Psychomusicology, 13, 28-59

SEMANTIC AND FORMAL CONGRUENCY IN MUSIC AND MOTION PICTURES: EFFECTS ON THE INTERPRETATION OF VISUAL ACTION

Valerie J. Bolivar
Dalhousie University, Halifax, Nova Scotia, Canada

Annabel J. Cohen
University of Prince Edward Island, Charlottetown, P. E. I., Canada

John C. Fentress
Dalhousie University, Halifax, Nova Scotia, Canada

It is proposed that the interpretation of action in film depends on the combination of semantic (i.e., meaning) and formal (e.g., temporal) information across auditory and visual channels. A series of experiments was conducted to examine the listener-viewer's use of semantic and formal audiovisual information. In all experiments, subjects were presented with 32 combinations of music excerpts and video clips. The audio and visual materials had been selected for their friendly or aggressive meaning as judged by independent groups of listeners or viewers. In Experiment 1, subjects judged the semantic congruency of the 32 audiovisual pairs. In Experiment 2, they assessed temporal congruency and in Experiment 3 they assessed friendly-aggressive video meaning. In Experiment 1, mean judged semantic congruency was predicted by the prior independent friendly-aggressive judgments of the audio and visual components. In Experiment 2, subjects agreed significantly on the degree of temporal congruency of audio and visual components although the judgment was also influenced by semantic congruency. In Experiment 3, ratings of the friendliness-aggressiveness of the video excerpts were predicted by multiple regression of the *a priori* semantic ratings for the independent visual and audio components. The results were discussed in terms of contrasting models of the integration processes for semantic and formal audiovisual congruency proposed by Marshall and Cohen (1988) and Boltz, Schulkind, and Kantra (1991).

Film music flourished during the era of the silent film from 1900-1930 (Anderson, 1988; Marks, 1979). The advent of the "talkies" in 1928 made speech and sound effects audible and reduced the special role for background music of providing meaning through an auditory channel. An exception to this is the case of "post-talkies" films which feature animals. In such presentations, as in silent films, the protagonist (e.g., *Old Yeller, The Bear, The Black Stallion, Phar Lap*) cannot communicate verbally. Much is still communicated via visual action. The present study uses film excerpts of animals as a source of visual material for examining effects of music in motion picture perception. This corpus of material stands midway between unrealistic animations and realistic films with human interaction and dialogue. With these materials, which offer both control and external validity, the present study explores the extent to which music can either enhance or detract from the visually expressed intent.

Studies in the psychology of animal behavior reveal that the movements of animals can systematically communicate complex social information (e.g., Golani, 1992; Havkin & Fentress, 1985; Moran, Fentress, & Golani, 1981). In such studies, it is often necessary for researchers to analyze hundreds of visual sequences to determine the common elements of a particular social exchange. The present research arose, in part, out of a serendipitous observation by one of the authors (J.F.), who found that background music from a radio seemed to "go with" videos of animal tapes which were being viewed in the laboratory. It then followed that if this randomly chosen music seemed to match the video, then perhaps it might also be influencing its interpretation.

Direct Effects of Background Music

The notion of the influence of background music on interpretation coincided with recent work in the psychology of film music. Some of this work has suggested that music directly affects the interpretation of video;[1] that is, the meaning of the audio systematically influences the meaning of the video—the exact meaning of the music is ascribed to the film. For example, background music can alter interpretations of abstract paintings (Lindner & Hynan, 1987). Similarly, Cohen and Szeto (1995) have shown that background music directly affects the meaning of a simple geometric figure. In their study, a computer-animated ball bounced at one of three heights (low-middle-high) and one of three tempos (slow-middle-fast) and a simple background melodic accompaniment was either low, middle or high in pitch and slow, moderate or fast in tempo. Judgments of the sadness-happiness of the ball in isolation were happier for greater heights and tempos. Similarly, happiness of the melody increased with increasing pitch and tempo. In combination, the resulting judgment of happiness of the ball was directly affected by the audio dimensions. In other words, a happy ball (e.g., high, fast bounce) was judged as less happy if presented with a slow, low melodic accompaniment.

Another demonstration of the direct effect of music meaning on film meaning comes from a study by Marshall and Cohen (1988). They presented subjects with a short animation featuring three geometric figures. The presentation was accompanied, for different groups of subjects, by one of two musical backgrounds or no musical background at all. Other groups of subjects heard only one of the two musical backgrounds in isolation. All subjects were asked to judge either the film or the musical backgrounds on semantic differential scales which provided measures of meaning on the Activity, Potency and Evaluative dimensions (following Osgood, Suci & Tannenbaum, 1957). It was observed that the meaning of the musical backgrounds differed from each other on all three dimensions. Differences on two of these dimensions, Activity and Potency, seemed to directly affect the change in ratings on the corresponding dimension of the film. In other

words, highly active music increased the Activity judgment of the film and high potency music increased the Potency judgment of the film. The influence on the Evaluative dimension, however, was not direct. Instead, this judgment was more likely based on a cognitive assessment of how well the music actually fit the presentations and on the quality of compositions. Thus, it appeared, that in judging a film overall, the meanings on the Activity and Potency dimensions rather simply contributed to the original meaning of the film on these two dimensions.

Direct effects of music on film meaning were also evident in another study reported by Cohen (1993). In this study, two short contrasting excerpts from a realistic film were presented. One excerpt depicted a fight between two men, the other showed a male and female interaction which was ambiguous with respect to aggressive or friendly intent. Two soundtracks contrasting in meaning were used as backgrounds. Subjects rated the video presentations on semantic differential judgments and judged the appropriateness of various titles for these same excerpts. Their judgments showed that the background music altered the meaning of the video presentation directly—in the direction of independently derived title choices and rating scales for the background music—although the effect was strong only for the more ambiguous excerpt.

These studies suggest that the meaning of background music adds to the meaning of the visual information. Although this can be explained through activation of temporally contiguous schemata, the situation is likely more complex. For example, although Marshall and Cohen (1988) observed that Activity and Potency dimensions of the music appeared to directly affect judgments of the film on those dimensions, this was not found for the Evaluative dimension. Moreover, a more complex picture emerged from results of a second task carried out by the subjects. In this part of the experiment, the subjects provided ratings for each of the three geometric characters in the film (a small triangle, a large triangle and a circle). If the effect of the music meaning was direct in this case, then all characters accompanied by the music with high Potency ratings would have been judged as having higher Potency ratings than the same characters accompanied by the music with low Potency ratings. This was not the case. Effects of the musical background on the meaning of the individual characters were nevertheless apparent. For example, with high Potency music, the small triangle received a significantly higher Activity rating than with low Potency music. To account for this result it was argued first that some aspect of the high Potency music directed attention to the small triangle. Second, since music also had a high Activity rating, this rating was consequently ascribed to the triangle. Thus, the music directly affected the meaning of *part* of the film, that part to which the music had drawn attention. Therefore, it was suggested that the music altered the meaning of the film and did so by first directing attention to certain parts of the film. But how did music direct attention to a part of the film?

Congruence-Associationist Model

It was presumed that the mechanism by which music altered attention was based on structural audiovisual similarities (Gestalt principle of grouping by similarity) and that the mechanism by which music altered meaning of the attended object was Associationist (principle of temporal contiguity). Marshall and Cohen (1988) therefore proposed a Congruence-Associationist theory to account for these results. Congruence dealt with the formal or structural information of the music which directed attention to particular parts of the film. For example, simultaneous audiovisual temporal accent would draw attention to that portion of the visual display which shared the auditory accent pattern. In this account, congruence preceded the Associationist processes and relied on the basic influences of amodal grouping principles which could account for differential visual attention.

Consistent with this Congruence-Associationist model is Spelke's (1979) research on infants. She observed longer viewing times and differential viewing directions in the presence of congruent visual and auditory structures. The role of audition in orienting visual attention is also supported by the research reviewed by Stein and Meredith (1994) in their book, *The Merging of the Senses*. For example, auditory reaction times are shorter than visual and evoked potentials are of higher amplitude for combined than individual modalities. The authors suggest that combined modalities increase salience and decrease ambiguity.

The Associationist (in other common terminology, Activation) component of the Marshall and Cohen's (1988) model dealt with the direct semantic effects. In this case, temporal congruity brought meaning from the music into association with that part of the visual display to which attention was directed. Marshall and Cohen (1988) differentiated three processes: (1) the generation of meanings by music, (2) selective attention prompted by formal congruencies between music and film, and (3) the association between meanings and the attended film items (see Figure 1a). With regard to the second process, Marshall and Cohen had emphasized the importance of *structural* (formal) similarities of music and film as the basis for selective visual attention.

