Audiovisual Interactions: the Impact of Visual Information on Music Perception and Memory

Author(s): Marilyn G. Boltz, Brittany Ebendorf and Benjamin Field

Source: *Music Perception: An Interdisciplinary Journal*, Vol. 27, No. 1 (September 2009), pp. 43-59

Published by: University of California Press

Stable URL: https://www.jstor.org/stable/10.1525/mp.2009.27.1.43

Audiovisual Interactions: The Impact of Visual Information on Music Perception and Memory

Marilyn G. Boltz, Brittany Ebendorf,
and Benjamin Field
Haverford College

Previous research has demonstrated that musical soundtracks can influence the emotional impact, interpretation, and remembering of visual information. The present research examines the reverse relationship and whether visual information influences the perception and memory of music. In Experiment 1, listeners were presented with affectively ambiguous tunes paired with visual displays varying in their affect (positive, negative) and format (video, montage), or a control condition (no visual information at all). After each, participants were asked to provide a set of perceptual ratings that evaluated different melody characteristics and qualities of the visual displays. Results showed that both the affect and format of visual information differentially influenced the way a melody was perceived. Experiment 2 extended these findings by revealing that the affect of visual displays distorted melody recognition in a mood congruent fashion. These results are discussed in terms of their theoretical implications for audiovisual processing.

Received September 10, 2008, accepted February 27, 2009.

Key words: audiovisual, music videos, musical soundtracks, video information, music/film

Music is an integral aspect of daily life and often is encountered within a visual context such as film, advertising, or an artistic performance. During the past several decades, a substantial amount of research has revealed various ways in which music can influence both the perception and memory of visual scenes (see Cohen, 2001, for a review). However, surprisingly few studies have examined the reverse relationship and ways in which visual information may influence the cognitive processing of music. The present research addresses this issue through the use of music videos and considers whether the mood and format of videos influence how a musical composition is heard and remembered.

The Influence of Music on the Film Experience

As noted by several music and psychological theorists (e.g., Juslin & Sloboda, 2001; Meyer, 1956), the meaning of music, to a large extent, arises from the moods and emotions it evokes within listeners. A number of studies have examined people's subjective reports, adjective ratings, physiological responses, or nonverbal expressions as they listen to various musical compositions and found a high degree of agreement on what particular emotions are elicited by a given musical selection (see Sloboda & Juslin, 2001, for a review). In addition, particular emotions are reliably associated with particular acoustical characteristics that generalize across different compositions. For example, sad melodies tend to be characterized by a slow tempo, lower amplitude, minor mode, and a narrow pitch range in the lower octaves. As one might expect, happy melodies tend to display a reverse set of structural characteristics (i.e., faster tempo, higher amplitude, wider pitch range, etc.; Gabrielsson & Lindstrom, 2001; Hevner, 1935a, 1935b; Juslin, 1997; Scherer & Oshinsky, 1977). Some individuals have described this relationship between emotional expression and musical structure as an iconic one in which the particular attributes corresponding to a given emotion bear a resemblance to those found in nonmusical entities. Given that the structural correlates of different emotions are remarkably invariant across a wide variety of everyday events, including those of music, the human voice (Cosmides, 1983; Scherer, 1979; Williams & Stevens, 1979), and walking gaits (Montepare, Goldstein, & Clauser, 1987; Sloman, Berridge, Homatidis, Hunter, & Duck, 1982), there is, in fact, ample opportunity to learn these types of associations. One illustration of emotional iconics comes from a study by Cohen (1993). Participants were asked to provide adjective ratings in response to musical patterns that varied in pitch height and tempo, as well as scenes of a bouncing ball that also varied in height and

Music Perception volume 27, issue 1, pp. 43–59, issn 0730-7829, electronic issn 1533-8312 DOI:10.1525/mp.2009.27.1.43

tempo. When each stimulus was judged independently, increases or decreases along both physical dimensions led to the same emotional judgment (happiness and sadness, respectively) and these ratings were either enhanced or attenuated when the two stimuli were paired in a congruent or incongruent fashion.

The concept of an iconic relationship becomes even more apparent in cinematic art in which musical soundtracks commonly appear within the background of the ongoing action. By using music that evokes a mood similar to the visual story, a film director can heighten the emotional impact of a particular scene and its effects upon listeners. Consider an example used in one study in which the aggressive or playful behavior of wolves (within a documentary film) was accompanied by music that had been independently described by listeners as aggressive vs. playful, respectively. Relative to the same scenes viewed in the absence of music, mood-congruent soundtracks increased the magnitude of the perceived aggressiveness and playfulness exhibited by these animals (Bolivar, Cohen, & Fentress, 1994). Mood-incongruent music not only attenuates the emotional impact of a scene but often is used to convey a more subtle meaning through "ironic contrast." For many, scenes from *A Clockwork Orange* readily will come to mind, in which the acts of murder, rape, and mayhem acquire a surreal meaning from the carefree beauty of Rossini's *The Thieving Magpie.*

In addition to influencing the emotional impact of scenes, musical soundtracks can be used to achieve other intended effects. One such function is to provide an interpretative framework for story comprehension. Especially in scenes that are relatively ambiguous, the emotive meaning of music often is relied upon to clarify and promote certain inferences about a character's actions and motivations (Boltz, 2001; Bullerjahn & Guldenring, 1994; Liu, 1976; McFarland, 1984; Parrott, 1982). These ideas are illustrated in a study by Boltz (2001) in which participants were presented with three ambiguous film clips paired with positive, negative, or no music. After viewing each, one group of participants was asked to extrapolate the film's ending and evaluate the personality, actions, and motivations of the main characters. A second group was asked to return a week later for a surprise recognition memory test of various objects within each film. The results showed that relative to the no-music condition, positive and negative music biased viewers' evaluations and subsequent remembering in a mood-congruent fashion. These findings, then, indicate that the emotive meaning of music provided an interpretative framework that influenced viewers' inferences and story comprehension. Music also guided attending toward certain aspects of a visual scene that were consistent with this interpretative framework and thereby influenced which particular items were later remembered. Related research has revealed that music's attentional highlighting of information leads to a more integrated memory code that enhances the ability to later recall a film (Boltz, 2003; Boltz, Schulkind, & Kantra, 1991).

The Influence of Visual Information on the Processing of Music

In the early 1980's, MTV heightened the popularity of a relatively new art genre, the music video, which immediately generated a debate about its predicted effectiveness. Some argued that an accompanying video would decrease the appeal of a song by dictating a particular interpretation that limits a listener's imagination (e.g., Zorn, 1984). Others claimed that videos would enhance the aesthetic appeal of music by clarifying a song's intended meaning (e.g., Sun & Lull, 1986). Although this debate is unresolved, there is evidence indicating that videos can help to prevent "wear out"—becoming tired of a song due to excessive exposure. Goldberg, Chattopadhyay, Gorn, & Rosenblatt (1993) investigated this phenomenon and found that videos are most likely to maintain listeners' interest when songs are relatively ambiguous and open to multiple interpretations.

More generally, the concept of music videos leads one to ask, in what ways might visual information influence music perception and appreciation? A review of the current literature reveals that surprisingly few studies have addressed this question and, of these, most involve music education. A primary issue of interest is whether the evaluation of musical performances varies in the presence of audiovisual vs. audio alone presentations. In general, the results have revealed that the overall quality of musicianship is judged more highly when evaluators are able to both see and hear a performance; an effect that applies to vocalists (Cassidy & Sims, 2006; Wapnick, Darrow, Kovacs, & Dalrymple, 1997), violinists (Gillespie, 1997; Wapnick, Mazza, & Darrow, 1998), and pianists (Ryan & Costa-Gioni, 2004; Wapnick, Ryan, Lacaille, & Darrow, 2004), and especially when these performers are highly experienced in their abilities. Several individuals have noted that the visual modality provides access to body movements that may convey musical intentions more clearly than music alone. In support, music expressiveness (Davidson, 1994), vibrato (Gillespie, 1997), phrasing (Marolt, 2006; Vines, Krumhansl, Wanderley, & Levitin, 2006) note duration (Schutz & Lipscomb, 2007), and the size of sung intervals (Thompson & Russo, 2007) are all better discerned in audiovisual vs. audio alone presentations.

