How Music Influences the Interpretation of Film and Video: Approaches from Experimental Psychology

ANNABEL J. COHEN
University of Prince Edward Island

The quantitative nature of experimental psychology enables researchers to test models of the mental interaction of musical and visual information presented in a film or video. One possibility is that associations (that is, meanings) between musical and visual elements simply add together to produce a composite meaning. However, as demonstrated in psychological experiments, this principle of additivity of associations is sometimes insufficient to account for the influences of music on the interpretation of film and video. In particular, Iwamiya (1994), Lipscomb and Kendall (1994a), and Marshall and Cohen (1988) have pointed to the factor of crossmodal (that is, audiovisual) temporal structure as a determinant of the transmission of associations arising from music in a multimedia context. Limitations on working memory, as reflected in models of educational multimedia (Mayer, Bove, Bryman, Mars, and Tapangeo 1996; Fisch 2000), also constrain the interaction of music and film meanings. Taken together, these factors can be represented by a multi-level Congruence-Associationist framework that provides a foundation for experimental-psychological research directed at understanding the influence of music on the interpretation of film and video.

Music often accompanies film, video, and other electronic multimedia. Music is considered an important factor contributing to the meaning of the production. The present article examines the influence of music in such productions and develops a framework for exploring how this influence is accomplished. (The support of the Social Sciences and Humanities Research Council of Canada (SSHRC) and the research assistance of Robert N. Drew and Susan M. Doucette are gratefully acknowledged.) The particular perspective taken is that of experimental psychology, just one of many approaches that together enable a more complete understanding of phenomena such as music and film and their interaction.

The Experimental Psychological Perspective

Psychology is the study of the mind, behavior, and the relationship between them (Sternberg 1996). It is also the science of behavior and the cognitive (or mental) processes (Baron, Earhard and Ozier 2000). As part of its scientific task, psychology aims to determine how perception of events or objects in the world directly depends on the properties of these events or objects which stimulate the senses. Such information is obtained through repeatable experiments in which human responses are measured under different controlled conditions of external stimulation. Experiments have led to precise accounts of the sensitivity of the eye to color and light, and of the ear to sound frequency and intensity. In addition, research has elucidated complex aspects of perception such as reading, speech, and more recently music. Whereas in the last century much progress has been made in explaining how the eye and ear represent simple and complex information that is either visual or auditory in nature, less is known about how the brain represents multisensory information (that is, auditory plus visual). Yet, the multisensory context is typical of everyday experience. Auditory and visual information co-occur naturally: a passing car produces both visual and auditory impressions; a breeze creates effects that are heard and seen, and a speaker's articulation provides correlated information to both the ear and eye. Film, video, and other electronic multimedia provide another common source of multisensory information although this audiovisual co-occurrence is not necessarily natural.

At the same time as experimental psychology begins to address multisensory integration of information as it naturally occurs (for example, speech sounds and corresponding visible changes of the articulatory facial musculature), there is also interest in understanding the perception of arbitrarily combined multisensory sources of information (such as music and film). Although these two domains—the natural and artificial combination of multisensory information—overlap considerably, there is an important difference. In the natural environment, the co-occurrence of auditory and visual information is subject to the rules of physics. For example, dropping a stone of particular shape and weight onto a hollow wood box of specific dimensions will create a sound having a particular intensity and quality. (Indeed this audiovisual regularity forms a basis for grading shellfish, such as mussels in Prince Edward Island.) In the artificial world of film and video, however, any recorded sound or music can arbitrarily accompany the impact of a stone with the surface below. Nevertheless, film-music scholarship (Gorbman 1987; Kalinak 1992) and the art of film-score composition (cf. Cohen 1994/1996, 2001; Davis 1999; Karlin 1994; Thomas 1997) reveal that the application of music in film is not arbitrary; certain combinations of film and music are more effective than others. From the experimental psychological perspective, the appropriate application of music in film reflects psychological rules. Critics of film music, composers of film scores, and audience members tacitly understand these rules. It is the job of the experimental psychologist to make such implicit rules explicit. Recent research in the psychology of film music has progressed in that direction. The present article reviews this research and provides a framework to assist future advances.

Music and Experimental Psychology: Structure and Meaning

The experimental psychological perspective has a long history of application to music. It recognizes that music depends on the interacting minds of composer, performer, and listener. In the last two decades, however, there has been an enormous increase of research in music cognition as represented in seminal books by Deutsch (1999), Dowling and Harwood (1986), Krumhansl (1990), and Sloboda (1985). This psychological literature focuses on many aspects of music and the human responses to them. For the purpose of examining the influence of music on film perception from a psychological perspective, it is useful to distinguish between two aspects: structure and meaning.

For the purpose of this article, structure refers to systematic relations among sounds that characterize the style or grammar of music originating within any time period or culture. Music theory often provides useful terminology for describing this kind of musical structure. For example, for music within the Western-European tradition, terms such as interval, triad, scale, tonality, and rhythm are useful. Descriptions based on the physical properties of sound (for example, frequency, intensity, and duration) can also be employed. Experimental psychological questions about music structure address such issues as perceptual grouping (for example, where do listeners naturally segment parts of an unfolding musical piece, Clarke and Krumhansl 1990; Frankland and Cohen 2004), tonality (for example, what tone, triad, or scale is most prominent in a section of a musical piece, Cohen 2000a; Krumhansl and Toivianinen 2001), or memory (for example, what structural characteristics of music facilitate melodic memory, Cohen, Trehub, and Thorpe 1989; Dowling, Kwak, and Andrews 1995). Perceiving the structure of music is necessary for music appreciation, but listeners are seldom directly cognizant of it; in the same way, speakers of a language are seldom cognizant of the rules of grammar that govern their use of language.

Most people are oblivious to the structure underlying the music they appreciate. It is the meaning of music that forms their predominant conscious experience. Meaning includes the emotional aspect as well as the associations that the music brings to mind (Sloboda 1985; Juslin and Sloboda 2001). Psychological questions about musical meaning focus on revealing common interpretations of particular musical passages (for example, Hevner 1936; Juslin 1997; Kamenetsky, Hill, and Trehub 1997; Krumhansl 1997), anecdotal details about the significance of particular musical experiences during one's life (Sloboda 1998; Sloboda and O'Neill 2001), and assessment of examples of music that gives rise to deep emotional experience (for example, Gabrielsson and Lindstrom 1993). Questions about musical meaning typically fall outside the realm of much musicological discourse. This situation arises because musicology has fostered a view of the autonomy of pure music, that music does not have meaning beyond the relations of the sounds themselves (Kivy 1990). But as film theorist Kassabian (2001, ch. 1) points out, such a view is inconsistent with the application of music in film, and with the experiments which show agreement among listeners on what a musical excerpt does and does not imply. Musicologist Nicholas Cook (1998) in *Analyzing Musical Multimedia* takes the stance, in contrast to the "music-alone" view, that "music is never alone." For Cook, music is a source of meaning in search of an object. Nowhere is its function clearer than its application in multimedia presentations of visual images to which musical meaning becomes aligned. In his book, however, Cook (1998) does not take the perspective of an experimental psychologist, but, as I have argued elsewhere, an experimental psychologist could test many of his implicit and explicit theoretical ideas about the integration of music and visual images (see Cohen 2000b for a review).

