas initially proposed by Torgerson (1952) and later expanded by Shepard (1962), that take as input similarity or dissimilarity judgments between all possible pairs of stimuli and output a mapping into a low dimensional space in which the magnitudes of the similarity judgments correspond to the distances between the stimuli in that space. Because the results of MDS are sensitive to the number and type of stimuli used in the initial similarity judgments, the number of different dimensional models for timbre is nearly as large as the number of studies conducted. That said, certain commonalities have arisen from the results of multiple studies. For example, the features of *attack time*, *spectral brightness*, and *spectral flux* seem to be important dimensions that explain much of the variation in human similarity judgments among different timbres (McAdams, 2006).

Attack time is a property of the loudness function of a waveform that measures how long it takes for the sound to reach peak volume after onset. All natural instruments have some amount of time between the onset of the sound and the loudness peak. For percussive instruments this is often very fast (<100ms), whereas for other instruments it may be slower or even variable according to playing style (Krumhansl, 1989). Spe*ctral brightness*, refers to the extent to which higher-frequency harmonics are present in a sound. A common measure of brightness is the spectral centroid, which is the average frequency of the harmonic components of a sound weighted by the magnitude of the different frequencies (Grey & Gordon, 1978). Finally, *spectral flux* is a measure of the change in the spectral distribution of a sound over the course of a single note. For example, in brass instruments the frequency spectrum tends to vary strongly with amplitude, which changes over time as a function of the loudness envelope of the sound. Strong dynamic changes in brightness give these instruments a high level of spectral flux compared to other orchestral instruments, and this is one of the acoustic features that people use perceptually to distinguish brass instruments from other musical sounds (Krumhansl, 1989; McAdams, 1995).

We were specifically interested in extending the line of research on cross-modal associations described in Palmer et al. (2013) and in Chapter 2 to instrumental timbres. Given the evidence that not only full-scale classical music (Palmer et al., 2013) but also more basic musical elements such as intervals and chords (Chapter 2) can be systematically mapped into different colors, a plausible next question is whether the same is true for other basic musical elements, such as the timbres of the same pitch played at the same loudness and same duration by different instruments. More specifically, we were interested in determining whether any of the acoustic features of timbre (attack time, spectral brightness, spectral flux, etc.) would be useful for predicting the visual associations of our participants, and whether common emotional associations might also play a role in their choices. Sound stimuli consisting only of timbral differences might seem to be so emotionally impoverished compared to full musical compositions that they would not be worth testing, but given the dominance of emotionally mediated judgments over low-level sensory mapping for musical compositions as well as for chords and intervals, it is possible that emotional mediation occurs for timbres as well.

Unlike intervals and chords, which are defined purely by harmonic relationships, timbre is identified by temporal as well as spectral components (e.g., Elliott, 2012). It is therefore not surprising that previous studies of audiovisual cross-modal matching have found that temporal features of sound, such as rhythm, can be intuitively matched to spatial features of images (e.g., Sherman, Grabowecky, & Suzuki, 2013). For this reason, we decided not to limit ourselves to color stimuli for these experiments, as we did in Chapter 2, but also to include stimuli that varied along spatial and temporal dimensions.

Experiment 6: Mapping Timbre to Color and Gaussian Blur

Methods

Participants. We recruited 29 undergraduate participants from the Berkeley Research Participant Pool. None had any color-vision vision defects, as tested using the Dvorine Pseudo-Isochromatic Plates. All gave informed consent and were naïve to the purpose of the study, and the Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

Stimuli. For the audio stimuli, we used a wide variety of naturalistic timbres recorded from actual musical instruments to approximate real-world listening experiences. These audio stimuli were generated using multi-sampled recordings of instruments from the Kontakt Orchestral Library. The 17 instrumental timbres included were recordings of a bassoon, celesta, cello, clarinet, English horn, flute, French horn, harpsichord, marimba, oboe, piano, electric piano, trombone, trumpet, tuba, and two recordings of a violin (one played in a *legato* style and the other played *staccato*). Each instrument was recorded playing a single ascending arpeggiated major triad beginning at middle C (261.6 Hz), at a rate of one note per second. Table 3.1 reports a number of relevant acoustic properties of these stimuli.

The visual stimuli consisted of two types: colored squares and white circles with varying degrees of blur. The color stimuli consisted of the 37 Berkeley Color Project colors displayed simultaneously as 100x100 colored squares against a neutral gray background (see Figure 2.7). The blur stimuli consisted of nine white circles of 60 pixel diameter displayed simultaneously against a gray background. Their edge blur/sharpness varied to different levels for each circle by blurring them using isotropic Gaussian filters with standard deviations of .1, .75, 1, 1.5, 2, 4, 8, and 10 pixels (see Figure 3.1).

Design

The color association task was a modified version of that used for the single-color experiments to study music-to-color mapping for interval stimuli described in Chapter 2 (Experiments 2, 3, and 4), as well as that used by Palmer, Schloss, Xu, and Prado-Leon (2013) to study music-to-color mapping for classical orchestral music. Participants were asked to listen carefully to each audio stimulus while viewing the matrix of 37 BCP colors and were asked to choose which three colors "went best" with that sound, starting with the best match, then the second best, and then the third best. They indicated each choice by clicking on the colored square using the computer mouse, at which point the color disappeared from the matrix in order to prevent it from being chosen twice. After indicating the colors that they felt "went best" with the sounds, the sound played again and participants were asked to chose which three of nine different levels of blur they felt "went best" with the sound. The sounds were then played again and participants were asked to chose which three of nine different levels of blur they felt "went worst" with the sound.

A separate group of 20 participants also rated all of the colors, timbres, and Gaussian blurs on line rating scales for the dimensions *happy/sad*, *calm/angry*, *active/passive*, *weak/strong*, and *harmonious/disharmonious* using a method similar to that described in Chapter 2.

Results

Average ratings for each timbre on each of the six emotional/semantic dimensions are plotted in Figure 3.2. These ratings differed significantly for all of the dimensions we tested with the exception of *harmonious/disharmonious*: *happy/sad* F(16,221)=9.81, p<.001; *not-angry/angry* F(16,221)=2.36, p<.01; *calm/agitated* F(16,221)=4.82, p<.001; *active/passive* F(16,221)=3.93, p<.001; *strong/weak* F(16,221)=5.99, p<.001; *harmonious/disharmonious* F(16,221)=1.62, p=.07.

To assess whether there are systematic associations between the timbres and the four dimensions of color appearance for the sounds chosen as going best/worst wit the sounds, we also calculated sound-color association (SCA) scores for each color dimension following the formulas described in Experiment 2. Average SCA scores for each timbre on each color appearance dimension are plotted in Figure 3.3. Statistical analyses of these data showed that different timbres significantly affected color choices for all color dimensions except *yellow-blue*: *saturated/unsaturated* F(16,221)=5.76, p<.001; *yellow/blue* F(16,221)=1.44, p=.116; *red/green* F(16,221)=2.31, p<.01; *light/dark* F(16,221)=2.29, p<.01 .

Next we investigated whether any of the major acoustic dimensions of timbre were predictive of participant color choices by computing correlations over timbres between the values of each acoustic feature and the SCA values for each color dimension. We found that participant color choices were moderately correlated with two acoustic dimensions: the average *yellow-blue* value of the chosen colors was correlated with timbre attack time, with quicker attack times being associated with yellower colors (*r*=.58, *p*<.05), and higher spectral centroid being associated with redder colors (*r*=.52, *p*<.05). The full correlation matrix is reported in Table 3.2. Given the moderate size of these correlations, the number of statistical tests performed (12), and the fact that the dimension of *yellow/blue* did not show significant differences between timbres in the previously reported ANOVA test, this result should probably be discounted.

Next we looked at whether acoustic features predicted average emotion ratings for the timbre stimuli. We found (prior to multiple-testing correction) that every emotion dimension we measured was significantly correlated with either attack time, brightness, or both, but spectral flux was not a significant predictor for any dimension. In order to correct for errors introduced by multiple testing, we ran a bootstrap simulation of the data with 1000 replications, and only retained correlations that were higher than the 95th percentile in the simulated data (absolute value r=.62).[4] After this correction, only the correlations between attack time and the dimension *happy/sad* and *strong/weak* were retained. The full set of correlational values for these dimensions is reported in Table 3.3.

We also calculated SCA scores for all of the emotion dimensions we tested. For all of them, the emotion ratings for colors chosen as going best/worst with a given timbre were significantly correlated with the emotion ratings of that timbre: *happy/sad* (r=.69, p<.01), *calm/angry* (r=.65, p<.01), *active/passive* (r=.65, p<.01), and *strong/weak* (r=.83, p<.001). Interestingly, unlike the results for intervals and chords in Experiments 1 and 2, in which the dimensions of *active/passive* and *strong/weak* had consistently lower SCA correlations, *strong/weak* actually produced the highest correlation in the present study. This fits well with the idea that different musical features, such as interval harmony and timbre, may be more strongly associated with different semantic and emotional dimensions.

We also calculated sound-blur associations (SBAs), analogous to SCAs (see Experiment 2) for the blur stimuli chosen as the three-best and three-worst fitting associations for the timbres, using the weighted average of the log of the Gaussian width for the selected images (Figure 3.4). In contrast to the color choices, choice of blur stimulus was reliably related to an acoustic parameter of the timbre's sound. In particular, blur level was significantly correlated with attack time (r=.81, p<.001), with quicker attacks being associated with sharper edges, but there was no relation for either spectral centroid (r=-.21, p=.41) or spectral flux (r=-.11, p=.67).

Several of the emotional/semantic dimensions were significantly correlated with blur, but none of them explained as much of the variance in blur choices as did attack time: (r=-.53, p<.05), *calm/angry* (r=-.43, p=.08), *active/passive* (r=-.60, p<.05), *strong/weak* (r=-.66, p<.01), and *harmonious/disharmonious* (r=.55, p<.05). We then investigated the contribution of different emotional/semantic and acoustic dimensions to participants blur choices by conducting a forward stepwise regression including the five emotional and the three acoustic features as possible predictors. The resulting model included only attack time as the first predictor (p<.001, partial r=.81) and spectral centroid as the second predictor (p<.01, partial r=.11), with a multiple r=.92 (85% of variance explained). None of the other emotional or acoustic predictors added significant variance beyond that explained by these two acoustic parameters, and a separate regression model using all five emotion dimensions as predictors only managed to explain 46% of the variance.

Discussion

The results of this experiment indicate that, when timbres are matched to best/worst fitting colors, the emotional mediation hypothesis extends beyond full-scale music (Palmer et al., 2013), chords and intervals (Chapter 2) to the instrumental timbres of individual notes. Notably, the dimension that showed the highest correlation (*strong/weak*) was not the same as those found to

---

[4] To bootstrap a significance threshold, we generated 1000 random data sets by repeatedly shuffling the values for the acoustic features of the timbres. For each of these data sets, we ran the full set of correlations and recorded the absolute value of the highest correlation found between emotional and acoustic dimensions. We then sorted the list of resulting correlation values, and took the 950th value in the list as the threshold for retention.

be the best predictors for chords and intervals (*happy/sad*, *calm/angry*, and *harmonious/disharmonious*). This result supports the idea that different musical features may differentially influence the various dimensions of emotional/semantic association present in the full musical compositions studied by Palmer et al. (2013).

In contrast, it appears that the best/worst fitting blur matches for timbres are better predicted by the acoustic features of attack time and spectral brightness than by emotional associations of the timbres. This result suggests that the Emotion Mediation Hypothesis does not extend to all types of cross-modal mappings between musical and visual features and may be specific to those visual features (such as colors) that have strong and consistent emotional associations that could be shared with musical stimuli. When people match complex visual images to music in the real world (such as in choreography or music visualization) they are probably relying on multiple strategies for making appropriate visual matches, including both emotional mediation and low-level sensory feature mapping.

Experiment 7: Dynamic Images

The results of Experiment 6 showed that the spatial feature of blur can be an effective way of representing the temporal dynamics of a timbre using a static visual image. But images do not have to be static and unchanging. Indeed, given that timbre is a dynamic auditory property, perhaps a better correlate for their temporal changes would be corresponding temporal changes in the image itself. We investigated this possibility by varying the temporal dynamics of a visual stimulus in a fashion analogous to the temporal dynamics of an individual musical note.

Methods

Participants. We recruited 30 undergraduate participants from the Berkeley Research Participant Pool. None had any color-vision vision defects, as tested using the Dvorine Pseudo-Isochromatic Plates. All gave informed consent and were naïve to the purpose of the study, and the Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

Stimuli. The auditory stimuli for this experiment consisted of single recorded notes from ten of the seventeen timbres listed above. These timbres were selected by conducting a k-means analysis of ten clusters on the timbre rating data from the previous experiment (*happy/sad*, *calm/angry*, *active/passive*, *strong/weak*, and *harmonious/disharmonious*), and then choosing the timbre from each cluster that was closest to the centroid of that cluster and therefore most representative of the cluster as a whole. When a cluster consisted of two timbres, such that both were equidistant from the centroid for that cluster, the timbre was chosen randomly from those two. The resulting set of timbres consisted of the bassoon, cello, clarinet, flute, English horn, harpsichord, piano, trumpet, tuba, and violin (staccato). In the audio stimuli, each instrument played a single note that was held for one second, with the full recording lasting two seconds in order to capture natural decay characteristics of the sound. These stimuli were played as a loop during each trial of the experiment.

