Psychomusicology, 13, 60-98

PERCEPTUAL JUDGEMENT OF THE RELATIONSHIP BETWEEN MUSICAL AND VISUAL COMPONENTS IN FILM

Scott D. Lipscomb
The University of Texas at San Antonio

Roger A. Kendall
University of California, Los Angeles

In this study, the authors investigate the relationship between the musical soundtrack and visual images in the motion picture experience. Five scenes were selected from a commercial motion picture along with their composer-intended musical scores. Each soundtrack was paired with every visual excerpt, resulting in a total of 25 audiovisual composites. In Experiment 1, the 16 subjects selected the composite in which the pairing was considered the "best fit." Results indicated that the composer-intended musical score was identified as the best fit by the majority of subjects for all conditions. In Experiment 2, the 15 subjects rated all 25 composites on semantic differential scales. A significant interaction ($p < .00005$) between audiovisual combination and the various semantic differential scales was found. Analysis of this interaction revealed that the composer-intended combination yielded higher mean scores in response to the 4 adjective pairs of the Evaluative dimension. Clustering the subject responses into 2 factor scores (Evaluative vs. a hybrid of Activity and Potency), confirmed these high Evaluative mean scores. In addition, the response contours of the Activity/Potency dimension remained relatively consistent, suggesting that music exercises a strong and consistent influence over the subject responses to an audiovisual composite, *regardless of visual stimulus*. The results corroborate previous research, indicating that a musical soundtrack can change the "meaning" of a film presentation. Comparison of the various soundtracks in music theoretical terms assisted in identifying musical elements that appeared to be relevant to specific subject ratings. These comparisons were utilized in the formulation of a model for music communication in the context of the motion picture experience.

Music has played an integral part in the motion picture experience almost since its inception.[1] Even so-called "silent films" were usually accompanied by musical performers. Considering the popularity of this artform and the fact that it has developed into a multi-billion dollar industry, it is quite surprising that there has been little empirical investigation into the role of film music. In the present study, the authors investigate the relationship between visual activity on-screen and the musical soundtrack. Two specific questions are of particular interest. First, can listeners reliably select the composer's intended soundtrack for a given visual scene from among several musical selections? Second, does a significant amount of variation occur in the perceptual response to a given scene when the visual stimulus remains constant and only the music is changed?

Related Literature

There has been much speculation about the interaction of music and the visual element in motion pictures (Eisenstein, 1947; Evans, 1975; Gorbman, 1987; Kracauer, 1960; Prendergast, 1977; Thomas, 1973; Thomas, 1991; Weis & Belton, 1985; Zettl, 1990). A search for related literature in this area, however, has shown a paucity of empirical investigations dealing explicitly with this interaction.

In the field of perceptual psychology, interaction between the aural and visual sensory modalities is well-documented. Radeau and Bertelson (1974) found that when a series of lights and tones is presented at a 15 degree spatial separation, the location judgments for both the lights and the tones are biased toward the location of the stimulus in the other modality. Staal and Donderi (1983) showed that introducing an aural stimulus lowered the interstimulus interval at which the subjects perceived continuous apparent motion of one light instead of partial motion or succession of motion between two lights. As a result, they concluded that the presence of sound may alter the perceived duration of a light flash by affecting visual persistence (see also Bermant & Welch, 1976; Massaro & Warner, 1977; Mershon, Desaulniers, Amerson, & Kiever, 1980; Regan & Spekreijse, 1977; and Ruff & Perret, 1976).

Three studies have utilized ecologically valid contexts in the consideration of the motion picture experience. Tannenbaum (1956), using Osgood, Suci, and Tannenbaum's (1957) three factors (i.e., Evaluative, Potency, and Activity) to collapse the bipolar adjectives used in semantic differential scaling,[2] found that music does influence the rating of dramatic presentations whether presented live on stage, in a studio taped version, or in a recorded version of the live performance. His results showed that the influence of music was most pronounced on the subject responses related to the Potency and Activity dimensions. The overall evaluation of the play did not change significantly. However, there are several problems concerning the musical aspect of his stimulus presentation. First, the musical selection was not composed for the specific scene that it accompanied. In a rather ambiguous explanation of the selection process, Tannenbaum explains that the piece was chosen by a person who "has considerable experience in this kind of work" and confirmed by a panel of four "experts" (Tannenbaum, 1956). The most serious problem, however, was the method employed to synchronize the audio and visual stimuli during the performances. A phonograph recording was played along with the visual image. As a result, the synchronization of the dramatic action and the musical accompaniment was left largely to chance. This procedure is not at all representative of the relationship that occurs in a well-edited motion picture. Finally, in an attempt to make the music seem more compatible with the visual action, during certain scenes the "volume" (meaning loudness) was manually increased "for dramatic impact" (p. 96). This is hardly an acceptable substitute for a soundtrack composed specifically for the scene under investigation.

In a second study of interest, Thayer and Levenson (1983) recorded five different physiological measurements during exposure to a 12 min black and white industrial safety film depicting three accidents. These measures included the subject's interbeat interval of the heart, general somatic activity, skin conductance level (SCL), finger pulse transmission times, and finger pulse amplitude. In addition, the subject was asked to provide a continuous self-report of anxiety level by turning a dial on which the extremities were labeled *extremely calm* and *extremely tense*. Two musical scores were composed for presentation with the film for comparison with the responses to a control (i.e., no music) condition. The *documentary music* is described as a mildly active progression of major seventh chords, purposely intended not to draw attention toward or away from any specific part of the visual scene. The *horror music* is described as a repetitive figure based on diminished seventh chords utilizing harsh timbres. In addition to the differences in musical style, placement of the music in the context of the film differed radically between the two music conditions. While the documentary music was present throughout, the horror music was edited so that it preceded the first accident by 20 s, the second accident by 10 s, and the final accident by 30 s. In each instance, the music ended approximately 10 s after the accident at a natural cadence point (both musically and visually). Although the film produced significant responses in all five of the physiological measures when compared with subjects' pre-exposure levels, only SCL differentiated the three film score conditions. From this result, the investigators claimed to have provided "preliminary experimental support for the efficacy of musical scores for manipulating the stressfulness of films" (p. 44). Recall, however, that the subjects' continuous self-reports of *perceived* anxiety level did not differentiate between the three film conditions. Therefore, although Thayer and Levenson may conclude from their data that the use of music caused either a heightened or reduced electrodermal response to the stressful stimuli, more evidence is needed to support a claim for the ability of a musical score to manipulate the stressfulness of a film in terms of the subjects' emotional responses.

In a third study, Marshall and Cohen (1988) selected a film utilizing abstract animation. They were interested in determining whether the information provided by a musical soundtrack would affect the judgments of personality attributes assigned by subjects to each of three geometric shapes presented as "characters" in the film. Marshall composed two distinct soundtracks (which were described as *strong* and *weak*) varying on a number of musical dimensions, for example, major/minor mode, fast/slow tempo, high/low pitch, and single-/multi-note texture. Each soundtrack consisted of three main "themes." Synchronization of the aural and visual elements was kept constant by editing the soundtrack directly onto the videotape. The authors provide a brief description of the action occurring at the point when each of the themes is introduced. However, apart from the beginning point, no information is provided concerning the specific interaction of the aural and visual stimuli . A second problem with these musical compositions is that

their extreme simplicity of content and excessively repetitive structures fail to provide an accurate representation of the highly developed craftsmanship evident in a typical movie score. Even using this limited musical vocabulary, the results produced were similar to those compiled by Tannenbaum (1956). In comparing five film conditions (i.e., film alone, weak music alone, strong music alone, weak music-film, and strong music-film), meaning of the music was found to be closely associated with the film on the Potency and Activity dimensions. Evaluative judgments, on the other hand, appeared to depend on a complex interaction of the musical and visual materials. However, in this particular investigation, the simplicity of the stimulus materials seriously limits generalizability of the results to motion pictures.

Significance of the Present Study

The significance of the present study lies in three main issues. First of all, unlike previous studies, this investigation uses "real" musical and visual examples. Hevner (1936) stated that

> Since we are looking for elements of *music* we must be sure that the material provided for observation represents real *music* and not merely *elements* trimmed down for experimental purposes to such an extent that all *music* has been left out (p. 248).

To the knowledge of the authors, the present investigation is the first to use actual excerpts from a major motion picture as stimuli for an experimental procedure of this type.[3] By improving the validity of the stimulus materials, the conclusions may be more confidently generalized to the "real world" from which they were abstracted.

Secondly, this study provides a method for quantifying compatibility between the aural and visual modalities. In the measurement of this compatibility, the motion picture experience is considered as a system. Judgments made within this context are not based on either modality (aural or visual) in isolation, but are made in reference to an audiovisual composite, taking into account cross-modal interactions and interrelationships.

Finally, the quantitative scores used in formulating this scale of compatibility will be made using the perceptual aesthetic judgment of each individual subject. The analytical methods utilized in the present study consider perception as the criterial frame of reference.

Quantization of perceptual response poses certain difficulties for empirical studies. In selecting semantic differential scaling, the purpose of the present investigation *was not to determine the connotative or denotative meaning* of various musical stimuli, but rather to observe indications of perceptual change based on the subjects' semantic differential responses to various audiovisual stimuli. Since musical communication does not require the transmission of an explicit connotative or denotative meaning, one might ask what information is gained from subject responses on a semantic differential scale to a musical stimulus. In this case, words must be assigned via the implicit perceptual/cognitive apparatus using past experience as a referent. Kendall and Carterette (1991) suggest that the process of assigning verbal attributes to musical sound involves:

> the superposition of (at least) two multidimensional spaces. One space is comprised of musical images, relationships among sound schemas, the other comprises referential, semantic (verbal) meanings. The mapping of one space to another is almost certainly not linear, and the resulting composite space is hybrid, neither "musical" nor "verbal." . . . Words are useful in communicating musical ideas, but they are not the ideas themselves (p. 391).

Therefore, within the context of this investigation, the responses are considered to be an indication of the subjects' changing evaluation of the audiovisual composite scaled in a semantic space. This semantic space is determined by the identification of a relationship or the lack of such a relationship between the present stimulus and an implicit model formed from past experience. Therefore, the semantic differential is merely an indicator, using words as a communicative medium, of perceptual change.

Basic Hypotheses

In Experiment 1, it was hypothesized that the highest percentage of responses on the "best fit" categorizations would match the composer's intended combination. It was also hypothesized that the ratings given by subjects on the semantic differential scale in Experiment 2 would vary significantly when the visual stimulus is kept constant and only the musical soundtrack is changed.

In comparing the responses in the matching procedure (Experiment 1) with those on the semantic differential scale (Experiment 2), it was hypothesized that determinate factors in the subjective decision of selecting "best fit" may be identified. By revealing such influences on this decision-making process, it was also hypothesized that insight could be gained into elements contributing to a successful symbiosis of audio and visual stimuli within the context of a motion picture.

Basic Methodology

The present study employed a post-test only repeated measures factorial design in which the musical soundtrack was manipulated as the independent variable. Subject responses (either categorization or scores on the semantic differential scale) were recorded as the dependent variable.

The visual excerpts were 35 to 40 s sequences of cinematic images excerpted from the motion picture *Star Trek IV: The Voyage Home* (see Appendix A for a brief description of the scenes used). All musical soundtracks were Leonard Rosenman's compositions, intended to accompany each of the visual scenes (reductions of the musical scores are provided in Appendix B). They were edited directly from a compact disc recording of the original soundtrack onto the video tape. An audiovisual composite is defined as any combination of a visual excerpt and a musical selection. Therefore, there were composer-intended composites and composites consisting of musical and visual combinations not intended by the composer. In Experiment 1, subjects provided "best fit" categorization responses, while subjects in Experiment 2 responded by providing ratings on a semantic differential scaling procedure.