Boltz, Schulkind, and Kantra (1991), in a study of effects of background music on memory for film took another view and suggested that it was congruent *meaning* in the music and film that directed attention to the formal features of the stimuli. In their study, music either accompanied the plot outcome or preceded the same scene. Music and video materials were selected so as to clearly represent either positive or negative plot outcomes. One group of listener-viewers received semantically congruent film and music, while another received semantically unmatched film and music. It was observed that recall of accompanied outcomes was superior when music and film were congruent in meaning, but the reverse was true for musically foreshadowed outcomes, in which violated expectancies led to improved memory. These effects were readily accounted for in terms of independent effects on selective attention consistent with principles in the so-

cial cognitive psychology literature. Boltz et al. (1991), however, proposed that shared *affective meaning* of the music and film could direct attention to *shared formal features* which in turn would "reinforce" and "engender" the affective meaning. In other words, in contrast to Marshall and Cohen's (1988) emphasis, shared audiovisual meaning directed attention to shared audiovisual structural characteristics–which in turn reinforced the shared meaning (see Figure 1b). The claim of Boltz et al. (1991) is consistent with the fact that structural correlates of mood are invariant across music, speech intonation, facial expression and walking gait. Thus, if both music and film mood are positive in affect, there is a likelihood that a fast tempo of both music and visual motion will be observed. On the basis of past experience, such information is expected and sought. When the structural information is found it reinforces the original affective meaning.

The two interpretations make different predictions of the effect of background music if the degree of temporal congruence of film and music is systematically varied. Following Marshall and Cohen (1988), the greater temporal congruence the greater the focus of visual attention to which the meaning of the music consequently can be ascribed. For Boltz et al. (1991), independent judgments of the meaning of film and music should inspire a search for particular formal features in the film and music. In spite of mis-

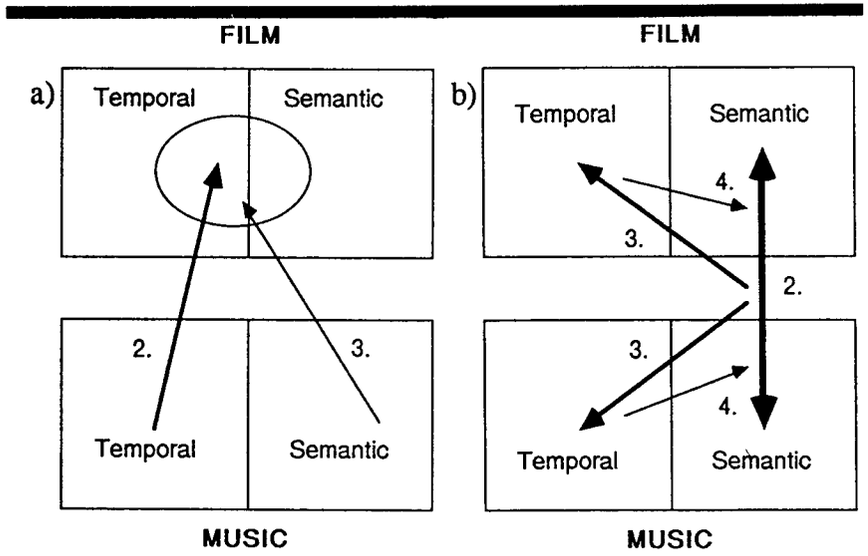

Figure 1. Schematic representation of Models of Semantic and Formal Feature Interaction. (a) Congruence-Associationist Model of Marshall and Cohen (1988). At 2, audiovisual temporal congruence establishes a visual (film) focus of attention (the oval) to which, at 3, the semantic meanings of the music are ascribed. (b) Shared Semantic Structural Search Model, extending ideas from Boltz, Schulkind & Kantra (1991). Semantic congruence at 2 elicits, at 3, a search for temporally congruent structure which, in turn at 4, reinforces the original semantic meaning.

alignment of temporal accent, these features should be equally accessible under the incongruent and congruent case and the effect of the music on film interpretation should be no greater under congruent than incongruent conditions.

The Present Study

The present study draws attention to the issues of information extraction and combination in complex dynamic audiovisual media. It sets out first to examine the ability to judge semantic congruence of audiovisual pairs. It secondly focuses on the ability of listeners to judge temporal congruence of audiovisual pairs. Finally, it attempts to predict the degree of shift of video meaning accompanied by music as a function of both semantic and temporal congruence.

The focus on congruency judgments might be regarded as taking a step backward from the theoretical advance of Boltz et al. (1991), who showed effects of semantic congruence on memory with the assumption that listener-viewers were capable of judging semantic congruence from the beginning. However, it is theoretically and practically significant to understand the processes underlying the semantic congruence judgments which have been claimed to influence memory. Whereas Boltz et al. (1991) provide new evidence of the effects of background music on memory, the process of extraction of meaning, on independent or combined dimensions, was incidental to their primary concern. They took effort to validate the positive or negative meaning of their excerpts with the assistance of three judges. Careful description of these materials was provided. Nevertheless, the processes by which the audiovisual meaning was assessed and combined was to some extent overlooked. It is also of interest to know the extent to which this information is available to awareness, particularly since it has been suggested by film theorists and from anecdotal observations that music in film often operates below the conscious level (Gorbman, 1987). The interpretation of movement can also occur with little conscious awareness of salient features (Moore & Yamamoto, 1988). The listener-viewer's ability to judge congruency would further attest to the ability to integrate information from the two sources.

The present study takes a step back also from a study by Lipscomb (1990) which showed that listener-viewers could judge which of several possible soundtrack excerpts "belonged" to particular visual excerpts of a feature film. Subjects agreed among themselves and for the most part this judgment coincided with that of the composer. Ratings of the sound and visual excerpts were not examined independently, so predictions based on independent meanings of the visual and audio information could not be made. In Lipscomb's (1990) study, both semantic and formal factors could potentially influence the judgment. The present study attempts to assess more carefully these two different aspects.

By manipulating both semantic and temporal congruency in the present study, we are able to compare Marshall and Cohen's (1988) versus Boltz et

al.'s (1991) views regarding whether formal (MC) or semantic (BSK) congruence guides selective attention. Audio and video excerpts were combined under two different temporal relations (judged in terms of temporal synchrony of accents in the animals' movements and the music) in order to answer two questions: Would listener-viewers agree on which of the pair was more temporally congruent and would the degree of temporal congruency influence the extent to which the music altered the visual meaning? Following the hypothesis of Marshall and Cohen (1988), it was predicted that audio components would influence video interpretation as a function of the degree to which the audio and visual components were temporally synchronized. In the event that such a finding was not observed, there would be support for the proposal of Boltz et al. (1991) that attention to formal features was directed by a priori assessment of audiovisual meanings rather than the other way around.

EXPERIMENT 1

Semantic Congruency

In the first Experiment we examined how well subjects could determine the semantic congruency between video excerpts and their musical accompaniments. In the study by Boltz et al. (1991), three individuals coded film and music excerpts with regard to their congruence in mood (i.e., happy plot outcome versus sad outcome). Agreement among these judges led to the selection of the stimuli in their main study of effects of music memory. The present study emphasized perception and interpretation rather than memory. Thus, the process of evaluating the congruence of meaning in the stimuli itself was of interest. The study was designed to test the hypothesis that judged audiovisual congruence could be predicted from a priori judgments of semantic meaning of the independent audio and visual components.

Method

Development of Materials

Visual. The visual selections were those used in a series of studies examining human perceptions of social interactions between pairs of wolves (Anderson, Hill, Ryon, & Fentress, in press; Anderson, Ryon, & Fentress, 1994). The 20 social interactions (2 to 13 s in duration) included close-ups of wolf faces as well as more distant views of chases, body slams, and play sequences. These interactions had been previously characterized as aggressive or friendly by individuals with at least two years of experience studying wolf social behavior. All 20 interactions (8 friendly, 8 aggressive and 4 ambivalent) were used in our initial pilot study. The displays were recorded in random order on VHS tape with 15 s pauses between interactions.

