

RUNNING HEAD: Immersion in the virtual environment...

Immersion in the virtual environment: The effect of a musical score on the video gaming experience

Scott D. Lipscomb¹⁾ and Sean M. Zehnder²⁾
Northwestern University, USA

1) Music Education and Music Technology programs, School of Music;

*2) CDMC*Digital-Kids Lab; Media, Technology and Society program, School of
Communication*

Word Count: 3,993, including footnotes and references

Contact information for primary author:

Dr. Scott D. Lipscomb
Northwestern University School of Music
711 Elgin Road
Evanston, IL 60201 USA
phone: (847) 467-1682
FAX: (847) 491-5260
lipscomb@northwestern.edu

Abstract (100-150 words)

This study provides one of very few experimental investigations into the impact of a musical soundtrack on the video gaming experience. Participants were randomly assigned to one of three experimental conditions: game-with-music, game-without-music, or music-only. After playing each of three segments of *The Lord of the Rings: The Two Towers* (Electronic Arts, 2002) – or, in the music-only condition, listening to the musical score that accompanies the scene – subjects responded on 21 verbal scales. Results revealed that some, but not all, of the verbal scales exhibited a statistically significant difference due to the presence of a musical score. In addition, both gender and age level were shown to be significant factors for some, but not all, of the verbal scales. Details of the specific ways in which music affects the gaming experience are provided in the body of the paper.

Word Count: 140

Keywords: cinema, film music, film sound, music perception, music cognition, sound design, soundtrack, video game, virtual environment

Immersion in the virtual environment: The effect of a musical score on the video gaming experience

Scott D. Lipscomb¹⁾ and Sean M. Zehnder²⁾
Northwestern University, USA

1) Music Education and Music Technology programs, School of Music

2) CDMC*Digital-Kids Lab; Media, Technology and Society program, School of Communication

Introduction

During the past three decades, dramatic technological developments have resulted in a significant enhancement of the aural component accompanying virtual experiences. In movie theaters worldwide, according to, ultra-high fidelity surround sound systems are present in nearly 100,000 theaters (Dolby Lab, 2004). During this same period, the sonic accompaniment for video games has progressed from the synthesized waveform beeps of *Pong* (Atari, 1975) to role-playing games (RPGs) that offer highly customizable lists of commercial music tracks (e.g., *Tony Hawk's Underground*; Activision, 2003) or incorporate complete specially composed orchestral scores (e.g., *Medal of Honor Frontline*; Electronic Arts, 2002; see Zehnder and Lipscomb, in press, for a more detailed discussion). The significantly faster data processing capabilities and greatly increased storage capacity for both game systems (Sony PlayStation 2, Microsoft Xbox, and Nintendo GameCube) and personal computers take great advantage of recently developed methods for compressing the audio and video components of these games, producing audio-visual (A-V) quality that, utilized to its fullest extent, begins to rival the quality of film and television. Given the amount of time spent by individuals using or interacting with media – an average of five and a half hours a day for children in the United States (Rideout, 2004) – and the serious attention given to the audio aspect of motion pictures and in the most recent generation of video games (Hill, 2003), the importance of investigating the role that music plays in this context is undeniable.

Since the mid-1980s, a significant number of investigations has begun to address the impact of the musical score upon the experience of motion pictures and animation (e.g., Boltz, 2001; Bullerjahn and Gldenring, 1994; Iwamiya, 1994; Iwamiya et al, 2001; Lipscomb, 1995; Lipscomb, in press; Lipscomb and Kendall, 1994; Marshall and Cohen, 1988; Thayer and Levenson, 1983; Thompson et al, 1994; for a more detailed review, see Cohen 2001 or Lipscomb and Tolchinsky, in press).

Few empirical studies have investigated the relationship of music to the video gaming

experience. One exception is a study by Yamada et al (2001; 2002), in which the presence of music is shown to have a negative effect on the performance in a video racing game, based on the score obtained. Using a semantic differential response scale (the same scales used by Iwamiya, 1997), participants were also asked to evaluate the gaming experience in conditions with and without music and also in a music-only condition, revealing a complex relationship between the presence of music and its overall impression and affect on driving performance. Yamada et al conclude that:

The positive correlation of the "neatness" axis vs. the "lightness," "calmness," and "performance score" vectors, shows that a "dark, agitated" excerpt provides a "mixed" impression to the game, and, in turn, results in a negative effect on driving performance. In order to obtain high performance scores, players must concentrate on manipulating the control unit. The negative effect on performance may be a result of music disturbing the players' concentration.