Structure and Meaning in Other Domains

For the purposes of this article, the breakdown of music into structure and meaning is useful because the same dichotomy can also apply to the visual aspect of a presentation. Again, structure refers to the formal characteristics of visual images, and meaning refers to the associations, feelings, or interpretations that the visual images bring to mind. For example, consider the video image of a person running over a hilly terrain. The structural description would include such aspects as the temporal patterning or rhythm of the motion pattern, the size and direction of the up and down excursions, the pattern of dark and light, and so on. These characteristics are independent of the individual moving agent and could indeed arise from many different agents, be they inanimate or animate, animal or human, male or female, and so forth. The meaning of the image (its interpretation), however, is closely tied to the particular identity of the moving agent, whether it is a car about to collapse due to rough terrain or a lost child wandering over unknown territory. The general distinction drawn here has been described in other theoretical frameworks using different though related terminology. For example, the experimental psychologist Garner (1962) distinguished internal and external structure. Internal structure referred to the relations within a set of events (for example, a matrix of black and white squares), and external structure referred to the relations between the entire event set and external objects (for example, the external object that could be represented by a black and white matrix pattern). Musicologist Leonard Meyer (1956) distinguished between internal (musical pattern) meaning and external musical meaning (associations to real-world events). These examples distinguish the two separate categories of structure and meaning as used in the present article. It is also possible to theorize a single, inclusive but more complex dimension of meaning as Kendall (1999; 2005) has done by introducing specially defined concepts of syntax and areferential vs. association/referential meaning. For present purposes, however, the terms structure and meaning as distinguished above serve our goals well by allowing us to readily refer to distinct properties that interact across media.

Indeed, the dichotomy of structure and meaning, so defined, apply to each of the five domains that comprise film, video, and other electronic multimedia (Stam 2000). Three of these domains are auditory: music, speech, and sound effects, and the other two are visual: the film image and written text. Whereas the present article focuses on the relation between music and the visual image, it is useful to appreciate that these two domains reside within a larger physical and mental context, yet a context that can be analyzed according to this dichotomy: structure and meaning. This is not to suggest materials, but, as will be argued here, it is one way that can help to elucidate the nature of the interactions between various media.

The discipline of experimental psychology entails many subdivisions that can apply to music or other aesthetic domains. Perception focuses on the basic processes underlying hearing and seeing, whereas cognition takes into account the mental processes representing long term experience and their impact on memory (Craik and Lockhard 1972), comprehension (Kintsch 1998) and consciousness (Baars 1997). Social psychology considers how attitudes develop and change (e.g., Bandura 2001; Heider and Simmel 1944). Another field of experimental psychology is that of emotion (for example, Juslin and Sloboda 2001). Whereas these separate fields overlap, an enormous literature and associated experimental methodology have been developed for each of them. A complete psychological theory of film-music would necessarily represent the subdisciplines of perception, cognition, social psychology, and emotion. In addition, such topics as lifespan development (how psychological principles and functions change with age), cross-cultural psychology, gender and individual differences would necessarily be included. One can predict a new subdiscipline of multimedia psychology in which the psychology of film-music would contribute an essential part.

Experiments on the Interaction of Music and Film

Three goals of the experimental-psychological perspective emphasized here are (1) to quantify film-music and film materials that send energy to the eye and ear (known as stimuli); (2) to quantify the mental effects of these stimuli independently and in conjunction; and (3) to describe the effects of the joint (film and music) stimuli in predictive models. These goals can be attained by conducting experiments that examine responses of audience members to specific musical and visual patterns (stimuli) in isolation and in conjunction.

To illustrate further, consider as experimental stimulus materials any two musical and visual excerpts. In one condition of the experiment, these examples would be presented independently (that is, music alone, and visual alone) to volunteer participants who would be asked to judge them for meaning. There is an infinite range of possibilities of music and visual examples. The experimenter would first decide what examples would be most appropriate for addressing the question of interest, and what human responses to these stimuli should be gathered from the participants. Basic principles can be revealed even with very simple materials. For example, Cohen (1993) reported experiments in which stimulus melodies of several seconds in duration were repeating tones and the visual patterns were computer animations of a single object moving up and down on a screen (bouncing ball). As shown in Figures 1 and 2 respectively, melodic notes varied in tempo and pitch height and the bounce of the ball varied in tempo and height. As a measure of meaning, participants used a five-point scale to rate the apparent happiness/sadness of the music and video examples separately presented.

The ratings showed that the judged happiness of the melody background increased for faster tempos and higher pitches, and that judged happiness of the bouncing ball increased for faster tempos of bounce and higher bounces. Thus, the meaning of the music and visual materials was systematically related to physical

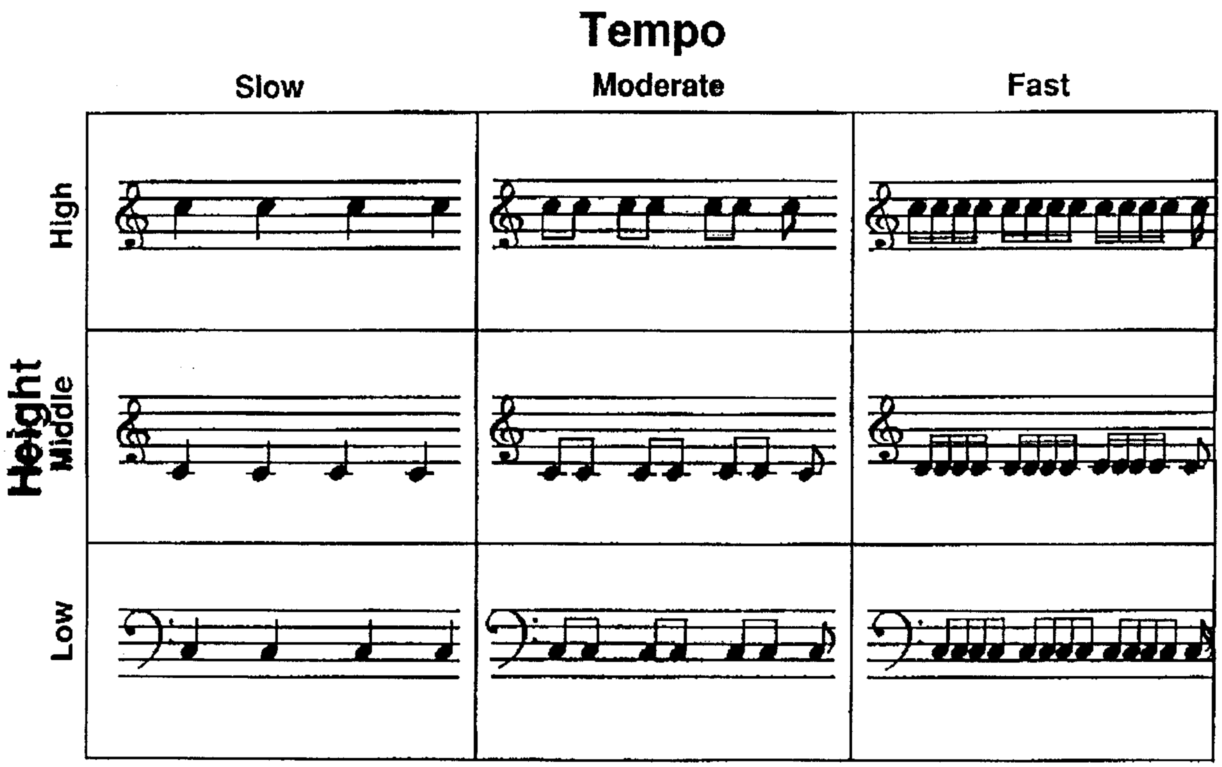

Figure 1. Representation of different musical soundtracks created from three tempos and three pitch heights of a repeating tone.