Although previous research (Anderson et al., in press; Anderson et al., 1994) has produced ratings of friendliness-aggressiveness for each of the 20 social interactions, it was deemed useful to obtain ratings for the visual selections by a sample of subjects from the same population as that which would participate in the main experiments. In the video pilot study, 31 university undergraduate subjects were instructed to watch the series of 20 wolf social interactions and to rate each one in terms of friendliness-aggressiveness, using a 5-point scale where 1 = very friendly, 2 = slightly friendly, 3 = neutral, 4 = slightly aggressive and 5 = very aggressive. Video playback was accomplished with an RCA VHS video cassette recorder and a Hitachi 21-inch screen monitor. Four sequences rated as most aggressive (mean ratings, 3.7, 3.8, 4.2, & 4.8) and four rated as most friendly (mean ratings, 1.4, 1.6, 1.8, & 2.0) were selected for use in the later experiments. These interactions included a variety of activity levels ranging from relatively inactive facial expression interactions to more active body slams, and play sequences. Further information about the stimuli has been provided in the appendix.

Audio. The music selections contained no lyrics and came from compact disc recordings of commercially used excerpts for broadcast referred to as stings, links, promos and jingles. To determine friendliness-aggressiveness ratings for these selections, a pilot study was conducted. In this pilot study subjects were instructed to listen to a series of 20 music selections (8 - 32 s) and to rate each one in terms of friendliness-aggressiveness using the same 5-point scale used by subjects when rating the wolf social interactions. The music selections were played by a Panasonic S-XBS portable stereo system. A total of 31 undergraduate students from an introductory psychology course at Dalhousie University participated in exchange for credit points. The five selections rated as most aggressive (mean ratings 4.1, 4.2, 4.3, 4.5, & 4.8) and five rated as most friendly (mean ratings 1.4, 1.4, 1.5, 1.5, & 1.6) were selected for use in the later experiments. Because of the need to meet temporal congruency constraints, there were two more audio than visual examples. Further information about the stimuli can be found in the appendix.

Combined audiovisual stimuli. For each of the eight video excerpts, two audio excerpts were chosen. One of these was semantically congruent (e.g., friendly audio/ friendly video) and the other was semantically incongruent (e.g., friendly audio/ aggressive video). The particular audio choice also satisfied the constraint of well-matched visual and auditory temporal structure (i.e., temporal accents of the visual action and the music matched). This resulted in 16 different visual-audio temporally congruent examples. For each of these combinations, a second combination was produced which altered the temporal relation (in terms of temporal synchrony) between the two components so as to represent a less temporally congruent pair. The process of combining the auditory and video materials was carried out with professional audiovisual editing equipment by an experienced research assistant/technician. Satisfaction of the temporal alignment constraint was accomplished through trial-and-error and patience and was based on the intuition and subjective judgment of the technician. With eight video excerpts, each paired in four ways with background music, a total of 32 audiovisual pairs were produced. Table 1, columns 1 through 5, represent the

structure of these materials and their a priori semantic ratings on the friendly-aggressive scale.

The visual and audio excerpts were recorded on VHS tape with 5 - 6 s pauses between selections. Five different random orders were created. The previously mentioned video cassette recorder and monitor were used for playback in this experiment.

Subjects

There were 36 (16 men and 20 women; mean age = 20.11, *SD* = 4.21) undergraduate students from an introductory psychology course at Dalhousie University who participated. Subjects were given extra course credit for their participation.

Procedure

Each subject was presented with all 32 audiovisual combinations in one of the five random orders. The subjects were tested in small groups (from 4 - 8) and were from 3.5 - 7.0 m from the screen. The sound level was comfortably audible. Subjects were told that they would be watching a series of wolf social interactions with accompanying music. Their task was to watch each interaction in its entirety and then rate on a 3-point scale (3 = yes [they match], 2 = cannot tell and 1 = no [do not match]) how well the mood of the music matched the mood of the social interaction between the two animals. A 3-point scale was used in this experiment instead of the 5-point scale used in the pilot studies, as it was a concern that subjects might be overly conservative in their ratings and might display a tendency to select ratings in the middle of the scale. It was expected that a 3-point scale might encourage use of all the choices. Testing took approximately 20 min.

Results

The mean semantic congruency ratings for the 32 audiovisual pairs are listed in Table 1, column 6. It can be seen that higher values are generally associated with a priori semantic congruency (column 2). If the mood of the background music was congruent with the mood of the video, subjects were more likely to report semantic congruency (M = 2.52, SD = .44) than if the two were incongruent (M = 1.71, SD = .45).

To test the significance of the effect of the experimental semantic congruency manipulation on judged semantic congruency, a 2 x 2 x 2 analysis of variance was completed with video mood (aggressive or friendly) and semantic congruency as the within subjects variables and sex as the between subjects variable. There was a significant main effect of semantic congruency, $F(1,34) = 240.31$, $p < .001$.

Of less interest is the observation that the mood of the video also had a significant but smaller effect on ratings of semantic congruency, $F(1,34) = 8.28$, $p < .01$. If the video was aggressive, subjects were more likely to say that the mood of the video and its background music were congruent (M = 2.20, SD = .53) than if the video was friendly (M = 2.03, SD = .66). This asymmetrical relation is reflected in the significant video meaning by semantic congruency interaction, $F(1,34) = 8.03$, $p < .01$, depicted in Figure 2.

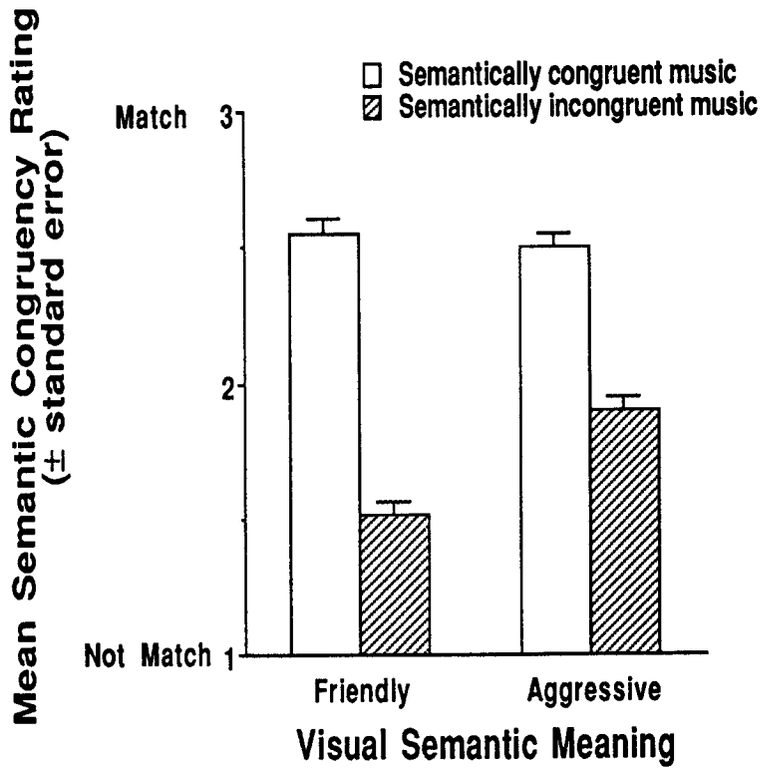

Figure 2. Judged semantic congruency ratings (± standard error) for the two types of video displays (friendly, aggressive) with either semantically congruent or incongruent background music.

Table 1

Summary Table of Means for Pilot Studies and All Three Experiments for all 32 Audiovisual Stimulus Combinations

Video Semantic Condition[a]	Audiovisual Semantic Congruency[b,c]	Audiovisual Temporal Congruency[b]	Pilot Video Rating & (Excerpt #) (V)	Pilot Audio Rating & (Excerpt ID) (A)	Expt 1 Semantic Congruency Rating	Theoretical Semantic Congruency (4-|V - A|)	Expt 2 Temporal Congruency Rating	Expt 3 Semantic Rating
F	+ (F)	+	1.4 (1)	1.5 (a)	2.39	3.9	1.45	1.86
F	+ (F)	-	1.4 (1)	1.5 (a)	2.00	3.9	1.30	1.71
F	+ (F)	+	1.6 (2)	1.4 (b)	2.64	3.8	2.28	1.59
F	+ (F)	-	1.6 (2)	1.4 (b)	2.61	3.8	2.25	1.44
F	+ (F)	+	2.0 (3)	1.4 (c)	2.83	3.4	2.70	1.54
F	+ (F)	-	2.0 (3)	1.4 (c)	2.83	3.4	2.68	1.37
F	+ (F)	+	1.8 (4)	1.4 (b)	2.64	3.6	2.42	1.63
F	+ (F)	-	1.8 (4)	1.4 (b)	2.44	3.6	2.00	1.68