The visual stimuli each consisted of a 60-pixel diameter white dot displayed against a gray background, that appeared and disappeared from the screen once every two seconds (see Figure 3.5). The dynamics of each timbre was matched to the dynamics of a corresponding brightness/contrast envelope of the dot that was calculated as follows. First, we took the absolute value of the loudness function for each audio sample, smoothed it with a 200ms Gaussian filter to remove flicker, and mapped it onto a brightness curve such that -100 dB (silence) was equivalent to a brightness of 128 (medium gray, the background color for the display) and -3dB (maximum loudness) was equivalent to a brightness of 255 (bright white, the maximum possible on our monitor). In the experiment itself, participants saw each brightness envelope paired with each timbre, with start and stop times fully synchronized. The participants' task on each trial was to rate how well the visual stimulus matched the sound of the timbre they were hearing. Both the audio and visual stimuli were displayed as a looping animation during the experiment.

Design. Before beginning the experiment, each participant completed an anchoring and instruction session in which they heard a piano tone and had to choose one of three dynamic visual images as being the best match, and one as being the worst. No feedback was given about which responses were "correct." The dynamic images were generated from the loudness envelopes of the piano tone (good match), the bassoon (intermediate match), and the violin legato (bad match). All participants reported the first image as being the best match and the last image as being the worst.

Participants were then exposed to all possible combinations of audio and visual stimuli, and asked to rate each one for the goodness of fit between the audio and the visual stimuli using a 400-pixel line rating scale. The ends of the scale were labeled "Good Fit" and "Bad Fit." Participants had to watch one full loop of a given stimulus pair before making their rating, and the stimuli looped continuously thereafter until the rating was made.

Results

We were interested in how participant ratings of good/bad fit could be predicted by both the overall correlation in envelopes between the two stimuli and by discrete features such as attack time, decay time, and spectral brightness. We created a single summary score for each audio-image pairing by averaging all of the participant fit ratings for that particular pairing. We found that these scores were well-predicted by the arctangent[5] of the correlation between the loudness envelope of the audio stimulus and the brightness envelope of the visual stimulus (r=.78, p<.05). However, we found that we could explain nearly as much variance simply using the difference in attack time between the stimuli (r=.74, p<.05), with the differences in other parts of the envelope producing statistically non-significant correlations (sustain, r=.49, p=.18; decay, r=.44, p=.23). Interestingly, difference in spectral brightness was also a marginally significant predictor of fit ratings on their own (r=.69, p<.05) although not spectral flux (r=.55, p=.09). However, spectral brightness didn't add significantly to a multiple regression model once attack time had already been taken into account (p=.38), indicating that its explanatory power derived only from being anti-correlated (r=-.39) with attack time in our stimuli.

Discussion

Although it seems that optimal fit ratings would be made by comparing the entire loudness envelopes of the audio and visual stimuli, the initial onset of the sound had a disproportionate influence, and the ratings were not independently affected by the spectral or spectro-temporal properties of the sound. There are a number of possible explanations for this interesting result. It might be due to a primacy effect in memory, but this seems unlikely because participants were presumably not relying on memory when making their judgments (Glenberg et al., 1980). Another possibility is that it arises from auditory forward masking, but the timescale for forward

---

[5] The arctangent transformation corrects for statistically problematic compression of correlation scores at extreme values.

masking effects is typically less than 50ms (Elliott, 1971), a span of time that is shorter than even the attack period alone for many of the timbres we used. A third possibility is that it may be related to findings on the phenomenon of echo suppression (Zurek, 1979). When presented with binaural sounds that include reverberation, the human auditory system is able to identify parts of the audio signal that are a result of echo based on their acoustic characteristics and specifically suppress this input, in order to better isolate the original non-echo components of the sound. This effect disappears with the use of monaural stimuli, as the auditory system is no longer able to effectively distinguish the echo components from the original signal. This presents a clear hypothesis for follow-up experiments in this area, where we would expect to see better matching results using monaural or anechoic stimuli.

Experiment 8: Multiple Simultaneous Timbres

Unlike the single instrumental timbres in Experiments 6 and 7, real-world music usually contains multiple instrumental timbres at the same time. We therefore asked how visual associations with multiple or blended timbres might relate to the timbres of the single instruments of which they are composed. For example, is it possible to predict participant responses to a complex mix of timbres based on their responses to those individual components? To investigate this problem, we looked at how participants make color matches to combinations of two and three timbres in a single stimulus and compared this to the choices they make for the individual component timbres.

Methods

Participants. We recruited 20 undergraduate participants from the Berkeley Research Participant Pool. None had any color-vision vision defects, as tested using the Dvorine Pseudo-Isochromatic Plates. All gave informed consent and were naïve to the purpose of the study, and the Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

Stimuli. We used a subset of 5 of the 17 timbres used in Experiment 6 (flute, harpsichord, marimba, trumpet, and violin legato). These specific timbres were chosen by conducting a series of k-means clustering analyses for five clusters based on the emotional/semantic rating data from Experiment 1 and using the best result out of 1000 replications. The five selected timbres were the ones closest to the centroid of the five resulting clusters, and therefore most representative of the parent cluster as a whole.

In the audio stimuli, these five timbres were combined in all ten possible combinations of three instruments (e.g., trumpet, violin and flute). Each of these combined stimuli consisted of the specified set of instruments playing the same arpeggiated major triad in unison, resulting in a blending or mixture of the component timbres. Even with the instruments all normalized to the same maximum volume, some instruments were nevertheless more audible in a mixture than others. To adjust for any imbalances, we conducted a pilot study in which 10 undergraduate participants listened to looped recordings of all possible combinations of two instruments. The presentation started with one instrument silent and the other at full volume. The participant then adjusted the volume ratio of the two sounds using a slider until they perceived the two instruments to be equally prominent in the mix. The initial volumes of the recordings were counterbalanced to eliminate the effect of starting position on participant choices, such that each participant heard the same instrument combination on two different trials: once with instrument-1 initially audible and instrument-2 silent, and once with instrument-1 initially silent and instrument-2 audible. The average volume chosen for an instrument across all trials and all participants was then used as the volume for that instrument in the subsequent study.

Design. Each participant listened to a total of 25 timbres, consisting of all of the two and three instrument mixtures as well as their component timbres, presented in a random order. Their task was to choose which three of the BCP 37 colors they thought went best and worst with each stimulus (in order). The same participants also subsequently rated all of the color stimuli, single timbres, and combinations of timbres on line rating scales for the dimensions *happy/sad*, *calm/angry*, *active/passive*, and *weak/strong*.

Results

The average emotional/semantic ratings for each multiple-timbre stimulus are shown in Figure 3.6. The combined timbre stimuli differed significantly on all of these dimensions except for *happy/sad* ($F(9,171)=0.80$, $p=.61$): *calm/angry* ($F(9,171)=5.65$, $p<.001$), *active/passive* ($F(9,171)=4.18$, $p<.001$), *strong/weak* ($F(9,171)=3.6,4$ $p<.001$), *harmonious/disharmonious* ($F(9,171)=5.24$, $p<.001$). We tried predicting these ratings using the average rating of the component timbres in the stimulus, and found that this simple model was a good fit for ratings of the multiple-timbre stimuli. The emotion ratings for the combined timbres were strongly correlated with the simple average of the emotion ratings of their component timbres for each of the emotion dimensions we tested: *happy/sad* ($r=.82$, $p<.01$), *calm/angry* ($r=.91$, $p<.001$), *active/passive* ($r=.96$, $p<.001$), *strong/weak* ($r=.86$, $p<.001$), and *harmonious/disharmonious* ($r=.86$, $p<.001$).

Using previously collected rating data for each of the colors for the dimensions *light/dark*, *saturated/unsaturated*, *red/green*, and *blue/yellow* (see Table 2.8 in Chapter 2), we were able to calculate sound-color association (SCA) values for each multiple timbre stimulus according to the equations specified in Experiment 2. The means for these data can be seen in Figure 3.7. We found that a simple averaging of the *light/dark* ratings of the component colors was a good predictor for the *light/dark* ratings of the colors chosen to go with a given mixture ($r=.71$, $p<.05$) and for the *saturated/unsaturated* ratings ($r=.88$, $p<.01$). Simple averaging was much less predictive, however, for the dimensions of *red/green* ($r=.23$, $p=.29$) and *blue/yellow* ($r=.20$, $p=.26$).

Discussion

The present results indicate that participants choose colors for a multiple-timbre stimulus whose emotional/semantic associations are similar to the average of the corresponding associations for its individual component timbres. In colorimetric terms, this relation seems to hold for the dimensions of lightness and saturation but not for the *red/green* and *yellow/blue* dimensions of hue. For pragmatic reasons arising from combinatorial considerations, we limited this study to a small number of component timbres. It thus remains possible that the specifics of these results might not replicate to a wider variety of timbres. However, the overall result is encouraging, as it implies that our findings about visual and emotional associates of timbre are likely to generalize not only to other individual timbres but also to complex mixtures of timbres such as those found in large-ensemble classical and popular music.

General Discussion

The results of the experiments described in this chapter show that people have consistent visual associations with timbres for several different dimensions of visual images. For color associations, we found evidence of emotional mediation, meaning that when asked to choose which colors went best with the sound of a given timbre, participants chose colors whose emotional associations were similar to those of the sound stimulus rather than matching according to low-level auditory features of timbre, such as attack time, spectral center, and spectral flux. For blur and onset/offset dynamics, however, we found that these visual matches were well predicted by the low-level, sensory, non-emotional factors related to the loudness envelope (attack time) and spectral qualities of the sound (spectral centroid). Additionally, it appears that participants match colors and make emotion ratings for combinations of timbres in a way that is predictable from their responses to single timbres. These findings, combined with the previously described work on intervals and chords, as well as Palmer et al.'s (2013) work on color matching with orchestral music, lends credence to the idea that emotional mediation is occurring at many different levels in music, from single sounds to entire compositions, and that the emotional associations at these different levels may build in systematic and predictable ways from lower-level components.

Figure 3.1: In Experiment 6, participants were given a choice among nine different levels of Gaussian blur for each audio stimulus. From left to right, the standard deviation of the Gaussian kernel for each circle (whose diameter was 60 pixels) was .1, .75, 1 1.5, 2, 4, 8 and 10 pixels.

	Attack	**Sustain**	**Decay**	**Centroid**	**Spectral Flux (Normalized)**
Bassoon	70ms	920ms	10ms	995 Hz	0.04
Celeste	50ms	0ms	950ms	859 Hz	0.04
Cello	350ms	490ms	160ms	1597 Hz	-1.38
Clarinet	270ms	590ms	140ms	918 Hz	0.67
English Horn	120ms	870ms	10ms	1572 Hz	-0.18
Flute	320ms	430ms	250ms	916 Hz	-2.16
French Horn	270ms	720ms	10ms	1015 Hz	0.87
Harpsichord	30ms	0ms	970ms	2358 Hz	0.34
Marimba	90ms	0ms	610ms	690 Hz	0.81
Oboe	130ms	860ms	10ms	1361 Hz	0.44
Piano	80ms	0ms	920ms	1033 Hz	0.17
Rhodes	20ms	0ms	980ms	1605 Hz	-2.23
Trombone	300ms	960ms	10ms	934 Hz	0.47
Trumpet	200ms	790ms	10ms	1601 Hz	1.13
Tuba	160ms	720ms	120ms	567 Hz	0.84
Violin (Legato)	900ms	60ms	40ms	1990 Hz	-0.06
Violin (Staccato)	120ms	850ms	30ms	1737 Hz	0.18

Table 3.1: Relevant spectral and temporal features of the audio stimuli used in Experiment 6, calculated using MatLab MIR Toolbox (Lartillot & Toiviainen, 2007).