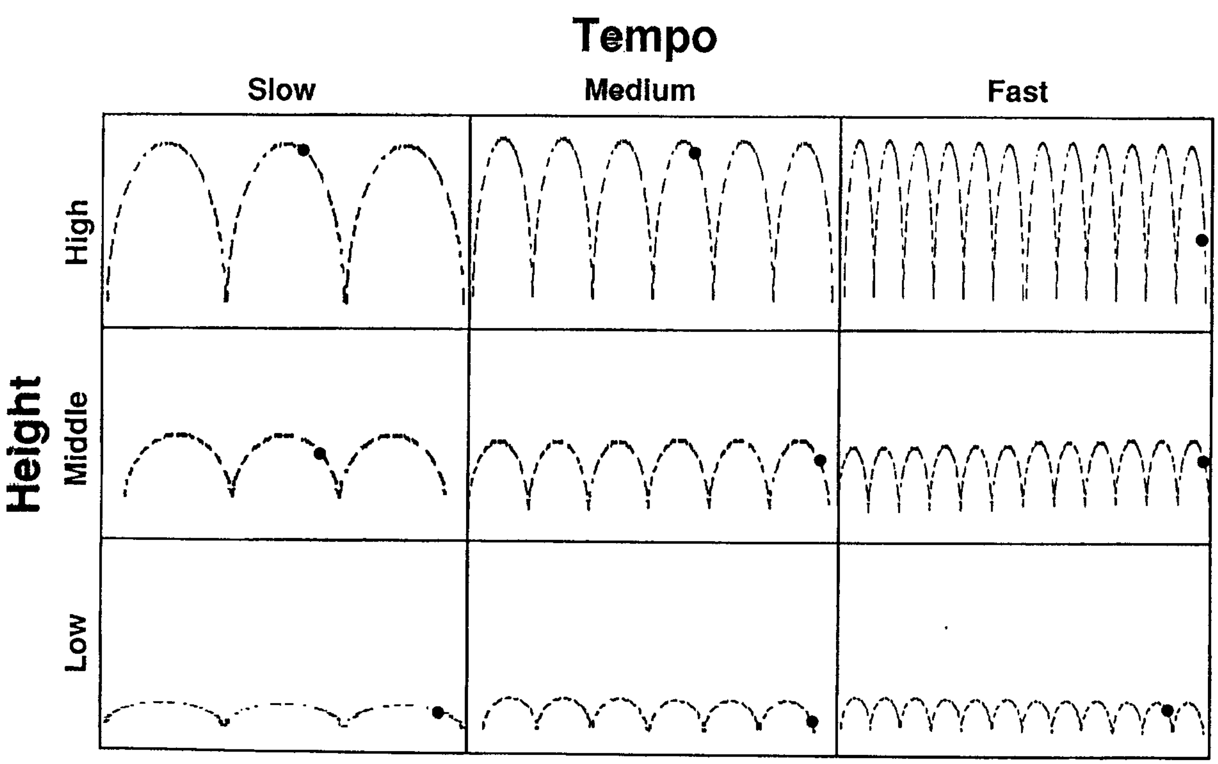

Figure 2. Representation of different video patterns created from three tempos and three heights of a bouncing ball.

characteristics of the sound and light patterns. Such information about the effect of tempo and pitch height on meaning of musical patterns had been previously revealed in early work of Hevner (1936) but had not been illustrated for visual patterns.

Having determined the meaning of the musical and visual stimuli separately, the next step was to determine the meaning of the musical and visual stimuli in combination. Thus, different participants judged many combinations of the musical and visual patterns (for example, the low, slow melody with the low, slow bounce, and the low, slow melody with the high, fast bounce, and so forth). The question of interest was whether the accompanying music could change the original judgments of the happiness/sadness of the bouncing ball. Indeed it was observed that a ball that bounced high and fast was judged as very happy if the background music was high pitched and fast but was judged less happy if the background music was low pitched and slow (see Figure 3). The statistical analysis of the mean ratings of happiness/sadness arising for the different bounce patterns under the different music backgrounds revealed an effect of both the tempo and the pitch height of the music on the judged happiness of the bouncing ball.

The final stage of the experimental process compared the results from the combined music-video condition to the results of the first stage of the experiment that had obtained meanings of the music and video independently. Recall that the combined results showed that happy music (high, fast tones) decreased the level of sadness (for example, low, slow bounce) of the bouncing ball, and conversely the sad music (low, slow tones) increased the level of happiness (for example, high, fast bounce) of the bouncing ball. Thus, the comparison of the unimodal (auditory *or* visual) and bimodal (auditory *and* visual) conditions suggested that the auditory and visual meanings (that is, associations) systematically combined to produce an overall happiness/sadness judgment.

Even an experiment with simple stimuli, such as those described above, can quickly become complex. For example, presenting a melody based on a major or minor triadic arpeggio rather than just a repeating tone creates yet another dimension of musical meaning for study. Indeed, melodies in the major mode add to the level of happiness of the bouncing ball and melodies in the minor mode decrease it. The temporal relation of the musical and visual patterns is yet another type of variation that might influence the degree to which musical and visual associations combine to create meaning. For example, if the temporal relations are in phase, meanings might combine more effectively than if the temporal relations are out of phase. Thus, studies with stimuli of greater complexity than that of a simple monophonic melody and bouncing ball cannot focus exhaustively on all the physical characteristics of the test stimuli. Careful descriptions of the experimental stimuli are often provided by the authors of these studies (for example, Boltz 2001; Lipscomb and Kendall 1994a) and urge us not to lose sight of the fact that the source of musical influence is ultimately the physical and structural characteristics of the music.

In the study of the bouncing ball just described, participants assessed only one kind of meaning (that on the happiness/sadness dimension). Other dimensions of meaning could of course be investigated. For example, Bolivar, Cohen, and Fentress (1994) asked participants to rate friendliness/aggressiveness of social interactions