(Continued)

F	- (A)	+	1.4 (1)	4.2 (f)	1.11	1.2	1.48	1.97
F	- (A)	-	1.4 (1)	4.2 (f)	1.31	1.2	1.25	1.95
F	- (A)	+	1.6 (2)	4.3 (i)	1.11	1.3	1.70	2.37
F	- (A)	-	1.6 (2)	4.3 (i)	1.14	1.3	1.50	2.13
F	- (A)	+	2.0 (3)	4.1 (g)	1.97	1.9	2.42	1.90
F	- (A)	-	2.0 (3)	4.1 (g)	2.17	1.9	1.70	2.02
F	- (A)	+	1.8 (4)	4.5 (j)	1.81	1.3	2.08	2.41
F	- (A)	-	1.8 (4)	4.5 (j)	1.53	1.3	1.92	2.59
A	- (F)	+	3.7 (5)	1.4 (c)	2.25	1.7	2.50	2.97
A	- (F)	-	3.7 (5)	1.4 (c)	2.36	1.7	2.05	3.11
A	- (F)	+	3.8 (6)	1.5 (d)	2.61	1.7	2.90	2.76
A	- (F)	-	3.8 (6)	1.5 (d)	2.39	1.7	2.78	2.78
A	- (F)	+	4.8 (7)	1.6 (e)	1.17	0.8	1.45	4.63

(Continued)

A	- (F)	-	4.8 (7)	1.6 (e)	1.00	0.8	1.45	4.44
A	- (F)	+	4.2 (8)	1.5 (d)	1.78	1.3	2.45	3.90
A	- (F)	-	4.2 (8)	1.5 (d)	1.61	1.3	1.78	3.75
A	+ (A)	+	3.7 (5)	4.2 (f)	2.31	3.5	2.30	3.68
A	+ (A)	-	3.7 (5)	4.2 (f)	2.14	3.5	2.02	3.51
A	+ (A)	+	3.8 (6)	4.1 (g)	2.64	3.7	2.80	3.67
A	+ (A)	-	3.8 (6)	4.1 (g)	2.00	3.7	2.38	3.40
A	+ (A)	+	4.8 (7)	4.8 (h)	2.58	4.0	2.85	4.59
A	+ (A)	-	4.8 (7)	4.8 (h)	2.83	4.0	2.82	4.71
A	+ (A)	+	4.2 (8)	4.8 (h)	2.78	3.4	2.98	3.97
A	+ (A)	-	4.2 (8)	4.8 (h)	2.72	3.4	2.90	3.98

[a]video excerpt categorization: F = friendly, A = aggressive
[b]congruency categorization: + = congruent, - = incongruent
[c]audio excerpt categorization: F = friendly, A = aggressive

If the moods were actually congruent (either both friendly or both aggressive) subjects were likely to rate them as congruent and the particular meaning had little bearing on this judgment. When subjects were presented with situations of semantic incongruency, however, they were more likely to say that semantic congruency existed (make an error) when the video was aggressive than when it was friendly.

In order to determine the extent to which the congruency judgment could be accounted for in terms of independently obtained a priori values of meaning for the audio and visual components, a theoretical measure of semantic congruency for each of the 32 audiovisual pairs was calculated (Table 1, column 7) from the pilot study data in which the audio and video components had been rated in isolation (Table 1, columns 4 and 5, respectively). The score for each friendly-aggressive rating of either an audio or a video selection could range between 1 and 5. If both video and audio excerpts received the same rating, the difference between them would be 0.0. The absolute value 4.0, is the maximum difference possible between audio and video components. So that higher theoretical score values reflected higher congruency and lower scores reflected lower congruency, the absolute difference between the ratings for an audio (A) and video (V) pair was subtracted from the maximum possible difference of 4.0. Thus for the case of minimally different (i.e., identical) ratings (e.g., $A = 2.0$, $V = 2.0$), the theoretical congruency value now would be 4.0 (e.g., $4.0 - |2.0 - 2.0|$); whereas for cases of maximally different ratings (e.g., $A = 1.0$ and $V = 5.0$), the theoretical value would be 0.0 (e.g., $4.0 - |1.0 - 5.0|$). Using the first line of Table 1 as an example, the video alone rating from the pilot study is 1.4 and the audio alone rating is 1.5. The absolute value of the difference between the two ratings was 0.1, which when subtracted from 4.0 equals 3.9 (the theoretical semantic congruency rating).

The Spearman correlation between the theoretical semantic congruency rating and the ratings from Experiment 1 was .7 ($p < .01$), showing that relative degree of semantic congruency can be predicted from independent audio and visual judgments.

Discussion

These results indicate that subjects can assess the degree of audiovisual semantic congruency. They are able to both identify and compare meaning from visual and auditory channels. Of greater significance is the fact that subjects' judgments of semantic congruency can in large part be predicted from independent semantic ratings of the audio and video components. Although subjects were able to detect semantic congruency, under some conditions they were more likely to make "errors." For instance, if the video was aggressive and the audio friendly, subjects were more likely to judge the moods as congruent than if the video was friendly and the music aggressive. In principle, this might have been explained if the theoretical congruency for incongruent aggressive videos had been higher on average than for incongruent friendly videos. The mean theoretical congruency values for

both of these conditions were, however, virtually identical. Perhaps the result may be accounted for in terms of several factors: greater salience of aggressive as opposed to friendly videos, the dominance of video over audio, and the tendency of audio meaning to be assimilated by salient video meanings. In any case, this complex effect was weaker than the main effect of semantic congruency to which listener-viewers showed sensitivity.

EXPERIMENT 2

Formal (Temporal) Congruency

Experiment 1 had shown that listener-viewers systematically judged the semantic congruency of audiovisual excerpts. In the present experiment we examined how well subjects could determine an aspect of the formal (temporal) congruency between the music and the video action. Previous research by Marshall and Cohen (1988) had led to the ad hoc suggestion that temporal audiovisual congruence between the musical background and action in part of the visual display would bring that part of the visual display to the focus of attention. This hypothesis was incorporated as an initial phase of the Congruence-Associationist hypothesis. Boltz et al. (1991) had hypothesized that because mood is typically associated with particular kinds of motion, mood itself would prime the perception of such specific motion types. In the Boltz et al. (1991) framework, temporal congruence plays no role. Thus, it might follow that listener-viewers could judge inconsistently the temporal congruence of audiovisual pairs. Boltz et al. (1991) suggest that since meaning, which is first extracted, guides the search for structure, it might follow that more salient meanings would guide the search for structure more strongly. Assuming that congruent meanings are more salient, they might therefore lead to higher ratings of temporal congruence.

The present study examined the sensitivity to temporal cross-modal congruence and tested the postulate fundamental to Marshall and Cohen's (1988) model that listener-viewers would show agreement in their judgments. At the same time, we tested the postulate, following Boltz et al. (1991), that congruent meaning indirectly would lead to high temporal congruence judgments.

Method

Subjects

Subjects were undergraduate students from an introductory psychology course at Dalhousie University ($N = 40$; 18 men and 22 women; mean age = 20.02, $SD = 4.56$). Each subject was given extra course credit for participating in the experiment. Subjects were tested in small groups of five or six individuals.

Materials

The same visual and auditory sequence combinations were used in this experiment as were used in Experiment 1. The same basic questionnaire was used to obtain subjects' congruency ratings. This time, however, subjects were asked if there was temporal congruency between the music and the animals' movements, that is, did the music and film go together with respect to timing and accent patterns.

Procedure

The procedure was virtually identical to that of the first experiment except that subjects were told to decide whether the timing of the music matched the timing of the animals' movements in the display. Testing time was approximately 20 min.

Results

Each audiovisual pair had been presented twice, each different from the other only on the basis of temporal congruency. Means were calculated for each of the 32 audiovisual combinations and for any particular combination the version with the highest temporal congruency rating was categorized as more temporally congruent, the other version as less congruent. (This method uses the subjects' judgments as the basis for the category placement rather than physical features or the technician's choice—neither of which seemed more viable in this situation.) Following this organization of temporal congruency, a 2 x 2 x 2 x 2 analysis of variance was completed with video mood, semantic congruency and formal (temporal) congruency as within-subjects variables and sex as the between-subjects variable. The mean temporal congruency ratings for all 32 audiovisual pairs are presented in Table 1, column 8.