There also is evidence that the presence of visual information influences the emotional experience of music listening. One strategy, adopted by Vines et al. (2006), asked participants to provide continuous judgments of emotional tension on a sliding potentiometer throughout a musical selection. Relative to an audio alone condition, more extreme judgments occurred in the audiovisual presentations in which visual information served to both enhance and reduce perceived tension at varying locations within a musical piece. A somewhat different approach was taken by Geringer, Cassidy, & Byo (1996, 1997), who assessed affective reactions to music that were accompanied by videos of cinematic scenes in lieu of the performing artists. In particular, listeners were presented with certain musical selections either heard alone or as an accompaniment to excerpts from the Walt Disney film, *Fantasia*. Immediately after each, participants described their emotional experience and those musical qualities that most held their attention. Results showed that relative to the music alone condition, the audiovisual format evoked greater emotional involvement that primarily was attributed to a composition's tempo, instrumentation, and dynamics. Although these effects were observed only for some songs and not others, they provide further evidence that musical appreciation appears to be enhanced by visual information.

The Current Study

The use of cinematic videos, as illustrated in the research by Geringer et al. (1996, 1997), provides the opportunity to examine how iconic relationships between music and visual information may be formed within listeners/viewers to influence cognitive behavior. One pathway of this iconic dyad has been well investigated. As noted earlier, music has been demonstrated to exert a number of effects on the movie-viewing experience by generating an affective framework from which to interpret characters' actions and motivations, and direct attending toward some aspects of a scene and away from others to influence remembering (e.g., Boltz, 2001). What is less clear is whether the reverse relationship applies wherein the affective meaning of visual information influences music cognition. Geringer et al. (1996, 1997) have shown that visuals enhance listeners' emotional engagement with a musical selection and direct attending toward certain musical qualities such as dynamics and instrumentation. But the question addressed here is whether visual displays extend their influence by altering the perceived acoustical parameters that typically convey emotions within music. For example, suppose that a musical composition is accompanied by a visual narrative that is positive in its affect. Relative to an audio-alone format, is this tune perceived to have a faster tempo, higher pitch, and greater amplitude—qualities that all convey positive emotions within the auditory modality (Gabrielsson & Lindstrom, 2001; Scherer, 1979)? Conversely, is this same tune perceived to be slower, softer, and lower in pitch when accompanied by a negative visual narrative? Experiment 1 was designed to test this hypothesis through a series of perceptual rating tasks, whereas Experiment 2 relied on a recognition memory paradigm as a converging operation.

The methodological strategy of both studies was similar to that of Boltz (2001). However, in lieu of pairing ambiguous film excerpts with tunes of a given affect, the reverse was done such that visual scenes pre-rated as positive or negative in their affect were paired with "ambiguous" tunes. An ambiguous tune was defined as one whose valence and activity level—the two dimensions that best differentiate different emotions (e.g., Russell, 1980)—had been prejudged as intermediate in their values. In addition to visual affect, the structural format of the accompanying visual scenes also was manipulated to reflect either a montage (i.e., slideshow) of still photos vs. a video of thematic scenes that smoothly transitioned from one to another. Although the particular affect of each should exert similar effects, visual format differentially may influence the perception of certain acoustical characteristics. The perceived rhythm and flow of a musical selection, in particular, are likely to differ in that the montage (vs. video) format is inherently more "choppy" due to the arrangement of discrete scenes.

During the actual experiment, participants were presented with five unfamiliar tunes, four of which were accompanied by visual scenes that varied in their affect and format. As a control condition, the remaining tune was presented in the absence of visual information. Immediately after each, all participants were asked to complete a set of perceptual ratings. The perceived musical qualities were of most interest and assessed those acoustical parameters that, based on the past literature, were likely to be sensitive to affective manipulations of the visual displays. In addition, ratings of the visual display and the stimulus as a whole were included to mask the main motivation of the experiment by directing attention to all aspects of the stimuli. These latter ratings have the added benefit of allowing one to assess the construct validity of the experimental stimuli.

Experiment 1

Method

DESIGN AND PARTICIPANTS

The design was a 2 × 2 repeated measures factorial with a "hanging" control group: all participants listened to a set of tunes in which the accompanying visual information varied in its overall affect (positive, negative) and format (video, montage of still photos). In addition, all participants listened to a control melody that was presented in isolation with no accompanying visual display.

Sixty participants from an introductory psychology course at Haverford College participated in the experiment for course credit. The majority of these were nonmusicians: although a small percentage (13%) were members of the campus choir, none were currently playing or had played a musical instrument within the past three years.

Stimulus Materials

A set of 28 musical selections was selected as potential stimuli. All were 90 s in duration, of equal amplitude, from the new age genre, and performed by a single instrument with no accompanying lyrics. Each tune was also selected for its neutral or ambiguous affect. Relying on those structural parameters identified in the past literature (Gabrielsson & Lindstrom, 2001; Levi, 1982; Scherer, 1979), each tune was based on a major mode and pitch values from the mid-range of the performing instrument, spanning no more than two octaves. In addition, each displayed a relatively predictable rhythm and a moderate tempo with an average beat duration of 0.7 beats per s. To assess the construct validity of these tunes, a pretest was conducted on a group of 10 independent participants who were asked to rate each melody, on 9-point Likert scales, for its degree of familiarity (1 = "very familiar"), affect (1 = "very positive"), and activity (1 = "very active"). Those five tunes consistently rated as highly unfamiliar (M = 7.75, SD = 1.75), of moderate activity (M = 5.00, SD = 2.25), and neutral in their overall affect (M = 5.20, SD = 2.01) were selected as experimental stimuli: *Mummer's Dance* (McKennitt, 1997), *Tiergarten* (Tangerine Dream, 1985), *The Chamber* (Oldfield, 1996), *Industrial Revolution Overture* (Jarre, 1988), and *Traverser le Temps* (Benza, 2004).

The visual materials were of two types, namely, a set of montages containing a series of still photos and a set of video clips. For the former, scenes were taken from the International Affective Picture System (Lang, Bradley, & Cutbert, 1997), which contains over 700 photos that have been pre-rated, on 9-point Likert scales, for their overall affect (1 = "very negative") and arousal (1 = "very calming"). Using the criteria that each photo be intermediate in arousal (6.5 < M > 3.5) but either positive (M > 5.5) or negative (M < 4.5) in affect, a total of 480 photos were selected as experimental stimuli. Examples of positive scenes included kittens, puppies, flowers, appetizing food, and smiling people, whereas those for negative scenes included garbage, injuries, the aftermath of natural disasters, and human suffering. Scenes of human injuries and/or death judged to be extremely disturbing were excluded. Relying on Microsoft Office Powerpoint software, the sets of positive and negative scenes were each arranged into two different slideshows to yield a total of four montages. Each contained a total of 130 photos of the same affect which meant that approximately 20 negative and 20 positive scenes appeared in both instances of their respective slideshows. The speed at which the montage progressed was always held constant at 0.7 s per photo in order to match the average beat duration of all melodies.

For the video clips, a total of fifteen, 90 s excerpts were selected from various films that contained no dialogue. These included *Baraka* (Fricke, 1992), *Chronos* (Fricke, 1985), *Koyaanisqatsi* (Reggio, 1982), *Powaqqatsi* (Reggio, 1988), and certain Bollywood productions. Although all excerpts contained an accompanying musical soundtrack within their original film, each was extracted with a muted audio. A pretest then was conducted on 10 independent participants who rated each excerpt, on 9-point Likert scales, for its degree of familiarity (1 = "very familiar"), overall affect (1 = "very positive"), and arousal (1 = "very arousing"). From these data, four videos were selected as experimental stimuli; all had been rated highly unfamiliar (M = 8.00, SD = 0.87) and moderately arousing (M = 5.20, SD = 2.50) but either strongly positive (M = 2.10, SD = 1.30) or negative (M = 8.40, SD = 1.75) in affect. One of the positive videos, from *Baraka* (Fricke, 1992) contained a nature scene of waterfalls, clouds, and tranquil bodies of water, whereas the other, *Darr: A Violent Love Story* (Chopra, 1993) depicted a romantic escapade between a young man and woman. The two negative videos were both taken from *Baraka* (Fricke, 1992). One showed a processing plant in which newborn chicks were being branded with a hot iron and in the other, scenes of homeless people in an urban setting. A more detailed description of each video clip is presented in Appendix A.

All stimuli were the same length (i.e., 90 s) that thereby allowed each melody to be paired with each of the eight visual displays. A total of 10 different presentation sets were created. Each included all five tunes, one paired with a negative montage (A or B), a positive montage (A or B), a negative video (A or B), a positive video (A or B), and no visual information (as a control condition). Across the 10 sets, the five tunes appeared in a different temporal order and were systematically paired with each of the different visual displays (or lack thereof) such that all stimuli occurred an equal number of times.