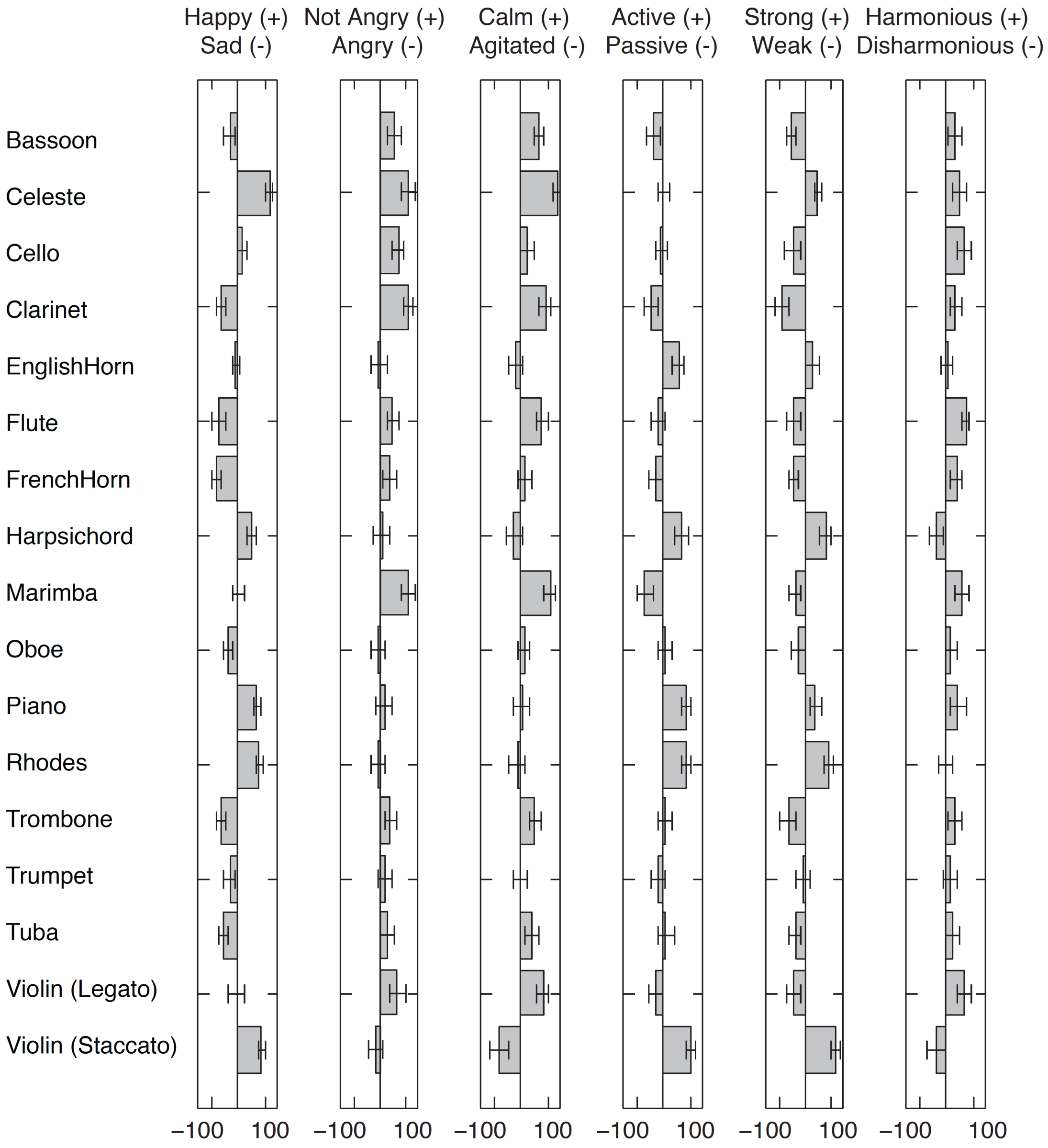

Figure 3.2: Average emotion ratings for the different timbre stimuli in Experiment 6. Error bars represent the standard errors of the mean (SEMs).

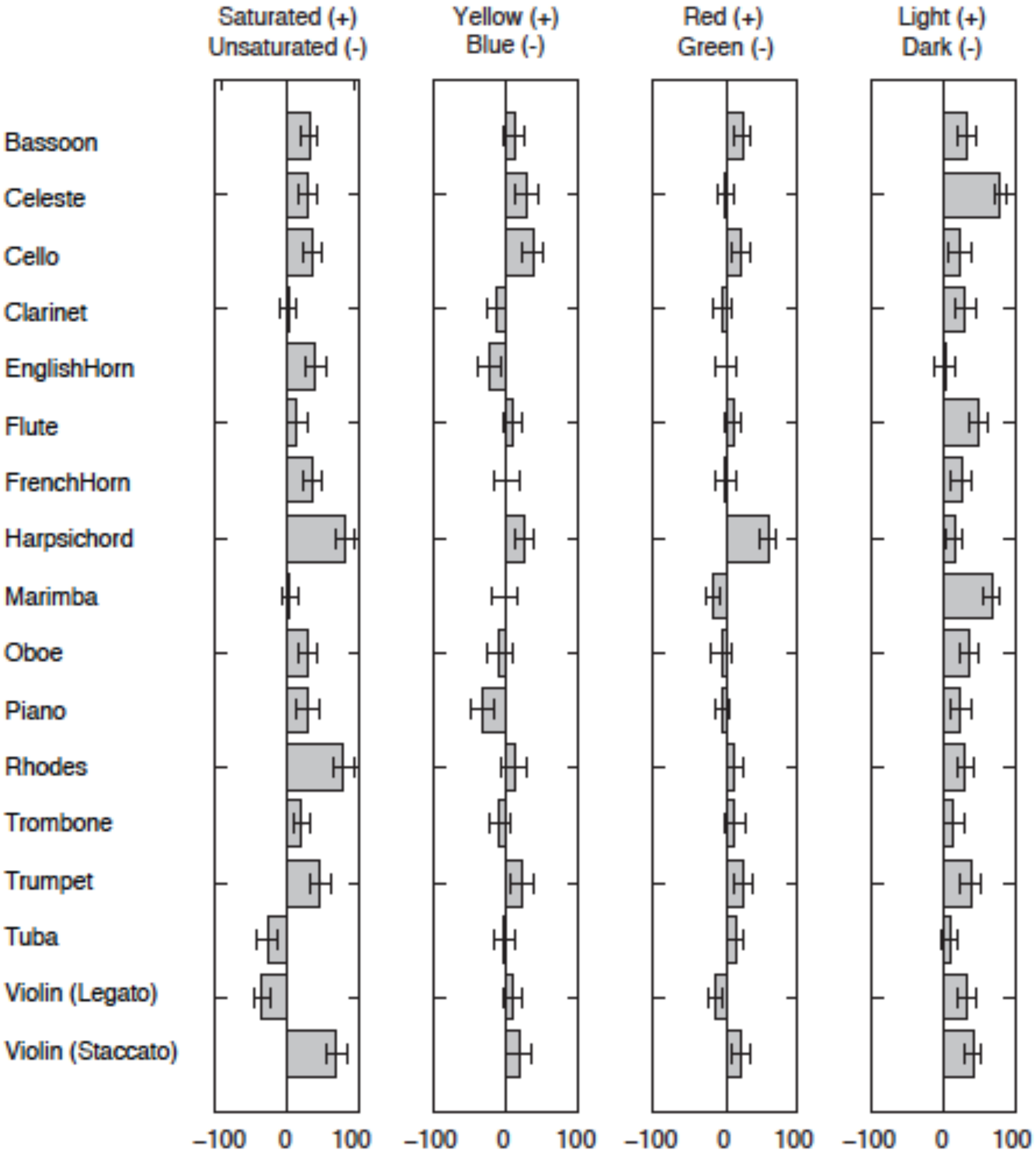

Figure 3.3: Average color appearance dimensions for the color-sound association (SCA) scores based on the colors chosen as going best/worst with timbre stimuli in Experiment 6. Error bars represent SEMs.

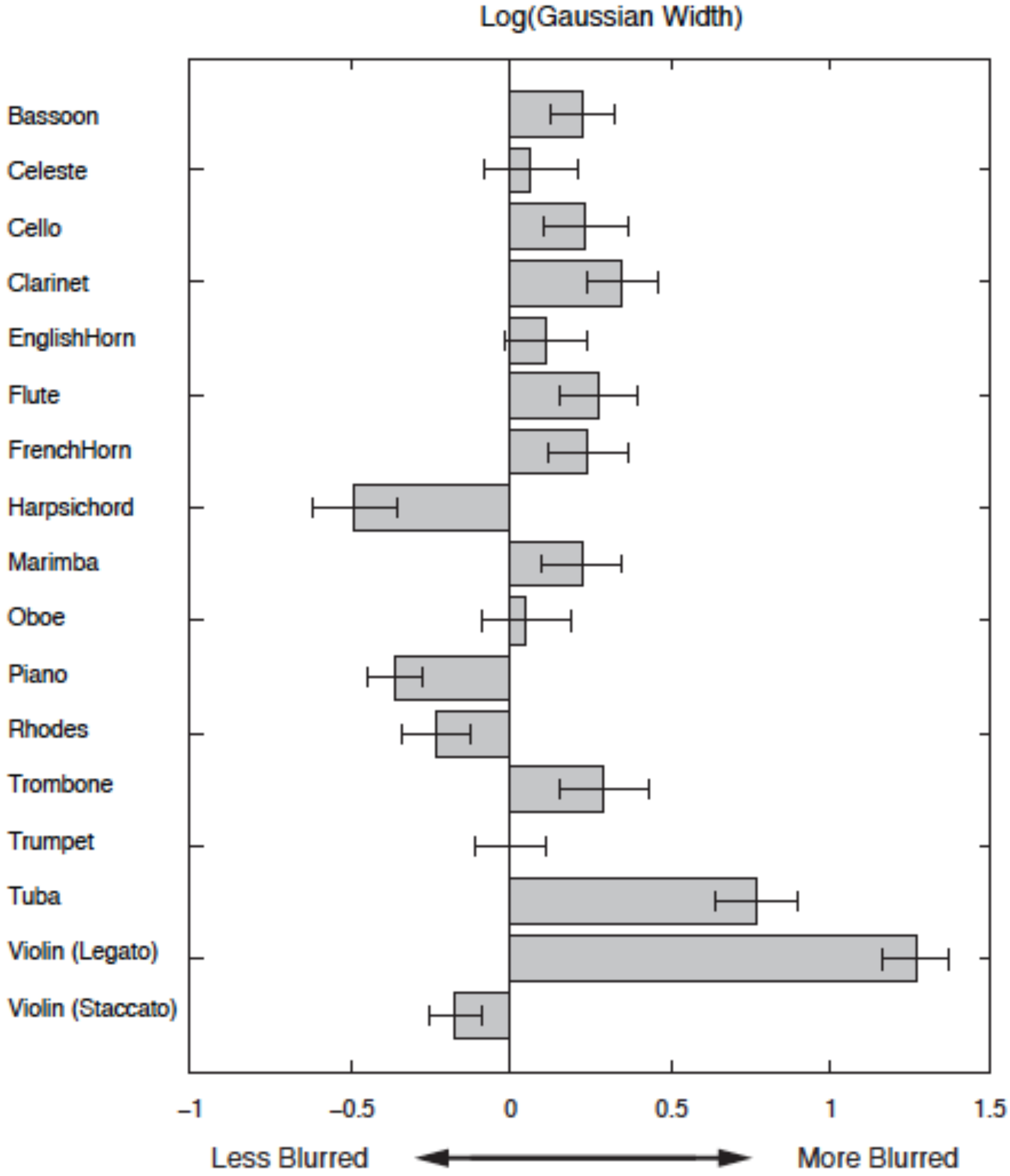

Figure 3.4: The log of the weighted average Gaussian width that was chosen as going best with each of the audio stimuli in Experiment 6. Error bars represent SEMs.

Correlations Between Color Dimensions and Acoustic Features of Timbres

	Log Attack Time	Spectral Centroid	Spectral Flux
Saturated (+) Unsaturated (-)	-0.35	0.17	0.24
Yellow (+) Blue (-)	-0.58	-0.09	0.16
Red (+) Green (-)	-0.28	0.52	0.13
Light (+) Dark (-)	-0.32	0.13	0.19

Table 3.2: Correlations of average SCA scores with acoustic features of the timbres studied.

Correlations Between Emotion Ratings and Acoustic Features of Timbres

	Log Attack Time	Spectral Centroid	Spectral Flux
Happy (+) Sad (-)	-0.62	0.37	0.18
Calm (+) Angry (-)	0.28	-0.48	-0.19
Active (+) Passive(-)	-0.52	0.47	0.26
Strong (+) Weak (-)	-0.68	0.50	0.32
Harmonious (+) Disharmonious (-)	0.56	-0.52	-0.01

Table 3.3: Correlations of participant ratings with acoustic features of the timbres studied.