The average rating for versions considered more temporally congruent of the 16 pairs was 2.30 ($SD = .55$) and for versions less temporally congruent was 2.05 ($SD = .55$). On a 3-point scale, where the range of ratings is 2.0 scale units, this difference represents 12.5% of the entire available scale. The difference between temporally congruent and temporally incongruent versions of the audiovisual displays was significant, $F(1,38) = 25.75$, $p < .001$.

Subjects, however, were also more likely to say that temporal congruency was present if the moods of the video and background music were congruent ($M = 2.38$, $SD = .49$) than if they were incongruent ($M = 1.96$, $SD = .55$), $F(1,38) = 40.63$, $p < .001$. In this case the difference represents 21% of the available scale.

The mood of the video display also influenced ratings of temporal congruency, $F(1,38) = 39.98$, $p < .001$. If the interaction between the wolves was aggressive, subjects were more likely to say that the audio and video were temporally congruent ($M = 2.40$, $SD = .47$) than if the video displayed a friendly interaction ($M = 1.94$, $SD = .56$).

A reanalysis of Experiment 1 was conducted to examine the effect of formal (temporal) congruency on ratings of semantic congruency. This analysis revealed a slight overall trend for temporal congruency to increase subject ratings of semantic congruency, $F(1,34) = 3.75$, $p < .061$. If the audio and video were temporally congruent, subjects were more likely to rate the moods

as congruent (M = 2.16, SD = .63) than if the audio and video were temporally incongruent (M = 2.07, SD = .58). This result was independent of whether or not the moods of the audio and video were actually congruent.

Discussion

These results indicate that individuals agree about whether an audio and video pair are temporally congruent or temporally incongruent. This supports a fundamental postulate of the Congruence-Associationist theory of Marshall and Cohen (1988). However, both the meaning of the video and the semantic congruency of the film and its background music influenced subjects' judgments of temporal congruency. In general, if the video was aggressive subjects were more likely to say the audio and video were temporally congruent. Subjects were also more likely to say that the film and music were temporally congruent when the moods were congruent. This finding is consistent with the Boltz et al. (1991) model. Mood congruence would lead to stronger searches for temporal features typically associated with that mood. Putting this another way, the salience of the semantic congruency influenced judgments of temporal congruency. The effect was asymmetric (i.e., semantic judgments in Experiment 1 were not significantly biased by temporal information, although this approached significance), but this may have been due to the small range of temporal incongruency represented by the stimulus pairs. Moreover, for several of these pairs both members received almost identical ratings.

EXPERIMENT 3

The Effects of Semantic and Formal Audiovisual Congruency on Semantic Judgment

Experiment 1 had shown that listener-viewers agreed on the degree of congruency in meaning of audiovisual combinations. This implied that when presented with an audio and visual pair, they were able to assess the independent meaning of each and to compare these meanings in order to provide the congruency judgment. This implication was supported by the positive correlation between the theoretical and actual measures of semantic congruency. Because it was clear that listener-viewers were sensitive to meaning from both audio and visual sources, the question remained as to whether the judged meaning of the visual source would be necessarily influenced by the presence of the audio source. In other words, we have shown that listener-viewers can attend simultaneously to both sources of information. Now we ask if this is necessarily the case; that is, can they resist the influence of audio and focus on visual meaning alone?

Experiment 2 had revealed that listener-viewers agreed on the degree to which audiovisual excerpts were temporally congruent. However, judged temporal congruency seemed to be contaminated by both meaning congruency and absolute visual meaning. That is, subjects tended to consider excerpts to be temporally congruent when they were congruent semantically

and when the visual information depicted aggression. Marshall and Cohen (1988) had proposed that temporal congruency could play a role in the impact of the meaning of film music on film interpretation. Let us assume that a video display contains several independently moving agents. Following Marshall and Cohen (1988), in such a display, background music may be temporally congruent with some of these moving agents and not with others. It follows that the temporal congruency might direct attention to these particular temporally congruent agents in the display, and that the meaning of the film music could be attributed to these agents rather than to the non-temporally congruent agents. It may also be that the specific impact of musical meaning in a film varies with the availability of a visual focus of attention. It could further be hypothesized that the influence of music on the overall meaning of the display (as opposed to the meaning of any particular agent) would be greater in the condition of greater temporal congruency. The present stimulus materials allowed us to examine the latter hypothesis. However, because temporally contrasting stimulus examples in the present study differed only by 12.5% of the available scale on average, the potential effect of the variable might be limited.

Experiment 3 was designed to determine the effects of both semantic and temporal information on semantic interpretation of video information. If viewers could not ignore the audio channel, then their judged degree of friendliness and aggression of the wolves in each excerpt would be accounted for in terms of independent meanings of both the visual and auditory information in the example. More specifically, semantically congruent visual examples should increase ratings (absolute values) of friendliness or aggressiveness whereas semantically incongruent visual examples should reduce ratings. Furthermore, the semantic effects of the audio information would be greater for more temporally congruent examples.

Method

Subjects

Sixty-three undergraduate students (25 men and 38 women; mean age = 21.06, SD = 6.72) enrolled in an introductory psychology course at Dalhousie University participated. Subjects were given extra course credit for their participation and were tested in groups of five or six individuals.

Materials

The videotape sequences of aggressive and friendly social interactions and background music were the same as those used in the previous two experiments. In this experiment the 5-point questionnaire (used in the pilot studies) was utilized again. In previous ratings of the wolf social exchanges, subjects used all five choices on the scale, and hence a more restricted scale was not necessary.

Procedure

The procedure was similar to that of the previous experiments. Subjects were shown the 32 audiovisual sequences in random order (each subject received one of five possible random orders) and were asked to rate each of the wolf social interactions (i.e., only the meaning of the visual information) on a scale from 1 - 5. Testing took approximately 20 min.

Results

The mean friendliness-aggressiveness ratings for each of the 32 audiovisual pairs can be found in Table 1, Column 9. A 2 x 2 x 2 x 2 analysis of variance was completed with the same between subjects and within subjects variables as those in Experiment 2.

As expected, the original friendly-aggressiveness meaning of the social exchange in the video had a main effect on the ratings, $F(1,61) = 318.56$, $p < .001$. Subjects viewing aggressive social exchanges tended to give higher ratings, that is, more aggressive ratings ($M = 3.74$, $SD = .51$) to those videos than they did to friendly ones ($M = 1.88$, $SD = .56$). Of greater importance, semantic relation between audio and video influenced the judgments, $F(1,38) = 65.55$, $p < .001$, as depicted in Figure 3. The overall effect depicted in Figure 3 was also evident for each of the eight video excerpts. In all four friendly videos, the semantically congruent music led to higher friendliness judgments than did the semantically incongruent music; and in all four aggressive videos, the semantically congruent music led to higher aggressiveness judgments than did the semantically incongruent music. Although overall semantic incongruency influenced ratings of the social interactions, aggressive exchanges were less influenced by friendly music than friendly exchanges were influenced by aggressive music. There was no main effect of temporal congruency on ratings of the wolf social interactions nor did it interact with any of the other variables.

To examine the relation between independent ratings of the wolf social exchanges and the ratings of these clips in the audiovisual context of Experiment 3, a Spearman rank order correlation coefficient was calculated ($r = .80$, $p < .01$). The same statistical test revealed a smaller yet significant correlation between the audio ratings alone and the ratings of the social interactions in Experiment 3 ($r = .50$, $p < .01$). The degree to which the audio and video components could predict the final ratings of the wolf social exchanges was examined using a multiple regression (StatView II). The outcome of the multiple regression accounted for approximately 90% of the variance associated with the ratings in Experiment 3, a statistically significant portion of the variance ($p < .001$). As can be seen in Table 2, audio and video components both emerge as significant predictors of friendliness-aggressiveness ratings in Experiment 3. When multiple regressions were computed on the 16 temporally congruent pairs and the 16 non-congruent pairs separately, these results did not change.