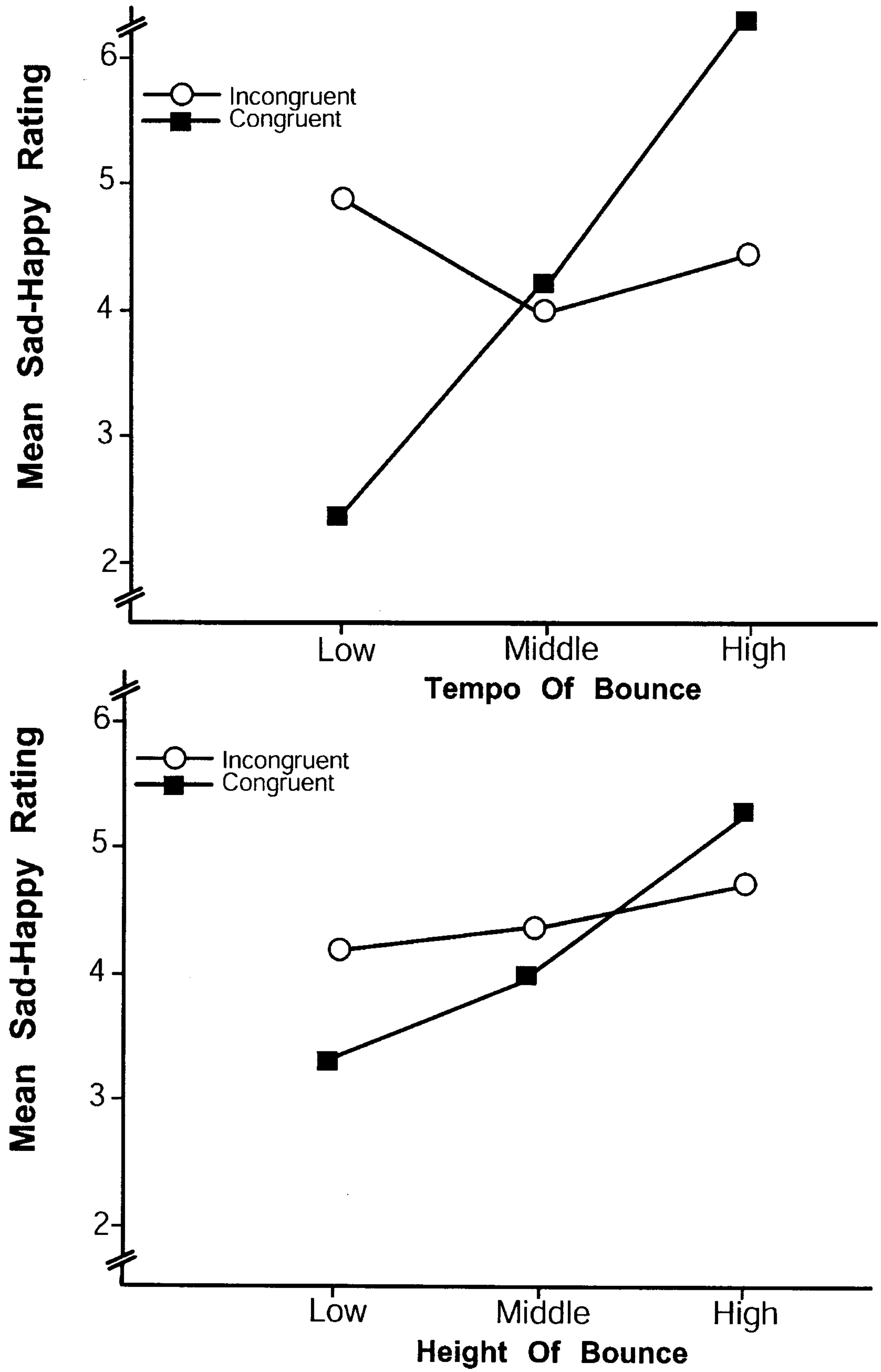

Figure 3. Mean rated happiness of the bouncing ball under congruent and incongruent audiovisual tempo and height combinations (data from Cohen 1993).

among wolves, the particular visual material under investigation. Fundamental work of Osgood, Suci, and Tannenbaum (1957) on the Semantic Differential showed that connotative meanings (that is, affective meaning as compared to dictionary definitions or denotative meaning) depend on three different dimensions: evaluation (for example, good-bad), potency (for example, strong-weak), and activity (fast-slow). The degree to which a bipolar adjective pair (for example, strong-weak) loads on each of the three dimensions has been empirically determined through the statistical method of factor analysis (other techniques such as cluster analysis can also provide similar information). A researcher can select from a published source (for example, Osgood et al. 1957) adjective pairs with previously determined loadings on the three dimensions so as to measure the importance of the three dimensions for the meaning of an audio, visual, or audiovisual excerpt. Thus, Marshall and Cohen (1988), Lipscomb and Kendall (1994a), and Sirius and Clark (1994) used three or four different bipolar adjective pairs to represent each of the three connotative dimensions of meaning.

Iwamiya (1994) took a different starting point and used 22 different bipolar adjective pairs that he felt could be applied to audio and visual stimuli drawn from commercially recorded materials. The rating data were then submitted to factor analysis to reveal a number of dimensions on which the connotative judgments were based. Rather than finding only the three traditional dimensions, his data led to five separate dimensions of meaning that he interpreted as tightness, evaluation, brightness, cleanness, and uniqueness. Similarly, Lipscomb and Kendall (1994a) carried out a cluster analysis of the ratings on ten bipolar adjective pairs and revealed two primary factors, one for evaluation and the other that they referred to as a combination of potency and activity.

It is not important that different dimensions or different numbers of dimensions appear in the studies of the diverse investigators. What is important is that within a study, the same dimensions of meaning can be tracked across different experimental conditions. Evidence has been observed both for additivity of audio and visual meaning along particular dimensions and for a more complex interaction between the audio and visual meanings on particular dimensions.

In the experiment of Marshall and Cohen (1988), a two-minute animation was presented that entailed a large and small triangle and a small circle. The animation had been developed within social psychology by Heider and Simmel (1944) to illustrate that people form stereotypes even of inanimate objects. People who see this short animation generally characterize the large triangle as a bully victimizing the two smaller geometric figures. The question that Marshall and Cohen posed was whether two contrasting examples of background music would alter observers' attitudes toward these figures. Thus baseline information about the meaning of the music and visual stimuli was first obtained on twelve bipolar adjective rating scales (for example, powerful/powerless). For the visual stimulus, the participants judged each of the three film characters (large and small triangle and small circle) and the film overall. In the second part of the study, other participants were presented with one combination of the music and film and judged the film and film characters on the same twelve bipolar adjective ratings. With this information, it was possible to determine whether the meaning of the visual material changed when the music was added to it, and, if so, whether the change was systematically related to the meaning of the music. Judgments of individual characters in the film led to a surprising finding. It appeared that the degree of meaning along a particular dimension ascribed to each of the film characters depended on the particular background music. In particular, the activity rating for the three geometric characters differed significantly under the two music backgrounds, although the overall activity rating of the two kinds of music did not differ.

In relating the change in meaning of the film characters to the different background music, Marshall and Cohen (1988) postulated that an interaction of the temporal structure of the film and music influenced visual attention, such that the focus of visual attention differed under the two different musical backgrounds. Thus, the attributes of the music were differentially ascribed to the visual attentional focus that differed for each of the two pieces of music. They argued that if music through structural similarity directs attention to a particular feature of the film and provides particular connotative information, then this particular connotation can become associated with the attended visual feature (see also Cook 1998: 69 for discussion). This joint process of the emergence of congruent music-visual temporal structure followed by ascription of meaning (associations) led to the postulation of the Congruence-Associationist Framework for understanding the effects of film music in film and video presentations.

Association

Although Marshall and Cohen (1988) focused on both temporal-structural congruence and association, in some cases, association alone is sufficient to account for effects of music in the interpretation of a film. In a further study reported by Cohen (1993), two one-minute feature-film excerpts (a fight scene between two men and an encounter between a man and woman) and two orchestrated music excerpts (suitably entitled "Conflict" and "Say hello to love") provided the stimulus materials. Semantic differential (that is, bipolar adjective rating) judgments were obtained to reflect the evaluation, potency, and activity dimensions for each of the music and video excerpts, and these judgments revealed large differences in meaning between the two music excerpts and between the two film excerpts. When the music and video excerpts were combined, there was a change in meaning for only one of the video excerpts. This excerpt showed an interaction between a man and a woman that could be interpreted as either a romantic or an aggressive encounter. The different background musical examples altered the interpretation systematically and raised or lowered the ratings on the semantic differential scales accordingly. The other visual excerpt of two men fighting was unambiguous, and the music had no strong impact on its meaning. Thus, when a visual excerpt is ambiguous, the associations from the background music can assist in establishing the context and disambiguating the interpretation.

Similarly, ratings of friendliness/aggressiveness of background music systematically influenced the judged friendliness/aggressiveness of the social interactions of wolves depicted in short film clips (Bolivar et al. 1994). Using music and film excerpts having various degrees of perceived finality or closure, Thompson, Russo, and Sinclair

(1994) showed a direct (linear) influence of musical closure on the perceived closedness or finality of a film excerpt. More recently, Boltz (2001) showed that positive or negative music significantly biased the interpretation of film clips drawn from Alfred Hitchcock mysteries. These examples are only some that show rather straightforward additivity of associations of music that influences the interpretation of a film.

Music-Visual Structural Congruence

In spite of the many examples whereby associations from music appear to combine directly with associations generated by the visual information, the connection between music and film may not necessarily be this straightforward in all, or even most, cases. The complication has been introduced earlier when we noted that, in the study by Marshall and Cohen (1988), two music excerpts having the same value on the activity dimension led to different ratings of activity of the three geometric "characters" in the presentation. It was suggested that similar temporal structure of the music and film altered the visual attention pattern and directed attention toward one character under one music soundtrack and to a different character under the other music soundtrack.