Apparatus

The Audacity software system was used to extract the tune excerpts from their original melodies and edit them such that each was 90 s in duration and ended at a phrase boundary marked by the tonic interval. Each tune was then inputted into the Sibelius software system, which displays a song's musical notation over time and thereby allows one to determine the average beat duration of each musical selection. Video editing and the synchronization of melodies with each of the eight visual displays were accomplished through iMovie. The five stimuli within each of the ten presentation sets were arranged into their respective temporal order and then burned onto a DVD. During the actual experiment, a set of stimuli was presented at a constant volume through Sony MDR-CD180 headphones on a set of four Dell Dimension 4500 computers containing an internal DVD player.

Procedure

Participants were tested in groups of two to four individuals and received a response booklet and written instructions for the experimental procedure. They were told they would be presented with a set of five audiovisual events and immediately after each, asked to complete four sets of judgments on a series of 7-point scales. For all participants, these sets of judgments were arranged in a consistent order to minimize any carryover effects, and to systematically direct attending from a holistic to more analytical levels. In rating Set I, participants were instructed to evaluate the stimulus as a whole for its overall *affect* (1 = "very positive" vs. "very negative"), *activity* (1 = "very active" vs. "very passive"), and *likeability* (1 = "strongly like" vs. "strongly dislike"). Next, they were given an array of 28 randomly arranged adjectives and told to circle all those that described the stimulus they had just encountered. These adjectives were adopted from Russell's (1980) Circumplex Model of Emotion which depicts a two-dimensional array of adjectives defined by both an activity (active, passive) and valence (positive, negative) axis. The seven adjectives from each of the four quadrants in Russell's model were included in the present study (see Appendix B). In rating Set III, participants were told to focus on the tune alone and evaluate its overall *affect* (1 = "very positive"), *activity* (1 = "very active"), *tempo* or "overall speed" (1 = "very fast" vs. "very slow"), *tonality* or "the extent to which the melody's notes seemed to belong or not belong to the piece" (1 = "very tonal" vs. "very atonal"), *rhythm* or "the extent to which the timing of notes was regular and predictable" (1 = "very regular" vs. "very irregular"), *loudness* (1 = "very loud" vs. "very soft"), *harmony* or "whether the sequence of notes appeared to unfold in a pleasing vs. discordant fashion" (1 = "very harmonious" vs. "very discordant"), and *flow* or "the way that notes seem to transition from to another" (1 = "very smooth" vs. "very staccato"). Lastly, in rating Set IV, attention was directed toward the visual display in which participants were asked to judge its *affect* (1 = "very positive"), *likeability* (1 = "very likeable"), the extent to which it *fit* the accompanying soundtrack (1 = "very well" vs. "very poorly"), and whether it appeared to contain a coherent *theme* that conveyed a meaningful message (1 = "very meaningful" vs. "very non-meaningful"). Each of these four sets of judgments was presented on a separate page of the response booklet and participants were asked to complete a given set of ratings before moving on to the next. Participants were also instructed that they had an unlimited amount of time to complete all ratings and when they had done so, to press the space bar on the computer keyboard to advance the next stimulus. An entire experimental session was approximately 35 min in duration.

Results

Separate analyses were conducted on each of the four sets of perceptual judgments and in each case, mean ratings were collapsed over the ten presentation sets. Given that ratings of the visual dimensions (Set IV) and the stimuli as a whole (Sets I, II) allow one to assess the construct validity of the experimental material, they are discussed first.

Visual Dimensions

These ratings directed participants' attending to the visual displays alone, which thereby excluded the control condition in which no visual information was presented. Table 1 depicts the set of mean ratings as a function of the affect and format of the visual displays.

TABLE 1. Mean Ratings of the Visual Dimensions in Experiment 1 as a Function of the Affect and Format of Visual Displays.

	Visual Display			
	Positive Affect		Negative Affect	
Visual Dimension	Video	Montage	Video	Montage
Visual Affect	1.82 (0.94)	1.60 (0.74)	5.90 (1.04)	6.05 (1.01)
Visual Likeability	2.36 (0.84)	2.20 (0.97)	5.25 (1.08)	5.75 (1.02)
Visual Theme	2.65 (1.09)	2.37 (1.08)	2.25 (1.11)	2.40 (1.13)
Visual/Musical Fit	2.87 (0.96)	3.01 (1.02)	2.75 (1.12)	2.83 (0.97)

Note: Each rating was on a 7-point scale in which a value of "1" corresponded to "very positive" for Affect, "very likeable" for Likeability, "very meaningful" for Theme, and "very good" for Fit. Standard deviations are shown in parentheses.

A 2 (visual affect) × 2 (visual format) repeated measures MANOVA revealed a significant main effect of visual affect on the combined set of ratings, $F(4, 56) = 77.63$, $p < .0001$, Wilks' Lambda = .11. Univariate analyses of the individual ratings, using an adjusted alpha level of .0125, showed that two of the judgments varied with visual affect, namely, rated affect, $F(1, 59) = 194.39$, $p < .0001$, and likeability, $F(1, 59) = 87.95$, $p < .0001$. Consistent with the results of the pretest, the affect of positive visual displays was rated significantly more positive than that of negative displays. Similarly, participants liked the positive displays more than the negative ones. The two remaining judgments, visual theme and fit, yielded null results. The means in all of these conditions were relatively low, indicating that the meaningfulness of the visual themes and the fit with paired melodies were judged to be quite high.

This overall pattern of results generalized across the montage and video formats of the visual displays: there was neither a main effect for format nor an interaction with visual affect in the initial MANOVA.

Music/Film Stimulus Dimensions

RATING DATA

This set of judgments required participants to evaluate each stimulus in a holistic fashion for its overall affect, activity, and likeability, and provides an additional opportunity to assess the construct validity of the visual manipulations. By subsequently comparing differences between the experimental and control conditions, these data also allow one to determine how ratings vary in the presence vs. absence of visual information. The set of mean ratings as a function of visual affect and format are shown in Table 2.

An overall 2 × 2 repeated measures MANOVA revealed that the combined set of ratings was significantly influenced by both a main effect for visual affect, $F(3, 57) = 94.36$, $p < .0001$, Wilks' Lambda = .11, and format, $F(3, 57) = 9.21$, $p < .001$, Wilks' Lambda = .56. Univariate tests of the visual affect manipulation, with an adjusted alpha level of .017, indicated that the rated affect, $F(1, 59) = 268.32$, $p < .0001$, likeability, $F(1, 59) = 118.98$,

TABLE 2. Mean Ratings of the Music/Film Stimulus Dimensions in Experiment 1 as a Function of the Affect and Format of Visual Displays.

	Visual Display				
	Positive Affect		Negative Affect		
Stimulus Dimensions	Video	Montage	Video	Montage	Control: No Visual
Stimulus Affect	1.92 (0.90)	1.45 (0.76)	6.27 (0.94)	6.48 (0.95)	3.45 (1.18)
Stimulus Likeability	2.25 (0.87)	2.00 (0.99)	6.35 (1.02)	6.10 (1.01)	3.50 (1.19)
Stimulus Activity	2.90 (0.98)	2.10 (0.88)	4.10 (0.93)	3.33 (0.86)	4.05 (1.48)

Note: Each rating was on a 7-point. scale in which values of "1" corresponded to "very positive" for Stimulus Affect, "strongly liked" for Stimulus Likeability, and "very active" for Stimulus Activity. Standard deviations are shown in parentheses.

$p < .0001$, and activity of the stimuli, $F(1, 59) = 64.11$, $p < .001$, were all influenced by this variable. Positive stimulus pairings were judged more positive in their affect, more likeable, and more active than negative stimulus pairings. For visual format, the univariate tests (with an adjusted alpha of .017) showed that only the rated activity of stimuli was influenced by this manipulation: visual montages were judged more active than video displays for both visual affect conditions.

In order to evaluate the rated stimulus dimensions in the presence vs. absence of visual information, a second one-way MANOVA was conducted that included the four experimental conditions as well as the single control condition. As before, a significant effect emerged for the set of combined ratings, $F(3, 57) = 52.98$, $p < .0001$, Wilks' Lambda = .04, and as indicated by the univariate analyses (with an adjusted alpha level of .017), judged affect, $F(4, 236) = 156.21$, $p < .0001$, likeability, $F(4, 236) = 60.25$, $p < .0001$, and activity level, $F(4, 236) = 9.38$, $p < .001$, were all influenced by the stimulus dimensions.

A set of Dunnett posthoc comparisons next was performed to contrast the experimental vs. control conditions for each type of rating. The results showed that relative to the control melodies presented alone, melodies accompanied by positive videos and montages were rated significantly more positive in their affect, but more negatively when accompanied by negative videos and montages. The same was true of the likeability ratings. For both format types, positive visual displays significantly increased the likeability of stimuli whereas negative displays decreased likeability relative to melodies presented by themselves. Lastly, a somewhat different pattern of results was found for the activity ratings. Relative to the control condition, both positive and negative montages as well as positive videos significantly enhanced the activity level of stimuli, whereas that of negative videos did not significantly change.