Stimuli

The first objective in selecting stimulus materials was to choose scenes that were intercategorical, that is, no two scenes could be so similar that the content would be easily confused. Using excerpts which were selected from the same motion picture allowed control of the production variables (e.g., composer, production quality, etc.).

Initially, 10 scenes were selected for use in a pilot study. Six student composer volunteers from a class in Composition for Motion Pictures and Television at The University of California at Los Angeles (UCLA) individually viewed a video tape of the 10 selected scenes exactly as they appeared in the film (i.e., including cinematic image, musical soundtrack, and sound effects). One scene was designated as the standard to which all others were compared; subjects responded on a 9-point Likert scale (*same/different*). Using cluster analysis with complete linkage (farthest neighbor), responses to the nine scenes were statistically clustered into four groups. One scene was selected from each of these clusters. The four chosen scenes, along with the standard, were used as the five excerpts for the main investigation.

The second objective in stimulus selection was to determine the method of aligning each of the soundtracks with every visual example so that even the most extreme mismatch results in the "best fit" possible (i.e., a potentially reasonable combination). The assistance of three student film composers was utilized in the actual editing process. In each composite, an attempt was made to synchronize, as well as possible, the aural and visual components. By trying several variations and discussing the results among the participants, a general consensus was reached concerning appropriate synchronization.

Since the ambient (i.e., non-musical) sound included on the movie soundtrack could have provided unwanted cues for the appropriateness of its pairing with a given visual scene, the original soundtrack was erased. The musical score from the compact disc soundtrack recording was then dubbed directly onto the videotape, eliminating this extramusical noise (e.g., car horns, water splashing, dialogue, etc.). The resulting composite stimulus materials consisted of only music and the cinematic image.

For Experiment 2, bipolar adjectives were selected for use in the attitude scaling procedure. Initially, adjectives were selected on the basis of factor loadings provided by Osgood et al. (1957). In addition, a review of past literature assisted in compiling a list of adjectives which had been utilized in previous studies. The final selection criterion was based on the semantic appropriateness of the adjective pair to the present study. As a result, the 10 adjective pairs listed in Table 1 (followed by factor loadings on the first three dimensions, where available) were chosen. The semantic differential scale utilizes bipolar adjectives from each of the three factors discussed by Osgood et al. (Evaluative, Potency, and Activity). A 10th adjective pair (effective/ineffective) was added as a means of providing a direct measure of the subjects' ratings on the audiovisual pairing.

Table 1
Bipolar Adjectives Used in the Semantic Differential Scaling Procedure

	I	II	III
Evaluative			
good/bad	1.00	.00	.00
beautiful/ugly	.52	-.29	-.02
interesting/boring	.40	-.09	.22
effective/ineffective	—	—	—
Potency			
strong/weak	.30	.40	.10
heavy/light	-.20	.48	-.02
tense/relaxed	.39	-.54	-.57
Activity			
active/passive	.17	.12	.98
fast/slow	.01	.26	.35
agitated/calm	-.15	.03	.26

The subject response procedure was slightly different from that utilized in most semantic differential studies. Rather than using a 5-, 7-, or 9-point scale, a continuous line was drawn between the adjective pair. The subject was asked to make a mark at the point on the line that was considered to provide the most appropriate response. Such an undifferentiated line scale maximizes the potential variation of subject responses and has been determined to be an important procedural modification (Schiffman, Reynolds, & Young, 1981). In order to quantify the responses, a ruler was used to measure from the zero point of each line to the mark provided by the subject. Each line segment was divided into a zero point and 25 equal divisions which could, in turn, be subdivided into four equal parts, providing a 101-point response scale.

Subjects

Using a random number table, 31 volunteers (both musicians and nonmusicians) from a Psychology of Music class at UCLA (Winter, 1989) were randomly assigned to one of two independent groups (Experiment 1, $N = 16$; Experiment 2, $N = 15$). The students were given the opportunity to participate in this investigation in exchange for three points of extra credit.

Apparatus

Stimulus presentation was made to one subject at a time using an Hitachi VT38EM Multi-System video cassette recorder and a 25-in Sony Trinitron color monitor. Since ratings on the combination of the aural and visual stimuli were the source of interest, the audio track was presented to the subjects without the use of headphones in an attempt to simulate a typical home video presentation. This environment avoids further separation of the two modes of perception under investigation. The entire procedure for an individual in either group required between 25 and 45 min to complete, depending on the length of time taken to review various combinations (Experiment 1) or to respond on the semantic differential scale (Experiment 2).

EXPERIMENT 1

Procedure

The subjects in Experiment 1 were assigned the task of selecting the musical score which provided the "best fit" for each visual scene. Subjects were presented with the first visual scene (randomly determined) in combination with each of the musical soundtracks, also randomly ordered. They selected the version that they considered to be the "best fit." After selecting a choice for the first visual scene, subjects moved on to the second visual scene in combination with each soundtrack, continuing until a match had been selected for each of the five visual examples. They were free to see each composite as many times as necessary in the decision-making process.[4]

Results

The responses resulted in the frequency counts provided in Figure 1, wherein the cells on the diagonal represent the composer's intended combinations. Notice that the total of each vertical column equals 16 (the number of subjects). Immediately evident is the fact that the composer-intended combinations were selected by the majority as "best fit" in every case. In fact, no other combination was selected by more than 25% of the subjects. This may be interpreted as confirmation of the effectiveness of Rosenman's compositional technique. However, no single composite was selected as the most appropriate combination by every subject. A closer look at Figure 1 reveals that Visual 5 is the most convincing example of agreement between the subject pool and the composer with Visuals 1 and 2 also reflecting a high level of agreement. Visuals 3 and 4 appear to have left the subjects less certain, as exemplified by the spread of their responses.

Considering these same results from the auditory dimension (horizontally, in Figure 1), notice that Audio 3 was not selected as an appropriate accompaniment to any scene other than that which was intended by Rosenman. Audio 4, however, seems to have been considered appropriate for a variety

VISUAL

AUDIO	1	2	3	4	5
A 1	*12*	1	0	1	0
U 2	1	*11*	0	3	0
D 3	0	0	*8*	0	0
I 4	3	4	4	*8*	3
O 5	0	0	4	4	*13*

Figure 1. Data matrix showing the number of subjects who selected each cell as "best fit."

of visual scenes, being selected as best fit *for each scene* by at least three subjects.

Although, overall, the responses appear to be heavily weighted toward the composer-intended combinations, the spread of responses to Visual 4 across both the audio and visual domains warrants some explanation. Visual 4 is the only visual excerpt in which no human figure appears. Also, due to miscommunication during an early conversation with the composer, Audio 4 is similar to, but not identical with, the composer-intended musical score. One possible explanation for the diffusion of responses to both the audio and visual components of Visual 4 is that, in the determination of a categorization response pairing musical stimuli and cinematic excerpts, two variables interact: abstractness of the visual image (in this particular instance, presence or absence of human interaction) and availability of a soundtrack that was specifically composed for the given scene. Further investigation will be necessary to test the following hypothesis.

In a cinematic excerpt such as Visual 4, because of its abstract imagery, the music takes on a more prominent role in determining the perceived meaning of the audiovisual composite. As a result, subjects might not necessarily judge contrasting musical scores as *incompatible* with a given visual image, but may simply render a *different interpretation* of the action occurring on the screen as a result of the variation of audio stimuli. In reference to the

other possible interactive variable (i.e., presence of a composer-intended soundtrack among the choices), responses of the subjects in Experiment 1 suggest that categorization decisions made between a set of options including the composer's intended musical score are much more homogenous than selections made when the composer-intended music is not among the possible choices, although further research is necessary to confirm this possibility.

Considering the results in Figure 1, two outcomes were possible other than the formation of a general consensus among the subjects, agreeing with the composer's intent. The majority of responses could have clustered into a cell other than that representing the composite utilizing the composer's intended score, or the responses could have been spread more equally across the cells of the matrix. The former would suggest that, though the composer's intent was not matched by the subject responses, there was general agreement upon which of the five soundtracks was considered most appropriate. In the latter case (which occurred in both the audio and visual dimensions for Visual and Audio 4), the fact that the responses are more spread out among the cells of the design implies that the visual image and the musical score are considered to be more ambiguous and, as a result, could be combined successfully with several stimuli of the other modality. Since the responses to Visual 4 and Audio 4 were spread across the cells (both vertically and horizontally) rather than clustered, ambiguity appears to have played a strong role in the subject responses to the composites utilizing either of these components.[5]

EXPERIMENT 2

Procedure

The subjects in Experiment 2 rated each of the 25 audiovisual combinations on a continuous-line semantic differential scale consisting of 10 pairs of bipolar adjectives. In an attempt to negate ordering effect, two independent presentation orders were randomly generated. The only restriction imposed upon ordering the stimuli was that no selection could follow a previous composite containing either the same musical or visual component. The bipolar adjectives were also randomly arranged into three configurations with respect to both presentation order and polar position (e.g., good/bad as opposed to bad/good). Using a random number table, each subject was assigned to one of the six subgroups (i.e., two composite presentation orders and three bipolar adjective configurations).

Results

Using only two within-subjects variables (audiovisual combinations and the semantic differential bipolar adjectives), an ANOVA was run as a separate statistical procedure on each visual scene in combination with all of the musical soundtracks (e.g., Visual 1 with Audios 1-5, Visual Two with Audios 1-5, etc.) and then keeping the audio constant and changing the visual image (e.g., Audio 1 with Visuals 1-5, Audio 2 with Visuals 1-5, etc.).[6] Therefore, in order to keep the overall confidence level well above 95%, the alpha level for each of the 10 individual ANOVAs was set *a priori* to $p <$ .005. The ANOVA results are provided in Tables 2a and 2b.

Table 2a
Analysis of Variance for the Condition when Visual Stimulus is Constant and Only the Musical Soundr\track Changes

Source	*df*	*SS*	*MSS*	*F*
SCENE ONE:				
Within Subjects	735	18994.85		
Audio	4	201.73	50.43	1.10
Semantic Differential	9	395.19	43.91	1.63
Audio/Differential	36	4560.79	126.69	8.10**
SCENE TWO:				
Within Subjects	735	18146.46		
Audio	4	1348.54	337.14	8.06**
Semantic Differential	9	1778.25	197.58	6.77**
Audio/Differential	36	2564.34	71.23	5.58**
SCENE THREE:				
Within Subjects	735	21308.42		
Audio	4	2428.97	607.24	8.08**
Semantic Differential	9	1839.43	204.38	7.40**
Audio/Differential	36	3253.40	90.37	7.47**
SCENE FOUR:				
Within Subjects	735	18707.47		
Audio	4	822.55	205.64	3.08
Semantic Differential	9	467.90	51.99	2.46
Audio/Differential	36	3625.82	100.72	6.87**
SCENE FIVE:				
Within Subjects	735	20513.23		
Audio	4	2017.20	504.30	7.36*
Semantic Differential	9	202.66	22.52	0.84
Audio/Differential	36	4249.57	118.04	8.72**

*$p < .0001$. **$p < .00005$.