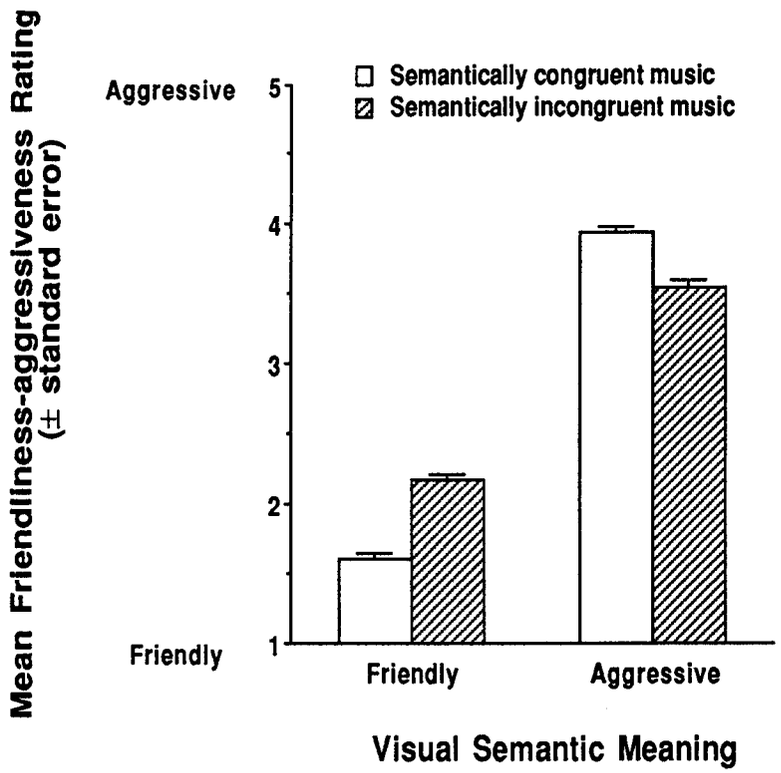

Figure 3. Friendliness-aggressiveness judgments (± standard error) for the two types of video displays (friendly, aggressive) with either semantically congruent or incongruent background music.

Table 2

Summary of Multiple Regression Analysis Predicting Friendliness-aggressiveness Ratings in Experiment 3 from Video and Audio Ratings in Pilot Studies (Beta Coefficient Table).

Variable	Coefficient	Standard Error	Standard Coefficient	*t*-value	Probability
Intercept	.1				
Video	.8	.042	.9	18.5	.0001
Audio	.2	.036	.2	4.8	.0001

Discussion

The results of Experiment 3 reveal the importance of background music when making judgments about the visual contents of a film. Music can influence subjects to modify their ratings of the mood of the animals' social exchanges. Of added significance is the fact that subjects had been instructed to rate the wolf social exchanges and had not been told to attend to the audio component. The affective meaning of music, as demonstrated by other researchers, including Hevner (1936), Rigg (1964) and Scherer and Oshinsky (1977), can influence the interpretation of information attended to via the other sensory channels. This finding is in agreement with those of Cohen (1993) and Marshall and Cohen (1988). The present study goes further by showing that one can predict the ratings of subjects when exposed to the audio and video in combination, given the ratings of both of these components in isolation. The present study also emphasizes that even when subjects were told to rate information via the video channel, information from the audio channel systematically interfered with their interpretations.

The effect of incongruent background music is less pronounced when the video is aggressive than when it is friendly. Perhaps the aggressive social interaction is more salient for the observers and information from other sensory channels that conflict with it either is ignored or is assimilated. The importance of monitoring visually mediated aggression in nature may "block" or even "draw in" counter information, hence making the music manipulation less effective in these instances. This phenomenon is similar to that observed by Cohen (1993), in which music had little influence on a fight scene, but had much influence on a scene ambiguous with respect to the degree of hostility or playfulness depicted. It appears that when a scene provides only one interpretation, the meaning of the music is not needed or called upon for clarification, especially if this one interpretation is associated with aggression.

General Discussion

Taken together, these experiments confirm systematic interactions between visual and auditory channels in the evaluation of visual information. Interpretations of the mood of wolf social interactions were influenced by the mood of the accompanying music. Although music has been widely used in the film and advertising industries, its effect on interpretations in a film has not often been studied scientifically. Only a few studies (Marshall & Cohen, 1988, Thayer & Levenson, 1983) have examined how characteristics of music can influence judgments of the film or its "characters." The results of the present experiments support their findings and also those of Lindner and Hynan (1987), who reported the influence of music on interpretations of paintings. Music does have the ability to modify the visual information attended to by subjects. Whereas past research has focused on effects of background music on interpretations of inanimate geometric objects at one extreme, or complex human interactions at the other, no previous study has examined its effects on interpretations of the mood of social

interactions between animals. Our methodology has the advantages of being more realistic than geometric figures and more controlled than interactions involving human actors. As well, similar effects of congruent and incongruent music were seen for each of the eight different animal actions shown.

The effect of temporal congruency on the interpretations is not as easily characterized by these results. In the second experiment, subjects showed agreement on judgments of temporal congruency; however, these judgments seemed to be influenced more by semantic congruency and degree of salience in the meaning than by temporal information. This may have occurred because of the low differentiation on the temporal synchrony dimension of the stimulus. However, the effect of temporal congruency did not resurface in the final experiment. So when asked to compare the audio and video portions subjects were influenced by the temporal incongruency; but when asked to rate the video exclusively they were not influenced by it. This finding is consistent with Boltz et al. (1991) but not with Marshall and Cohen (1988). As described earlier, Boltz et al. (1991) suggest that congruent meaning elicits a search for structurally consistent features which when found, will reinforce the original meaning. The outcome of this search would be unaffected by audiovisual temporal congruence. Marshall and Cohen (1988) suggested that audiovisual grouping by temporal similarity will focus attention on particular visual aspects to which the audio meaning is then assigned. Because temporal congruence forms the basis for visual attention, the degree of temporal congruence would control the effect of music on the video interpretation. Perhaps, however, there is a broad temporal window in which effects can occur and where subjects often compensate for audio and video components that are "slightly off." Future research could examine the size of this window and determine at what point timings of the audio and video become far enough apart that the subject's congruency and semantic judgments are influenced. It is also important in future research to address the issue of the various types of temporal congruence and their effects on the ratings of the accompanying video. For instance, temporal congruence could be altered in terms of rhythm and tempo as well as temporal synchrony in order to determine which manipulation has the greater effect and under what circumstances. Perhaps through the use of music written specifically to match the animal's movements one can manipulate characteristics of the musical selections more specifically. The issue of temporal congruence and its effects on cognitive processes is an important one remaining to be systematically addressed. Assistance from new digital techniques for the control of the synchrony/asynchrony and accent/non-accent dimensions should also facilitate this endeavor.

Finally, several effects of the visual meaning require some explanation. In Experiment 1, for incongruent semantic conditions, aggressive videos were associated with higher judgments of congruency than friendly videos. In Experiment 2, aggressive videos led to higher judgments of temporal congruency and in Experiment 3, aggressive videos tended to be less influ-

enced by incongruent music. A general construct of *semantic salience* may help to account for these diverse findings. Information may differ in terms of the extent to which it captures attention. In the present study, only the friendly-aggressive dimension of meaning was assessed. A wider set of semantic differential scale judgments on the present stimuli might have provided evidence for variation in salience. Other visual stimuli which were high in Potency, Activity and Evaluation may have been less influenced by music overall. The dominance of certain meanings may also be accompanied by the dominance of visual over the audio sense. In the present study the dominance of the visual depiction of aggression is at least consistent with the weaker effects of audio background in these instances. The high, although unfounded congruency judgments of incongruent audio information for aggressive videos may have resulted from an effect of visual salience which "captures" auditory information and assigns it similar temporal and semantic characteristics. Visual capture is a well-known phenomenon which accounts for the localization of a voice track with the visual image in a film (e.g., Radeau & Bertelson, 1974). The present dominance of aggressive videos may be explainable in terms of extensions of this notion. To test this, it would be necessary to ask listeners to rate the audio track in the presence of the video.

In conclusion, our research findings illustrate how one can effectively predict a priori the effect of music on interpretation of video. Our approach to the study of the effects of background music on interpretation of film events relies on the independent ratings by human subjects for video and audio excerpts in isolation instead of predictions based on specific structural characteristics. Ultimately, one would want to make predictions based on the analysis of specific physical properties of visual and musical form which account for the ratings. In the case of animal social interactions the movements are very complex temporally, making such analysis very difficult. Detailed analyses of animal movements are currently being perfected which will enable detailed examination of temporal aspects of the movements (Bolivar, Coscia, Danilchuk, Fentress & Manley, 1991; Bolivar, Manley, Fentress & Danilchuk, 1995; Fentress, 1992), and perhaps these techniques will enable more refined video analyses and hence enable the creation of more specifically structured musical accompaniments.

References

Albersnagel, F. A. (1988). Velten and musical mood induction procedures: A comparison with accessibility of thought associations. *Behavior Research Theory, 26*, 79-96.