ADJECTIVE SELECTION DATA

Recall that for each stimulus presented, participants were asked to select all relevant adjectives that varied along the two dimensional axes of stimulus affect and activity. In order to evaluate these data, the percentage of adjectives selected from each of the four quadrants ($n/7 \times 100$) was determined for each participant and analyzed through a 2 (visual affect) × 2 (visual format) × 2 (affect axis) × 2 (activity axis) repeated measures ANOVA. These means are shown in Table 3.

The most important finding was a significant four-way interaction, $F(1, 59) = 46.01$, $p < .001$. A set of Bonferroni posthoc comparisons (p set at .05) confirmed that for both types of visual formats, a significantly greater percentage of positive (vs. negative) adjectives was selected in the presence of positive visual displays, whereas the opposite was true for negative displays. These selections, however, also depended on format. Video stimuli were significantly more likely to result in the selection of passive adjectives and montages with active adjectives, and this applied to both positive and negative material.

To assess how this set of results compared to the control condition, a 5 (stimuli) × 2 (affect axis) × 2 (activity axis) repeated measures ANOVA was performed. The three-way interaction was once again significant, $F(4, 236) = 27.81$, $p < .001$, and a set of Dunnett posthoc comparisons (p set at .05) was conducted to isolate the pattern of differences. In the control condition, all melodies had been pretested as neutral in affect and, in fact, the percentage of selected adjectives was relatively low, ranging from 10% to 19%. However, relative to these melodies, the selection of positive and negative

TABLE 3. Mean Percentage of Adjectives Selected in Experiment 1 from the Activity × Affect Dimensions of Russell's (1980) Circumplex Model as a Function of the Affect and Format of Visual Displays.

	Visual Display				
	Positive Affect		Negative Affect		
Adjective Dimensions	Video	Montage	Video	Montage	Control: No Visual
Positive/Active	33% (7)	63% (8)	1% (1)	1% (1)	14% (6)
Positive/Passive	57% (7)	27% (5)	2% (1)	0% (0)	19% (8)
Negative/Active	0% (0)	0% (0)	24% (5)	72% (8)	10% (5)
Negative/Passive	3% (1)	1% (1)	56% (6)	31% (4)	17% (6)

Note: Standard deviations are shown in parentheses.

adjectives significantly increased in response to positive and negative displays, respectively.

As a set, then, these data along with the ratings of the visual displays and stimuli as a whole all confirm the construct validity of the experimental manipulations.

Musical Dimensions

The question of most interest in this first experiment is whether visual displays influence the perception of their accompanying melodies. To address this, a 2 × 2 repeated measures MANOVA first was conducted to assess the potential effects of visual affect and format on the musical dimensions of perceived affect, activity, tempo, tonality, rhythm, loudness, harmony, and flow. This set of means is depicted in Table 4.

The results indicated that the combined ratings were influenced by significant main effects for both visual affect, $F(1, 52) = 18.98$, $p < .0001$, Wilks' Lambda = .11, and format, $F(1, 52) = 9.13$, $p < .0001$, Wilks' Lambda = .19. For the visual affect manipulation, univariate analyses of the individual ratings (with a Bonferroni adjusted alpha level of .006) revealed that five of the eight dimensions were influenced by the mood of the visual display. Relative to tunes accompanied by negative scenes, those heard in the context of positive scenes were judged faster, $F(1, 59) = 18.05$, $p < .001$, more rhythmic, $F(1, 59) = 16.27$, $p < .001$, louder, $F(1, 59) = 15.98$, $p < .001$, more active, $F(1, 59) = 19.11$, $p < .001$, and more positive in overall affect, $F(1, 59) = 30.72$, $p < .001$. Music perception also varied with visual format. Univariate tests of this manipulation (adjusted alpha of .006) showed that relative to videos, tunes accompanied by montages were heard as faster, $F(1, 59) = 32.69$, $p < .001$, more rhythmic, $F(1, 59) = 11.74$, $p < .001$, more active, $F(1, 59) = 23.61$, $p < .001$, and more staccato in their flow, $F(1, 59) = 13.72$, $p < .001$. The two remaining musical dimensions, tonality and harmony, yielded null effects and were not influenced by either visual format or affect.

To assess whether music perception was altered by the presence vs. absence of visual information, a one-way repeated measures MANOVA was next performed that included the control condition along with the experimental ones. It was significant, $F(4, 52) = 11.07$, $p < .001$, Wilks' Lambda = .13, and univariate analyses of the individual ratings (with an adjusted alpha level of .006) indicated that four of the musical dimensions were influenced by the stimulus dimensions, namely, tempo, $F(4, 236) = 21.05$, $p < .001$, rhythm, $F(4, 236) = 13.06$, $p < .001$, loudness, $F(4, 236) = 12.26$, $p < .001$, and activity, $F(4, 236) = 17.57$, $p < .001$. A set of Dunnett comparisons between the control and experimental conditions further revealed that melodies paired with any type of visual scene were judged significantly faster, more rhythmic, louder, and more active than those in the control condition. For the dimension of flow, control melodies did not significantly differ from those accompanied by a video format, but were perceived as having a smoother flow of notes than those accompanied by a montage. Lastly, this set of comparisons revealed that the overall affect of a tune was enhanced by visual information. Relative to tunes heard in isolation, melodies were judged significantly more

TABLE 4. Mean Ratings of the Musical Dimensions in Experiment 1 as a Function of the Affect and Format of Visual Displays.

	Visual Display				
	Positive Affect		Negative Affect		
Musical Dimensions	Video	Montage	Video	Montage	Control: No Visual
Tempo	3.67 (0.92)	3.01 (0.89)	4.11 (1.07)	3.50 (1.02)	4.75 (1.09)
Tonality	2.75 (1.47)	2.61 (1.29)	2.83 (1.37)	2.50 (1.28)	3.03 (1.67)
Rhythm	1.90 (0.74)	1.45 (0.69)	2.47 (0.77)	2.00 (0.80)	3.00 (1.06)
Loudness	3.31 (0.82)	3.25 (0.83)	3.87 (0.87)	3.90 (0.78)	4.63 (1.04)
Harmony	2.50 (1.32)	2.33 (1.39)	2.45 (1.33)	2.60 (1.40)	2.87 (1.59)
Flow	2.50 (0.75)	3.22 (0.82)	2.60 (0.78)	3.40 (0.83)	2.67 (1.25)
Affect	3.00 (0.57)	3.17 (0.60)	5.83 (0.62)	5.90 (0.59)	4.67 (1.22)
Activity	2.65 (0.72)	2.12 (0.76)	4.20 (0.81)	3.72 (0.86)	4.75 (1.31)

Note: Each rating was on a 7-point scale in which values of "1" corresponded to "very fast" for Tempo, "very tonal" for Tonality, "very regular" for Rhythm, "very loud" for Loudness, "very harmonious" for Harmony, "very smooth" for Flow, "very positive" for Affect, and "very active" for Activity. Standard deviations are shown in parentheses.

positive and negative when accompanied by positive and negative scenes, respectively.

Discussion

This first experiment demonstrates that visual displays influenced music perception in a variety of ways. At the most general level, the mere presence of visual information, regardless of its affect or format, enhanced certain musical dimensions such that melodies were heard as faster, more rhythmic, louder, and more active than the same melodies heard by themselves. One will immediately notice that these qualities are all associated with positive affect in the auditory modality (Gabrielsson & Lindstrom, 2001; Hevner, 1935a, 1935b; Juslin, 1997; Scherer & Oshinsky, 1977). However, affect cannot be the mediating factor since the phenomenon also occurred in the presence of negative scenes. Similarly, stimulus likeability is unlikely to be at play since only the positive pairings were judged more likeable than the control melodies—negative pairings were judged less likeable. An alternative possibility is that the presence of visual information renders an accompanying melody more interesting and/or meaningful, which in turn enhances the salience of its acoustical qualities. Sun & Lull (1986), for example, have argued that the meaning supplied by visual scenes is one of the main factors contributing to the popular appeal of music videos among young listeners. Relative to listening to a melody by itself, an accompanying sequence of scenes provides an interpretation of the song's meaning and the underlying message it is attempting to convey. This provides the basis for further elaboration and the opportunity to relate the song to associated concepts. In his book, *Analysing Music Multimedia*, Cook (1998) has made a similar argument and succinctly states: ". . . music interprets the words and pictures" (p. 22). Although none of the perceptual judgments directly assessed this idea, ratings of the visual scenes were judged high in meaningfulness and to fit well with their respective tune. If the meaning of scenes generalized to the accompanying music to heighten its perceived meaningfulness and interest level, then the types of acoustical changes observed here would be expected.