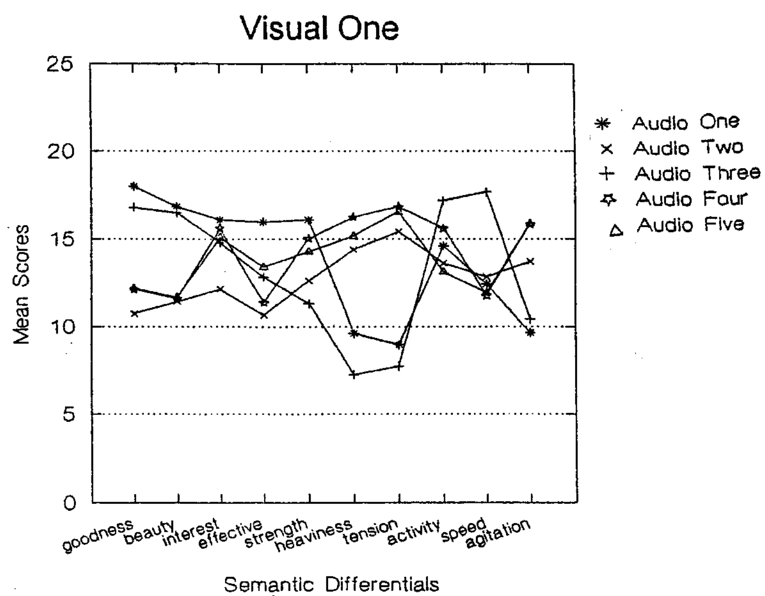
Table 2b
Analysis of Variance for Condition When the Musical Soundtrack is Kept Constant and Only the Visual Stimulus Changes

Source	*df*	*SS*	*MSS*	*F*
AUDIO ONE:				
Within Subjects	735	20623.41		
Scene	4	434.97	108.74	1.65
Semantic Differential	9	1755.31	195.03	6.64**
Scene/Differential	36	3245.96	90.17	5.83**
AUDIO TWO:				
Within Subjects	735	16460.46		
Scene	4	1568.25	392.06	8.40**
Semantic Differential	9	915.27	101.70	2.71
Scene/Differential	36	1323.13	36.75	3.49**
AUDIO THREE:				
Within Subjects	735	23421.17		
Scene	4	1628.25	407.06	7.51*
Semantic Differential	9	5696.19	632.91	15.50**
Scene/Differential	36	2385.10	66.25	6.04**
AUDIO FOUR:				
Within Subjects	735	15766.76		
Scene	4	261.40	65.35	1.40
Semantic Differential	9	3329.92	369.99	11.02**
Scene/Differential	36	769.38	21.37	2.35**
AUDIO FIVE:				
Within Subjects	735	17831.11		
Scene	4	670.81	167.70	2.73
Semantic Differential	9	2587.35	287.48	7.37**
Scene/Differential	36	929.73	25.83	2.46**

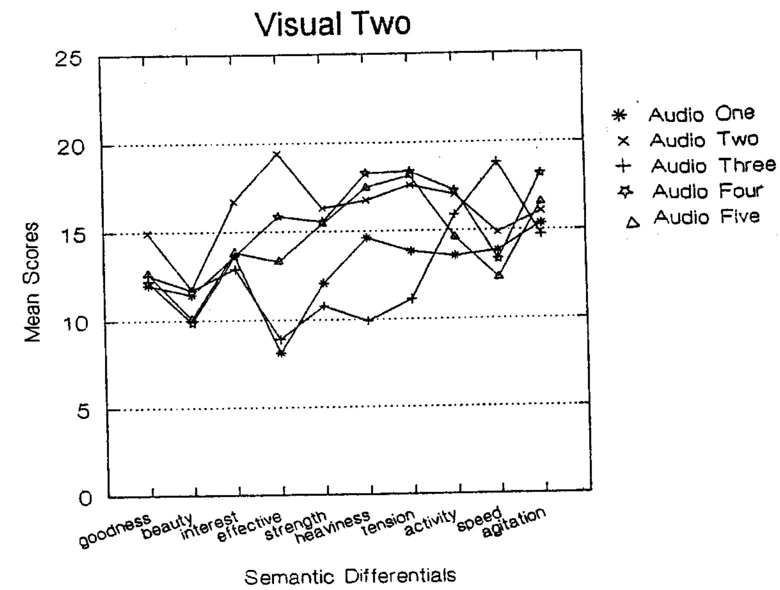
*$p < .0001$. **$p < .00005$.

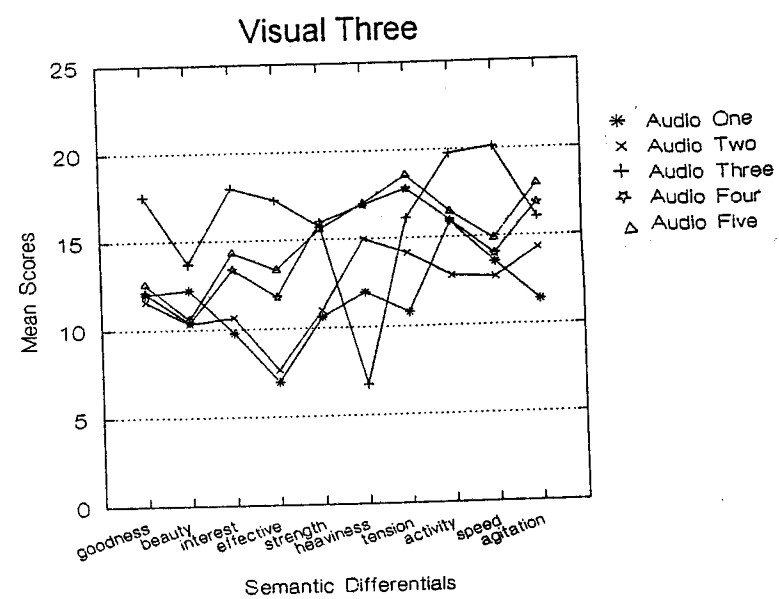
The most important result of these multiple ANOVAs is the consistency of the significant difference shown in the interaction of the audiovisual combination and the ratings on the differentials ($p < .00005$).[7] Using this information, the null hypothesis can be rejected. Quantitative analysis has shown, with an extremely high level of confidence, that the subject ratings vary significantly when keeping the visual stimulus constant and changing only the musical score.

In an attempt to further discern the relationship between the audiovisual combinations and the differential ratings, the means were plotted on overlayed line graphs. First, keeping the scene constant, the mean ratings were plotted for each musical score. Then the soundtrack was kept constant and the mean scores were plotted for each of the various cinematic images. These graphs (Figure 2a and 2b) were formatted so that the higher a point is plotted on the ordinate, the more a particular quality is present in relation to its antipode. From left to right on the graphs, these qualities are goodness, beauty, interest, effectiveness, strength, heaviness, tension, activity, speed, and agitation.[8]

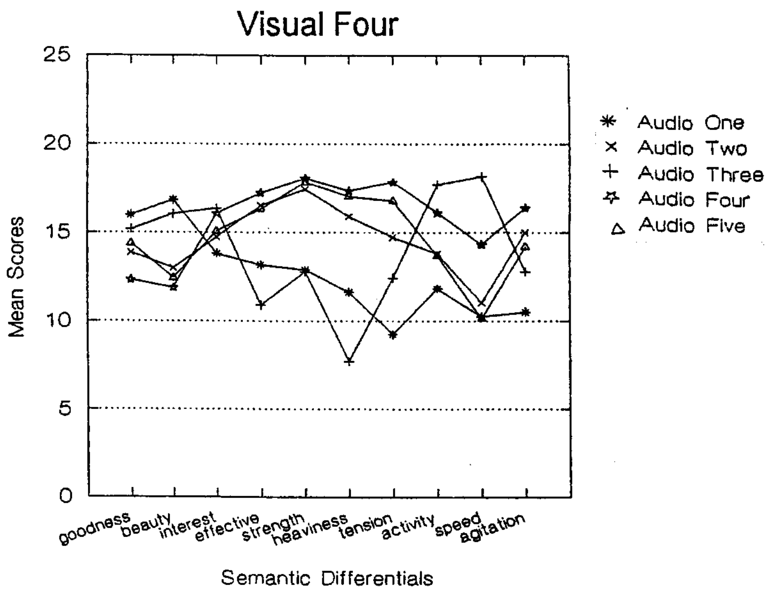


(Continued)





(Continued)



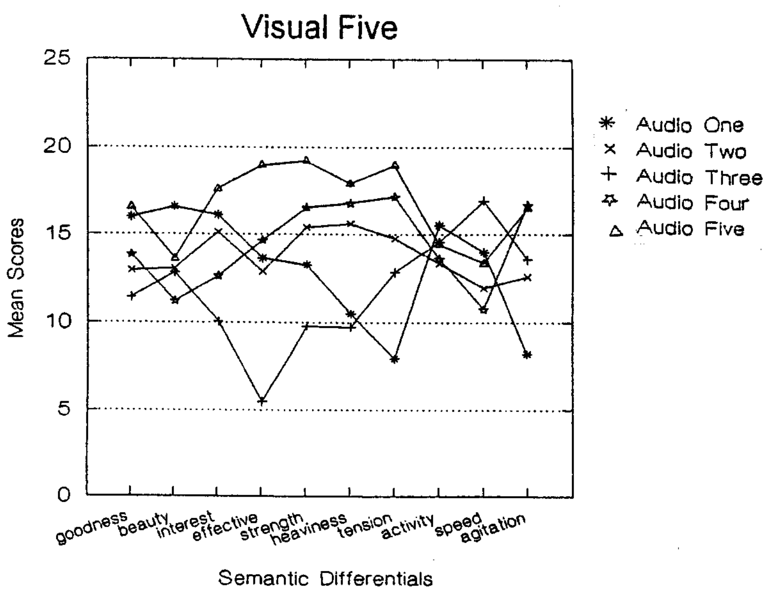
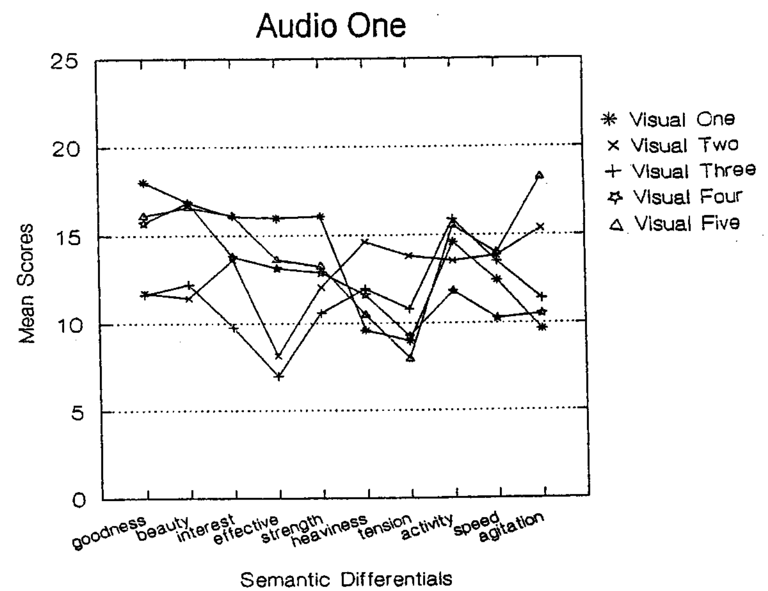
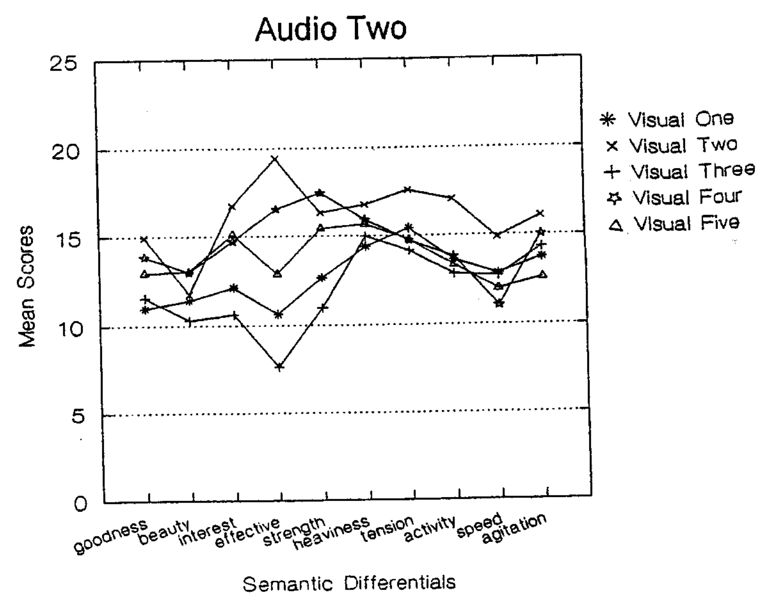
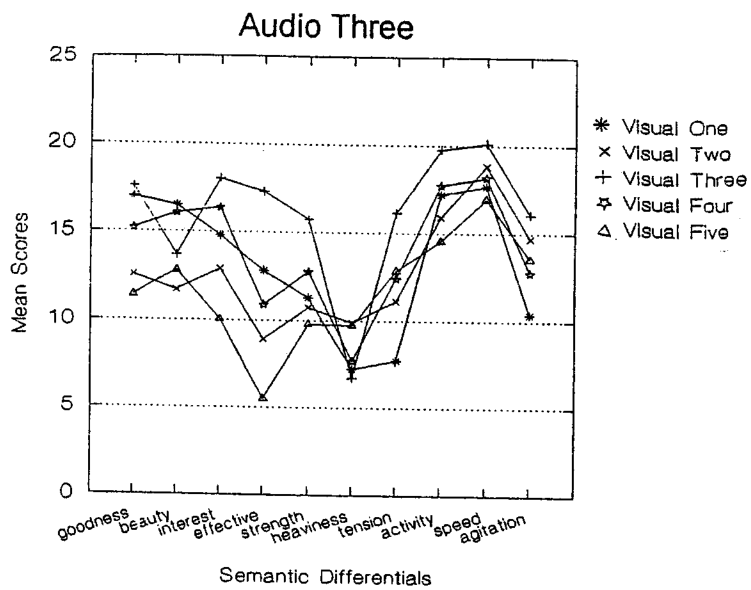