The musical excerpts used by Yamada et al (2001, 2002) were selected from a wide range of varying musical styles for use in the experiment based on a music-only semantic differential rating task. As a result, the results of the study are likely to be significantly different from those in which a musical score has been specifically composed to accompany events as a dramatic narrative unfolds. To increase the ecological validity of results reported herein, the present study incorporates the music of Academy Award-winning composer Howard Shore, as it was intended to accompany the video gaming experience for *The Lord of the Rings: The Two Towers* (LR:TT; Electronic Arts, 2002).

The present study will focus solely on the impact that the musical score has on the video gaming experience, without regard for the improvement or degradation of performance scores. Of particular interest are the following research questions:

1. Do players perceive a game differently due to the presence of music?

2. If so, in what ways does music impact the aesthetic experience of gaming?
3. Do attribute variables (e.g., gender or age level) significantly impact participant responses?

We predicted that responses to a number of verbal scales – particularly those associated with Osgood et al's (1957) "potency" dimension – would be positively enhanced by the presence of the musical score and would intensify the overall gaming experience for the player.

Method

Participants included 76 students, 63 enrolled at Northwestern University (19 males and 44 females) and 13 high school students in Evanston, Illinois (six males and seven females).

Demographic information, including age level and gender, was collected from each participant using a pre-experimental survey form. Each subject was randomly assigned to one of three experimental conditions: game-with-music, game-without-music, and music-only.

Creating experimental stimuli for the first two experimental conditions was trivial, since *LR:TT* allows players the option of turning the musical soundtrack on or off without affecting the presence of ambient sounds. The stimuli for the third condition, however, posed a certain challenge, since it is not possible to turn off the non-musical sound unless the auditory component is eliminated completely. In order to create these materials, therefore, it was necessary to extract segments of the in-game musical soundtrack multiple times, digitally sampling the audio output from the PlayStation 2 with a MOTU 828 audio interface connected to a Macintosh G4 Powerbook running Logic Audio Platinum5. These excerpts were then edited to ensure that all non-musical sounds were eliminated and reconstructed into an intact, natural sounding musical soundtrack. Lengths of the music-only stimuli for the three video game segments were 54, 86, and 85 seconds in duration, respectively. In comparison, scenes chosen for the game playing conditions were approximately 120 seconds each.

Participants in each of the first two conditions played three segments (*Weather Top*, *Moria*, and *Amon Hen*) from *LR:TT* on the Sony PlayStation 2, controlling the character Aragorn. The characters, objectives, and musical score for each segment are briefly described below.

Weather Top. In this second-level segment, the player's primary objective is to protect the character Frodo from menacing creatures known as Ringwraiths. The nighttime setting is high atop a mountain ruin, surrounded by forest. The player must repeatedly re-light and wield a torch in order to fend off the enemy. The musical score consists

of a driving rhythm with string glissandi at the beginning, gradually replaced by a low string ostinato with a prominent presence of a vocal choir, singing words in one of the languages of Middle Earth. This excerpt maintains a clear sense of tonality throughout.

Moria. In this segment encountered later in the second level, the player must fight through a marsh while unhuman creatures (Orcs) attack from the water and arrows fly from surrounding cliffs. This scene is also very dark, as if at night, and the player experiences periods of relative quiet, punctuated by an occasional attack from Orcs hidden in the marsh. The musical score for this excerpt is highly dissonant, loud, and brash, initiated at the outset by a series of ascending chromatic motifs played by brass, woodwind, and string instruments, followed by a brief section of pounding percussion rhythms that is interrupted briefly by the same chromatic ascending motifs, then returning to the pounding rhythm. The tonal clusters in this excerpt result in an atonal harmonic context.

Amon Hen. In this excerpt from the seventh level of the video game, the player must fight ferociously to make a path through the forest, as numerous fierce fighters (Uruk-Hai, a stronger and larger breed of Orc) are encountered. The daytime scene is set in a forest and includes a large number of enemies in close combat situations. The musical score for this segment is majestic in tone featuring a steady, syncopated rhythm in the percussion and low strings with brass instruments in the melodic foreground over a clearly tonal harmonic structure. Eventually, following the sound of a hunter's horn, the brass instruments are supplanted by strings in a chordal homophonic texture, and then, the brass return to the majesty of the opening section.