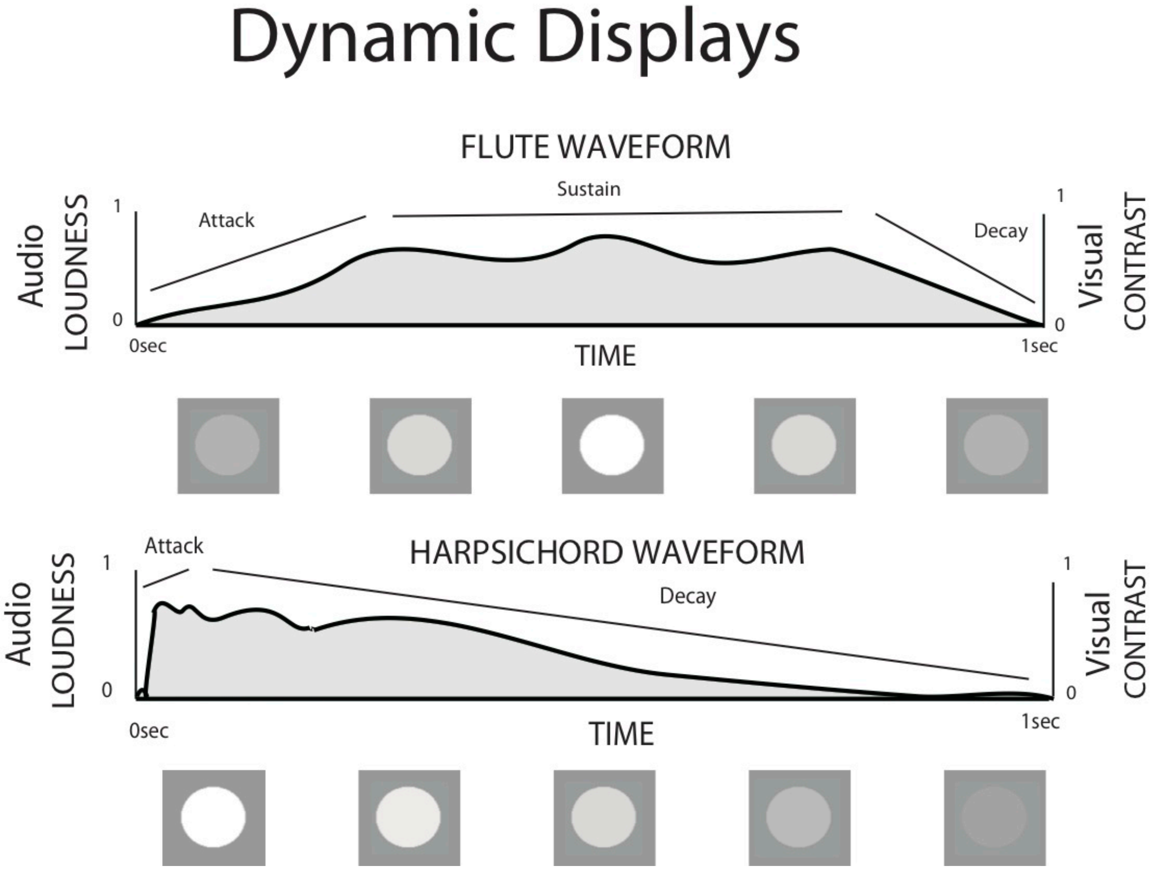

Figure 3.5: An illustration of the relationship between the loudness envelopes of two timbre stimuli and the brightness of envelopes of the corresponding dynamic visual images which were generated from them. The flute has a long attack time and an extended sustain period before a relatively short decay. The harpsichord is a percussive instrument with a short attack time, long decay period, and essentially no sustain. The corresponding brightness envelopes maintain these features, but convey them using differences in visual contrast rather than loudness.

	Bassoon	Cello	Clarinet	English Horn	Harpsichord	Piano	Trumpet	Tuba	Violin (Staccato)
Bassoon	135.23	75.09	77.70	115.87	34.27	29.37	95.75	133.71	5.78
Cello	91.06	107.86	109.56	73.96	-59.95	-55.52	73.75	66.19	-114.44
Clarinet	64.65	105.45	106.04	93.79	-60.54	-72.86	69.75	90.53	-132.56
English Horn	108.05	75.62	85.25	129.15	41.22	28.26	113.53	103.36	-24.23
Harpsichord	65.81	-36.90	-56.38	40.06	101.79	121.07	50.46	37.70	74.11
Piano	49.98	-57.48	-75.81	47.29	107.36	104.67	36.40	9.73	87.45
Trumpet	113.64	95.46	102.42	133.93	50.54	38.55	120.58	125.34	-12.03
Tuba	121.90	90.68	99.11	129.01	15.95	33.53	129.23	133.73	-46.29
Violin (Staccato)	-0.92	-120.77	-129.01	-0.04	88.34	66.26	-38.64	-12.94	126.24

Table 3.4: Average fit ratings from Experiment 7, on a scale from -200 to +200. The rows in the table represent the different audio stimuli, and columns represent the corresponding dynamic visual stimuli (as defined in the text).

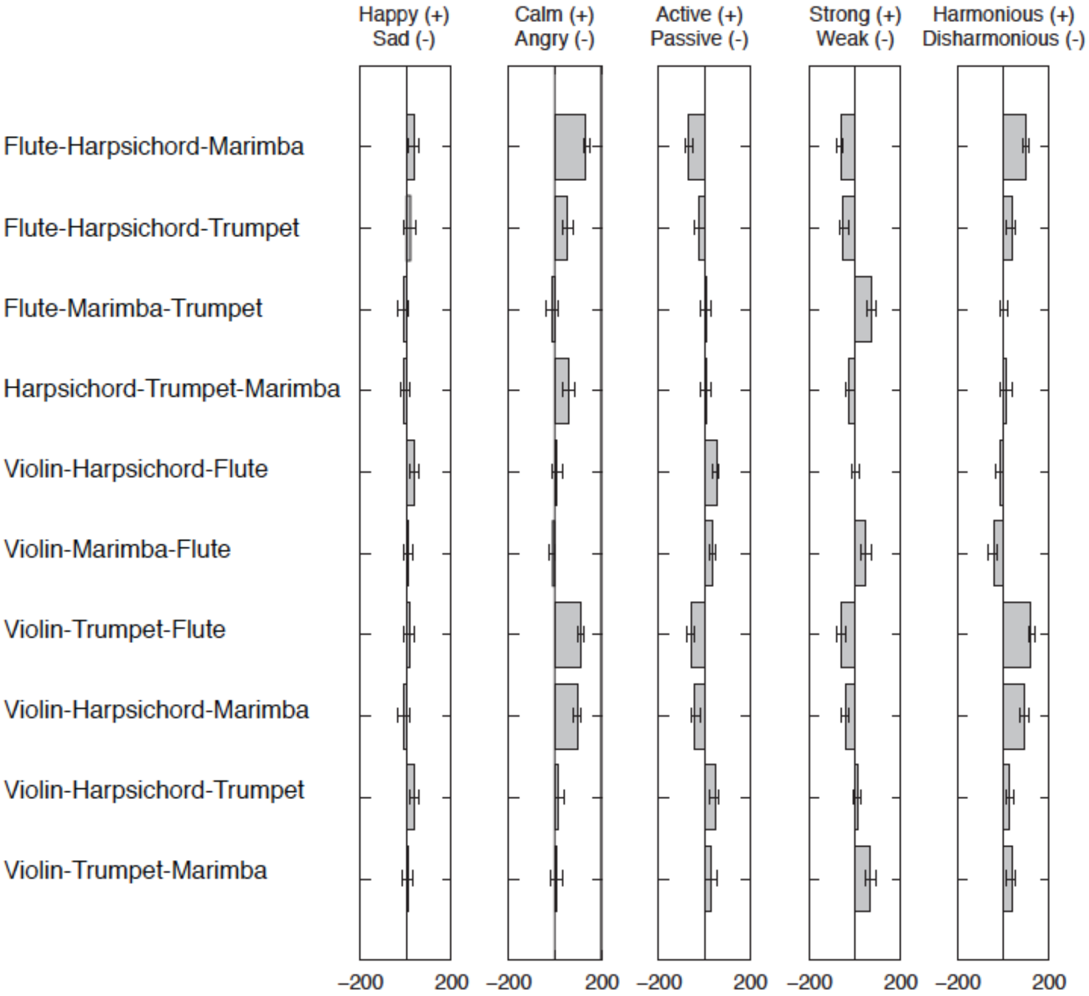

Figure 3.6: Average emotional/semantic ratings for combined timbre stimuli used in Experiment 8. Error bars represent SEMs.

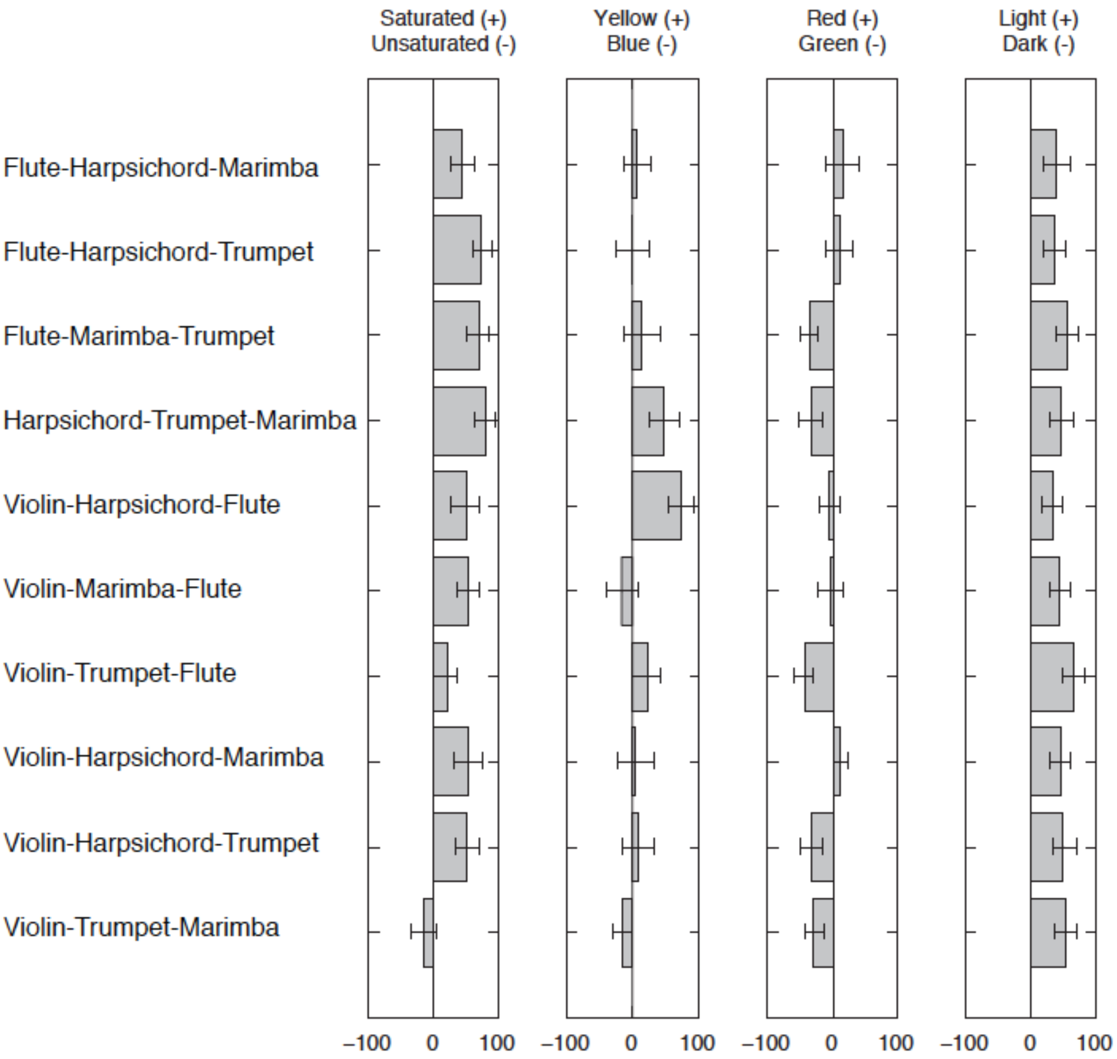

Figure 3.7. Average color appearance values for combined timbre stimuli used in Experiment 8. Error bars represent SEMs.

Chapter 4: Audio-Visual Cross-Modal Associations in Synesthetes and Non-synesthetes

Although many people have strong associations between music and color, it's uncommon for individuals to automatically *see* color in response to hearing music. A person for whom this happens is said to have "synesthesia," a rare condition in which the person experiences a percept in one modality or sensory domain when they are presented with a stimulus in another modality or sensory domain. For example, individuals with music-to-color synesthesia sometimes experience colored fields, shapes, and textures when listening to specific musical sounds. Figure 4.1 shows a reproduction of drawings done by two such synesthetes, illustrating their typical visual experiences in response to music. Many other types of synesthesias have been documented, although the most common and by far the best studied form of synesthesia is grapheme→color synesthesia, in which individuals experience uncolored letters and numbers as looking as if they were colored, in some sense (Cytowic and Eagleman, 2006; Day, 2005).

There has been a debate for several decades about what the relationship might be between synesthetic experiences and non-synesthetic cross-modal associations. Based on the finding that both synesthetes and non-synesthetes consistently make consistent matches between luminance and loudness, T. Hubbard (1996) argues that both synesthesia and cross-modal mapping are likely to be semantically mediated, and that the key difference between the two phenomena lies in the involuntary and explicitly sensory nature of synesthetic associations. While both synesthetes and non-synesthetes can imagine colors in their minds eye that seem to intuitively match a sound, only synesthetes actually *experience* this type of visual imagery consistently and automatically. Some researchers have also theorized that the two phenomena simply represent different points on a unified spectrum of multimodal experiences (see for example, Martino & Marks, 2001). Certainly, some music→color correspondences, such as mapping loudness or pitch height to color lightness, do seem to be structurally similar in both music→color synesthetes and non-synesthetes (Ward, 2006), but there is to date little hard evidence to support this idea.

The question of how differences at the neurological level might relate to the differing experiences of synesthetes and non-synesthetes is a complicated issue, but it is important to consider when interpreting the results of behavioral studies. Currently the two most prominent neural theories of synesthesia are *incomplete pruning* (E. Hubbard et al., 2011; Ramachandran & E. Hubbard, 2001) and *failure of inhibition* (Grossenbacher & Lovelace, 2001). The incomplete pruning hypothesis is based on the fact that many of the neural connections present during infancy do not survive into adulthood. This selective loss of neural connections has been hypothesized to occur to a lesser extent in synesthetes than in the general population, resulting in greater local connectivity between synesthetically related sensory modalities. In contrast, the failure of inhibition theory hypothesizes that synesthete's brains could be structurally the same as non-synesthetes, but functionally different, due to reduced inhibitory signaling from multisensory and other higher level processing areas.