Beyond effects due to the mere presence of visual information, there also was evidence that visual affect influenced music perception in a mood-congruent fashion. Consistent with the findings of Geringer et al. (1996, 1997), the affect of visual scenes enhanced the emotional impact of a tune. Melodies that had been prerated as neutral in overall affect were judged to have a positive affect in the presence of positive scenes but a negative affect in the presence of negative scenes. More interestingly, the acoustical qualities of these tunes were influenced in a corresponding manner. Melodies in a positive visual context seemed louder, faster, more rhythmic, and more active than those in a negative context. The mood of the film, then, not only biased the affective interpretation of an accompanying tune, but once this interpretation was adopted, changed the way the tune was perceived in a mood consistent manner.

It is noteworthy that only the temporal qualities of tunes were biased by film affect, but the pitch qualities of tonality and harmony were not. Two processes may be at play. In an extensive review of the literature, Gabrielsson & Lindstrom (2001) noted that among the various factors specifying mood in the auditory modality, tempo appears to be the most influential (e.g., Hevner, 1937; Juslin, 1997; Scherer & Oshinsky, 1977). In the Scherer and Oshinsky (1977) study, for example, participants rated the emotional tone of auditory sequences that systematically had been manipulated along a given dimension, and tempo was found to be the strongest determinant of any change in perceived affect. More generally, positive (or negative) events within everyday life tend to be accompanied by increases (or decreases) in energy, and the iconic equivalents of this may be more clearly expressed through tempo, loudness, and the extent to which there are rhythmic and predictable patterns of change. A second process potentially at play is that the concepts of tempo, loudness, and rhythm may seem more familiar to listeners and therefore more unambiguous than those of tonality and harmony. Although these latter concepts were defined for participants, they are more specific to music than the former, and nonmusicians in particular may be more uncertain about their acoustical correlates. The degree of variability for these two judgments was, in fact, higher than that of the other ratings.

The last main finding was the effect of visual format on perceptual ratings. Both the adjective selection and musical rating data showed that in contrast to videos that were judged smoother and more passive, visual montages seemed more active. These terms, in fact, aptly describe the structure of each format. The videos used here displayed a relatively slow rate of action in which thematically related scenes smoothly transitioned from one to another. Montages, on the other hand, contained approximately 130 still photos that changed every 700 ms to match the average beat duration of its accompanying tune. Moreover, even though the photos within each montage displayed a positive or negative affect that participants viewed as related to an overarching theme, the content of each was relatively independent of the immediately surrounding ones. Results from the musical rating data showed that relative to a video format, these various characteristics enhanced the perceived rhythm, rate, and activity of the accompanying tunes and led to a seemingly more staccato flow of notes.

Experiment 2

Experiment 2 was designed with two related goals in mind. The first was to determine whether the effects of visual information on music cognition generalize to a different methodological context. Although perceptual ratings provide a useful means in which to assess how audiovisual events are experienced, these evaluations are very subjective in nature and can be influenced by demand characteristics and/or response bias effects. To confirm the reliability of those effects observed in Experiment 1, it is therefore useful to assess behavior on a very different type of task. The particular one adopted in Experiment 2 is that of recognition memory that not only serves as a converging operation, but also enables one to assess whether visual displays influence other types of cognitive behavior beyond perceptual experience.

In an investigation of the effects of music on film, recall that the study by Boltz (2001) manipulated the affect of musical soundtracks in the presence of ambiguous film clips. In addition to effects on the film's interpretation, results from a second group of participants revealed that music biased selective attending and subsequent remembering in a mood-congruent fashion. Both the remembering of old items and the confabulation of new items varied with the affect of the accompanying soundtrack. The question addressed here is whether the reverse effect occurs such that variations in visual affect influence melody recognition in a corresponding manner. If visual displays actually do alter the perceived affect and acoustical qualities of a song, then this should be reflected through memory distortions. In the presence of positive scenes, for example, melody recognition may be biased in a mood-congruent fashion such that tunes are falsely recognized as faster, louder, more rhythmic, and higher in pitch. Similarly, tunes may be misremembered as having the opposite set of characteristics in the presence of negative scenes.

These ideas were investigated in an experimental design that consisted of two phases. The first was identical to that of Experiment 1 in which participants listened to a set of five tunes, four of which were accompanied by visual scenes that varied in affect and format, and one appearing in the absence of visual information. Immediately afterwards, all participants were then given a surprise recognition memory task. On a given trial, they were required to discriminate an old melody in its original form from a new version that systematically had been transformed along one of three structural dimensions that conveys emotional information in the auditory environment (e.g., Gabrielsson & Lindstrom, 2001). In particular, each tune was increased or decreased in either tempo, pitch, or both pitch and tempo together to reflect a more positive or negative affect, respectively. In contrast to the qualities of tonality and harmony evaluated in Experiment 1, a different pitch dimension was assessed here, namely, absolute pitch in which the frequency (Hz) of all notes within a melody was uniformly increased or decreased. Given that this type of pitch change frequently is associated with emotional changes in music and the human voice (Juslin & Sloboda, 2001; Scherer, 1979), it may be susceptible to mood bias effects arising from the visual displays. In all conditions, the main measure of interest is the pattern of false recognitions that is predicted to reflect biases from visual affect.

Method

Design and Participants

The experimental design for phase one was identical to that of Experiment 1: all participants listened to four melodies accompanied by visual information that varied in its format (montage, video) and affect (positive, negative). As before, a control melody was presented in isolation. During a subsequent recognition phase, all participants re-encountered their respective set of melodies along with distracter versions that were more positive or negative in their overall affect. The type of melody transformation displayed by distracters (tempo, pitch, tempo + pitch) varied as a between-subjects variable. The resulting design was a 2 (visual affect) $\times$ 2 (visual format) $\times$ 2 (affect of distracter melodies) $\times$ 3 (melody transformation type) mixed factorial with a hanging control group.

A different group of 60 participants from an Introductory Psychology course at Haverford College participated for course credit. Once again, the sample primarily consisted of nonmusicians. None were currently playing or had played a musical instrument within the past three years and only 8% were members of the campus choir.

Stimulus Materials

Although material for the presentation phase remained the same, an additional set of melodies was developed for the subsequent recognition memory phase. Using digital audio editing software (Audacity), each melody from Experiment 1 systematically was altered along the acoustical dimensions of pitch, tempo, and pitch + tempo. For tempo, each melody was accelerated and decelerated by 15% while holding pitch constant. The pitch transformation involved increasing and decreasing the absolute frequency (Hz) of all notes within a melody by 15% while holding tempo constant. For the third type of transformation, the tempo and pitch frequencies of all melodies were jointly increased and decreased by 15%.[1] As indicated in the past literature (e.g., Gabrielsson & Lindstrom, 2001), these types of increases and decreases in pitch and/or tempo result in more positive and negative musical affect, respectively.

From these melodies, three different melody sets were formed for the recognition memory phase of the experiment. Each contained 20 trials and all trials contained melodies alone (with no accompanying visual information) incorporating only one type of transformation, namely, pitch increases/decreases, tempo increases/decreases, or pitch + tempo increases/decreases. On a given trial, two versions of a given tune were presented, one in its original form (old) and another in an altered form (new); the order of these was evenly balanced across all trials. Within a given set, each of the five melodies was presented on four occasions, two in which the altered version represented an affectively more positive tune and two representing an affectively more negative tune. Lastly, each of the three sets was arranged into two different orders of randomized trials to yield a total of six recognition memory sets.

Procedure

The experimental procedure consisted of two phases. In the first, participants were presented with one of the 10 presentation sets from Experiment 1, each containing four music/visual pairings along with a control melody, and instructed to listen carefully because they later would be asked questions about them. Immediately following this, participants were randomly assigned to one of the six melody recognition sets and given a surprise recognition memory task. On each trial, a warning tone signaled the trial's onset and was followed 1 s later by two versions of a given melody. During a 5 s response period that followed, participants were asked to indicate on a response sheet which of the two melodies was identical to one heard earlier. Participants were tested in groups of one to four individuals, and an entire experimental session was approximately 45 m.

Results

Given that participants were required on each experimental trial to discriminate between a melody's original vs. altered version, hit ("old"/old) and false alarm ("old"/new) rates were inversely related to one another. False alarm rate was assessed in the overall ANOVAs since certain predictions have been offered as a function of the melodies' transformations. An initial ANOVA assessed performance in the experimental conditions alone in which melodies were accompanied by visual information. A subsequent ANOVA was conducted that contrasted performance between the experimental and control conditions.