The importance of temporal structural congruence has also been addressed in several other studies of effects of the film music. Iwamiya (1994), for example, shifted by 500 ms the audio and visual channels of short excerpts drawn from commercially available laser discs (other bases of mismatch were applied to other excerpts in the materials examined). He asked participants to rate their perceived degree of match of the original and altered versions of the excerpts. Not surprisingly the degree of perceived match of the music and visual information was enhanced in the synchronized as compared to the asynchronized condition. The other form of mismatch that Iwamiya explored entailed presentation of a new music background with the original visual material and this also reduced the degree of matching. Participants also rated audiovisual stimuli independently and in conjunction on 22 rating scales previously mentioned. The influence of higher order factors (for example, cleanness and uniqueness) was direct only when the music and visual materials were well matched; however, for the lower order factor, brightness, the influence on meaning was direct regardless of degree of matching. The results of this study are consistent with a three-stage process in which an analysis of the meaning of the separate audio and visual stimuli is first assessed. This process itself has more than one level, in that some features of the stimuli are simple (for example, brightness) whereas others (for example, uniqueness) require comparisons with past experience for their complete analysis. A second process entails comparison of the properties of the audio and visual stimuli, the outcome of which is an assessment of the degree of matching. A third and final process ascribes information from the audio channel to the visual focus of attention. Information from lower-order audio features, such as intensity, is passed to the output of the visual analysis regardless of the degree of match between the music and visual materials. Information from higher-order auditory analyses involving memory and cognition is passed to the visual analysis only if the music and visual material match each other well. Thus, association alone cannot explain how the music influences the visual interpretation. Audiovisual incongruence (both structural and associationist) produces a bottleneck.

Like Iwamiya (1994), Lipscomb and Kendall (1994a) focused on the degree of matching of music-visual pairs. Their stimuli were derived from the motion picture *Star Trek IV: The Voyage Home*, and all soundtracks were taken from the film score composed by Leonard Rosenman. All possible combinations of five video and music excerpts were created making special effort to temporally align the accent patterns of the video and audio, especially in the twenty combinations that were novel. For each visual excerpt, participants in the experiment were presented with each of the five music soundtrack excerpts and they were asked to select the soundtrack that was the best-fit for the excerpt. Almost all participants chose the excerpt that had been created originally by the composer for the excerpt. Participants and composer obviously perceived the visual and music materials according to the same psychological principles and could detect that the originally intended soundtrack made the best match to the visual excerpt.

Because the authors took special effort to temporally align the soundtracks with the visual accent pattern in the visual clip, it is unlikely that temporal synchrony of the original pair provided the basis for matching. Had the film clip and audio track been deliberately desynchronized, as in the case of Iwamiya (1994), this would have provided a basis for the higher rating of the composer's original soundtrack. Thus, we are left with the possibility that the associations brought to mind by the music originally composed for the excerpt blended better with the associations elicited by the visual excerpt than did the associations elicited by the four other music excerpts arbitrarily assigned to each visual clip. The best match of course might not necessarily result from identity of music and visual associations but rather from the complementarity of the two sources of information.

To further explore the basis of the perceived matching, Lipscomb and Kendall (1994a) asked other participants to provide ratings for the five visual and five music excerpts on ten bipolar adjective dimensions. Mean ratings on the scales differed significantly for different music soundtracks when each film excerpt was held constant. The evaluation judgments were typically highest for the original audio and visual combination confirming the previously observed "best-fit" results. An analysis of the structure of the music suggested that meaning on particular dimensions relied on particular features for the music (for example, high potency ratings arose for complex harmonic structure, rubato, and complex rhythm; high activity ratings arose for allegro tempo and steady pulse). Consideration of all of this information led Lipscomb and Kendall (1994a: 91) to postulate a multistage process to account for the contribution of film music to the overall film production, and more specifically the role of music/visual congruence on attentional focus. They posit the importance of an initial comparison process that examines the relation of the accent structure of the visual and audio material. If these two accent structures match, then attention will be directed to the composite. If the focus of attention is on both audio and visual materials due to this joint accent structure, then associations from the audio source will flow through to the visual source. If there is no match between the audio and visual accent structure, then the direction of attention will shift from the audiovisual pair and associative information from the audio source will not be transmitted further.

Marshall and Cohen (1988), Lipscomb and Kendall (1994a), and Iwamiya (1994) all have focused on the importance of music-visual temporal-pattern matching on the ascription of associations of music to the visual source. Additional studies have explored the role of shared musical-video temporal congruencies. Bolivar et al. (1994) used two examples of each music-visual stimulus excerpt of wolf social interactions. Each film clip lasted only several seconds. Temporal match was greater in one music-visual pair than the other. An increased degree of audiovisual synchrony of these very short excerpts, however, did not increase the influence of the background music on the meaning of the film. Lipscomb (1998, 1999; Lipscomb and Kendall 1994b) reported that synchrony played a role for very simple audiovisual patterns but that as the materials became more complex and realistic, the role of synchrony decreased. Sugano and Iwamiya (1998) have also shown the sensitivity of a perceiver to the synchrony of audio-visual periodicity using very simple computer generated displays.

Capacity-Limited Working Memory

The research reviewed in this article suggests that much of the influence of music on the interpretation of film may be explained by association and by structural congruence. The review, however, also makes clear that not all musical information that might influence the interpretation of film exerts such an influence. The brain is limited in the amount of information it can handle at any one time. In making sense of a film or video, mental processes must relate incoming information to past experience stored in long-term memory. This task requires working memory, of which there is a limited supply (Baddeley 1992; Snyder 2000). Although the focus of the present article is primarily on entertainment media, two education-theoretic models provide a precedent for considering the concept of capacity limitation in the development of a framework for understanding multimedia representation. The first model is provided by Mayer and his colleagues in educational psychology at the University of California at Santa Barbara. The second model comes from Fisch at the Children's Television Workshop in New York.

Mayer's cognitive theory of meaningful multimedia learning incorporates elements from several other cognitive-psychological theories: dual-coding theory (Paivio 1990), cognitive-load theory (Mousavi, Low, and Sweller 1995), and generative theory (Wittrock 1989). Meaningful learning requires that the learner engage in active cognitive processes such as selecting words, selecting images, organizing words, organizing images, and integrating words and images (compare Mayer et al. 1996: 64–66). All of these processes entail building representations be they visual or verbal (auditory) or building connections between the different types of representations or creating logical sequences of them.

Most of the work of Mayer and his colleagues has focused on visual and verbal (auditory) material (for example, Mayer 1989; Mayer and Anderson 1991; Mayer and Moreno 1998). However, Moreno and Mayer (2000) explicitly examined the role of music and sound effects (auditory adjuncts) in the design of multimedia instructional messages. They evaluated two competing hypothetical effects of the role of background music, sound effects, or both on learning. The first hypothesis was that entertaining auditory adjuncts, like music, increase both arousal level and the overall level of attention and, consequently, the amount of processing capacity. The second hypothesis was that extraneous (incoherent) sound requires processing capacity and consequently depletes processing resources. Thus, the hypotheses differ in their effect on working memory and processing capacity.

Moreno and Mayer (2000) examined the effect of adding music and environmental sound to multimedia instructional material presented to students who were learning about lightning or the braking system of a car. Presentation of music as part of the multimedia instruction did not enhance learning as measured by memory for verbal material and transfer of information. The results were interpreted as inconsistent with the arousal hypothesis and as consistent with the coherence hypothesis, that is, that extraneous entertaining audio materials depleted processing resources. The music, however, had been arbitrarily selected and played throughout the entire narrative. This is not the way that music is typically used in film. For film, music is specifically chosen or composed to complement the narrative and it is spotted in only at specific points in the drama. Hence, the negative results of Moreno and Mayer (2000) may well apply to irrelevant sounds, and the authors themselves state that it is necessary to conduct additional studies in which the coordination of the sounds is directly manipulated. They conclude that problems arise due to auditory adjuncts because auditory adjuncts "do their damage by limiting the amount of relevant verbal material the learner selects for processing in working memory and by reducing the learner's resources for building connections between verbal and visual representations of the to-be-learned material" (Moreno and Mayer 2000: 124). Clearly, in their view, music and sound effects that are poorly integrated with the narrative can deplete working memory capacity. At the same time, it can be argued, following Boltz (2001), that music can also assist the efficient use of working memory capacity by providing useful context, or "advance organizers" in the words of Fisch (2000), to be discussed below.