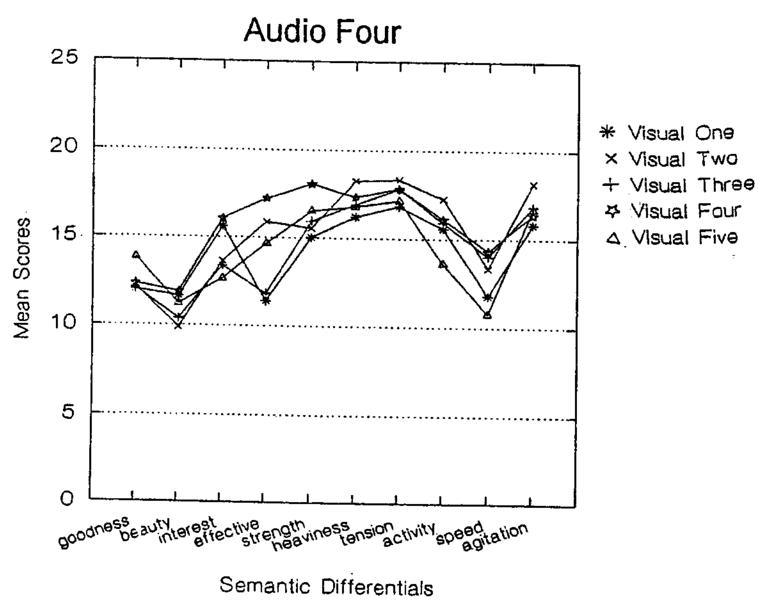
Figure 2a. Line graphs representing mean scores on the semantic differential scales across audio/visual composites while keeping the visual stimulus constant.





(Continued)





(Continued)

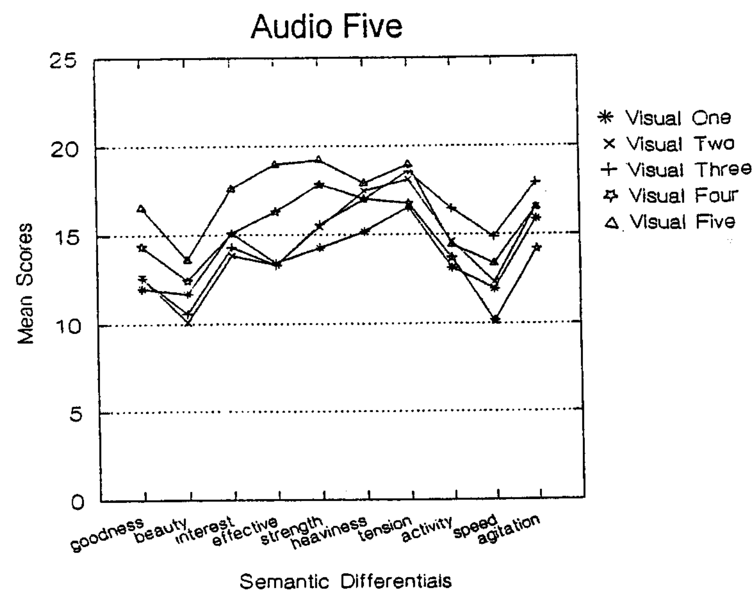


Figure 2b. Line graphs representing mean scores on the semantic differential scales across audio/visual composites while keeping the audio stimulus constant.

Notice that the semantic differential mean ratings for the Evaluative dimension (e.g., goodness, beauty, interest, and effectiveness) are higher for the composer's intended combination than for any other composite. In general, the range of mean scores representing a single bipolar adjective pair (vertical spacing on the graphs) is much greater in the graphs representing the five musical soundtracks across a single visual excerpt (Figure 2a) than in the graphs representing a single musical selection with each visual excerpts (Figure 2b). For example, observe the relative contour similarity and the clustering of mean scores in Audio 4 and Audio 5 of Figure 2b when compared with the Visual scenes represented in Figure 2a. The largest range of scores on any differential in Figure 2b is consistently on the effective/ineffective scale. The fact that this effectiveness rating is consistently variable as a function of musical score supports Marshall and Cohen's (1988) statement that Evaluative judgment is affected by the appropriateness of the pairing of musical and visual components in the context of an audiovisual composite. Also notice that in Figure 2a, in which the visual stimulus is kept constant, Audios 1 and 3 form a contrasting (and sometimes opposing) response contour to that generated by Audios 2, 4, and 5. The graphs for Visual 1 and Visual 4 in Figure 2b provide the clearest example of this clustering.

To confirm this observation statistically, Pearson cross-correlations were calculated for all audiovisual composites across 9 of the 10 semantic differential mean scores. The effective/ineffective rating was not included due to the large range of response means across all conditions. The resulting triangular matrix was subjected to multidimensional scaling (MDS). A two-dimensional solution accounted for 98.3% of the variance (R^2). The first dimension alone accounted for 91.8% of the variance. Therefore, cluster analysis was selected as a more appropriate means of analyzing this data. The resulting cluster using complete linkage (farthest neighbor) is presented in Figure 3 and shows a clear separation into two distinct groups.

Cluster analysis clearly separates the composites according to musical soundtrack. All of the composites located in the bottom cluster utilize either Audio 1 or Audio 3. In contrast, the composites in the top cluster incorporate either Audio 2, Audio 4, or Audio 5. The single exception to this clustering is the combination of Visual 2 and Audio 1 (the lowermost composite in the upper cluster of Figure 3). This is the only visual scene selected for use in the present investigation that incorporates one-on-one verbal interaction between characters. The intensity of the characters' facial expressions and body movements may have overridden the effect of the musical score in this composite. Specific musical parameters will be suggested in the General Discussion section to account for response differences between the two clusters observed in Dimension 1.

Clustering Differentials into Factor Scores

Since an average of judgments on a number of related categories provides a more reliable indicator of meaning than any individual judgment on a single rating scale, Tannenbaum (1956) and Marshall and Cohen (1988) both combined their adjective pairs into factor scores based on the work of Osgood *et al.* (1957). Similarly, after analyzing the statistical results on the large set of raw data compiled in the present study, as described above, the individual semantic differentials were combined into factors so that the entire data set could be analyzed as a single ANOVA. Although the adjectives used in the present study were selected using factor loading data (Osgood et al., 1957) as one of the selection criteria, the investigators did not presume a priori that the data would factor cleanly into Osgoods et al.'s three dimensions (i.e., Evaluative, Potency, and Activity). Instead, we chose to analyze the data collected in the context of this particular investigation in an effort to determine if such a method of data reduction would provide reliable categorization.

Using complete linkage, the adjective pairs were statistically clustered into two distinct dimensions. Two differentials (strong/weak and fast/slow) alternated between these two divisions showing no particular tendency to associate consistently with one or the other. Therefore, these two differentials were eliminated from consideration, leaving the groups as given in Table 3. By looking at the adjective pairs contained in each group, the dimensions were determined to be Evaluative (Dimension 1) and a hybrid factor combining Activity and Potency (Dimension 2). Using cluster analysis as a data reduction technique, the research design can be modified from the initial 250 cells (25 audiovisual composites and 10 semantic differential scales) to a much more manageable 50 cells (25 audiovisual composites and 2 factors).

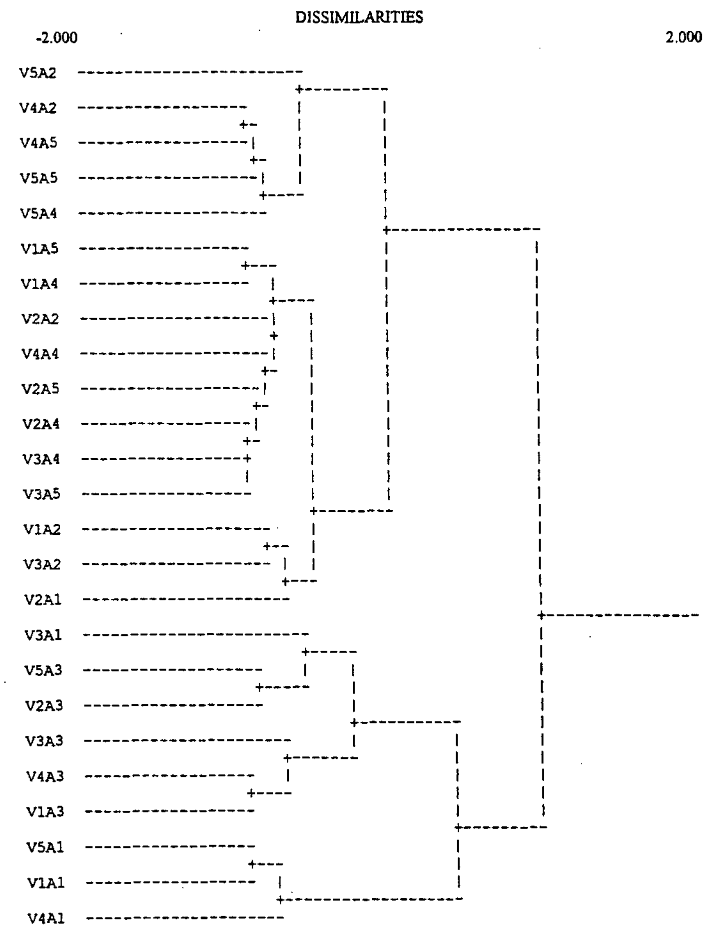


Figure 3. Cluster diagram representing every audiovisual composite across 9 of the 10 semantic differential scales.