For the first analysis, the data were collapsed over the ten presentation sets from phase one and the six melody sets from phase two such that the resulting statistical design was a $3 \times (2 \times 2 \times 2)$ mixed factorial. Melody transformation type (pitch, tempo, pitch + tempo) was the single between-subjects factor whereas visual affect (positive, negative), visual format (montage, video), and affect of the new distracter melodies (positive, negative) were the repeated measures variables. These means are depicted in Table 5. When significant differences emerged, the set of experimental means was further analyzed through Bonferroni posthoc comparisons (p set at .05).

The overall ANOVA revealed three significant effects. First, a main effect for melody transformation type, $F(2, 57) = 5.38$, $p < .007$, revealed that relative to melody sets containing pitch ($M = 15\%$) or pitch + tempo transformations ($M = 15\%$), error rate was significantly higher for melodies displaying tempo transformations ($M = 25\%$).

[1]The 15% magnitude of change used in the melody transformations of Experiment 2 was determined through a pretest in which eight participants, all nonmusicians, independently listened to three blocks of the five melodies. Within a given block, all melodies displayed the same type of transformation (i.e., pitch, tempo, or both) and each was presented twice for a total of 30 trials. On a given trial, a melody was initially presented in its unaltered version (0 change) and immediately followed by six additional versions that had been transformed by ±5%, ±10%, ±15%, ±20%, ±25%, and ±30% in pitch and/or tempo. The participant's task was to indicate which transformed version seemed to just noticeably differ from the initial melody. The results showed that a change of ±15% was most frequently selected as the jnd for tempo (81%), pitch (75%), and pitch + tempo (83%).

TABLE 5. Mean Percent False Alarm Rate ("Old"/New) in Experiment 2 as a Function of Visual Affect, Affect of Distracter Melodies, and Type of Melody Transformation.

	Visual Display						
	Positive Affect		Negative Affect			Control: No Visual	
Type of Melody Transformation	Positive Distracters	Negative Distracters	Positive Distracters	Negative Distracters	*Means*	Positive Distracters	Negative Distracters
Pitch	21% (7.1)	10% (2.2)	9% (3.7)	19% (7.3)	*15%*	18% (6.4)	17.5% (6.2)
Tempo	42.5% (5.2)	11% (2.0)	14% (1.6)	34% (4.1)	*25%*	18% (5.8)	15% (6.1)
Pitch + Tempo	31% (9.3)	1% (0.5)	5% (0.75)	24% (8.2)	*15%*	10% (3.4)	15% (5.6)
Means	*31.5%*	*7%*	*9%*	*26%*		*15%*	*16%*

Note: Standard deviations are shown in parentheses.

Second, there was a significant visual affect × distracter melody affect interaction, $F(1, 57) = 32.81$, $p < .001$, that can be seen in the column means of Table 5. Notice that the false alarm rate for positive distracters is significantly higher than that of negative distracters when tunes were initially presented in a positive visual context. Conversely, negative (vs. positive) distracters were more likely to be misidentified when they initially appeared in a negative visual context.

Table 5 also reveals a significant three-way interaction between melody transformation type, visual affect, and distracter melody affect, $F(2, 57) = 11.72$, $p < .001$. Within each of the three transformation conditions, mood congruency effects were observed in that the false alarm rate for positive vs. negative distracters was higher in the positive visual affect condition, whereas the reverse was observed in the negative visual affect condition. This pattern of mean differences, however, was significantly more pronounced with melodies displaying tempo transformations than pitch + tempo transformations, which in turn was more pronounced than melodies displaying pitch transformations alone. Lastly, unlike Experiment 1, visual format exerted no influence on recognition memory performance: the results generalized across both the video and montage formats.

To assess recognition memory performance in the presence vs. absence of visual information, a secondary ANOVA was performed that included the control condition along with the experimental ones. The resulting statistical design was a 3 (melody transformation type) × 2 (affect of the distracter melodies) × 5 (stimuli) mixed factorial. A significant main effect for melody transformation type, $F(2, 57) = 4.10$, $p < .009$, once again emerged, and a set of Dunnett posthoc comparisons (p set at .05) indicated that the error rate for tempo transformations was significantly higher in the presence of visual displays (M = 25%) than in their absence (M = 16.5%). Null differences between the experimental and control conditions were observed for both the pitch and pitch + tempo transformations.

The two-way interaction between stimuli and distracter melody affect also was significant, $F(4, 57) = 15.5$, $p < .001$. Dunnett comparisons showed that in the No Visual condition, a comparable percentage of false recognitions was observed for positive and negative distracters. However, relative to each of these values, the added presence of positive and negative displays significantly increased the false recognition of positive and negative distracters, respectively.

Lastly, as in the initial analysis, the three-way interaction between stimuli, distracter melody affect, and melody transformation type was significant, $F(4, 57) = 8.63$, $p < .001$. A set of Dunnett comparisons indicated that in the three No Visual conditions, the false alarm rates for positive vs. negative distracters were comparable to one another. Although the percentages of these false alarms significantly increased in the presence of mood congruent visual information, this only applied to melodies with tempo or pitch + tempo transformations. No significant differences between the control vs. experimental conditions were observed with melodies displaying pitch transformations alone.

Discussion

The present set of results converges with those of Experiment 1 by showing that music cognition was influenced by the presence of visual information. In lieu of perceptual judgments that are subjective in nature, the recognition memory task used here was an incidental one that participants were not expecting and was therefore less vulnerable to demand characteristics. As such, it represents a more implicit measure of cognition.

The most notable finding is that visual affect distorted melody recognition in a mood-congruent fashion. Relative to control melodies that produced a comparable false alarm rate across positive and negative distracters, negative displays increased the false recognition of negative distracters and decreased that of positive distracters. The reverse effect occurred in the presence of positive displays. This phenomenon, however, only appeared in distracter melodies whose tempo had been systematically increased or decreased to reflect more positive or negative emotion, respectively. The added presence of pitch transformations reduced the mood congruency bias, and it was eliminated entirely by pitch transformations alone. This too is consistent with the results from Experiment 1 in which tempo and the more dynamic qualities of a tune were most influenced by visual affect. In contrast to tonality or harmony assessed in the previous study, the type of pitch characteristic addressed in Experiment 2 consisted of pitch increases or decreases—emotive changes commonly experienced in speech and other auditory events. Nonetheless, pitch was not influenced by the visual manipulations and the general meaning of this finding is unclear. Within the past literature, tempo often is found to have the greatest impact on musical affect when multiple factors are examined (Hevner, 1937; Juslin, 1997; Scherer & Oshinsky, 1977), and perhaps the present set of findings is another illustration of this phenomenon. Alternatively, the null effects of pitch may be due to the methodology that was used. For example, perhaps absolute pitch was not transformed to a sufficient magnitude and/or is dependent upon other pitch characteristics (e.g., mode, timbre, pitch range, melodic contour) in order to influence perceived affect. A third possibility is that despite the results of the pretest[1] showing that a 15% change in pitch can be discriminated, perhaps participants did not encode the absolute pitch intervals of melodies. Previous research has shown that except for those who have this rare ability, most individuals are not able to remember absolute pitch values unless the tunes are highly familiar (Halpern, 1989; Schellenberg & Trehub, 2003). Given the unfamiliarity of the musical selections and a sample of nonmusicians, this too may account for the results. Additional investigations are needed to clarify these issues and determine whether visual displays can, in fact, influence the perception and remembering of certain pitch qualities.

Lastly, unlike the previous study, visual format exerted no significant impact on recognition performance. The observed pattern of results generalized across both the video and montage formats and it was only the mood of the visual display which led to distortions in melody recognition.

General Discussion

In many ways, the present set of results parallel the demonstrated effects of musical soundtracks on the cognitive processing of visual information. In both cases, the mood of music or visual displays provides an interpretative framework that then biases the nature of perception and memory. In the context of film, the mood of music influences inferences about why people are exhibiting the behaviors they are, what items from a scene are best remembered and what's likely to be confabulated, and how well the film later persists in long-term memory (see Cohen, 2001, for a review). When one examines the reverse relationship and the effects of visuals on music, the two experiments conducted here revealed that visual affect, at the very least, led to a general emotive interpretation that was either positive or negative in nature. This in turn biased the perception of a melody's affect and corresponding acoustical qualities in a mood congruent fashion. These effects were not only found in an explicit task of perceptual judgments but an implicit one that relied on recognition memory. Memory distortions occurred such that melodies were falsely recognized as faster or slower than their actual tempi in the presence of positive and negative displays, respectively.