Fisch (2000) proposes a model that focuses on allocation of working memory resources while watching television. The model consists of a theoretical construct with three basic components: processing of a narrative, processing of educational content, and distance, that is the degree to which the educational content is integral or tangential to the narrative. Fisch (2000) developed a model to represent how information about the narrative and the educational content can be simultaneously extracted.

The primary assumption of the model is that comprehension of television draws on the limited capacity of working memory. Working memory is required to encode and retain information from the production and integrate it with knowledge stored in long-term memory. Demands on working memory are reflected in longer response times and poorer performance in secondary tasks that are concurrent with viewing. According to the model, narrative and educational content use the same processing resources. Competition for resources increases with decreasing connection of the story and the educational content. When the educational content is integral to the story, then the parallel processes become complementary rather than competitive, and the comprehension of the educational content is likely to be strengthened.

Based on his capacity model, Fisch (2000) predicts several ways in which television program characteristics would result in greater comprehension and retention of educational content. It is suggested here that several of these characteristics can be implemented via music. Music can be used to foreshadow (Boltz 2001; Boltz, Schulkind, and Kantra 1991) and provide what Fisch refers to as "advance organizers" of both the narrative and the educational content or deeper message. Such advance organizers decrease demands for processing both narrative and educational content respectively. By associating the same music with different aspects of the content (be it narrative or educational) the distance between diverse content can be reduced and consequently decrease competition between resources. Music can increase the interest of a presentation and thereby increase transmission of content (although, the work of Mayer and his colleagues earlier described cautions against such a role for irrelevant music). Fisch (2000) also distinguishes between two kinds of capacity limitations in watching television: those related to more peripheral perceptual and attentional mechanisms, and those related to controlled processing (that is, the ability to thoughtfully process and store information). It can be argued that music can play a role in both, and this role would need to be clarified through future research. To this end, empirical investigations of music and film from a memory perspective have begun to define independent and integrative aspects of the auditory and visual modalities (see Boltz 2004; Cohen 1995).

Congruence-Associationist Framework of the Mental Representation of Multimedia

The present article has reviewed studies that have examined the role of film music on the interpretation of a film or video presentation. It has emphasized two principles: association and structural congruence. Association accounts for the direct transfer of meanings elicited by music to the film context, setting the mood, or disambiguating plot. On the other hand, structural congruence, through principles of grouping, influences attention to specific visual information. It was hence argued that, under certain circumstances, the impact of associative principles was to some extent governed by the structural congruence of audio and visual materials. In other words, associations would be applied to the focus of attention that was under control of structural congruence. However, it was also noted that the musical and visual channels operate within the broader context of other domains such as printed or written text, sound effects, and speech. Drawing on the literature from educational multimedia, the processing of multimedia relies on limited-capacity working memory and feeds back from and forward to long-term memory. This literature also emphasizes that narrative is primary, and the audience member is actively engaged in constructing a narrative. A final assumption, developed elsewhere (Cohen 2001 with respect to film music) is that vision generally predominates over audition. Thus, the narrative created is primarily a visual one (for example, according to film theorist Mitry 1990: 235, "the logical development and principle significations are based on the development of images, not on verbal associations"). Putting all of this together, a framework for representing the processing of multimedia is proposed below as shown in Figure 4.

The framework consists of five parallel channels representing each of the five domains that contribute to multimedia presentations: printed/written text, speech, music, sound effects, and visual images (Stam 2000). Each channel represents a separate but interacting information processing system. At the first stage, surface (physical) information is received by the sense organs and is analyzed at the next stage into structural characteristics (for example, accent patterns, contours, motions) and meaning characteristics (associations brought to mind). At this stage, interactions among the five domains may take place. For example, if cross-domain temporal accents coincide, then attention will be focused at these loci of information. Given the dominance of vision, coincident accents may focus visual attention on that part of the visual scene

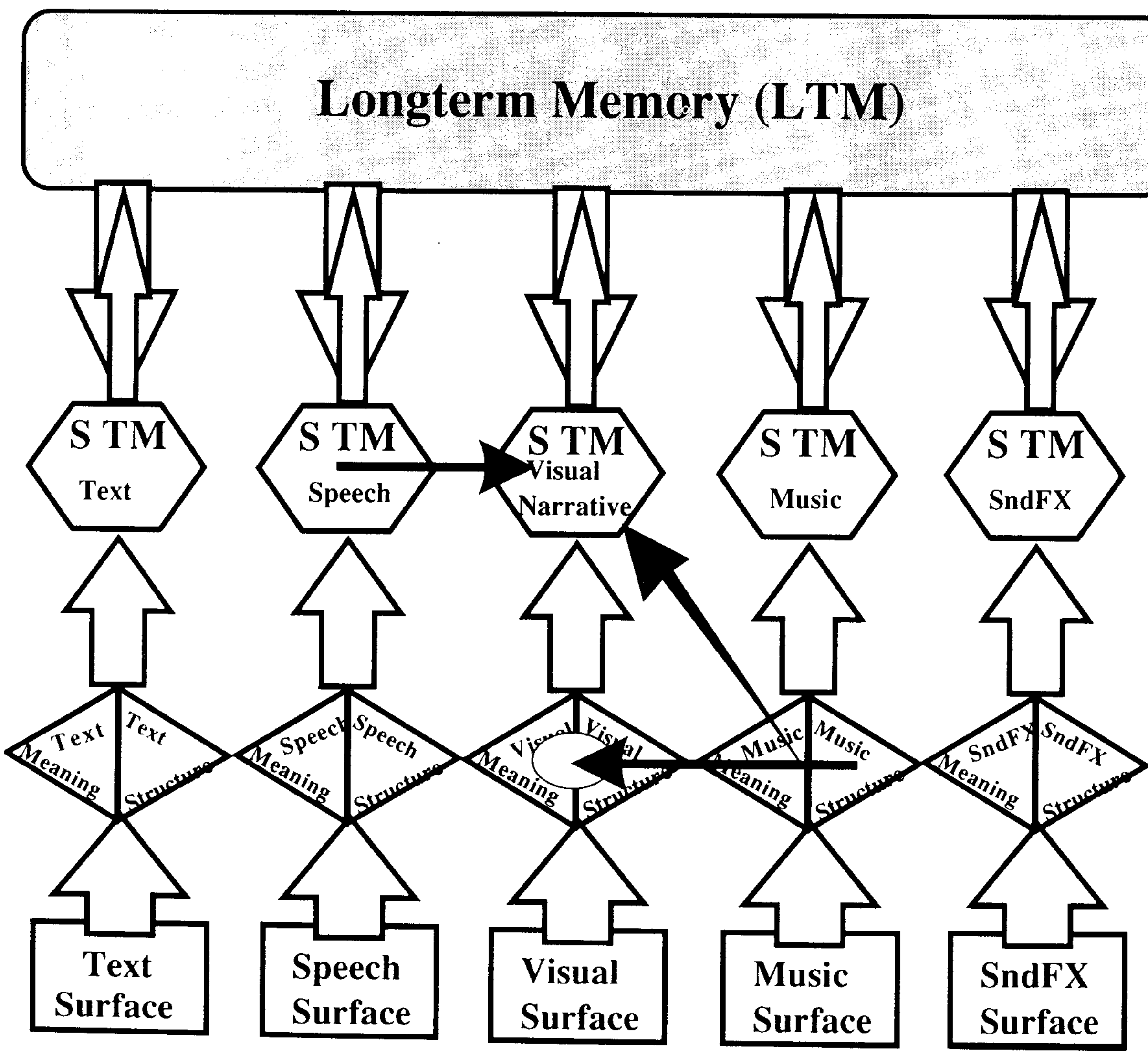

Figure 4. Congruence-Associationist Framework for understanding how music influences the interpretation of film and video.

that is structurally congruent with the music. An example of this visual focus is shown by the arrow in figure 4 from music structure to the visual channel, highlighting here a portion of the total visual information depicted in the shaded oval. At the next short-term memory (STM) stage, preattended and attended information is processed in working memory and to some extent leaks through to long-term memory (LTM) above. LTM is the repository of knowledge gained through lifelong experience. It is the source of inferences and contexts that an individual actively generates in order to make sense of the external world, known initially through the surface information. The inferences generated via LTM are sent down to working memory. Working memory also receives information sent up from the surface. It is here at the site of working memory that the matching process occurs. The best match between the bottom-up information from the surface and the top-down information from LTM enters consciousness (Cohen 2001; Grossberg 1995) and contribute to the *working narrative*. The narrative plays out primarily in the visual domain and information from the other channels feed into it. The working narrative is the audience member's representation of the story based on the cues from the separate media domains that are deciphered with the help of LTM.