Following each segment of game play, participants responded on a series of 21 verbal scales, using a computer interface and a scroll bar that recorded the responses on a scale from 0 to 100. Those assigned to the music-only condition simply listened to the musical score, then provided ratings on the same verbal scales. Rather than using the semantic differential scale which consists of approximate opposites (bad-good), the present study utilized verbal attribute magnitude estimation (VAME; Kendall and Carterette, 1993) in which true opposites are used instead (not good-good). The attributes used in the present study included the following terms: active, annoying, bright, busy, cold, colorful, dangerous, exciting, fast, gentle, good, high, intense, labored, loud, masculine, pleasant, powerful, relaxed, simple, and strange.

Results

Due to space limitations associated with the Rapid Communications format, results of the present research paper will be limited to those focusing specifically on statistically significant

differences related to experimental condition and interactions involving this same between-subjects variable. In addition, only statistically significant main effects and two-way interactions will be discussed herein.¹

A repeated measures analysis of variance procedure was utilized to examine subject responses and determine which differences were statistically significant. In this context, three between-subjects variables were of interest: experimental condition, gender, and age level (high school vs. college). The single within-subjects variable of interest was the video game excerpt (*Weather Top*, *Moria*, or *Amon Hen*) immediately preceding the collection of VAME responses. An initial repeated measures ANOVA, incorporating VAME responses as a second within-groups variable, revealed no main effects or interactions when considering the between-subjects variables. However, highly significant differences emerged when considering the within-subjects variables and their interactions.

Research participants responded differently to the various VAME scales ($F_{(20,1280)}=22.606$, $p<.0005$). Participants from high school responded differently to the verbal scales than university students ($F_{(20,1280)}=1.760$, $p=.020$). Significant differences were found for the main effect of game segment ($F_{(40,2560)}=1.464$, $p=.031$) and also in the interactions between gender and video game segment ($F_{(40,2560)}=1.657$, $p=.006$) and age level and game segment ($F_{(40,2560)}=1.622$, $p=.008$). Most important to the present study are the statistically significant main effect of experimental condition ($F_{(40,1280)}=1.522$, $p=.020$) and the interaction between VAME scales and experimental condition ($F_{(80,2560)}=1.330$, $p=.028$), confirming that the verbal ratings provided by participants do appear to be significantly influenced by the presence or absence of music.

To look at the relationship between the verbal scales and the experimental conditions more closely, a repeated measures ANOVA was computed for each verbal scale independently. Several VAME scales revealed no significant difference in subject responses involving either a main effect or interaction with experimental condition. Organized into Osgood et al's (1957) dimensions, these terms include evaluative concepts (cold, exciting, good, and pleasant), potency terms (gentle, high, intense, loud, and powerful), and a single scale representing the

activity dimension (active). Statistically significant main effects and two-way interactions will be discussed below.

Group main effect. Four VAME scales revealed a significant main effect of experimental condition (music only, game without music, and game with music): colorful ($F_{(2,64)}=4.204$; $p=.019$), dangerous ($F_{(2,64)}=4.875$; $p=.011$), relaxed ($F_{(2,64)}=3.344$; $p=.042$), and simple ($F_{(2,64)}=6.844$; $p=.002$). As shown in Figure 1, experimental condition affected the mean subject ratings in different ways. For the "colorful" rating, the music-only condition received a higher rating (45.85) than the game-without-music condition (39.49). The game-with-music condition revealed an *additive* relationship, since the combination of music and visual was rated higher (50.81) than either of the other two conditions. For the "relaxed" rating, however, an opposite (*subtractive*) relationship emerges, since the game-without-music condition receives the highest mean rating (38.35), the music-only condition is significantly lower (32.17), and the game-with-music condition results in the lowest mean rating (25.75).² The "dangerous" rating reveals yet another type of relationship, which might be referred to as *musical transference*. In this case, a moderate mean rating is given to the game-without-music condition (48.25) and a much higher mean rating is given to the music-only condition (61.31). When the audio and visual components are combined, the mean rating is nearly identical to that of the music-only condition (61.20), as if the rating on this scale is simply transferred from the musical stimulus onto the A-V combination. Finally, the "simple" rating illustrates yet another type of relationship. Notice that both the game-without-music and music-only conditions result in nearly identical mean ratings (55.59 and 55.76, respectively), while the game-with-music condition receives a much lower mean score (39.76), a result that will be referred to as *combinatorial negativity*. It is worth noting that, because of the VAME scale used, a lower rating is associated with a lower level of simplicity or, conversely, greater complexity. Therefore, in this context, the combination of a moderately simple game with no musical soundtrack and a moderately simple musical score is perceived as a less simple – or more complex – A-V stimulus.