Although visual affect exerted a greater and more reliable impact than visual format, the latter nonetheless influenced music perception. The two formats used here were intentionally selected for both their common usage and structural distinctiveness from one another. In contrast to the smooth transition of scenes within videos, montages displayed a set of still scenes that transpired at a rapid rate. These structural characteristics biased perception in a congruent fashion in that melodies in the presence of montages were judged faster, more rhythmic, more active, and more staccato than the same tunes in either the presence of videos or the absence of any visual information at all. Visual format, however, did not contribute to the types of memory distortions investigated in Experiment 2; namely, pitch and tempo. Nonetheless, it is possible that visual format may influence the remembering of other acoustical qualities, such as rhythm, and future research is needed to investigate this issue.

The overall results observed in this research were reliable ones that not only applied to different tasks, but a majority of participants in each experiment and a set of five different tunes that, across different presentation sets, were paired with the entire array of visual material. Moreover, the tunes themselves were ambiguous in the sense that each displayed a neutral affect and moderate level of activity. By pairing the same set of melodies with scenes of different affects and structural formats, this allowed one to examine the effects of the experimental manipulations alone while holding melodies constant across all experimental conditions. Lastly, an attempt was made to disguise the purpose of each experiment from participants, and thereby reduce demand characteristics. Participants in Experiment 1 were asked to rate all dimensions of the music-film pairs, and to perform an incidental (surprise) recognition memory task in Experiment 2.

From a theoretical perspective, one can argue that the effects of visual affect and format on music cognition reflect a type of stimulus overgeneralization effect. This was apparent with visual format in that melodies were perceived to have the same qualities as an accompanying montage or video, and visual affect exerted a similar influence via a mood congruent bias. In both cases characteristics of the visual display were attributed to their accompanying melodies such that, perceptually, they appeared to be structurally and affectively congruent with one another. In a theory of crossmodal perception, Welch (1999) has claimed that this sort of phenomenon is inherent to the cognitive system and illustrative of the "unity assumption." The basic idea is that humans are motivated to maintain congruence in their perceptual world such that any physical discrepancies (of a reasonable magnitude) are reduced in order to attain an integrated and unitary percept. This not only provides individuals with a sense of harmony and well-being, but also ensures a more efficient use of cognitive resources. In lieu of dividing one's attention among separate sources of information, perceptual unification allows one to direct attention toward a single event that entails less effort. Welch primarily has discussed the unity assumption relative to intersensory conflicts such as the ventriloquism effect or viewing a prismatically displaced hand. In a context such as the present one, where one modality is affective and the other is ambiguous, an intersensory bias toward structural congruence is even more likely to occur given the relatively small magnitude of discrepant information.

These ideas are consistent with a Congruence-Associationist Model (Cohen, 1999, 2009) specifically developed to address music-film interactions. This model envisions a set of hierarchical stages that operate in both a top-down and bottom-up fashion. At an initial level, different sources of film information (i.e., music, visual activities, speech) are processed independently for their physical features. Patterns of these features are then compared for any crossmodal structural congruencies at level B and depending on the intended purpose of the art form (e.g., cinematic art to depict a visual story vs. music videos to highlight a song), attention will be biased toward the visual or auditory modality. At level C, information enters short-term memory and, hence, consciousness where people attempt to construct and make sense of the working narrative. This comprehension process is facilitated by top-down contributions from level D, which contains pre-stored knowledge about how people typically act in different situations.

The types of perceptual biases observed in the present research would presumably occur at the pre-attentive level of the model in which physical features from the audio and visual modalities are compared (Level B) before entering short-term memory for narrative construction and comprehension (Level C). In cases in which one modality is ambiguous but an accompanying modality has clear emotive meaning, the features of the latter appear to distort the perception and remembering of features within the former to achieve structural congruence and unity. Such effects previously have been observed in the perception of ambiguous visual material accompanied by positive or negative music (e.g., Boltz, 2001; Bullerjahn & Guldenring, 1994; Liu, 1976; McFarland, 1984; Parrott, 1982), and were similarly observed here in the perception and remembering of ambiguous melodies accompanied by positive or negative visual displays. These parallel results suggest that the relationship between music and visual information is a reciprocal one that results in a unified percept. Although such unifications may result in perceptual and memory distortions, as seen here, they also can enhance the emotional impact of stimuli and, as argued by Welch (1999), serve to reduce cognitive processing effort.

The research reported here is, in many ways, preliminary in nature and suggests several avenues of future research. One question involves situations in which visual displays and music each convey a strong and unambiguous emotion but ones which conflict with one another. This technique is commonly used in film-making and is known as "ironic contrast." Does film continue to bias music cognition in any way, including judged affect, memory, or the perception of acoustical qualities? More generally, does the impact of visual information on music produce the same magnitude of intersensory bias as the impact of music on visual information, or does one dominate the other? In addition, it may be useful to examine the potential influence of other visual dimensions beyond affect and structural format. For example, the particular content of a display along with its spatial properties may be of relevance, as may the amount and potency of activity. Lastly, visual information may influence other components of music cognition beyond those considered here, including the long-term memorability of a tune or the interpretation of a song's lyrics. The recognition memory paradigm of Experiment 2 offers a particularly useful means in which to investigate other types of memory distortions that may arise from audiovisual interactions, and whether these vary with certain types of individuals such as those with varying degrees of musical experience. In short, a multitude of questions arise when one considers the various ways in which music and visual information may reciprocally influence behavior. Although the past literature seems to tacitly reflect a visual bias by extensively investigating ways in which music can influence the processing of visual mediums, the reverse relationship is of equal interest and ecological validity. Investigations of the latter and parallels with the former will ultimately enhance our understanding of those processes mediating intersensory perception that is so typical of everyday cognition.

Author Note

This research was supported by a Faculty Research Grant from Haverford College and presented at the 2008 Meeting of the Psychonomic Society in Chicago, IL. The authors thank Allison King and Carly Giffin for their assistance with data collection.

Correspondence concerning this article should be addressed to Marilyn Boltz, Department of Psychology, Haverford College, Haverford, PA l9041. E-MAIL: mboltz@haverford.edu

References

BENZA, M. (2004). Traverser le temps. On *Benza!* [CD] Malibu, CA: Academy of Sound, Color, and Movement.

BOLIVAR, V. J., COHEN, A. J., & FENTRESS, J. C. (1994). Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action. *Psychomusicology, 13,* 28-59.

BOLTZ, M. G., SCHULKIND, M., & KANTRA, S. (1991). Effects of background music on the remembering of filmed events. *Memory and Cognition, 19,* 593-606.

BOLTZ, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception, 18,* 427-454.

BOLTZ, M. G. (2003). The cognitive processing of film and musical soundtracks. *Memory and Cognition, 32,* 1194-1205.

BULLERJAHN, C., & GULDENRING, M. (1994). An empirical investigation of effects of film music using qualitative content analysis. *Psychomusicology, 13,* 99-118.

CASSIDY, J. W., & SIMS, W. L. (1991). Effects of special education labels on peers' and adults' evaluations of a handicapped youth choir. *Journal of Research in Music Education, 39,* 23-34.

CHOPRA, Y. (Director/Producer). (1993). *Darr: A violent love story.* [Motion picture]. Mumbai, India: Yash Raj Films.

COHEN, A. J. (1993). Associationism and musical soundtrack phenomena. *Contemporary Music Review, 9,* 163-178.

COHEN, A. J. (1999). The functions of music in multimedia: A cognitive approach. In S. W. Yi (Ed.), *Music, mind, and science* (pp. 53-69). Seoul, Korea: Seoul National University Press.

COHEN, A. J. (2000). Film music: Perspectives from cognitive psychology. In J. Buhler, C. Flinn, & D. Neumeyer (Eds.), *Music and cinema* (pp. 360-377). Middlebury, VT: Wesleyan University Press.

COHEN, A. J. (2001). Music as a source of emotion in film. In P. Juslin & J. Sloboda (Eds.), *Music and emotion* (pp. 249-272). Oxford, UK: Oxford University Press.

COHEN, A. J. (2009). Music in performance arts: Film, theatre, and dance. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford handbook of music psychology* (pp. 441-451). New York: Oxford University Press.

COOK, N. (1998) *Analysing music multimedia.* Oxford, UK: Oxford University Press.

COSMIDES, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 864-881.

DAVIDSON, J. W. (1994). What type of information is conveyed in the body movements of solo musician performers? *Journal of Human Movement Studies, 6,* 279-301.

FRICKE, R. (Director/Producer). (1985). *Chronos.* [Motion picture]. Universal City, CA: Miramar Films.