Music plays a role in creating the working narrative (in the diagram, the visual narrative) in several ways. It directs attention to certain features in the visual image domain (see the horizontal arrow from music structure), it feeds information directly to the working narrative (see the diagonal arrow directed from music meaning to the STM Visual Narrative), and it provides associations that establish inferences in LTM. For the sake of clarity, only a few of the many possible interactions between domains are shown. Given the limitations of working memory, only some of the available information ever reaches consciousness and the working narrative.

Elements of the present framework were first presented by Marshall and Cohen (1988) and Cohen (1990), who focused primarily on the influence of music-visual structural congruence on attention and the consequent directing of musical meaning to a visual focus. Cohen (1999, 2000a, 2001) expanded this framework to reflect the contribution of long-term memory on the establishment of inferences and the development of a working narrative based on matching between short-term and long-term memory. The present framework presents a more complete context, extending now to the additional domains of scripted text and sound effects. The review of the educational multimedia literature further justifies the emphasis of this framework on the limited capacity of working memory and the role of active cognitive processes in creation of a working narrative.

In conclusion, experimental psychology research on the effects of film music is in its infancy. The present article has reviewed some of the progress that has been made and provides the Congruence-Associationist framework for considering several challenging issues that lie ahead. The framework illustrates an orderly complexity that underlies the influence of music on the interpretation of film, video, or other electronic multimedia. The framework may help both to stimulate new questions and to interpret results of experiments aimed toward understanding the ways in which music enhances communication between the mind of a director and the mind of an audience member.

REFERENCES

Baars, B. J.
1997 *In the Theater of Consciousness: The Workspace of the Mind.* New York: Oxford University Press.

Baddeley, A. D.
1992 "Working Memory." *Science* 255: 556–559.

Bandura, A.
2001 "Social Cognitive Theory of Mass Communication." *Media Psychology* 3: 265–299.

Baron, R.A, Earhard, B., and Ozier, M.
2000 *Psychology*. 2d ed. Scarborough: Allyn & Bacon.

Bolivar, V., Cohen, A. J., and Fentress, J.
1994/1996 "Semantic and Formal Congruency in Music and Motion Pictures: Effects on the Interpretation of Visual Action." *Psychomusicology* 13: 28–59.

Boltz, M.
2001 "Musical Soundtracks as a Schematic Influence on the Cognitive Processing of Filmed Events." *Music Perception* 18: 427–455.
2004 "The cognitive processing of film and musical soundtracks." *Memory and Cognition* 32: 1194–1205.

Boltz, M., Schulkind, M., and Kantra, S.
1991 "Effects of Background Music on the Remembering of Filmed Events." *Memory and Cognition* 19: 593–606.

Clarke, E. F. and Krumhansl, C. L.
1990 "Perceiving Musical Time." *Music Perception* 7: 213–251.

Cohen, A. J.
1990 "Understanding musical soundtracks." *Empirical Studies of the Arts* 8: 111–124.
1993 "Associationism and Musical Soundtrack Phenomena." *Contemporary Music Review* 9: 163–78.
1995 "One-trial memory integration of music and film: A direct test." Paper presented at the Annual Meeting of the Canadian Acoustical Association, Quebec City.
1994/1996 Ed. "The Psychology of Film Music [Special two-issue volume]." *Psychomusicology* 13.
1999 "The Functions of Music in Multimedia: A Cognitive Approach." In S.W. Yi, ed., *Music, Mind, & Science*. Seoul: Seoul University Press, pp. 53–69.
2000a "Film Music: A Cognitive Perspective." In D. Neumeyer, J. Buhler, and C. Flinn, eds., *Music and Cinema*. Middletown, CT:Wesleyan University Press, pp. 360–377.
2000b "Musicology Alone." Review of *Analysing Musical Multimedia*, by Nicholas Cook (Oxford University Press). *Music Perception* 17: 247–260.
2001 "Music as the Source of Emotion in Film." In P. Juslin, and J. Sloboda, eds., *Music and Emotion*. Oxford: Oxford University Press, pp. 249–272.

Cohen, A. J., Trehub, S. E., and Thorpe, L. A.
1989 "Effects of Uncertainty on Melodic Information Processing." *Perception and Psychophysics* 46: 18–28.

Cook, N.
1998 *Analysing Musical Multimedia*. Oxford: Clarendon Press.

Craik, F. I. M., and Lockhardt, R. S.
1972 "Levels of Processing: A Framework for Memory Research." *Journal of Verbal Learning and Verbal Behavior* 11: 671–684.

Davis, R.
1999 *Complete Guide to Film Scoring*. Boston: Berklee Press.

Deutsch, D.
1999 *Psychology of Music*. 2d ed. N.Y.: Academic.

Dowling, J., and Harwood, D.
1986 *Music Cognition*. N.Y.: Academic.

Dowling, W. J., Kwak, S. Y., and Andrews, M.W.
1995 "The Time Course of Recognition of Novel Melodies." *Perception & Psychophysics* 57: 136–149.

Fisch, S. M.
2000 "A Capacity Model of Children's Comprehension of Educational Content on Television." *Media Psychology* 2: 63–91.

Frankland, B. W., and Cohen, A. J.
2004 "Parsing of Melody: Quantifying and Testing Group Preference Rules of Lerdahl & Jackendoff's (1983) *Generative Theory of Tonal Music*." *Music Perception* 21: 499–543.

Gabrielsson, A., and Lindstrom, S.
1993 "On Strong Experiences of Music." *Musik Psychologie*, Band 10: 118–139.

Garner, W. L.
1962 *Uncertainty and Structure as Psychological Concepts*. N.Y.: Wiley.

Gorbman, C.
1987 *Unheard Melodies: Narrative Film Music*. Bloomington: Indiana University Press.

Grossberg, S.
1995 "The Attentive Brain." *American Scientist* 83: 438–49.

Heider, F., and Simmel, M.
1944 "An Experimental Study of Apparent Behavior." *American Journal of Psychology* 57: 243–259.

Hevner, K.
1936 "Experimental Studies of the Elements of Expression in Music." *American Journal of Psychology* 57: 243–259.

Iwamiya, S.
1994/1996 "Interaction Between Auditory and Visual Processing when Listening to Music in an Audiovisual Context: 1. Matching 2. Audio Quality." *Psychomusicology* 13: 133–154.

Juslin, P.
1997 "Emotional Communication in Music Performance: A Functionalist Perspective and some Data." *Music Perception* 14: 383–418.

Juslin, P., and Sloboda, J. A.
2001 *Music and Emotion*. Oxford, UK: Oxford University Press.

Kalinak, K.
1992 *Settling the Score*. Madison: University of Wisconsin Press.

Karlin, F.
1994 *Listening to the Movies*. N. Y.: Schirmer.

Kassabian, A.
2001 *Hearing Film: Tracking Identifications in Hollywood Film Music*. N.Y.: Routledge.

Kamenetsky, S. T., Hill, D. S., and Trehub, S. E.
1997 "Effect of Tempo and Dynamics on the Perception of Emotion in Music." *Psychology of Music* 25: 149–160.

Kendall, R.
1999 "A Theory of Meaning and Film Music." Joint meeting of the Acoustical Society of America and the European Acoustics Association, Berlin. Invited paper.
2005 "Empirical Approaches to Musical Meaning." *Selected Reports in Ethnomusicology*, Vol. 12: 69–102.