Table 3
Clustering of Semantic Differential Adjectives Into Factors

Dimension 1	Dimension 2
good/bad	heavy/light
beautiful/ugly	tense/relaxed
interesting/boring	active/passive
effective/ineffective	agitated/calm

An ANOVA was run on the factor scores with the confidence level set a priori at 95% ($p < .05$). The results revealed a significant difference between ratings on the audiovisual combinations ($p < .00005$) and a close to significant difference between the factor scores ($p < .0525$). Most importantly, as in all of the 10 ANOVAs previously discussed, the interaction between the audiovisual combinations and factor scores was significant ($p < .00005$). As a result, the null hypothesis of Experiment 2 was once again rejected.

As seen in the bar graphs in Figure 4a, the Evaluative factor scores vary in a consistent manner. In every graph of the Evaluative dimension except Visual 4, the composer's intended soundtrack is rated higher than any other combination. One possible reason for the homogeneity of the response means for Visual 4 may be the fact that, as discussed earlier, none of the musical scores was actually Rosenman's intended composition, although the soundtrack used in this procedure may be described as a different version or musical variation of the composer-intended music. It is also possible that the degree of abstraction (i.e., lack of specific intent) of the hovering spacecraft in Visual 4 left the subjects uncertain of the context of the scene, allowing five different, but equally valid, interpretations of the action on-screen. A discussion of the concept of "varying degrees of abstraction" will follow the presentation of experimental results. The spread of responses shown for Visual 4 and Audio 4 in Figure 1 and the homogeneity of mean scores represented in the bar graph of the Evaluative dimension for Visual Four in Figure 4a provide an example of subject responses when none of the composites utilize the composer's intended musical score and the intent of the visual object is uncertain. Further research is needed to determine the role that the degree of abstraction (in both the visual and aural domains) plays in perceptual judgment of a motion picture. Despite the exception of Visual 4, these results confirm that the audio track was a significant factor in the ratings given on the Evaluative dimension.

In the vertical bar graphs representing the mean score for each musical soundtrack on the Activity/Potency factor while keeping the visual scene

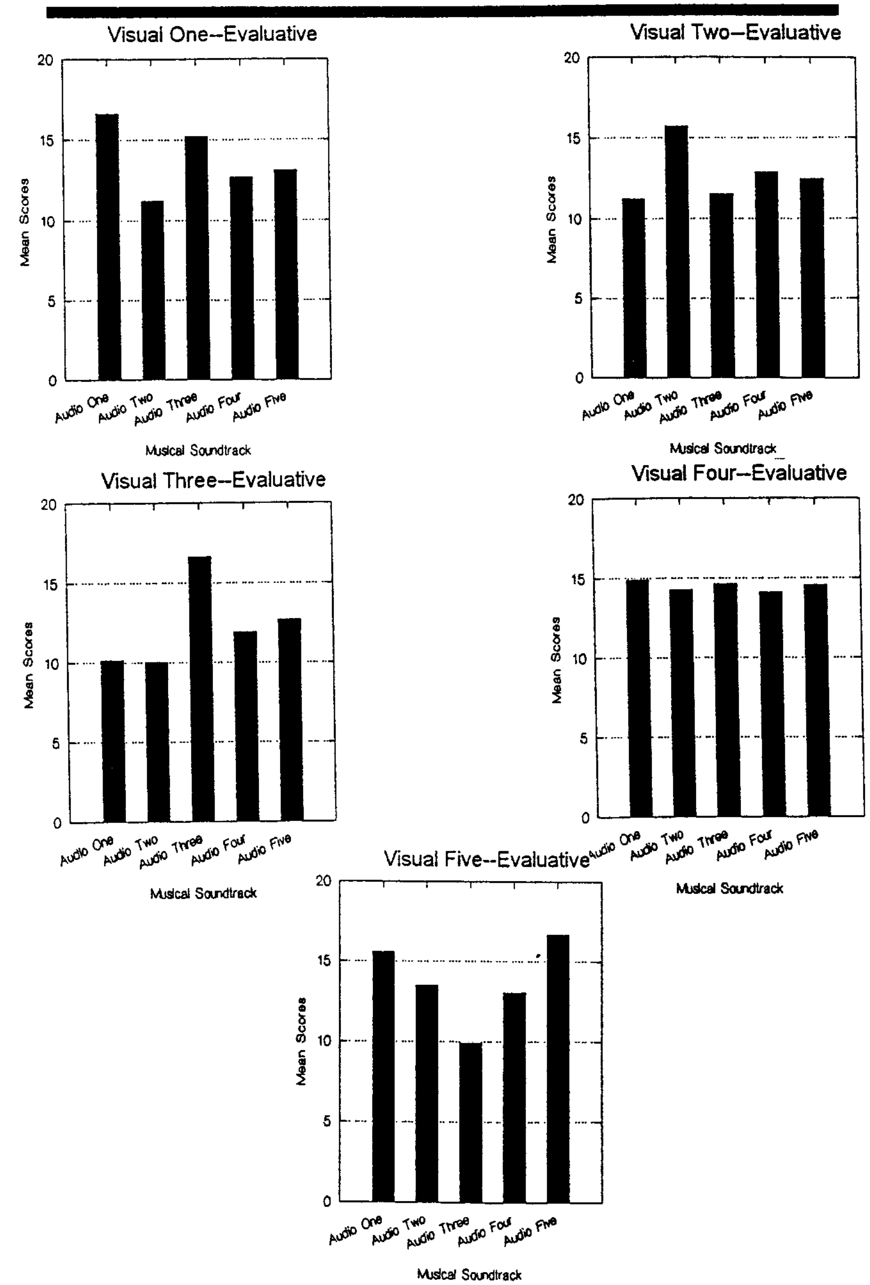


Figure 4a. Vertical bar graphs depicting the mean scores for the Evaluative factor while keeping the visual excerpt consistent and varying the musical soundtrack.

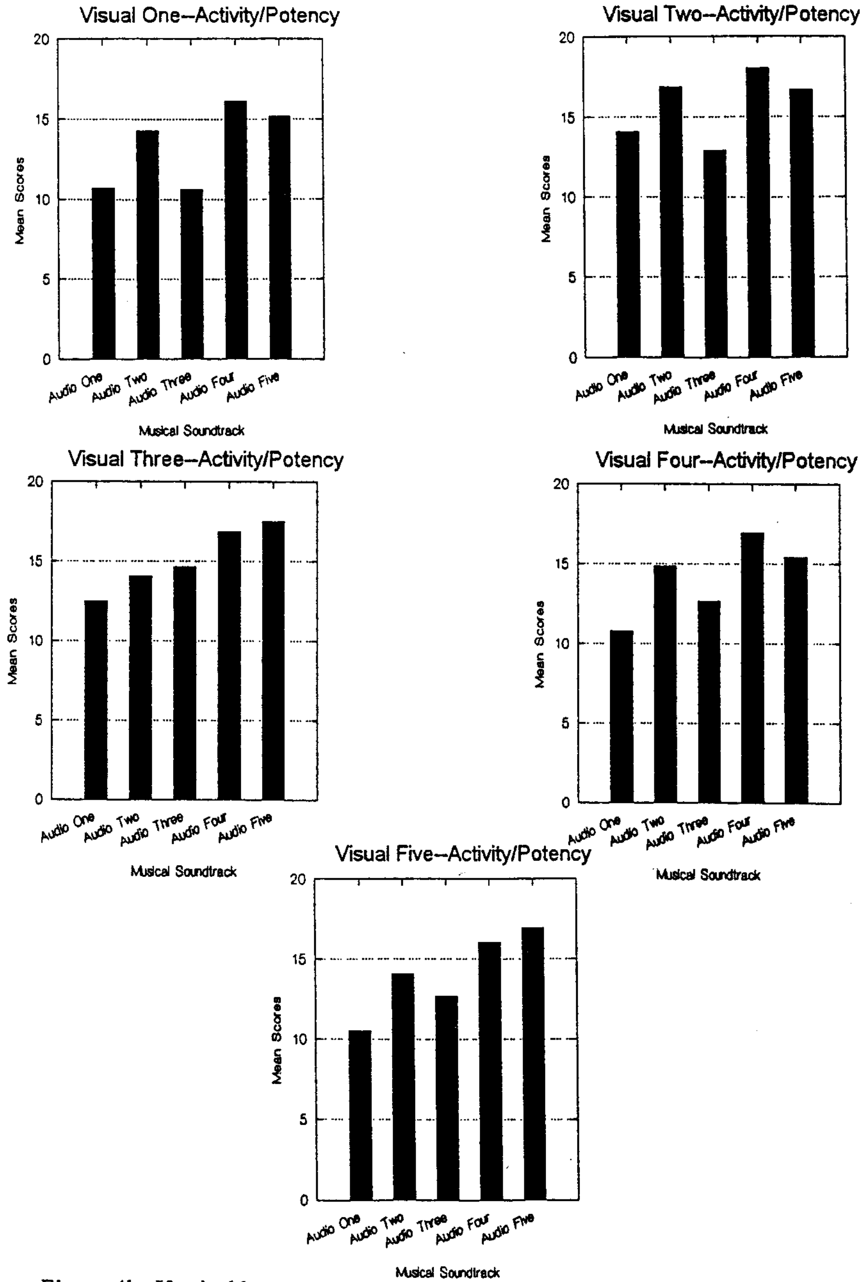


Figure 4b. Vertical bar graphs depicting the mean scores for the Activity/Potency factor while keeping the visual excerpt consistent and varying the musical soundtrack.

constant (Figure 4b), the contour similarity is striking. Note the consistency of the relatively low ratings for Audios 1 and 3 across all graphs with a slight peak on the mean responses for Audio 2. Audios 4 and 5 alternately attain the highest mean score. These two audios consistently receive the two highest mean ratings throughout the scores on the Activity/Potency dimension (except in Visual 2 when the composer-intended audio claims the second highest rating). The only score keeping this bimodal contour from being identical across all five bar graphs is the presence of a slightly higher mean score for Audio 3 on Visual 3, representing the composer-intended combination. The consistent contour exemplified in these Activity/Potency graphs (Figure 4b) suggests that the music exercises a strong and consistent influence over the subject responses provided to an audiovisual composite associated with this factor, regardless of visual stimulus.

In addition, the individual semantic differential ratings (Figures 2a and 2b) were more consistent across the aural domain (especially Audios 3, 4, and 5) than the visual. Confirmation for this statement is provided in Table 4. A comparison of the average response ranges (i.e., the vertical spacing of Figures 2a and 2b) was made for every semantic differential mean score within each visual excerpt and then recalculated within each audio example. The results were used to determine the average mean range for the visuals (6.51) and the average mean range for the audios (4.78). These data suggest that, in the context of the present investigation, not only was the musical soundtrack influential in the subject ratings, but it appears to have been a more consistent influence on the subject responses than the visual images themselves. This previous statement must be made with some qualification because of the variation apparent in the overlayed line graphs for Audios 1 and 3. Overall, however, statistical analysis has shown that the subject ratings do vary significantly when the musical soundtrack is changed.