FRICKE, R. (Director/Producer). (1992). *Baraka: A world beyond words.* [Motion picture]. Orland Park, IL: Mpi Home Videos Inc.

GABRIELSSON, A., & LINDSTROM, E. (2001). The influence of musical structure on emotional expression. In P. Juslin & J. Sloboda (Eds.), *Music and emotion* (pp. 223-248). Oxford, UK: Oxford University Press.

GERINGER, J., CASSIDY, J., & BYO, J. (1996). Effects of music with video on responses of nonmusic majors: An exploratory study. *Journal of Research in Music Education, 44,* 240-251.

GERINGER, J., CASSIDY, J., & BYO, J. (1997). Nonmusic majors' cognitive and affective responses to performance and programmatic music videos. *Journal of Research in Music Education, 45,* 221-233.

GILLESPIE, R. (1997). Ratings of violin and viola vibrato performance in audio-only and audiovisual presentations. *Journal of Research in Music Education, 45,* 212-220.

GOLDBERG, M. E., CHATTOPADHYAY, A., GORN, G. J., & ROSENBLATT, J. (1993). Music, music videos, and wear out. *Psychology and Marketing, 10,* 1-13.

HALPERN, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory and Cognition, 17,* 572-581.

HEVNER, K. (1935a). The affective character of the major and minor modes in music. *American Journal of Psychology, 47,* 103-118.

HEVNER, K. (1935b). Expression in music: A discussion of experimental studies and theories. *Psychological Review, 47,* 186-204.

HEVNER, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology, 49,* 621-630.

JARRE, J. M. (1988). Industrial revolution overture. On *Revolutions.* [CD]. Brooklyn, NY: Dreyfus Records, Inc.

JUSLIN, P. (1997). Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy. *Musicae Scientiae, 1,* 225-256.

JUSLIN, P., & SLOBODA, J. (2001). *Music and emotion*. Oxford, UK: Oxford University Press.

LANG, P. J., BRADLEY, M. M., & CUTBERT, B. N. (1997). *International affective picture system: Instruction manual and affective ratings.* Gainesville, FL: University of Florida, The Center for Research in Psychophysiology.

LEVI, D. (1982). The structural determinants of melodic expressive properties. *Journal of Phenomenological Psychology, 13,* 19-40.

LIU, A. (1976). Cross-modality set effect on the perception of ambiguous pictures. *Bulletin of the Psychonomic Society, 7,* 331-333.

MAROLT, M. (2006). Listeners' emotional engagement with performances of a Scriabin Etude: An explorative case study. *Psychology of Music, 34,* 481-510.

MCFARLAND, R. A. (1984). Effects of music on emotional content of TAT scores. *Journal of Psychology, 116,* 227-234.

MCKENNIT, L. (1997). Mummer's dance. On *The book of secrets.* [CD]. Stratford, Ontario, Canada: Quinlan Road.

MEYER, L. B. (1956). *Emotion and meaning in music.* Chicago, IL: University of Chicago Press.

MONTEPARE, J., GOLDSTEIN, S., & CLAUSEN, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior, 11,* 33-42.

OLDFIELD, M. (1996). The chamber. On *The songs of distant earth.* [CD]. Burbank, CA: Reprise Records.

PARROTT, A. C. (1982). Effects of paintings and music, both alone and in combination, on emotional judgments. *Perceptual and Motor Skills, 54,* 635-641.

REGGIO, G. (DIRECTOR/PRODUCER). (1982). *Koyanisquatsi: Life out of balance.* [Motion picture]. Los Angeles, CA: MGM Home Entertainment, Inc.

REGGIO, G. (DIRECTOR/PRODUCER). (1988). *Powaqqatsi: Life in transformation.* [Motion picture]. Los Angeles, CA: MGM Home Entertainment, Inc.

RUSSELL, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161-1178.

RYAN, C., & COSTA-GIOMI, E. (2004). Attractiveness bias in the evaluation of young pianists' performances. *Journal of Research in Music Education, 52,* 141-154.

SCHELLENBERG, E. G., & TREHUB, S. E. (2003). Good pitch memory is widespread. *Psychological Science, 14,* 262-266.

SCHERER, K. (1979). Acoustical concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In S. Weitz (Ed.), *Nonverbal communication* (pp. 249-253). New York: Oxford University Press.

SCHERER, K., & OSHINSKY, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1,* 331-346.

SCHUTZ, M., & LIPSCOMB, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception, 36,* 888-897.

SLOBODA, J. A., & JUSLIN, P. N. (2001). Psychological perspectives on music and emotion. In P. Juslin & J. Sloboda (Eds.), *Music and emotion* (pp. 71-104). Oxford, UK: Oxford University Press.

SLOMAN, L., BERRIDGE, M., HOMATIDIS, S., HUNTER, D., & DUCK, T. (1982). Gait patterns of depressed patients and normal subjects. *American Journal of Psychiatry, 139,* 94-97.

SUN, S. W., & LULL, J. (1986). The adolescent audience for music videos and why they watch. *Journal of Communication, Winter,* 115-125.

TANGERINE DREAM (1985). Tiergarten. On *LeParc.* [Record]. Santa Monica, CA.: Relativity Records.

THOMPSON, W. F., & RUSSO, F. A. (2007). Facing the music. *Psychological Science, 18,* 756-757.

VINES, B. W., KRUMHANSL, C. L., WANDERLEY, M. M., & LEVITIN, D.J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101,* 80-113.

WAPNICK, J., DARROW, A., KOVACS, J., & DALRYMPLE, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education, 45,* 470-479.

WAPNICK, J., MAZZA, J. K., & DARROW, A. (1998). Effects of performer attractiveness, stage behavior, and dress on violin performance evaluation. *Journal of Research in Music Education, 46,* 510-521.

WAPNICK, J., RYAN, C., LACAILLE, N., & DARROW, A. (2004). Effects of selected variables on musicians' ratings of high-level piano performances. *International Journal of Music Education, 22,* 7-20.

WELCH, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Musseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 371-387). New York: Elsevier Press.

WILLIAMS, C., & STEVENS, K. (1979). Emotions and speech: Some acoustical correlates. In S. Weitz (Ed.), *Nonverbal communication* (pp. 233-248). New York: Oxford University Press.

ZORN, E. (1984, February 13). Memories aren't made of this. *Newsweek, 13,* 6.

APPENDIX A. Summary of the video sequences used in Experiments 1 and 2.

Positive Video A
A waterfall tumbles down a rock with mist rising to the sky; a rainbow gradually appears. In the next scene, a different and more dynamic waterfall is shown through both panoramic and close-up views. The film ends with an aerial shot above a peaceful and large body of water with clouds reflecting from its surface. From Fricke (1992), *Baraka: A World Beyond Words.*

Positive Video B
The film opens with a young woman playfully dancing in front of a campfire in the mountains before her husband comes to steal her away. The action jumps to the woman standing on a ladder and making some repairs to their house. She suddenly falls but he is immediately below to catch her. They then share a romantic embrace on the pier beside their house, which overlooks a beautiful pond in the mountains. The final scene depicts a flirtatious moment in a bathtub before the couple embraces once again on a bed. From Chopra (1993), *Darr: A Violent Love Story.*

Negative Video A
Live baby chickens in a processing plant are squeezed together and fall helplessly from one conveyor belt to another. Workers inspect their wings and toss them into a funnel where other workers brand their beaks with a heated metal instrument. From Fricke (1992), *Baraka: A World Beyond Words.*

Negative Video B
A homeless couple sleeps under a bridge as cars pass by. The scene shifts to another homeless person in a lean-to and yet another sleeping on a sidewalk. A mother covers her children with rags as they sleep, while another child in the background unsuccessfully begs for money from passers-by. The film ends with a group of homeless individuals beneath a highway viaduct, seeking shelter from the cold and wind. From Fricke (1992), *Baraka: A World Beyond Words.*

APPENDIX B. Materials from the adjective selection task of Experiment 1. These particular adjectives were selected from Russell's (1980) Circumplex Model of Emotion, which depicts a two-dimensional array defined by both an activity (active, passive) and valence (positive, negative) axis.

Positive Affect/High Activity	Positive Affect/Low Activity
Exciting	Satisfying
Delightful	Serene
Happy	Calm
Pleasing	Soothing
Passionate	Relaxing
Triumphant	Tranquil
Cheerful	Sleepy
Negative Affect/High Activity	**Negative Affect/Low Activity**
Frustrating	Boring
Annoying	Gloomy
Distressing	Mournful
Tense	Melancholy
Tragic	Sad
Angry	Depressing
Frightening	Miserable