There is currently some evidence from diffusion-tensor imaging studies supporting the incomplete pruning hypothesis in the case of color→grapheme synesthesia (e.g., Rouw & Scholte, 2007), but very little is known about the neural basis other types of synesthesia. It is worth noting that unisensory synesthesias (such as color→grapheme synesthesia), which involve communication between cortically neighboring areas, may well be be the result of very different neural processes from multisensory synesthesias (such as music→color synesthesia), which necessarily involve communication between anatomically distant areas of cortex. It is also an open question to what extent any of these connections involve direct communication between sensory areas as opposed to mediated activation through processing in other brain areas.

Somewhat surprisingly, no one has ever systematically studied how music-color synesthetes compare to non-synesthetes in their responses to complex musical features such as timbre or interval and chord harmony. In this chapter I directly address the nature of the relationship between the patterns of associations from music to color in non-synesthetes and the patterns of visual experiences evoked by music in music-to-color synesthetes. By collecting data from music→color synesthetes as well as non-synesthetes using the same musical and color stimuli and similar tasks, it should be possible to determine the extent to which these groups show similar structural, semantic, and/or arbitrary mappings between musical features and colors while listening to identical musical stimuli. If cross-modal mappings are dependent on the same underlying processes as synesthesia, we should find the same pattern of responses for synesthetes and non-synesthetes when averaged across individuals, despite the two groups' differing phenomenology. Additionally, if color→music synesthesia is a result of direct communication between sensory areas, we would expect to see mappings based on low-level sensory features rather than high-level semantic attributes.

Experiment 9: Sound-Color Associations and Synesthesia

Methods

Participants. We recruited 15 participants from a variety of sources, all of whom met the criteria for timbre→color synesthesia on the Eagleman Synesthesia Battery (Eagleman et al., 2007). This battery tests for synesthesia using a consistency measure, based on the finding that synesthetic color associations are remarkably consistent, even in experiments where the same participant is tested in sessions that are years apart (Baron-Cohen et al, 1993). Additionally, even over a shorter time span of a few minutes, non-synesthetic participants have a hard time remembering precise color choices for the same stimulus, whereas synesthetes do not have to remember at all: they are reminded anew by their central nervous systems every time they experience the same stimulus (Asher et al., 2006).

The Eagleman battery discriminates synesthetes from non-synesthetes in a single session by presenting the participants with a set of audio stimuli three times, in random order, and then calculating the variability of their color choices across different presentations of the same stimulus. Difference scores are calculated from the R (red), G (green), and B (blue) values of the chosen colors separately and then normalized and summed such that the minimum possible difference score is zero and the maximum is 3.0. These scores are then averaged across stimuli to achieve a single test score for each participant for each synesthesia test. The standard threshold for synesthetes is a score of 1.0 or lower, and non-synesthetes typically score around 2.0. The highest score for any of the synesthetic participants in our experiment on the timbre→color test was .88.

Additionally, three of our participants met the criteria for pitch→color synesthesia, and one met the criteria for chord→color synesthesia. For a summary of individual participants scores, see Figure 4.2.

For a comparison group, we also recruited an equal number of age-, gender-, and experience-matched non-synesthetic controls. These non-synesthetic volunteers were not asked to complete the full Eagleman battery, but simply to indicate on a Likert-scale questionnaire whether they experience visual sensations in response to musical chords, pitches or timbres, with the response options being "never," "rarely," "sometimes," "often," or "always." Participants were disqualified from the control group if they responded "sometimes" "often" or "always" to any of these questions.

Audio Stimuli. Two sets of audio stimuli were presented to every participant: one consisting of 24 different two-note intervals and the other consisting of 17 instrumental timbres. For the interval stimuli we used the same set of stimuli that were used in Experiment 1: the set of all possible intervals in a chromatic scale within the span of the octave that starts at middle C, and that same set transposed down an octave. All of these intervals were generated from recordings of a grand piano timbre and were played first sequentially (the lower note first and then the higher note) and then simultaneously, with each note sounding for one second. The second set of stimuli were the same as those used in Experiment 6: a set of arpeggiated major chords played using the timbres of 17 different instruments (bassoon, celesta, cello, clarinet, English horn, flute, French horn, harpsichord, marimba, oboe, piano, electric piano, trombone, trumpet, tuba, violin-staccato and violin-legato). All stimuli were rendered digitally using timbres from the Kontakt Orchestral Library.

Visual Stimuli. The visual stimuli in this experiment consisted of the 37 Berkeley Color Project colors displayed onscreen simultaneously, identical to those presented in the single-color experiments described in Chapters 2 and 3.

Procedure. All parts of this experiment were completed within the context of a five-session series of experiments designed to compare the synesthetic experiences of music→color synesthetes with the corresponding color associations of non-synesthetes. The nature of the other tasks will not be described here except to say that they all concerned color responses to more complete musical stimuli and that most of those tasks were completed before the ones reported here.

In the first session using timbres and intervals, we asked non-synesthetic participants, to choose the three colors (in order) that they felt "went best" with the audio stimulus they were hearing and, later, to choose the three colors (in order) that "went worst" with the audio stimulus. Their task was essentially identical to that for the non-synesthetes in Experiment 2 (for intervals) and Experiment 6 (for timbres) described above. The instructions for our synesthetic participants were as similar as possible, except that we asked them to choose the three colors that were *most similar* to their synesthetic experiences and, later, to choose the three colors that were *least similar* to their synesthetic experiences while they were listening to the audio stimuli. They were also given the option to respond "no color" at any time if they did not experience a synesthetic response to a sound, or if they had already exhausted the set colors they had synesthetically experienced in response to the sound on that trial.

In the second session of the experiment, both groups of participants rated all of the audio stimuli along six emotional/semantic dimensions (*happy/sad*, *not-angry/angry*, *agitated/calm*, *strong/weak*, *active/passive*, and *harmonious/disharmonious*) using a 400 pixel line rating scale, similar to those used in the experiments described in Chapters 2 and 3 of this dissertation. In the third session of the experiment, the participants rated all of the color stimuli along the same dimensions and completed a questionnaire about their experience with music and art as well as their history of synesthetic experiences (see Appendix for a full list of questions) as well as a semi-structured interview about the history and details of their synesthetic experiences. We also gave each participant the opportunity to illustrate the visual appearance of a typical synesthetic experience using colored markers.

In the fourth session we asked both groups to complete another task in which they chose colors that best fit the emotional qualities of the sound stimuli, effectively asking them to perform a cross-modal matching task based explicitly on emotional mediation. In this task, participants listened to each sound stimulus and chose the three colors that "best matched" and "worst matched" their emotional associations with that stimulus. We added this task specifically to investigate whether synesthetes and non-synesthetes have similar emotional associations with all of the stimuli. If synesthetes and non-synethetes performed this task in the same way, then any differences in the emotional effects in the initial color choice task are not likely to be due to differences in their emotional associations to either music or colors.

Results

All of our synesthetes reported seeing one or more colors in response to every audio stimulus. For a summary of their average choices in the interval stimulus task, see Figure 4.3. Some interesting patterns in their color associations are immediately apparent, such as the fact that lower pitches are associated with darker colors and higher pitches are associated with lighter colors, which has been found in previous studies of both audiovisual synesthetes and non-synesthetes (e.g., Ward, 2006).

In order to assess the level of emotional mediation that might be present in synesthetic color experiences, we calculated Sound-Color Association (SCA) scores for all six of the rated emotion dimensions, using the formulas described in Experiment 3 of Chapter 2 for single color SCA values. The synesthetic participants showed significant correlations between SCA values and emotion ratings for a number of dimensions on the interval stimulus task, including *happy/sad* (r=.55, p<.01), *not-angry/angry* (r=.66, p<.001, *weak/strong* (r=.64, p<.001) and *harmonious/disharmonious* (r=.73, p<.001) but not *calm/agitated* (r=.30, p=.07) or *active/passive* (r=.29, p=.08). Our matched control subjects also showed evidence of emotional mediation on the interval stimulus task. For the interval stimuli, SCA scores and emotion ratings were significantly correlated for the dimensions *happy/sad* (r=.74, p<.001), *not-angry/angry* (r=.66, p<.001), *calm/agitated* (r=.57, p<.01), and *harmonious/disharmonious* (r=.68, p<.001), but not for *active/passive* (r=.31, p=.14) or *strong/weak* (r=.39, p=.06), largely replicating the results described in Chapter 2.

For the timbre stimulus task, synesthetes showed a significant relationship between rated emotion and CSA for the dimensions *happy/sad* (r=.62, p<.01), *not-angry/angry* (r=.56, p<.05), *active/passive* (r=.67, p<.01), *strong/weak* (r=.60, p<.01), and *harmonious/disharmonious* (r=.48, p<.05) but only a marginally significant effect for *calm/agitated* (r=.41, p=.05). For the timbre stimuli, non-synesthete CSA scores and emotion ratings were significantly correlated for all of the emotion dimensions tested, including *happy/sad* (*r*=.63, *p*<.01), *not-angry/angry* (*r*=.56, *p*<.05), *calm/agitated* (*r*=.57, *p*<.05), *active/passive* (*r*=.67, *p*<.01), *strong/weak* (*r*=.68, *p*<.01) and *harmonious/disharmonious* (*r*=.53, *p*<.05). It is worth noting that on average the CSA correlations for non-synesthetes on in this experiment were somewhat weaker than those obtained in similar experiments described in Chapters 2 and 3. It is possible that this overall decrease is due to fatigue, as the studies were part of a larger set of experiments that was spaced out across multiple days, or is due simply to the fact that the numbers of participants were greater in the previously described studies.

When we asked synesthetes and non-synesthetes to make color judgments explicitly based on emotional associations rather than their synesthetic experiences, the relevant CSA correlations increased on average but only by a small amount which failed to reach significance (average difference was +.08, *t*(11)=1.66, *p*=.12). For a full visual comparison, see Figures 4.7 and 4.8. If we directly compare the synesthetes and controls for the emotional ratings of colors they chose by correlating the two SCA values from the two groups for each audio stimulus, we find that synesthetes and non-synesthetes are actually choosing colors with fairly similar emotional associations on both tasks (overall *r*=.76 for interval CSA scores across all five dimensions, and *r*=.55 for timbre CSA scores across all five dimensions). The full set of correlations with all of the dimensions is reported in Table 4.1. These results further support the idea that the cross-modal mappings between sound and color are mediated to some degree by by emotional/semantic associations in both groups and that the strength of these emotional mediation effects are relatively similar for synesthetes and non-synesthetes.

Discussion

One interesting and unexpected result of our survey measures was that 60% of our synesthetes described themselves as having color→emotion synesthesia (see Figure 4.9). Presumably, this means that they experience colored visual percepts in association with the experience of specific emotions. This percentage is even greater than that for grapheme→color synesthesia, which has been consistently found to be the most common type of synesthesia in other studies (Day, 2005). In addition, three participants who did not experience direct color-emotion synesthesia reported emotional modulation of their audiovisual percepts, wherein an incongruent emotional experience could attenuate their synesthetic percept. We have not yet tested any of the processing implications that might be entailed by the presence of emotion-to-color synesthesia, however.

These facts, along with the strong evidence for emotional mediation in the colors that synesthetes experience in response to musical intervals and timbres, promote an interesting hypothesis. Perhaps audiovisual synesthesia is not purely the result of low level feature-to-feature binding, but actually involves complex networks of semantic and emotional processing in addition to purely sensory information. This possibility has clear implications for the study of the neuroscience underlying synesthesia. For example, it is inconsistent with reduced pruning of local auditory-to-visual connections as the sole explanation of synesthetic experiences in the case of music→color synesthesia, as this theory would posit that synesthesia is unrelated to non-sensory semantic or emotional processes. On the other hand, it may also represent an important distinction between intermodal synesthesias, such as music→color synesthesia, and intra-sensory synesthesias, such as grapheme→color synesthesia. Non-grapheme synesthesias have generally been understudied in the literature, and it is safe to say that not enough attention has been paid to the way in which different varieties of synesthesia may represent fundamentally different underlying neural mechanisms.

Figure 4.1: Drawings by two of our synesthetic participants illustrating typical visual experiences in response to music. Participants describe these images as being overlaid over their vision or seen in their "mind's eye."