Anderson, G. B. (1988). *Music for the silent films: 1894-1929. A guide.* Washington: Library of Congress.

Anderson, R. E., Hill, B. L., Ryon, J., & Fentress, J. C. (in press). Attitudes and the perception of wolf social interactions: Some implications for public information programs. In L. N. Carbyn, S. H. Fritts, & D. Seip (Eds.), *Ecology and conservation of wolves in a changing world (Proceedings of the second North American symposium on wolves).* Edmonton: Canadian Circumpolar Institute.

Anderson, R. E., Ryon, J., & Fentress, J. C. (1994). *Friendly and agonistic wolf social interactions: The eye of the naive beholder.* Manuscript in preparation.

Bolivar, V. J., Coscia, E., Danilchuk, W., Fentress, J. C., & Manley, K. (1991, November). *Development of movement patterns in neurological mutant mice.* Poster presented at the 21st Annual Meeting of the Society for Neuroscience, New Orleans, LA, U.S.A.

Bolivar, V.J., Manley, K., Fentress, J. C., & Danilchuk, W. (1995). *Measurements of early motor development in weaver mutant mice: A quantitative approach to movement analysis.* Manuscript in preparation.

Boltz, M., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory and Cognition, 19*, 593-606.

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

Cohen, A. J., & Szeto, K. (1995). *Children's audiovisual integration in an emotional rating task.* Manuscript in preparation.

Fentress, J. C. (1992). Emergence of pattern in the development of mammalian movement sequences. *Journal of Neurobiology, 23*, 1529-1556.

Golani, I. (1992). A mobility gradient in the organization of vertebrate movement: The perception of movement through symbolic language. *Behavioral and Brain Sciences, 15*, 249-308.

Gorbman, C. (1987). *Unheard melodies: narrative film music.* Bloomington: Indiana University Press.

Havkin, Z., & Fentress, J. C. (1985). The form of combative strategy in interactions among wolf pups (*Canis lupus*). *Zeitschrift fur Tierpsychologie, 68*, 177-200.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48*, 246-268.

Lindner, D., & Hynan, M. T. (1987). Perceived structure of abstract paintings as a function of structure of music listened to on initial viewing. *Bulletin of the Psychonomic Society, 25*, 44-46.

Lipscomb, S. D. (1990). *Perceptual judgment of the symbiosis between musical and visual components in film.* Unpublished master's thesis. University of California, Los Angeles.

Marks, M. (1979). Film music: The material, literature, and present state of research. *Notes, 36*, 282-325.

Marshall, S. K., & Cohen, A. J. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception, 6*, 95-112.

Moore, C. L., & Yamamoto, K. (1988). *Beyond words. Movement observation and analysis.* Philadelphia: Gordon & Breach Science Publishers.

Moran, G., Fentress, J. C., & Golani, I. (1981). A description of relational patterns of movement during 'ritualized fighting' in wolves. *Animal Behaviour, 29*, 1146-1165.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *Quarterly Journal of Experimental Psychology, 26*, 63-71.

Rigg, M. G. (1964). The mood effects of music: A comparison of data from four investigators. *Journal of Psychology, 58*, 427-438.

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

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*, 626-636.

Stein, B. E., & Meredith, M. A. (1994). *The merging of the senses.* Cambridge, MA: MIT Press.

Thayer, J. F., & Levenson, R. W. (1983). Effects of music on psychophysiological responses to a stressful film. *Psychomusicology, 3*, 44-52.

Footnote

[1]The present paper considers musical semantics from the point of view of the connotative and denotative information provided by the musical stimulus. It is also thought that music systematically alters mood; that is, how a person feels (e.g., Albersnagel, 1988). The relation between mood induction and affective communication is not clear. Exposure to pleasant music may produce associations of pleasantness, thereby invoking a more positive demeanor. On the other hand, someone who is in a poor mood may listen to happy music and understand that this music conveys by convention the message of happiness. Nevertheless, this may have no impact on how the listener actually feels. The impact of music on demeanor, although possibly relevant to the present study, is not addressed here.

Author Note

This research was supported by a grant from the Social Sciences and Humanities Research Council to A. J. Cohen, a grant from the Natural Sciences and Engineering Council to J. C. Fentress, and a doctoral fellowship from the Social Sciences and Humanities Research Council to V. J. Bolivar.

We would like to thank Rita Anderson and three anonymous reviewers for their helpful comments, Jenny Ryon for providing us with the original wolf social interaction tapes, Debora Dunphy for her many hours of technical assistance, Kevin Manley and Wanda Danilchuk for graphics assistance, and Richard Parncutt for discussion of music selections.

Correspondence concerning this article should be sent to Annabel Cohen, Department of Psychology, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada, C1A 4P3; Email: acohen@upei.ca.

(Manuscript received March, 1993; revision accepted October, 1994.)

Appendix: Description of Stimulus Materials

Video

Three experts, with at least four years of experience interpreting animal behavior patterns, rated each excerpt with regard to activity level (low, middle, high), and tempo (slow, moderate or fast). They also indicated whether the video clip was dynamic or static, and whether it was of the face or body of the wolves. The excerpts were presented in one random order. There was no disagreement among the raters. Figure 4 provides the summary of these results, collapsing over the four presentations of each excerpt. Aggressive excerpts were more active, faster in tempo and more dynamic. Verbal descriptions of each excerpt are also provided in Table 3.

Audio

Three listeners with musical expertise beyond entrance requirements for a university music degree program assessed the selections. The excerpts were presented in a random order from one of the test tapes and were repeated from two to four times. The listener was asked to complete a three-point rating scale regarding the following categories: activity level, tempo, static/dynamic, syncopation, instrumentation (1 instrument - many instruments) pitch range, consonance, number of successive notes, number of simultaneous notes. For all dimensions except static/dynamic, the mean ratings differentiated the friendly and aggressive music as shown in Figure 5. It is interesting to note that the direction of the difference is opposite to that found for video on two of the same dimensions examined for both modalities. Moreover, static/dynamic differentiated the video but not the audio excerpts. Musical excerpts judged as friendly were characterized by high consonance and high pitch, high rate of successive notes, and faster tempo than were pieces judged as aggressive, however, there was no simple formula for relating the musical structure to the meaning. This was also seen in the orchestral sketches of each excerpt and the verbal descriptions provided by each listener (see Table 4).

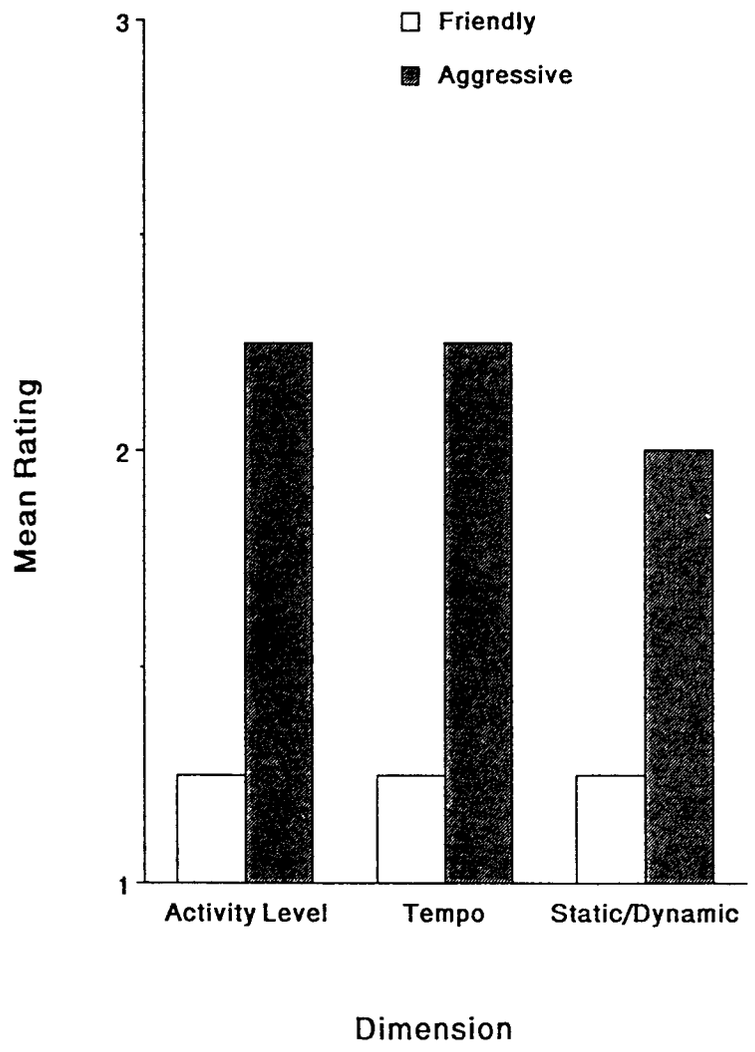

Figure 4. Mean ratings of the friendly and aggressive visual excerpts on the dimensions of activity, tempo and static/dynamic.