[insert Figure 1 about here]

Group x Gender interaction. Two VAME scales revealed a statistically significant interaction between group assignment and gender: dangerous ($F_{(2,64)}=3.269$; $p=.045$) and busy ($F_{(2,64)}=8.586$;

¹ For the sake of completeness, all three- and four-way interactions involving the Group between-subjects variable that were revealed to be statistically significant are provided: Group x Scene x School [annoying ($F_{(3,693,118,165)}=2.623$; $p=.042$) and labored ($F_{(3,947,126,305)}=3.947$; $p=.050$)], Group x School x Gender [bright ($F_{(2,64)}=4.053$; $p=.022$)], Group x Scene x Gender [masculine ($F_{(3,734,119,489)}=3.572$; $p=.010$) and fast ($F_{(3,852,123,277)}=3.043$; $p=.021$)], and Group x Scene x School x Gender [busy ($F_{(3,974,127,157)}=2.457$; $p=.049$)].

² This interpretation may be slightly misleading, however. Rather than considering this a subtractive relationship on the "relaxed" scale, it could be argued that this is, by interpolation, an additive relationship of "less relaxed."

$p < .0005$). For the “dangerous” rating, the manner in which males and females rate game segments is quite different (Figure 2). There is little variation in the mean ratings provided by females in response to the three experimental conditions, with the mean rating for the game-without-music and game-with-music conditions almost identical and the music-only mean rating just slightly higher. In contrast, ratings provided by males vary widely as a result of experimental condition. An extremely high mean rating emerges in response to the music-only condition, an extremely low mean rating is given for the game-without-music condition, and the rating for the game-with-music condition is between the two. In contrast to the *additive* and *subtractive* relationships described previously, this kind of *averaged* relationship suggests that the response to the A-V combination is a compromise between the rating given to the audio and visual components in isolation.

For the “busy” rating scale, there is a highly significant difference in the ratings given by both males and females to the game-without-music condition. However, the manner in which the difference emerges is exactly opposite between the groups. As shown in Figure 3, the music-only and game-with-music conditions are rated very similarly by both males and females, though the mean scores given by males are significantly higher than those given by females. Responses to the game-without-music condition reveal a highly significant difference between the two groups. For males, this condition is rated much lower than either of the other two conditions. Ratings provided by females, however, are significantly *higher* for the game-without-music condition than for the other two conditions. It appears, therefore, that verbal ratings for “busy” in the context of the gaming environment change significantly with the addition of music, though the specific manner in which they change is dramatically different when considered by gender.

[insert Figure 2 about here]

[insert Figure 3 about here]

Group x Scene interaction. Two VAME scales revealed a statistically significant interaction between group assignment and video game segment: annoying ($F_{(3,693,118,165)} = 3.383$; $p = .014$) and strange ($F_{(3,893,124,577)} = 2.571$; $p = .043$).³ For the “annoying” rating (Figure 4), all scenes are rated very similarly for both the music-only and game-with-music conditions. However, the game-without-music condition reveals that both *Weather*

Top and *Moria* are significantly less annoying when the music is removed, while the *Amon Hen* scene is judged to be more annoying when the music is absent. The former relationship could be due to the fact that the music transfers annoying quality – when it exists – onto the visual image when the two are combined, while the latter suggest that music with a less annoying quality than the visual image can, in fact, reduce the level of aggravation perceived when the music and visual are combined. Both of these examples provide instances of *musical transference*, as described previously).

For the “strange” rating (Figure 5), the music-only condition for each scene is rated significantly less strange than the game-without-music condition. As a result, in all scenes except *Weather Top*, the music serves to suppress the strangeness of the visual scene and activity. The precise relationships illustrated in this figure are worthy of elaboration. In the *Weather Top* segment, the music-only condition is rated quite low, while the game-without-music condition receives a higher rating (for this scene only). As a result, the slightly higher rating for the game-with-music condition exemplifies an *additive* A-V relationship. In contrast, the ratings for the *Amon Hen* segment illustrate a *subtractive* relationship, since the game-without-music condition receives a moderately high mean score, the music-only condition is rated lower, and the game-with-music condition receives the lowest rating of all. VAME ratings for the *Moria* scene exemplify an *averaging* relationship, since the lack of strangeness judged in the music-only condition is counteracted by the high level of strangeness perceived in the game-without-music condition, resulting in a moderate mean rating for the game-with-music condition.