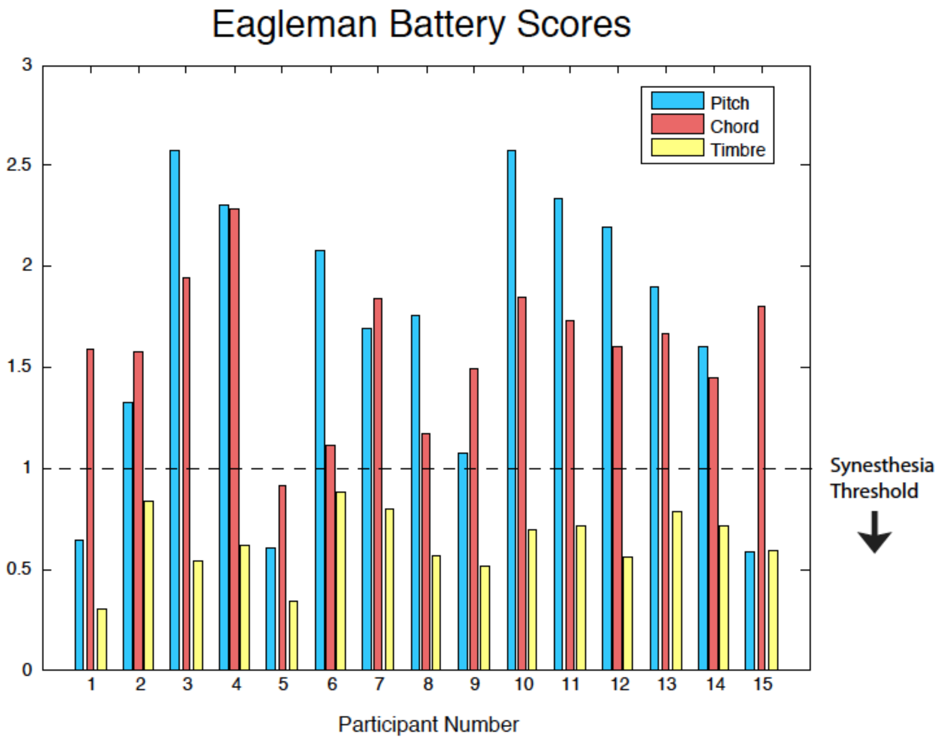

Figure 4.2: Eagleman Synesthesia Battery scores for all 15 synesthetic participants on the three music→color tests. The cutoff for inclusion in the synesthetic group of our study was a score of 1.0 or less on one of the tests. All 15 participants scored below 1.0 for timbre-color synesthesia. Three additional participants scored below 1.0 for pitch-color synesthesia (#1, 5, and 15) and one for chord-color synesthesia (#5).

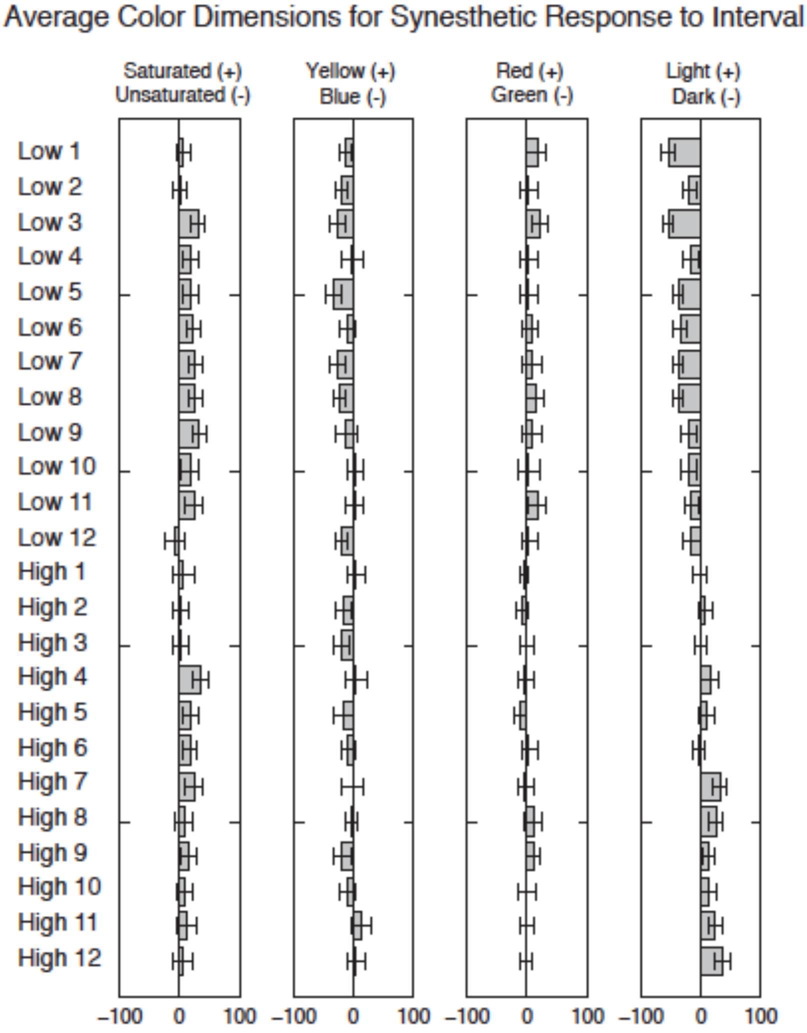

Figure 4.3: Average color appearance dimensions for the colors reported by synesthetes while listening to interval stimuli.

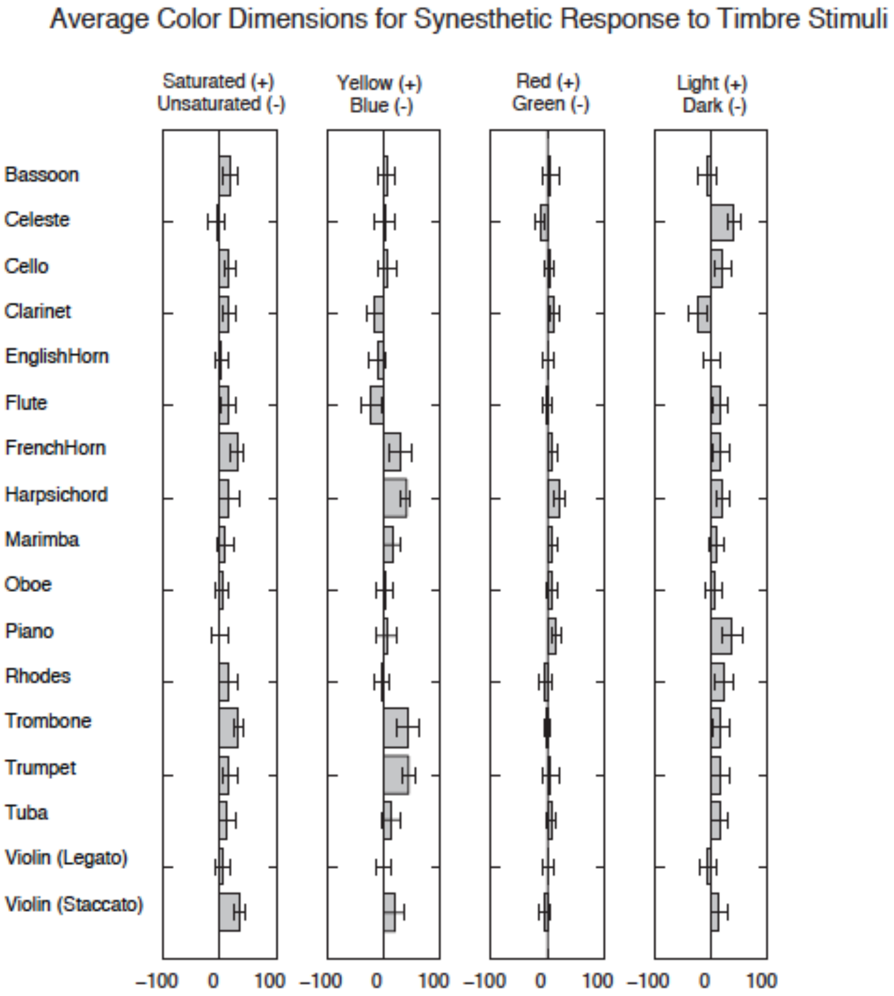

Figure 4.4: Average color appearance dimensions for the colors reported by synesthetes while listening to timbre stimuli.

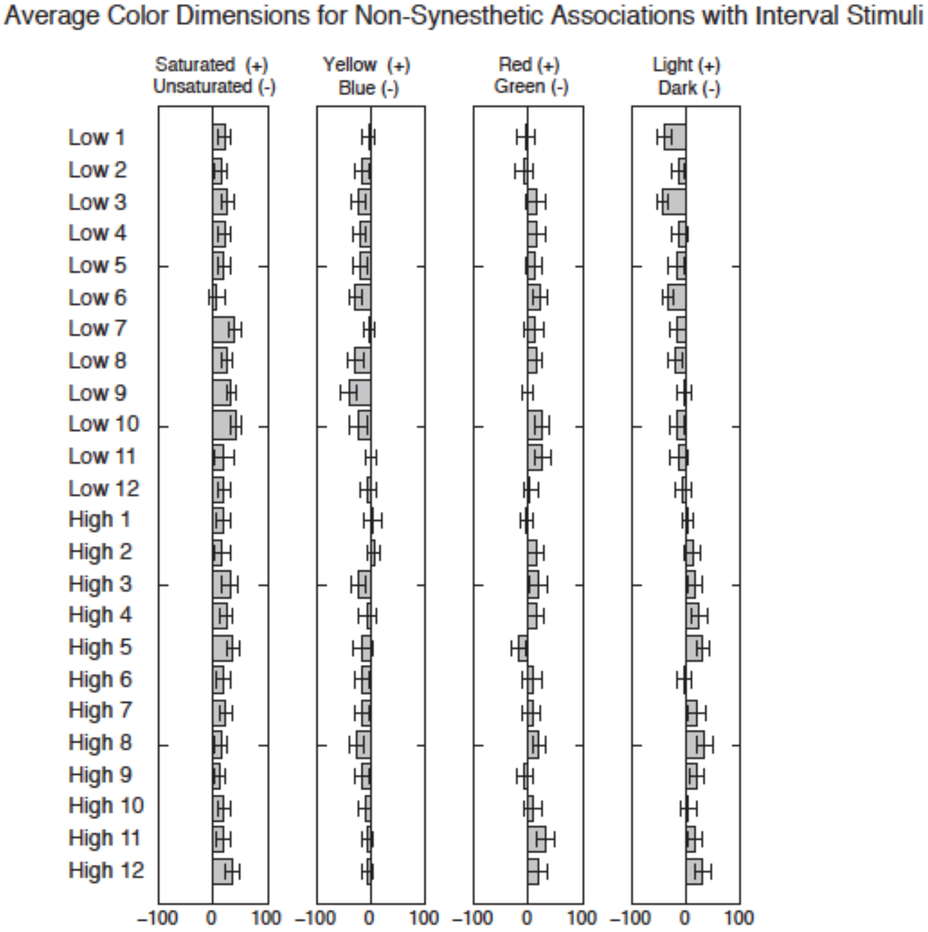

Figure 4.5: Average color appearance dimensions for the color associates chosen by non-synesthetic participants while listening to interval stimuli.

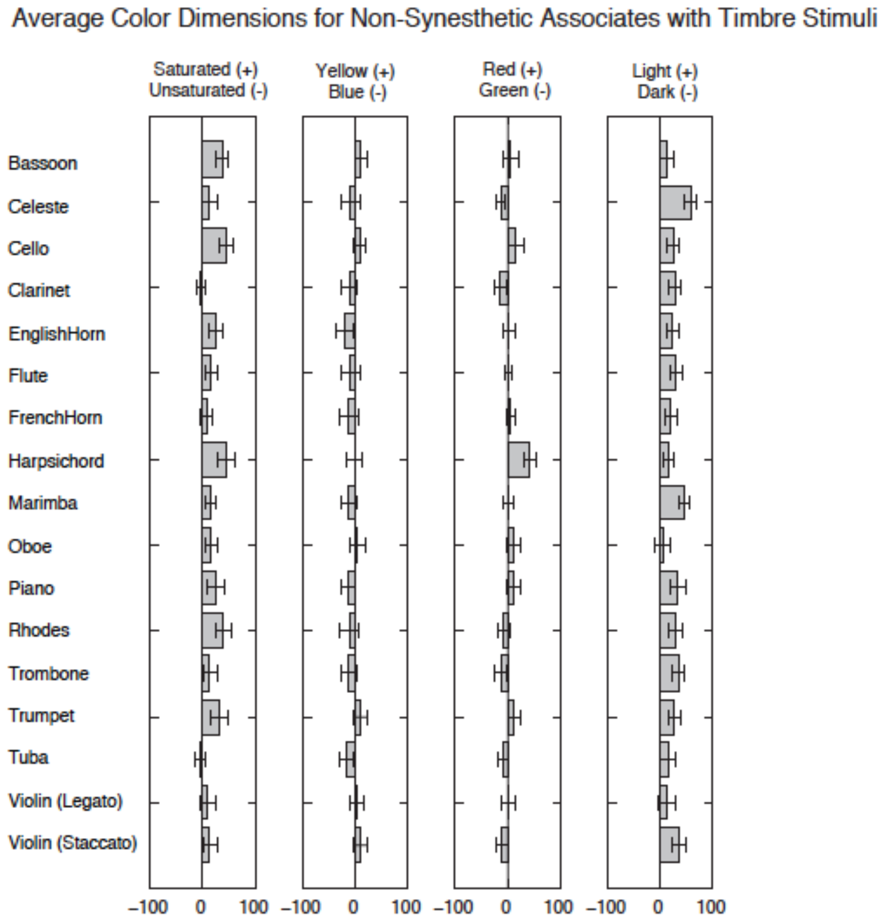

Figure 4.6: Average color appearance dimensions for the color associates chosen by non-synesthetic participants while listening to timbre stimuli.