Table 4
Average Range of Response Means for Each Semantic Differential Mean Score Within Each Visual Excerpt and Within Each Audio Example

Visual	Means	Audio	Means
Visual One	6.06	Audio One	5.94
Visual Two	5.51	Audio Two	4.86
Visual Three	7.23	Audio Three	6.23
Visual Four	6.06	Audio Four	2.93
Visual Five	7.71	Audio Five	3.92
Average = 6.51		Average = 4.78	

Discussion

Determination of Musical Parameters

As suggested by Felix Cohen (1929), the questions that we ask and the method of investigation determine, to a great extent, the answers we receive (cited in Langer, 1942, p. 4). Therefore, meticulous consideration must be given in the determination of those musical aspects selected for comparison, because the parameters chosen will be highly influential in the results obtained. Since this selection process is considered to be crucial to the conclusions of the present study, a review of relevant literature concerning possible musical parameters was deemed necessary (Hevner, 1935, 1936; Kendall & Carterette, 1990; Lomax, 1962, 1976; Lundin, 1985; Nakamura, 1987; Seashore, 1938; Senju & Ohgushi, 1987).

Analyzing 400 sets of recordings from about 250 different cultural areas, Lomax (1962) created categories in an attempt to identify what are, in many cases, extremely slight structural variations. Lomax's categories were utilized as a starting point for the music theoretical portion (i.e., analysis of the music itself) of the present study, resulting in a metric which provided responses at a much finer resolution than the broad dimensions suggested in much of the other literature cited above. Granted, not all of Lomax's musical "traits" are relevant in the context of the present investigation, but by utilizing the ones that *are* appropriate, the focus of each category will be narrowed significantly. For example, the musical parameter of melody has been divided into 10 related observations: melodic activity, contour, melodic repetition, melodic range, dominant intervallic width, embellishment, articulation, phrase length, number of phrases, and position of final tone. Using these categories resulted in a much more precise analysis of the musical examples than a general discussion of the broad music theoretical categories of melody, rhythm, and harmony.

Comparison of Subject Responses and Musical Parameters

Every musical soundtrack was analyzed by one investigator (S.L.). Table 5 provides the classification determined for each of the audio tracks. Although the notation was consulted and used as a guide, all decisions were made on the basis of the investigator's aural perception of each musical example. The entire list is provided for consideration in future investigations. However, several of the categories were eliminated from consideration in the present study because of the homogeneity between selections or irrelevance in the consideration of excerpts removed from their formal context.

Musical commonalities can be identified within the response contour clusters discussed previously in relation to the overlayed line graphs shown in Figure 2a and the cluster diagram represented in Figure 3. Audios 1 and 3 received higher ratings in Activity (e.g., active and fast) and were generally considered more beautiful, while Audios 2, 4, and 5 were rated higher across the Potency dimension (e.g., strong, heavy, and tense) and more agitated. In an effort to determine what elements in the aural domain might

Table 5

Music Parameter Classification for the 5 Audio Tracks

Music Parameter	Audio One	Audio Two	Audio Three	Audio Four	Audio Five
Tonality	tonal	chromatic	tonal	ambiguous	atonal/ single tonal chord
Mode & Chroma	G-Major	G-Phrygian	C Major	E(?)	G Major (?)
Pitch height	EE-g3	FF-e3	FF-g3	EE-f3	GG-f#3
Loudness					
average	*f*	*mp*	*mf-f*	*f*	*ff-p*
dynamics	little variation	varied	little variation	greatly varied	2 discrete levels
accent	consistent level	large variation	consistent level	heavily accented	sustained
Timbre					
dominant instruments	strings & brass	winds, brass, & strings	winds, brass, & percussion	winds, strings, & brass	electronic, strings, & winds
mass	high volume	medium volume	high volume	high volume	high volume
Tempo					
metronomic marking	180	66	140	62	52
meter	triple	duple	duple	duple	no pulse
deviation	slight broadening	heavy rubato	steady pulse	rubato	no pulse
Music type	melodic	melodic/rhythmic	mixed (m/(h)/r)[9]	rhythmic/harmonic	harmonic

(Continued)

Stylistic consistency	consistent	consistent	consistent	consistent	2 juxtaposed styles
Compositional approach	horizontal/vertical	horizontal	horizontal/vertical	vertical/horizontal	vertical
Melody					
activity level	high	moderate	high	moderate	low
contour	undulating/ascending	undulating	descending	undulating (asc. motives)	N/A
repetition	through-composed	through-composed	strophic (AAB)	through-composed	N/A
range	f#2-g3	Bb-e3	b1-g3	f1-e3	N/A
dominant intervallic width	highly scalar	scalar/chromatic	scalar/arpeggiated	scaler/chromatic	N/A
embellishment	simple	simple	simple	simple	N/A
articulation	mixed	slurred	articulated	mixed	N/A
phrase identification	recognizable	difficult to discern	easy (symmetrical)	recognizable	N/A
phrase length	relatively long	extremely short	medium	short	N/A
Harmony					
complexity	simple diatonic	complex	simple diatonic	complex	complex/simple
support	contrapuntal	solo/accompaniment	solo accompaniment & contrapuntal	solo/accompaniment	homophonic
Rhythm	counterpoint	accompanying/ accent	accompanying	accompanying/ homophonic	homophonic

have influenced these results, within-cluster commonalities were identified among the musical parameters given in Table 5. These similarities have been assembled into Table 6.

Using Table 6 and the information provided in the preceding paragraphs, several influential musical characteristics can be identified in the determination of semantic differential ratings across Osgood's three factors. Although Evaluative responses do, in fact, appear to depend on the appropriateness of the pairing of the audio and visual components, the subject ratings for beautiful/ugly provide an exception to this rule. The subjects responded with higher ratings on this differential for musical soundtracks with a clear sense of tonality, consistent dynamic level, steady pulse, and a diatonic melody above a supporting tonal harmonic structure. In contrast, correlations between the subject responses and specific musical parameters can be identified with a relatively high level of consistency across the other two dimensions. The Potency of a given scene (e.g., strength, heaviness, and tension) is positively influenced by complexity of the harmonic structure, absence of a clear tonal center, dynamic variation, presence of sforzandi, variation of the temporal fabric (rubato), complexity of the rhythmic structure, and melodic construction based on motivic fragments rather than flowing phrases. Higher ratings across the Activity dimension result from an allegro tempo with a steady pulse. This factor also appears to be positively affected by high melodic activity which continues at a consistent level, as opposed to an assemblage of fragmented motives presented within the context of a fluctuating temporal framework.

Table 6
Musical Parameters Considered Influential on Subject Ratings

Audios 1 & 3	Audios 2, 4, & 5
clear tonal center	chromatic, atonal, or ambiguous
diatonic (harmonically simple)	complex harmonic structures
little dynamic variation	varied dynamic level, accented
allegro tempo	slow tempo
tempo giusto (steady pulse)	rubato
organization by phrase	motivic organization
high melodic activity	less melodic activity

Proposed Models Of Film Music Perception: Future Directions

Does a subject's perception of a film sequence change appreciably when the musical soundtrack is varied? The preceding experimental procedure has answered this research question in the affirmative. But *how* and *why* does this change take place?

Continua of Musical and Visual Referentiality

Leonard Meyer (1956) suggested a classification of musical meaning into two major categories. *Referentialism* emphasizes the importance of the fact that musical sound can refer to extramusical objects, ideas, and actions. In contrast, *expressionism* implies that musical meaning is essentially intra-musical (i.e., areferential). Both referential and expressionist aspects of meaning are important to the film composer.[10]

In contrast to the privileged status often given to associational (i.e., referential) aspects of film music, the present authors suggest that a model of motion picture perception must focus upon the *source* of meaning within both the auditory and visual modalities. In this context, meaning is imposed by the perceiver upon experiences representing varying degrees of abstraction. It is possible to imagine a *continuum of referentiality* that stretches from "concrete" to "abstract" in either perceptual modality. Musical meaning, for example, can be highly referential (i.e., pointing to something outside of the music itself) or just as highly areferential (i.e., the meaning is inherent in the musical tones themselves and their relations one to another).[11] All other levels of meaning may be considered artifacts upon this continuum from areferential to referential (abstract to concrete, respectively). Similar attributes exist in the visual domain.[12]

Considering motion pictures as audiovisual composites, it is possible to place any given composite in a space representing the degree of abstraction of the musical dimension and the degree of abstraction of the visual dimension. A graphic representation of this two dimensional space is provided in Figure 5. Two examples from the film repertoire are represented by an "X" within this area. Norman McLaren's "Synchrony" (1971) is a piece of experimental animation in which abstract shapes appear on the screen as identical images pass over the photoelectric cell on the "sound track" portion of the film celluloid. The resulting tones are not intended to have a conventional association for the typical viewer. Therefore, the degree of abstraction is high in both the visual and the musical domains, since the "meaning" is a function of the synchronization of the sound and images. The second example in Figure 5 represents one of the final scenes from Edward Zwick and James Horner's "Glory" (1989) in which a troop of soldiers prepares to march to war as the themes that have become associated with the various characters and situations throughout the film are heard on the soundtrack. This scene is extremely "concrete" (i.e., low degree of abstraction) in its representation of human drama while the music is referential both thematically and stylistically.

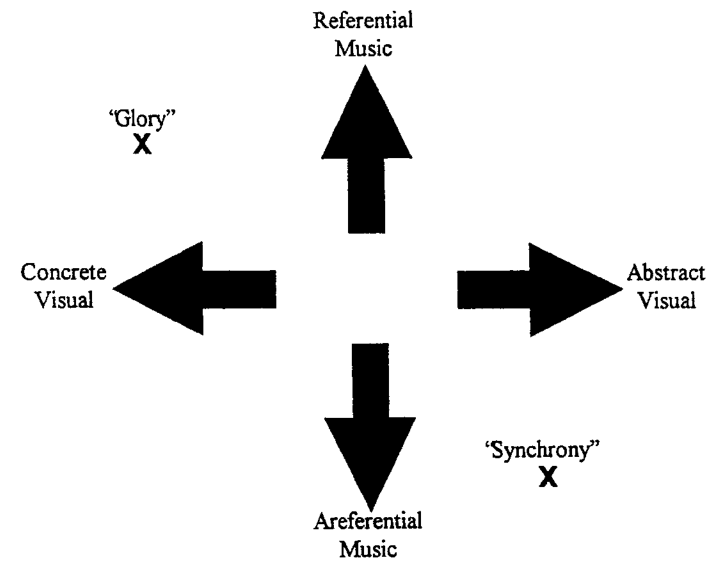


Figure 5. Two dimensional space representing the continuum of referentiality in both the auditory and visual components of the motion picture experience. Two specific cinematic examples are mapped into this space (see text).

Musical and visual periodicity. In addition to considering the aspect of associational (i.e., referential) meaning, one must also take into account the syntactical (i.e., expressionist) alignment of accent structure between the aural and visual strata in the context of a motion picture. There are many examples that illustrate the film composer's use of periodicity in the musical structure as a means of heightening the effect of recurrent motion in the visual image.[13] In the process of perception, an individual seeks out such periodicities (i.e., recurrent motion or a repeated melodic motif) in order to facilitate data reduction.[14]

Yeston (1976) proposed three possible relationships between metrical (i.e., regularly occurring) sources of accent: consonance, dissonance, and out-of-phase consonance.[15] Figure 6 provides an example of each of these three relationships. Positive "pulses" in the diagrams represent an emphasized event. Time is represented as flowing from left to right along the horizontal axis. Each of these representations should be considered as an extremely simplified example of a theoretically infinite number of possibilities. In a motion picture, it is not necessary for the music and visual images to be in perfect synchronization for the composite to be considered "appropriately aligned." The degree to which the two strata must be misaligned before perceived synchronicity breaks down is a matter for future research.