Kintsch, W.
1998 *Comprehension: A Paradigm for Cognition*. Cambridge: Cambridge University Press.

Kivy, P.
1990 *Music Alone: Philosophical Reflections on the Purely Musical Experience*. Ithaca, NY: Cornell University Press.

Krumhansl, C. L.
1990 *Cognitive Foundations of Musical Pitch*. N.Y.: Oxford University Press.
1997 "An Exploratory Study of Musical Emotions and Psychophysiology." *Canadian Journal of Psychology* 51: 336–352.

Krumhansl, C. L., and Toivianinen, P.
2001 "Tonal Cognition." In R. J. Zatorre, and I. Peretz, eds., *The Biological Foundations of Music*. N.Y.: New York Academy of Sciences, pp. 77–91.

Lipscomb, S. D.
1998 "Synchronization of Musical Sound and Visual Images: Issues of Empirical and Practical Significance in Multimedia Development." Abstract in *Journal of the Acoustical Society of America* 104: 1780.
1999 "Cross-Modal Integration: Synchronization of Auditory and Visual Components in Simple and Complex Media." Abstract in *Journal of the Acoustical Society of America* 105: 1274.

Lipscomb, S. D., and Kendall, R.
1994a/1996 "Perceptual Judgment of the Relationship Between Musical and Visual Components in Film." *Psychomusicology* 13: 60–98.

1994b "Sources of Accent in Musical Sound and Visual Motion." In I. Deliege, ed., *Proceedings 3ICMPC*. Liege: Belgium, pp. 451–452.

Marshall, S. E. and Cohen, A. J.
1988 "Effects of musical soundtracks on attitudes toward animated geometric figures." *Music Perception* 6: 95–112.

Mayer, R. E.
1989 "Systematic Thinking Fostered by Ilustrations in Scientific Text." *Journal of Educational Psychology* 81: 240–246.

Mayer, R. E., and Anderson, R. B.
1991 "Animations Need Narration: An Experimental Test of a Dual-Coding Hypothesis." *Journal of Educational Psychology* 83: 484–490.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., and Tapangeo, L.
1996 "When Less is More: Meaningful Learning from Visual and Verbal Summaries of Science Textbook Lessons." *Journal of Educational Psychology* 88: 64–73.

Mayer, R. E., and Moreno, R.
1998 "A Split-Attention Effect in Multimedia Learning: Evidence for Dual Processing Systems in Working Memory." *Journal of Educational Psychology* 90: 312–320.

Meyer, L.
1956 *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Mitry, J.
1990 *The Aesthetics and Psychology of the Cinema*. (Abridged edition by Mitry). C. King (trans/1997; original unabridged French ed. 1963). Bloomington, IN: Indiana University Press.

Moreno, R., and Mayer, R. E.
2000 "A Coherence Effect in Multimedia Learning: The Case for Minimizing Irrelevant Sounds in the Design of Multimedia Instructional Messages." *Journal of Educational Psychology* 92: 117–125.

Mousavi, S. Y., Low, R., and Sweller, J.
1995 "Reducing Cognitive Load by Mixing Auditory and Visual Presentation Modes." *Journal of Educational Psychology* 87: 319–334.

Osgood, C.E., Suci, G. J., and Tannenbaum, P.H.
1957 *The Measurement of Meaning*. Urbana: University of Illinois Press.

Paivio, A.
1990 *Mental Representations: A Dual-Coding Approach*. 2d ed. N.Y.: Oxford University Press.

Sirius, G., and Clarke, E. F.
1994/1996 "The Perception of Audiovisual Relationships: A Preliminary Study." *Psychomusicology* 13: 119–132.

Sloboda, J.
1985 *The Musical Mind: The Cognitive Psychology of Music*. New York: Oxford University Press.
1998 "Everyday Uses of Music Listening." In S. W. Yi, ed., *5th ICMPC Proceedings*. Seoul: Seoul National University, pp. 55–60.

Sloboda, J., and O'Neill, S.
2001 "Emotions in Everyday Listening to Music." In P. Juslin, and J. Sloboda, eds., *Music and Emotion*. Oxford: Oxford University Press, pp. 415–429.

Snyder, B.
2000 *Music and Memory*. Cambridge, MA: MIT.

Stam, R.
2000 *Film theory*. Malden, MA: Blackwell.

Sternberg, R. J.
1996 *Cognitive Psychology*. Ft. Worth, TX: Harcourt Brace.

Sugano, Y., and Iwamiya, S.
1998 "On the Matching of Music and Motion Picture Using Computer Graphics and Computer Music." *Proc. 5ICMPC*: 465–468.

Thomas, T.
1997 *Music for the Movies*. 2d ed. Los Angeles: Silman-James.

Thompson, W. F., Russo, F. A., and Sinclair, D.
1994/1996 "Effects of Underscoring on the Perception of Closure in Filmed Events." *Psychomusicology* 13: 9–27.

Wittrock, M. C.
1989 "Generative Processes of Comprehension." *Educational Psychologist* 24: 345–376.

Perspectives in Systematic Musicology

Editors
Roger A. Kendall & Roger W. H. Savage

Department of Ethnomusicology
University of California, Los Angeles

Cover Painting: Vasily Kandinsky, Composition 8, July 1923, Solomon R. Guggenheim Museum, New York, © 2005 Artists Rights Society (ARS), New York/ADAGP, Paris.

Cover Design: Robin Weisz/Graphic Design

ISBN: 0-88287-056-4

Selected Reports in Ethnomusicology is a refereed series published by the Department of Ethnomusicology, University of California, Los Angeles.

Address inquiries to:
Ethnomusicology Publications
Department of Ethnomusicology
University of California
2539 Schoenberg Music Building
Los Angeles, CA 90095-1657

Contributors

Paul Attinello is a lecturer in the International Centre for Music Studies at the University of Newcastle upon Tyne; he has also taught at the University of Hong Kong. His 1997 dissertation from UCLA considered the aesthetic implications of European avant-garde vocal music from the 1960s. He has published in the *Journal of Musicological Research, Musik-Konzepte, Musica/Realtá, MLA Notes,* the revised *New Grove,* and several collections.

Annabel Cohen (B.A. McGill; M.A., Ph.D. Queen's University) is a professor of psychology at the University of Prince Edward Island and an adjunct professor at Dalhousie University in Halifax, Nova Scotia. Her research focuses on the effects of music in multimedia (sponsored by SSHRC) and the acquisition of musical grammar (sponsored by NSERC). She serves on the editorial boards of *Music Perception, Musicae Scientiae, Psychomusicology, Psychology of Music,* and is a Fellow of the Canadian Psychological Association.

Frank Heuser is the director of the music education program at UCLA where he teaches methods courses, supervises student teachers, and conducts an outreach program that brings violin instruction to elementary students and provides Music Department undergraduates with early field experiences. His research includes work in the motor activation patterns in the facial muscles of musicians playing brass instruments, in the relationship between inner hearing and performance skills, and in music teacher education. He serves as a guest conductor and clinician for public school ensembles and is a wind and brass instructor for the Idyllwild Arts summer music festival.

Roger A. Kendall is a professor in the systematic musicology program in the Department of Ethnomusicology at UCLA. He has published numerous scientific research articles and book chapters in music perception and cognition, psychoacoustics, and acoustics. His computer program, the Music Experiment Development System (MEDS), permits flexible and real-time design of experiments in music science and is used internationally.

Scott D. Lipscomb, Ph.D., is an associate professor in the School of Music at Northwestern University, jointly appointed to both the music education and music technology programs. In addition to his primary research interest in film music perception, he is currently collaborating on a variety of investigations related to the surround sound presentation of cinema and musical sound, the affect of music in video game contexts, tonality judgment in elementary school children, and the development of interactive instructional media to enhance the music learning experience. Dr. Lipscomb is currently serving his second term as president of the Association for Technology in Music