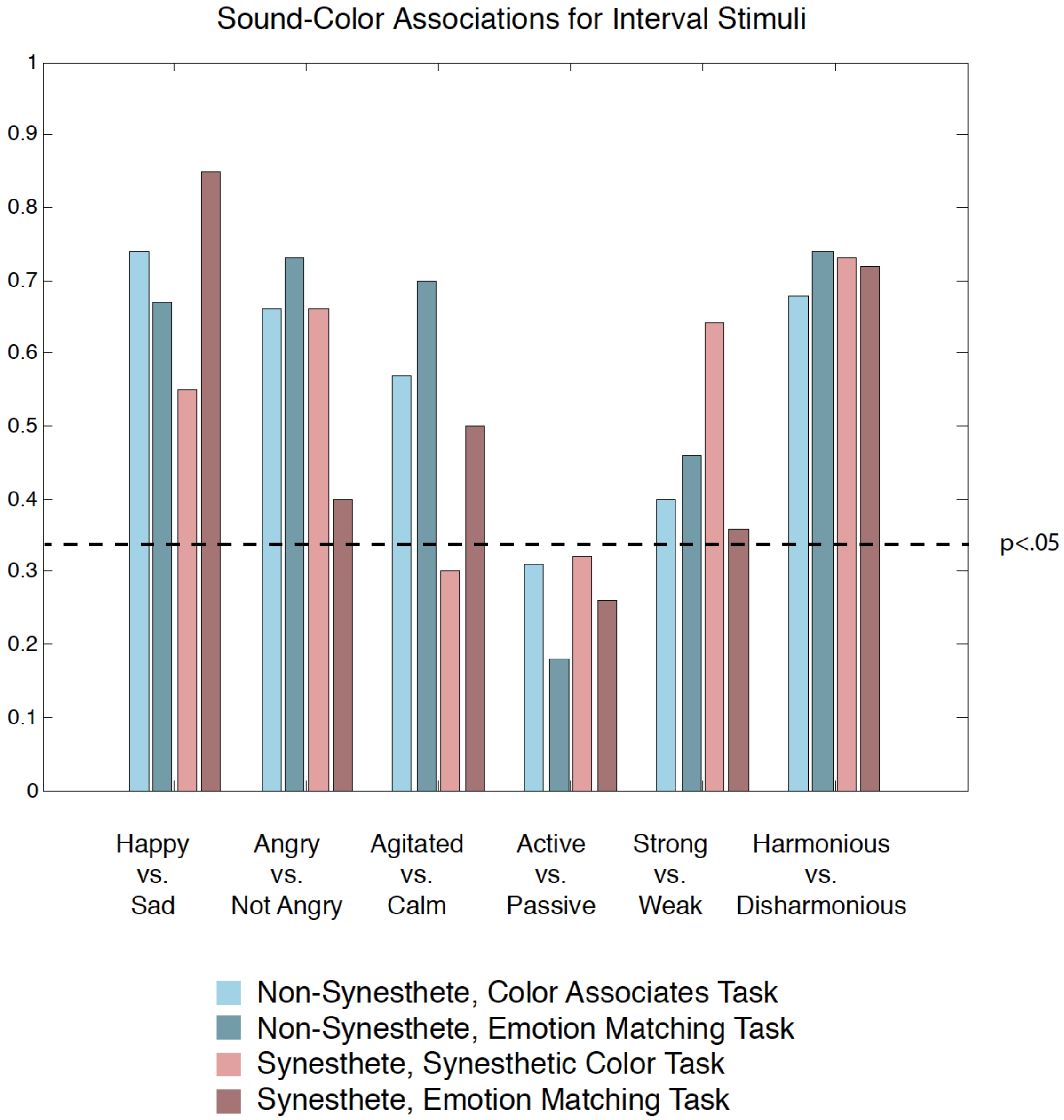

Figure 4.7: Correlation between SCA scores and rated emotion for interval stimuli. Synesthetes are coded as red and non-synesthetes as blue; the lighter colored bars represent the colors chosen in the synesthetic percept/color association task, and the darker colored bars show scores from the comparison condition in which we asked both sets of participants to explicitly match colors to sounds based on emotional meaning.

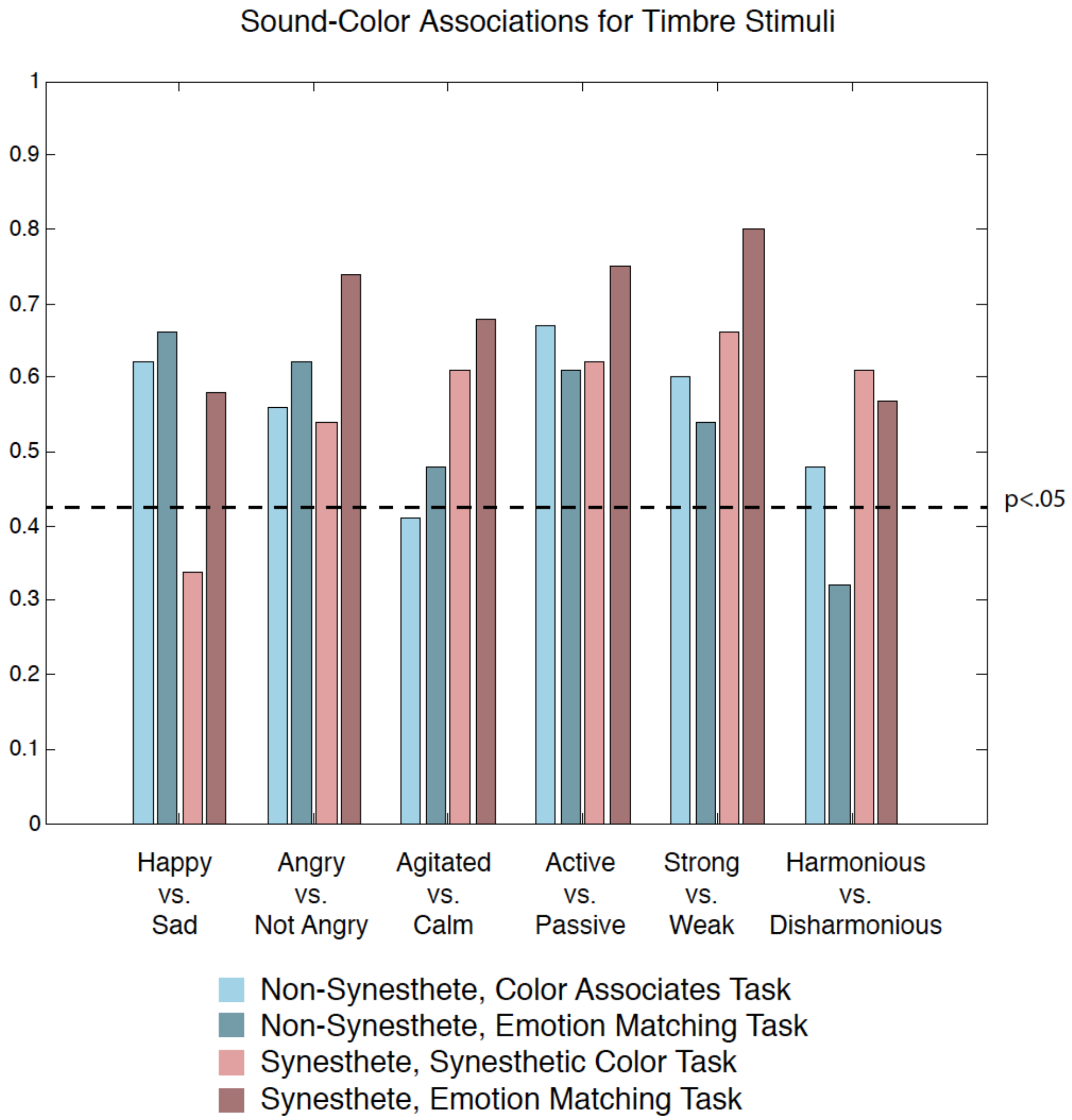

Figure 4.8: Correlations between average emotion ratings of audio stimuli and colors chosen to go with those colors in several tasks.

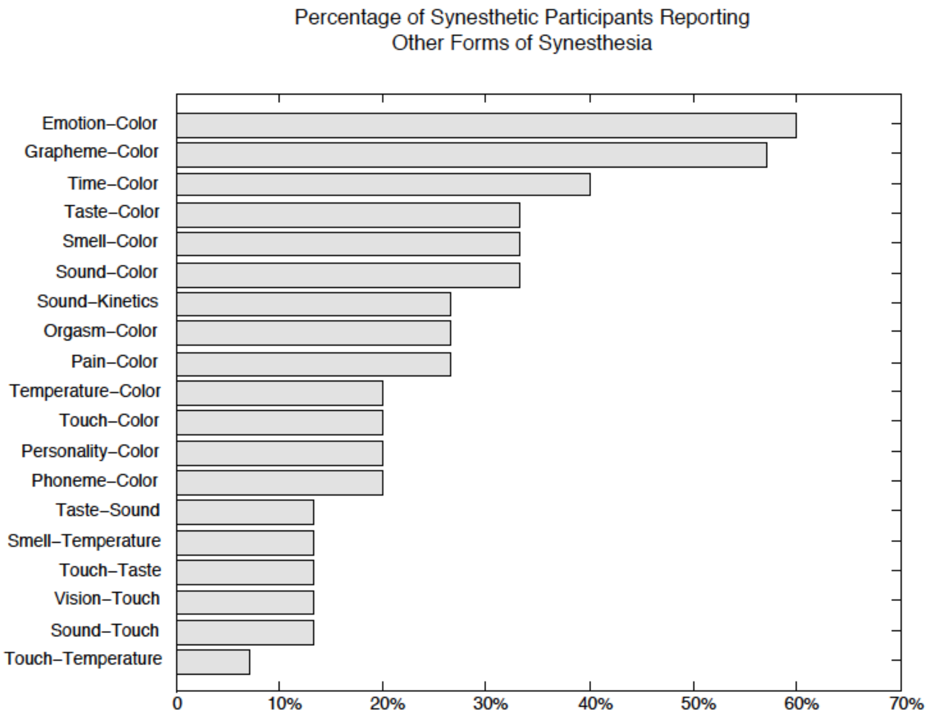

Figure 4.9: The percentage of our 15 synesthetic participants who reported experiencing other types of synesthesia in addition to music→color synesthesia, based on a self-report questionnaire.

Correlations between Synesthete and Control Group Stimulus Ratings

	Interval Emotion Ratings	Timbre Emotion Ratings	Color Emotion Ratings	Interval Color Associate Emotions	Timbre Color Associate Emotions
Happy **Sad**	0.93***	0.79***	0.91***	0.59**	0.56*
Angry **Not Angry**	0.83***	0.85***	0.91***	0.39	0.52*
Calm **Agitated**	0.87***	0.91***	0.95***	0.33	0.47
Active **Passive**	0.50*	0.94***	0.95***	0.27	0.30
Strong **Weak**	0.39	0.83***	0.93***	0.47*	0.56*
Harmonious **Disharmonious**	0.96***	0.82***	0.79***	0.67***	0.66**

Table 4.1: Comparison of emotion ratings from the synesthete and control groups for different stimulus categories. Columns 1-3 show correlations for average emotion ratings of interval, timbre, and color stimuli, where a high correlation indicates high inter-group agreement about the emotional meaning of the stimuli. Columns 4-5 show the correlations for the SCA emotion scores for interval and timbre stimuli, where a high correlation indicates that both groups chose colors with similar emotional meanings to go with the stimuli.

Chapter 5: Conclusion

In the introduction to this dissertation, I posed three questions that arise from previous research on multisensory perception and music:

1) Do untrained, non-synesthetic individuals make consistent cross-modal mappings from basic musical units of harmony (e.g., intervals and chords) to colors?

2) Do non-synesthetic individuals show consistent cross-modal mappings from instrumental timbres to colors?

3) What is the relationship between these various cross-modal mappings and the visual experiences evoked in synesthesia in comparison to corresponding cross-modal associations in non-synesthetes?

In the course of this research, we were able to provide some preliminary answers to each of these questions, as well as a host of others that arose in the process of conducting the experiments themselves. For example, we found that typical non-synesthetic volunteers have consistent cross-modal associations from both intervals and chords to colors, and that these effects are robust to different methods of testing, including two-alternative forced choice and free choice, as well as to sequential or simultaneous presentation of stimuli. We tested the emotional mediation hypothesis by calculating Sound-Color Association scores, that represent the average emotional quality of colors chosen for a particular sound, and found that these are strongly correlated with the corresponding emotional ratings of the sound itself.

We also found that participants have consistent cross-modal associations from individual instrument timbres to colors. These associations show evidence of being emotionally mediated, but cross-modal matching for other dimensions of visual stimuli, such as edge contrast and temporal dynamics, is less strongly mediated by emotion, and instead is better explained by matching between low-level sensory features such as attack time and spectral brightness. We also showed that, for color matching tasks, participants respond to multiple timbres in a way that is highly predictable based on their responses to individual timbres.

Finally, we looked at how these kinds of multisensory matching tasks relate to the perceptual experiences of individuals with music-to-color synesthesia. We found that synesthetes and matched non-synesthetic controls show a similar level of emotional effects in their color choices for intervals and timbres and that they actually both choose colors with similar emotional/semantic qualities for a given sound. This provides evidence for the idea that music-to-color synesthesia may not be merely sensory in nature, but actually reflects a much more widespread network of brain processes that interact with emotional and semantic processing (**reference others here?).