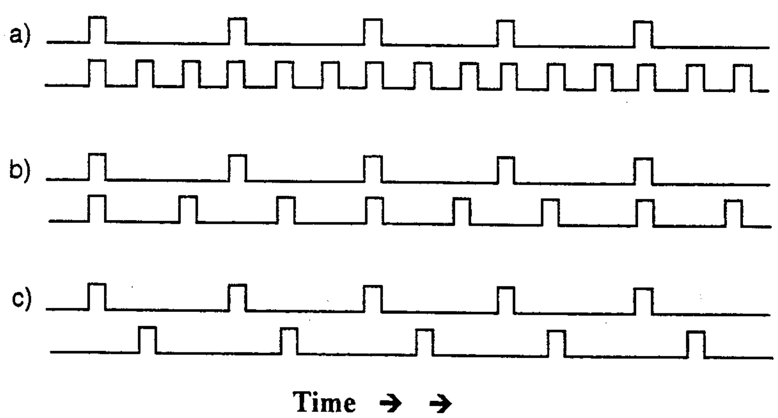


Figure 6. Visual representations of relationships between sources of accent. In each pair, the upper pulse train represents the musical stratum and the lower pulse train represents the visual stratum.

Film music paradigm. Taking into account both the referential and syntactical aspects of the motion picture experience, the following model of film music perception is proposed for future investigation. An effective film score, in its interactive association with the visual element, need not attract the audience member's attention to the music itself. In fact, the most successful film composers have made a fine art of manipulating audience perception and emphasizing important events in the dramatic action without causing a conscious attentional shift. In fact, when watching a film, it is quite possible that perception of the musical component will remain at a subconscious level (Lipscomb, 1989). A possible explanation for this lack of attentional focus on the musical score is that there are two implicit judgments made during the perceptual processing of the motion picture experience: an association judgment and a mapping of accent structures. The association judgment (i.e., referential meaning) relies on past experience as a basis for determining whether or not the music is appropriate for a given context (e.g., legato string lines for "romantic" scenes, brass fanfares for a "majestic" quality, or low frequency synthesizer tones for a sense of "foreboding"). Such judgments are made implicitly, based on an individual's past experience perceiving similar stimuli.

The second implicit judgment (i.e., mapping of accent structures) matches the emphasized points in one perceptual modality with those in the other, as discussed above. The present investigators propose that, if the associations identified with the musical style are judged appropriate and the relationship of the aural and visual accent structures is consonant (or, possibly, a simple ratio dissonance; e.g., 1:2, 1:3, 2:3, etc.), attentional focus will be maintained on the symbiotic composite, rather than on either modality in isolation. Figure 7 presents a model of attentional focus as a function of audiovisual congruence. Each of the small boxes contained in the Implicit Processes box represents a decision concerning the appropriateness of the audiovisual pairing. As can be seen by the arrows along the pathway, if the composite is determined to be inappropriate at either level, the subject is more likely to separate the unit into its two constituent perceptual modalities in an explicit analytical attempt to determine a reason for the incongruence. There are no absolute answers to the appropriate/inappropriate response. This decision will be determined by the subject's past experience and is, as a result, individual-specific.

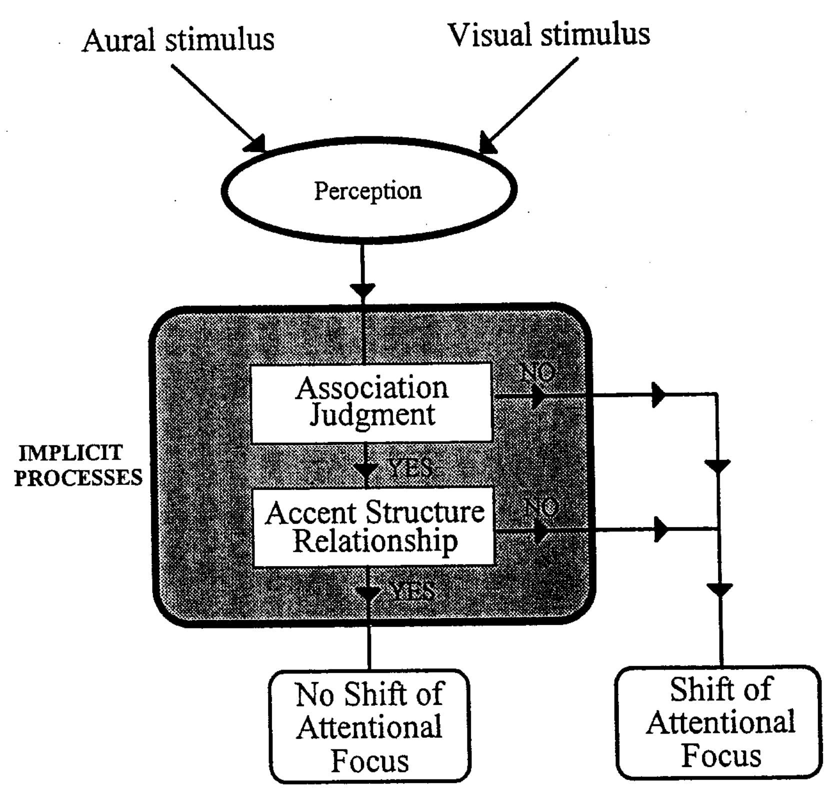


Figure 7. Model of attentional focus as a function of audio/visual congruence.

Conclusion

The preceding study has provided evidence of the powerful effect a musical soundtrack has on the motion picture experience. Recall that, in every case, a majority of the subjects in Experiment 1 was able to select the composer's intended soundtrack for five visual scenes. In Experiment 2, the subject ratings given on 10 semantic differential scales to these same five scenes varied significantly depending on the musical score utilized.

The sociological significance of film as an artform within contemporary American society (as well as many other cultures around the world) warrants thorough research. In the Music Perception and Cognition laboratory at UCLA there is ongoing research into the interaction of the perceptual modalities at work when viewing a motion picture. At present, we are investigating possible visual correlates for the musical parameters of pitch, loudness, timbre, and duration. In addition, we are looking systematically at the contour strata in both the audio and visual domains and their interrelationship.

References

Bermant, R. I., & Welch, R. B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Perceptual and Motor Skills, 43,* 487-493.

Boltz, M. *Time estimates of filmed narratives: Effects due to the relative placement and number of commercial breaks.* Unpublished manuscript.

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

Cohen, F. (1929). What is a question? *The Monist, 31,* 350-64.

Eisenstein, S. (1947). *The film sense.* New York: Harcourt Brace Jovanovich.

Evans, M. (1975). *Soundtrack: The music of the movies.* New York: Da Capo Press.

Gorbman, C. (1987). *Unheard melodies.* Bloomington, IN: Indiana University Press.

Hevner, K. (1935). Expression in music: A discussion of experimental studies and theories. *Psychological Review, 42,* 186-204.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48,* 246-69.

Jones, M. R., Boltz, M. G., & Klein, J. M. (1993). Expected endings and judged durations. *Memory & Cognition, 21,* 646-65.

Kendall, R. A., & Carterette, E. (1990). The communication of musical expression. *Music Perception, 8,* 129-64.

Kendall, R. A., & Carterette, E. C. (1991). Perceptual scaling of simultaneous wind instruments timbres. *Music Perception, 8,* 369-404.

Kohler, W. (1929). *Gestalt psychology.* New York: Liveright Publishing Company.

Kracauer, S. (1960). *Theory of film.* London: Oxford University Press.

Langer, S. K. (1942). *Philosophy in a new key: A study in the symbolism of reason, rite, and art.* Cambridge, MA: Harvard University Press.

Lipscomb, S. D. (1989, March). *Film music: A sociological investigation of influences on audience awareness.* Paper presented at the Meeting of the Society of Ethnomusicology, Southern California Chapter, Los Angeles.

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, CA.

Lomax, A. (1962). *Song structure and social structure. Ethnology, 1,* 425-51.

Lomax, A. (1976). *Cantometrics.* Berkeley: University of California Press.

Lundin, R. W. (1985). *An objective psychology of music* (3rd ed.). Malabar, FL: Robert E. Krieger.

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

Massaro, D. W., & Warner, D. S. (1977). Dividing attention between auditory and visual perception. *Perception & Psychophysics, 21,* 569-74.

Mershon, D. H., Desaulniers, D. H., Amerson, T. C. (Jr.), & Kiever, S. A. (1980). Visual capture in auditory distance perception: Proximity image effect reconsidered. *Journal of Auditory Research, 20,* 129-36.

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

Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics, 41,* 576-600.

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception and Psychophysics, 41,* 525-33.

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

Prendergast, R. M. (1977). *Film music: A neglected art.* New York: W.W. Norton.

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

Regan, D., & Spekreijse, H. (1977). Auditory-visual interactions and the correspondence between perceived auditory space and perceived visual space. *Perception, 6,* 133-8.

Ruff, R. M., & Perret, E. (1976). Auditory spatial pattern perception aided by visual choices. *Psychological Research, 38,* 369-77.

Schiffman, S. E., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling.* New York: Academic Press.

Seashore, C. E. (1938). *Psychology of music.* New York: Dover Publications.

Senju, M., & Ohgushi, K. (1987). How are the player's ideas conveyed to the audience? *Music Perception, 4,* 311-24.

Staal, H. E., & Donderi, D.C. (1983). The effect of sound on visual apparent movement. *American Journal of Psychology, 96,* 95-105.

Tannenbaum, P. H. (1956). Music background in the judgment of stage and television drama. *Audio-Visual Communications Review, 4,* 92-101.

Thayer, J. F. & Levenson, R. W. (1983). Effects of music on psychophysiological responses to a stressful film. *Psychomusicology, 3,* 44-54.

Thomas, T. (1973). *Music for the movies.* New York: A.S. Barnes.

Thomas, T. (1991). *Film score: The art & craft of movie music.* Burbank, CA: Riverwood Press.

Weis, E., & Belton, J. (1985). *Film sound: Theory and practice.* New York: Columbia University Press.

Yeston, M. (1976). *The stratification of musical rhythm.* New Haven, CT: Yale University Press.

Zettl, H. (1990). *Sight sound motion: Applied media aesthetics.* Belmont, CA: Wadsworth.

Footnotes

[1] Throughout this study, the words *film*, *motion picture*, and *video* will be used interchangeably. We acknowledge the distinction between the three terms and the types of media referenced.

[2] Developed during research on synesthesia, the semantic differential scale is one method employed to identify patterns in the stimulus/response chain of human perception and cognition for the purpose of inferring a cause-and-effect relationship. Semantic differentiation is defined by Osgood et al. (1957) as "the successive allocation of a concept to a point in the multidimensional semantic space by selection from among a set of given scaled semantic alternatives." In reference to the concept being evaluated (music, in the present study), he identifies two properties of this point in semantic space: direction from the origin (determined by the bipolar adjectives selected) and distance from the origin (i.e., extremeness of the subject response on the rating scale). These two properties of the semantic differential are assumed to be related to conceptual mediators evoked in the subject and the intensity level of this evocation, respectively (Osgood et al., 1957, pp. 26-30). Osgood, et al. divided the semantic differential scale into three factors: Evaluative, Potency, and Activity. Examples of specific bipolar adjectives which load heavily on each of these factors are provided in the body of the text.

[3] Marilyn Boltz and her colleagues have, however, used television series episodes in her work on memory and perceived temporal duration (Boltz, unpublished; Boltz, Schulkind, & Kantra, 1991; Jones, Boltz, & Klein, 1993).

[4] Several subjects took advantage of this at various times, but never were more than three scenes reviewed for a single visual excerpt. In most cases, the decision had been made by the end of the fifth combination for each scene.

[5] In discussing this data set, trends apparent in the data are presented rather than results of a statistical analysis, because the number of subjects does not meet the minimal criteria for number-per-cell when using Chi-square.

[6] Ideally, the entire design would have been run as a single ANOVA, but neither of the available statistical software packages could handle the amount of data required.