Table 3
Verbal Descriptions of the Wolf Social Interaction Excerpts

Excerpt ID*	Description
1. (Friendly)	Face - gentle face bite, recipient does not jump away
2. (Friendly)	Face - muzzling, muzzle licking, one wolf places its paw over the other wolf's neck
3. (Friendly)	Body - playful body slam (one wolf slams its body against the other wolf's body), both initiator and recipient have neutral ear and tail positions
4. (Friendly)	Body - one wolf (displaying erect tail and ears) slowly approaches the other wolf
5. (Aggressive)	Body - one wolf engages in rough play initiation by biting another wolf on the side
6. (Aggressive)	Body - body slam (with lateral-lateral contact) - initiator wolf maintains dominant postures (erect ears and tail) and recipient displays subdominant postures (tail tuck and ears back)
7. (Aggressive)	Face - facial expressions during eating, bared teeth and snapping jaws
8. (Aggressive)	Face - aggressive face bite, recipient jumps away

*The identifying code (1 - 8) and descriptors (friendly or aggressive) are consistent with Table 1.

Table 4
Description of the Musical Excerpts

Excerpt ID[a]	Musical Description[b]	Descriptions by three musicians
a. Friendly *Silly Sketches*	vibraphone, piano, bass, drums MM = 106 syncopation begins on major triad harmony repeating tonic to dominant chord progression harmonies in sixths and thirds	(1) light-hearted happy, happy Muzak (2) happy, a light, standard shuffle feel (3) calypso
b. Friendly *Playroom*[c]	flute, strings MM = 60 (3/8 note) beginning features orchestral bells and flutes end features string finale fast waltz, major, lilt end on major ascending melody and descending bass arpeggio	(1) light, orchestral, happy, final (2) very Mozart, 18th century, light texture (3) light, cheerful
c. Friendly *Happy Go Lucky*	piano, trombone, drums MM = 75 syncopated 16th notes, accidentals, phrase begins in minor, ends in major	(1) honkytonk, dixieland, light-hearted, comical (2) show-tune, bouncy, jolly, ragtime (3) jaunty, carefree

(Continued)

d. Friendly *Bowled Out*[c]	brass, mallets, bass, drums MM = 120 calypso, rhythm with a melody line	As Beginning: (1) Caribbean, steel drum, cruise music (2) Caribbean, holiday, horns feature, pseudo latin feel (3) calypso, bright As Ending: (1) Caribbean, steel drum, cruise music (2) digital sampling disco/pop/reggae, island getaway commercial (3) the end is signalled, the end of an action
e. Friendly *Daytime Live*	brass, xylophone, bass, drums MM = 126 melody harmonized in sixths and thirds syncopation ending begins with an upbeat to a sustained half note, followed by two 16th notes and an accented quarter note, all in the melody line	(1) dance, happy, light (2) fake, high production sounding TV commercial sampling (3) neutral, upbeat
f. Aggressive *Stalker*	synthesizer, marimba, bass, cymbal schwing MM = 140 high sustained harmonic perfect fifth interval against chromatic, non-diatonic dynamic bass line 4 eighth notes, followed by 4 matching rests, followed by a passage of 7 consecutive eighth notes	(1) dark, sinister (2) spy movie music, suspenseful, diminished 5th sound (3) ominous

(Continued)

g. Aggressive *High Performance*	synthesized string ensemble, bass, drums MM = 138 minor/modal prominent tritone interval, syncopation	(1) techno-dance, militaristic (2) processed, digital sampling, cold, pseudo-bossanova beat mix (3) calypso, industrial
h. Aggressive *Conflict*	brass, tympani, snare drum MM = 60 descending minor line, mi-re-do decreasing division of the beat (triplets, duple, half-note)	(1) dark, ominous, foreboding, conclusive (2) 20th century movie music, dramatic, dark (3) driving, relentless
i. Aggressive *Purple Suspense*	organ, pizzicato strings, tympani MM = 118 sustained harmonic tritone interval, with addition of minor 7th and flat 5th chord tones	(1) dark, sinister (2) pizzicato strings, gloomy, foreboding (3) ominous
j. Aggressive *Dramatic Chord*	synthesized strings MM = not applicable as selection is only one note sustained chord (minor 7th, flat 5th, flat 9th)	(1) ominous, gloomy, static (2) single symphonic chord, firmata (3) orchestral chord, dissonant, suspenseful, tense

[a]Identifying code (a - j) is consistent with the title (as provided on the original recording) and with the use of music as friendly or aggressive in Table 1.

[b]Musical Description includes primary instrumentation, metronome setting (MM) for a quarter note (unless otherwise stated) and other musical features or characteristics.

[c]Only part of the material from this selection was used in each pair of excerpts. One pair included the beginning, but not the end, and vice versa.

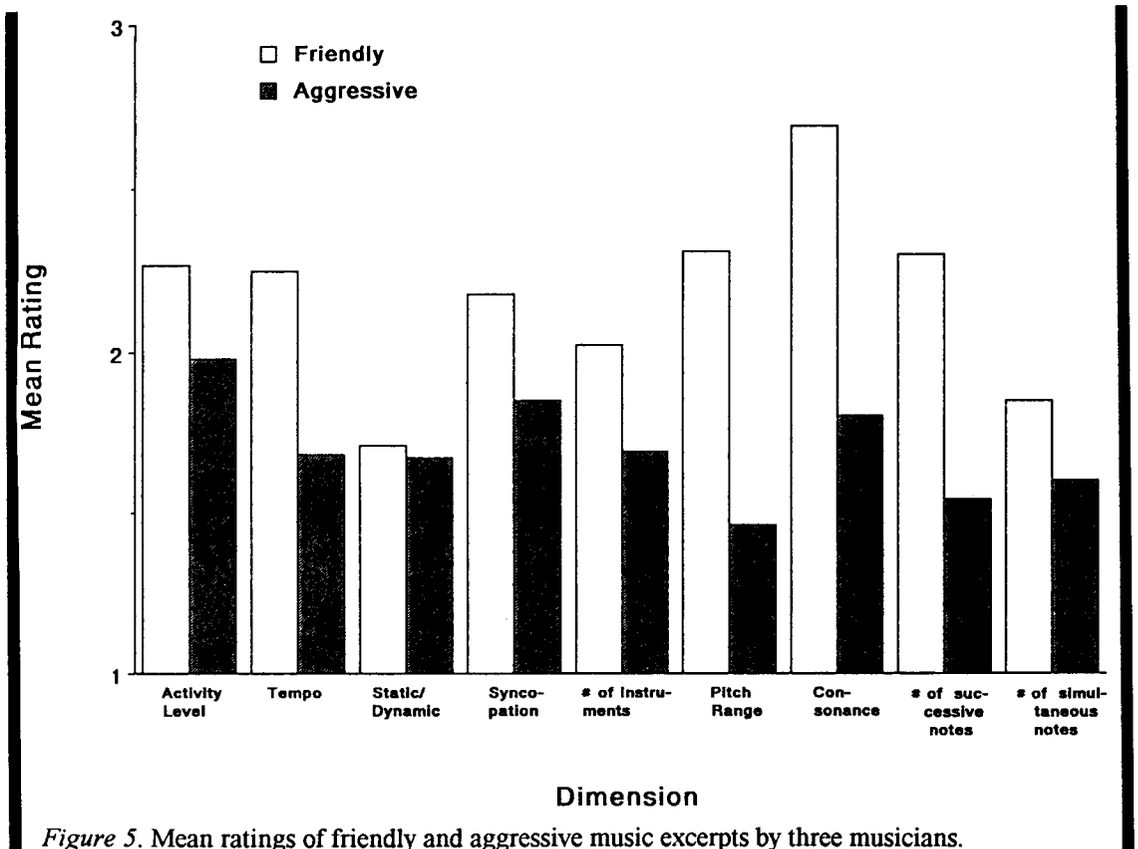

Figure 5. Mean ratings of friendly and aggressive music excerpts by three musicians.