Of course, there are many questions that are still open, some going far beyond what is possible to cover in this dissertation. For example, we still do not know whether the emotional effects we see in synesthetes' color choices for musical intervals and timbres are "online" processes in which emotional experiences influence color experiences at the moment of hearing the sounds or whether they are fixed associations that were either innate or learned earlier in life, yet still affect synesthetic experiences in adulthood. Indeed, we do not know in general whether cross-modal associations between music and color are learned or innate, nor whether they vary across different musical and visual cultures. It also remains unclear what might be so special about music and color that makes their matches so strongly mediated by emotion, if indeed other sensory matches are not. And, of course, despite the large amount of research conducted over the past several decades, synesthesia still remains largely a mystery. Can one neural explanation really account for both unisensory and multisensory synesthesias, or do we need to start treating them as separate phenomena? And what is the underlying source of the relationship between emotion and music in synesthesia? These are all projects for future research. Happily, there is no end to interesting questions in the study of multisensory perception.

References

Asher, J. E., Aitken, M. R., Farooqi, N., Kurmani, S., & Baron-Cohen, S. (2006). Diagnosing and phenotyping visual synaesthesia: a preliminary evaluation of the revised test of genuineness (TOG-R). *Cortex*, *42*(2), 137-146.

Baron-Cohen, S., Harrison, J., Goldstein, L. H., & Wyke, M. (1993). Coloured speech perception: Is synaesthesia what happens when modularity breaks down? *Perception*, *22*, 419-419.

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology 87*(2), 241.

Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics*, *6*(6), 337–339.

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, *126*(2), 165–172. doi:10.1016/j.cognition.2012.09.007

Costa, M., Bitti, P., & Bonfiglioli, L. (2000). Psychological connotations of harmonic musical intervals. *Psychology of Music*, *28*(1), 4–22.

Day, S. (2005). Some demographic and socio-cultural aspects of synesthesia. In L.C. Robertson and N. Sagiv (Eds.) *Synesthesia: Perspectives from Cognitive Neuroscience*. Oxford: Oxford University Press.

Driver, J., & Spence, C. (1998). Crossmodal attention. *Current Opinion in Neurobiology*, *8*(2), 245–253.

Cytowic, CW, and Eagleman, D.M. (2009). *Wednesday is Indigo Blue*. Cambridge: MIT Press.

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods*, *159*(1), 139–145. doi:10.1016/j.jneumeth.2006.07.012

Elliott, L. L. (1971). Backward and forward masking. *International Journal of Audiology*, *10*(2), 65-76.

Elliott, T. M., Hamilton, L. S., & Theunissen, F. E. (2013). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *The Journal of the Acoustical Society of America*, *133*, 389.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6–6. doi:10.1167/10.1.6

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., Gretz, A. L., ... & Turpin, B. M. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 355.

Grey, J. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, *61*(5), 1270–1277.

Grey, J. M. and Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, 1493-1500.

Grossenbacher, P. G., & Lovelace, C. T. (2001). Mechanisms of synesthesia: cognitive and physiological constraints. *Trends in Cognitive Sciences*, *5*(1), 36–41.

Helmholtz, H. V. (1885). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Aberdeen University Press.

Hubbard, E. M., Brang, D., & Ramachandran, V. S. (2011). The cross-activation theory at 10. *Journal of Neuropsychology*, *5*(2), 152–177. doi:10.1111/j.1748-6653.2011.02014.x

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology*, 219–238.

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *The Journal of the Acoustical Society of America*, *94*, 2595.

Karwoski, T. F., Odbert, H. S., & Osgood, C. E. (1942). Studies in synesthetic thinking: II. The role of form in visual responses to music. *The Journal of General Psychology*, *26*(2), 199–222.

Kohler, S. (1969). *Gestalt Psychology*. Princeton: Princeton University Press.

Krumhansl, C. L. (1979) Quantification of tonal hierarchy within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, 579-594.

Krumhansl, C. L. (1989). Why_Is_Musical_Timbre_so hard_to_understand, 1–11.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. *Developmental Psychology*, *16*(6), 597.

Maher, T. F. (1980). A rigorous test of the proposition that musical intervals have different psychological effects. *The American Journal of Psychology*, 309–327.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology*, 173–188.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384.

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and Weak. *Current Directions in Psychological Science*, *10*(2), 61–65. doi:10.1111/1467-8721.00116

Maurer, D., Pathman, T., and Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316–322.

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192.

McAdams, S. (2006) A meta-analysis of acoustic correlates of timbre dimensions. *Journal of the Acoustical Society of America*, 3275-3276.

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (2010) Individual differences reveal the basis of consonance. *Current Biology*, 1035-1041.

Meyer, L. (1956) *Music and Emotion.* Chicago: Chicago University Press.

Oelmann, H., & Laeng, B. (2009). The emotional meaning of harmonic intervals. *Cognitive Processing*, *10*(2), 113–131.

Osgood, C.E. (1957) *The Measurement of Meaning.* Urbana: University of Illinois Press.

Osgood, C. E. (1981). The cognitive dynamics of synesthesia and metaphor. *Review of Research in Visual Arts Education*, 56–80.

Ou, L.-C., Luo, M. R., Woodcock, A. E., & Wright, A. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *COLOR Research and Application*, *29*(3), 232–240. doi:10.1002/col.20010

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, *107*(19), 8877.

Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences*, *110*(22), 8836–8841. doi:10.1073/pnas.1212562110/-/DCSupplemental/pnas.201212562SI.pdf

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.

Rouw, R., & Scholte, H. S. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nature Neuroscience*, *10*(6), 792–797. doi:10.1038/nn1906

Schloss, K. B., & Palmer, S. E. (2010). Aesthetic response to color combinations: preference, harmony, and similarity. *Attention, Perception, & Psychophysics*, *73*(2), 551–571. doi:10.3758/s13414-010-0027-0

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27(2), 125-140.

Sherman, A., Grabowecky, M., & Suzuki, S. (2013). Auditory rhythms are systemically associated with spatial-frequency and density information in visual scenes. *Psychonomic Bulletin & Review*, *20*(4), 740–746. doi:10.3758/s13423-013-0399-y

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. doi:10.3758/s13414-010-0073-7

Spence, C. and Driver, J. (2004) *Crossmodal space and crossmodal attention.* Oxford: Oxford University Press.

Stevens, J. C., & Marks, L. E. (1965). Cross-Modality Matching of Brightness and Loudness. *Proceedings of the National Academy of Sciences of the United States of America*, *54*(2), 407–411.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, 17(4), 401-419.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-Colour Synaesthesia: to What Extent Does it Use Cross-Modal Mechanisms Common to us All? *Cortex*, *42*(2), 264–280. doi:10.1016/S0010-9452(08)70352-6

Wessel, D. (1979). Timbre space as a musical control structure. *Computer Music Journal*, *3*(2), 45–52.

Wicker, F. W. (1968). Mapping the intersensory regions of perceptual space. *The American Journal of Psychology*, *81*(2), 178–188.

Zurek, P. M. (1979). Measurements of Binaural Echo Suppression. *The Journal of the Acoustical Society of America*, 1750-1757.

Appendix:
Synesthesia Project – Color and Music Background Questionnaire

A) Music Experience:

1. How much experience do you have with singing or playing musical instruments? (Rate from 1-5)

1	**2**	**3**	**4**	**5**
None		High school band/chorus		Professional musician

2. How much experience do you have with music theory and composition? (Rate from 1-5)

1	**2**	**3**	**4**	**5**
None		Music Theory 101		Pro composer / theorist

3. Have you taken classes in music appreciation or music history? (Rate from 1-5)

1	**2**	**3**	**4**	**5**
No		A few		It is my primary academic focus

4. If you play a musical instrument:

Instrument	**Age Started**	**# of Years Playing**	**Type of Training (circle)**	**Proficiency** (1 = Not at All, 5 = Very Much)	**Currently playing?**
			Private setting Group setting Self-taught	1 2 3 4 5	Yes No
			Private setting Group setting Self-taught	1 2 3 4 5	Yes No
			Private setting Group setting Self-taught	1 2 3 4 5	Yes No
			Private setting Group setting Self-taught	1 2 3 4 5	Yes No

5. On average, how many hours per day do you listen to music? ____________

6. What type(s) of music do you usually listen to? (check all that apply)

Genres			
Bach	Country	Progressive House	Soundtrack
Mozart	Classic Rock	Trance	Arabic
Stravinsky	Progressive Rock	Funk	Balkan Folk
Big Band	Alternative	Reggae	Gamelan
Dixieland	Psychobilly	Ska	Hindustani Sitar
Jazz	Heavy Metal	Folk	Irish
Smooth Jazz	Dubstep	Hip Hop	Salsa
Blues	Electronic	Indie	Other: ________
Bluegrass	Eighties Pop	Piano	________________

7. In the above table, circle your favorite types of music.

8. To the best of your knowledge:

a.	Are you tone deaf?	*Yes*	*No*
b.	Do you have absolute or perfect pitch?	*Yes*	*No*
c.	Can you identify a note relative to a reference note (i.e. relative pitch)?	*Yes*	*No*

B) Color Experience Questionnaire:

9. How many hours per week do you spend considering aspects of color:

In your job or at school: ___________________________________
In picking out clothes to wear (or buy) for yourself: ___________
In considering the clothes other people wear (or buy): __________
In creative outlets (painting, photography, etc.): ______________
In other activities: ______________ If more than zero, please give examples here:____

10. Compared to what you consider to be an average person:

a. How **aware** do you consider yourself to be about colors?

1	**2**	**3**	**4**	**5**
Not at all				Very Much

b. How **sophisticated** do you consider yourself to be about colors?

1	**2**	**3**	**4**	**5**
Not at all				Very Much

c. How much **formal training** have you had in color?

1	**2**	**3**	**4**	**5**
None		High School Art		Very Much

11. If you have formal training in visual arts (painting, photography, etc.):

Type	**Age Started**	**# of Years**	**Type of Training (circle)**	**Proficiency** **(1 = Not at All, 5 = Very Much)**	**Currently practicing?**
			Private setting Group setting Self-taught	1 2 3 4 5	Yes No
			Private setting Group setting Self-taught	1 2 3 4 5	Yes No
			Private setting Group setting Self-taught	1 2 3 4 5	Yes No
			Private setting Group setting Self-taught	1 2 3 4 5	Yes No

C) Synesthetic Experience:

12. What percentage of the time (0% to 100%) do you experience color while you are listening to music, on average?

Percentage: _______ %

13. Does this percentage vary for different types of music?
Yes *No*

If so, how? ____

14. How do you feel about your synesthesia? Do you like it? Or do you dislike it?

1	**2**	**3**	**4**	**5**
I dislike it strongly	I dislike it	I am indifferent about my synesthesia	I like it	I like it strongly

15. If you are listening to music purely for pleasure, do you try to enhance or suppress your synesthesia?

1	**2**	**3**	**4**	**5**
Suppress		Neither		Enhance

16. If there's music in the environment while you're engaged in some other activity (e.g., grocery shopping), do you try to enhance or suppress your synesthesia?

1	**2**	**3**	**4**	**5**
Suppress		Neither		Enhance

17. Are there times when you respond differently to your synesthesia?
Yes *No*

If yes, when do you respond differently? ________________________________

18. Do you have control of your synesthesia? Can you change the colors you experience?
Yes *No*

If so, how? ____

19.) Have you always had synesthesia?
Yes *No*

a. If not, at what age did you develop synesthesia? _______________
b. Did any specific event (e.g. a concussion) trigger your synesthesia?
Yes *No*
c. If yes, explain: ____

20. To the best of your knowledge, do you have other types of synesthesia?
Yes *No*

If so, please check all that apply:

Types of Synesthesia			
Graphemes/Colors	Sound/Touch	Touch/Tastes	Musical notes/Tastes
Time units/Colors	Temperatures/Colors	Smells/Sounds	Personalities/Touch
General sounds/Colors	Vision/Tastes	Sounds/Kinetics	Smells/Tastes
Phonemes/Colors	Sounds/Smells	Sound/Temperatures	Smells/Temperatures
Smells/Colors	Vision/Sounds	Tastes/Touch	Tastes/Sounds
Tastes/Colors	Orgasm/Colors	Kinetics/Sounds	Tastes/Temperatures
Sound/Tastes	Emotions/Colors	Personalities/Smells	Temperatures/Sounds
Pain/Colors	Vision/Smells	Touch/Sounds	Touch/Temperatures
Personalities/Colors	Vision/Touch	Touch/Smell	Other: ______________
Touch/Colors	Smells/Touch	Vision/Temperatures	___________________ ______

Note to experimenter: Play “synesthesia visualizations.m3u” on white mac with **no sound**.

21.) How similar is this visualization to your experience of synesthesia? (Rate from 1 - 5)

1	**2**	**3**	**4**	**5**
Not at All				Very Much

a. *If so, how is it similar? If not, how is it different?*________________________

____

Note to experimenter: Play “synesthesia visualizations.m3u” again on white mac **with sound**. Please use the headphones from Booth C. Allow participant to adjust the volume so that it is at a comfortable listening level.

22.) How similar is this visualization to your synesthetic response to this particular piece of music? (Rate from 1-5)

1	**2**	**3**	**4**	**5**
Not at All				Very Much

a. *If so, how is it similar? If not, how is it different?*________________________

____