[insert Figure 4 about here]

[insert Figure 5 about here]

Group x School interaction. Only a single VAME scale resulted in a statistically significant difference in the interaction between group assignment and age level (high school vs. college): labored ($F_{(2,64)} = 2.571$; $p = .042$). High school students rated the music-only condition significantly higher on this verbal scale than college students, while both groups provided a moderate mean rating for the game-without-music condition, though the college students provided a slightly lower mean score. The most dramatic difference between the two groups is that the high school students found the game-with-music much less labored than the college students. In fact, the *subtractive* relationship evident in the high school ratings results in the game-with-music condition receiving the lowest rating of all, while the *additive* relationship exhibited in the college ratings result in

³ Throughout the presentation of results, when the assumption of sphericity is violated as revealed by Mauchly's test, F -ratio and p -values will be stated using the Greenhouse-Geisser method of adjustment.

the game-with-music condition receiving the highest mean score, a dramatically different result.

Returning to the Research Questions posed at the outset, the results of the present investigation confirm that some, but not all, of the VAME scales exhibited a statistically significant difference due to the presence of a musical score. Details of the specific ways in which music affects the gaming experience were provided above. In addition, both gender and age level were shown to be significant factors for some, but not all, of the VAME scales.

[insert Figure 6 about here]

Discussion

Results of the present investigation suggest that the relationship between a specially-composed orchestral musical score and involvement with an RPG video game is highly complex and varied. Specific relationships that emerged as a result of examining subject VAME responses included *musical transference*, *additive*, *subtractive*, and *averaging*. Though not exhibited in any of the analyses provided herein, there is likely to be a fifth category of *visual transference*, in which the quality of the visual image and/or game situation is imposed upon the musical score when both components are combined into an A-V experience.

The statistical analyses reported herein reveal a set of highly complex relationships, as did Yamada et al (2001, 2002). This is certainly understandable given the added narrative complexity inherent in RPGs when compared to racing games and the presence of a specially composed orchestral music score, rather than arbitrarily selected compositions used on the basis of subject verbal ratings.

Future research will be needed to determine whether the relationships emerging from this investigation can be generalized beyond the present experimental context using the specific scenes selected from *LR:TT*. A series of studies is warranted in which the same method is applied to a variety of video games, including – but not limited to – RPGs, to determine to what extent these emergent relationships appear in other contexts.

Due to the widespread use of video games as a source of entertainment and as a passive form of social activity, it is imperative that future investigations provide a clear understanding of the contributions to the experience made by each of the auditory and visual components, as well as the level of interactivity exhibited. There is much work yet to be done.

Acknowledgements: We would like to acknowledge the invaluable support of Northwestern University, the School of Music, and the School of Communication for use of facilities, equipment, and computing resources necessary to

complete the research project reported herein. The authors would also like to acknowledge the assistance of Mr. Louie Igoe, who assisted with the early stages of this project, including data collection for a pilot study. This research was supported by a grant from the National Science Foundation to the Childrens Digital Media Center (Grant # 0126014).

References

- Boltz M (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception*, 18(4): 427-454.
- Bullerjahn C, G ldenring M (1994). An empirical investigation of effects of film music using qualitative content analysis. *Psychomusicology*, 13: 99-118.
- Cohen AJ (2001). Music as a source of emotion in film. In Juslin PN, Sloboda J eds. *Music and emotion: Theory and research*. New York: Oxford University Press, 249-272.
- Dolby Laboratories (2004, May 19). Worldwide statistics. Retrieved June 14, 2004, from <http://www.dolby.com/stats/>.
- Hill D (2003, July). Four producers reinvent video-game sound design. *Remix*, 38-44.
- Iwamiya S (1994). Interactions between auditory and visual processing when listening to music in an audio visual context: 1. Matching 2. Audio quality. *Psychomusicology* 13: 133-153.
- Iwamiya S (1997). A computer-controlled experiment on the interaction between music and motion picture: The effects of various musical factors on the impression of audio-visual products. *Journal of Music Perception and Cognition* 3: 25-32.
- Iwamiya S, Johgetsu Y, Sugano Y (2001). The effects of musical and image factors on the impression of audio-visual contents. *Proceedings of Mechatronics '01, the 5th Franco-Japanese Congress and 3rd European-Asian Congress* (Besan on, France, October 9-11, 2001).
- Kendall RA, Carterette EC (1993). Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck adjectives. *Music Perception*, 10(4): 445-467.
- Lipscomb SD (1995). Cognition of musical and visual accent structure alignment in film and animation. Unpublished doctoral dissertation, University of California, Los Angeles.
- Lipscomb SD (in press). The perception of audio-visual composites: Accent structure alignment of simple stimuli. *Selected Reports in Ethnomusicology* 12.
- Lipscomb SD, Kendall RA (1994). Perceptual judgment of the relationship between musical