[7] For a more detailed discussion of these results, see Lipscomb (1990).

[8] A detailed comparison of these line graphs is provided in Lipscomb (1990).

[9] melodic, (harmonic) & rhythmic

[10] General conventions (e.g., "romantic" string lines, "majestic" brass, etc.) assist the composer in creating music considered appropriate for a given setting. By using these conventions, the composer must rely on the listener to provide the appropriate emotional response, based on past experience. Similarly, the relationship between the musical sounds themselves is important in supplying a means for the musical expression of psychological drama (e.g., tension expressed musically as dissonant chord clusters) and in providing a means of unification (e.g., the recurring theme or motive that becomes associated with a character or idea).

[11] *The Star Spangled Banner* is an example of the former, while the pieces in J.S. Bach's *Das Wohltemperirte Klavier* are representative of the latter. It is imperative to understand, however, that association of the national anthem to the country which it represents does not nullify the meaning inherent in its tonal relationships. Likewise, simply because a piece of music does not have a strong associational meaning, it is far from the truth to suggest that the piece has *no* meaning. Rather, the meaning is of a different type (i.e., syntactic instead of semantic, dealing with relationships rather than labels).

[12] Compare, for instance, the abstraction inherent in the paintings of Jackson Pollack (e.g., "One" [1950]) to the realistic portrayals of Gustave Courbet (e.g., "The Stone Breakers" [1849]) or contrast the films produced by the experimental animators of the 1930s and 40s (e.g., Oskar Fischenger's "Allegretto" [1936] and Norman McClaren's "Dots" [1943]) to the classic mainstream films of approximately the same era (e.g., Michael Curtiz and Max Steiner's "Casablanca" [1943] or William Wyler and Hugo Friedhofer's "The Best Years of Our Lives" [1946]).

[13] The galley rowing scene from Miklos Rosza's score composed for *Ben Hur* (1959) is an excellent example of the mapping of accent structures, both in the pitch and tempo of the musical score. As the slaves pull up on the oars, the pitch of the musical motif ascends. As they lean forward to prepare for the next thrust, the motif descends. Concurrently, as the Centurion orders them to row faster and faster, the tempo of the music picks up accordingly, synchronizing with the accent structure of the visual scene. A second illustration may be found in John Williams' musical soundtrack composed for *ET: The Extraterrestrial* (1982). The bicycle chase scene score is replete with successful musical emulation of the dramatic action on-screen. Synchronization of the music with the visual scene is achieved by inserting 3/8 patterns at appropriate points so that the musical accents of the metrical structure remain aligned with the pedaling motion. The fact that *every* rowing or pedaling motion is not perfectly aligned with the musical score is not perceived by the average member of the audience (even if attention were somehow drawn to the musical score).

[14] The Gestalt psychologists proposed that humans seek organization, imposing order upon situations that are open to interpretation utilizing the principles of good continuation, closure, similarity, proximity, and common fate (Köhler, 1929).

[15] See Monahan, Kendall, & Carterette (1987) for an extension of Yeston's concept of stratification to pitch-time strata in melody.

Author Note

Portions of this investigation were carried out in partial fulfillment of the requirements for Scott Lipscomb's Master of Arts Degree in the Department of Ethnomusicology and Systematic Musicology at the University of California, Los Angeles (Lipscomb, 1990). Financial assistance was provided to Mr. Lipscomb in the form of fellowships from the UCLA Department of Ethnomusicology and Systematic Musicology and the UCLA Affiliates' Elaine Krown Klein Fine Arts Scholarship. Roger Kendall also received a research grant from the UCLA Academic Senate. Part of this study was presented at the 2nd International Conference on Music Perception and Cognition (UCLA; February, 1992).

We extend our gratitude to Leonard Rosenman for his interest and willingness to discuss various aspects of his compositional process, to Eldridge Walker and Bob Borenstein in the Music Division of Paramount Studios for allowing access to the musical notation of the musical excerpts utilized in the present investigation, and to Bob Knight of Famous Music for allowing reductions of the musical scores to appear in Appendix B. We also gratefully acknowledge Sue Carole DeVale, Paul Reale, Andrea Halpern, and the reviewers for their commentary on earlier drafts of this article. Paul Attinello, Alexandra Harwood, Amy Rawlings, and Doug Scott assisted in making decisions concerning the alignment of the audio track in each composite used in this study. We also express our sincere appreciation to the University of California, Los Angeles for the use of its facilities.

Request for reprints should be sent to Scott D. Lipscomb at the University of Texas at San Antonio, Division of Music AR 3.01.58A, 6900 N. Loop 1604 West, San Antonio, TX 78249-0645 or to Roger Kendall, Program in Systematic Musicology, Department of Ethnomusicology, UCLA, 405 Hilgard Ave., Los Angeles, CA 90024.

(Manuscript received January, 1993; revision accepted May, 1994)

Appendix A

The following is a description of the stimulus materials used.

Video timings were taken from the readout of the video cassette recorder. The counter was reset to 0:00:00 when the "Paramount Pictures Presents" title appears following the "Paramount 75th Anniversary" logo. All visual stimulus time references (in *hours:minutes:seconds* format) specify the beginning and ending of each individual excerpt.

Visual One (1:46:52 to 1:47:22): The crew of the Enterprise splashes happily in the water as two whales swim to freedom.

Visual Two (1:17:20 to 1:17:57): Jillian discovers the whales missing and expresses her anger to her supervisor. This is the only visual scene utilized in the present study in which dialogue (i.e., one-on-one verbal interaction between characters) occurs in the excerpt selected from the original film.

Visual Three (1:25:59 to 1:26:39): The crew members rescue Chekov in the hospital chase scene.

Visual Four (0:16:38 to 0:17:15): The probe hovers over the Earth, reeking havoc on atmospheric conditions.

Visual Five (0:31:39 to 0:32:15): The final segment of the "time travel" sequence in which the image of a falling body splashes into water and the scene fades into a silhouette of Captain Kirk.

Audio timings were taken from the display of the compact disc player, playing the Original Motion Picture Soundtrack recording of *Star Trek IV: The Voyage Home* (MCAD-6195). Audio stimulus timings are given in *minutes:seconds* format.

Audio One (Track 4; 7:30 to 8:00)—"Whale Fugue"

Audio Two (Track 6, 0:00 to 0:37)—"Gillian Seeks Kirk"

Audio Three (Track 9, 0:00 to 0:40)—"Hospital Chase"

Audio Four (Track 10, 0:00 to 0:37)—"The Probe;" see discussion concerning this Audio excerpt.

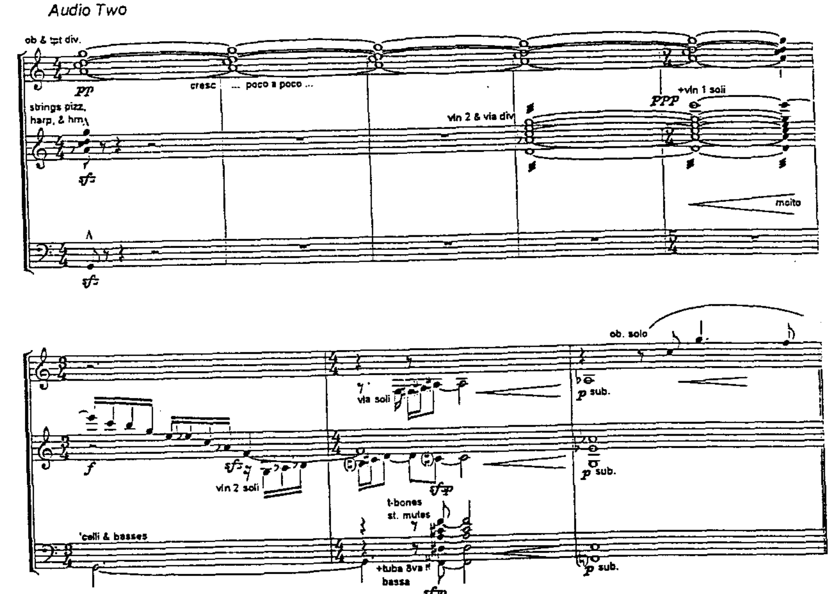
Audio Five (Track 8, 0:46 to 1:22)—"Time Travel;" this music was replaced by sound effects and speech sounds in the final production phase. However, the composer confirmed that this was the intended music for the sequence.

Appendix B

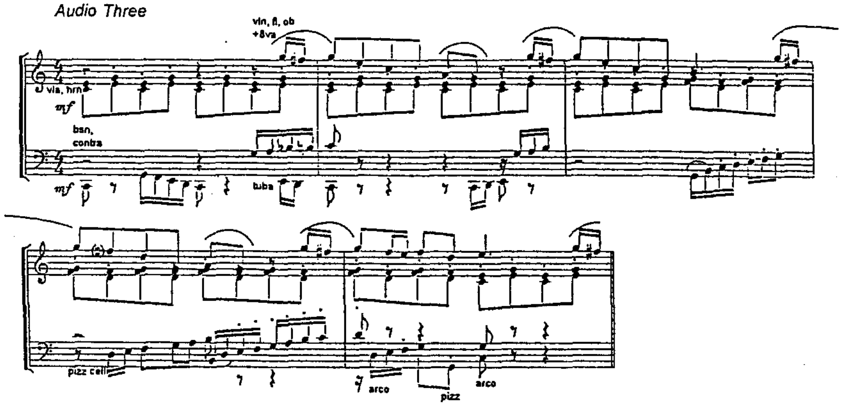
The following are the author's (R. K.) piano reductions of the opening measures for each musical excerpt used in the present investigation. Music by Leonard Rosenman (1986 Famous Music Corporation, [ASCAP]). All rights reserved—used by permission.

Audio One

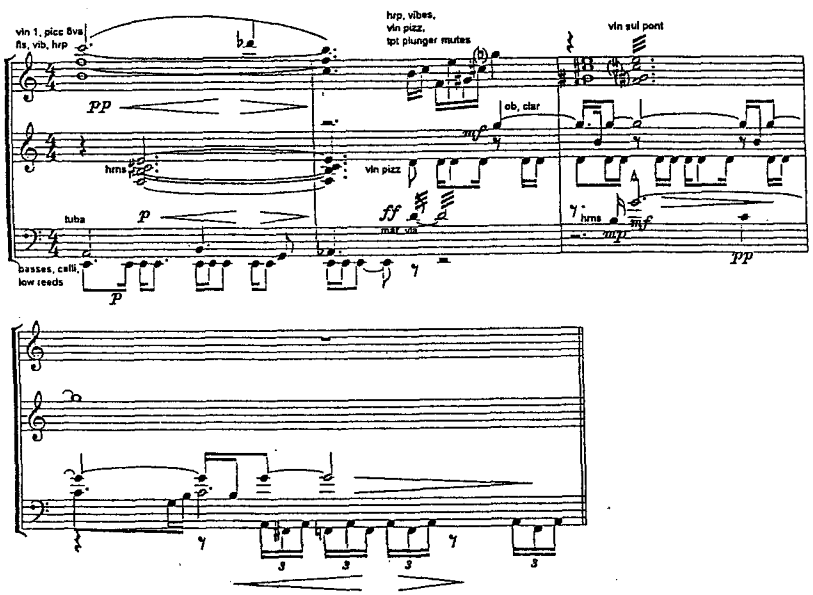
Audio Two



Audio Three



Audio 4



Audio Five

The musical score for Audio Five was not available. Rosenman's orchestrated score fades out prior to the beginning of the scene used, as an electronic-sounding timbre appears. This soundtrack may be heard on the compact disc recording (see Appendix A).