- and visual components in film. *Psychomusicology*, 13(1): 60-98.
- Lipscomb SD, Tolchinsky DE (in press). The role of music communication in cinema. In Miell D, MacDonald R, Hargreaves D eds. *Musical communication*. Oxford: Oxford University Press.
- Marshall SK, Cohen AJ (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception*, 6(1): 95-112.
- Osgood CE, Suci GJ, Tannenbaum PH (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Rideout VJ (2004, March). [Summary of testimony]. Presented before the United States Senate Subcommittee on Competition, Infrastructure, and Foreign Commerce. Retrieved September 1, 2004 from <http://www.kff.org/entmedia/entmedia030204tst.cfm>.
- Thayer JF, Levenson RW (1983). Effects of music on psychophysiological responses to a stressful film. *Psychomusicology*, 3(1): 44-52.
- Thompson WF, Russo FA, Sinclair D (1994). Effects of underscoring on the perception of closure in filmed events. *Psychomusicology*, 13: 9-27.
- Yamada M (2002). In Stevens C, Burnham D, McPherson G, Schubert E, & Renwick J eds., *Proceedings of the 7th International Conference on Music Perception & Cognition*. Adelaide: Causal Productions, 340-343.
- Yamada M, Fujisawa N, & Komori S (2001). The effect of music on the performance and impression in a video racing game. *Journal of Music Perception and Cognition*, 7(2): 65-76.
- Zehnder SM, Lipscomb SD (in press). The role of music in video games. In Vorderer P and Bryan J eds. *Playing computer games: Motives, responses, and consequences*. Lawrence Erlbaum Associates, Inc.

Figures

Figure 1.

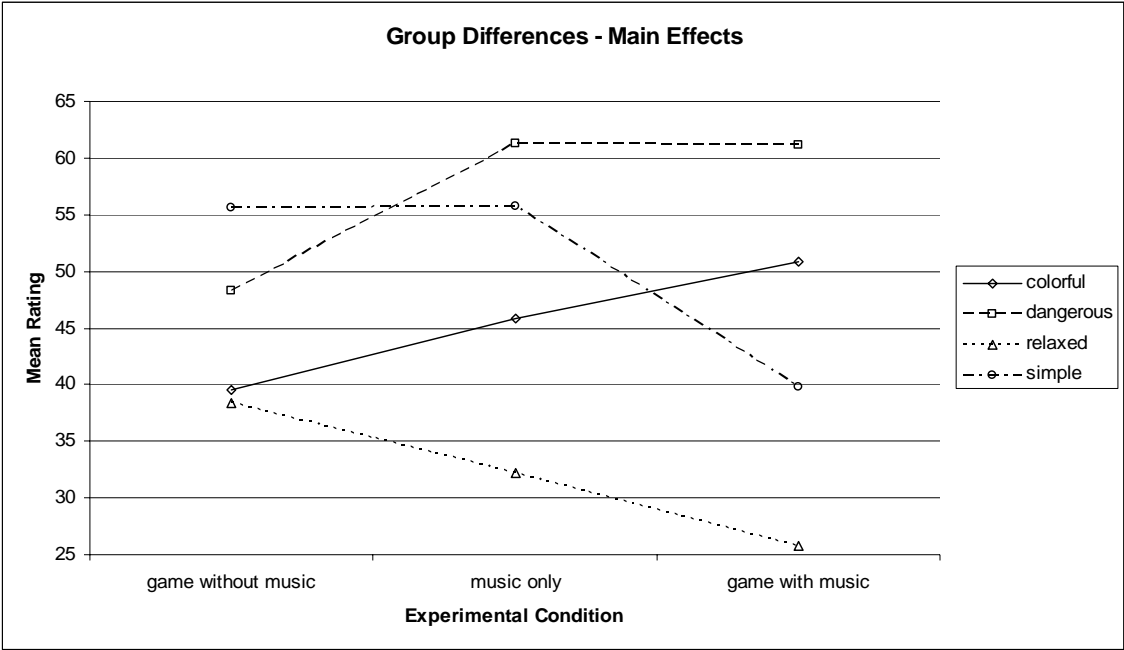


Figure 2.

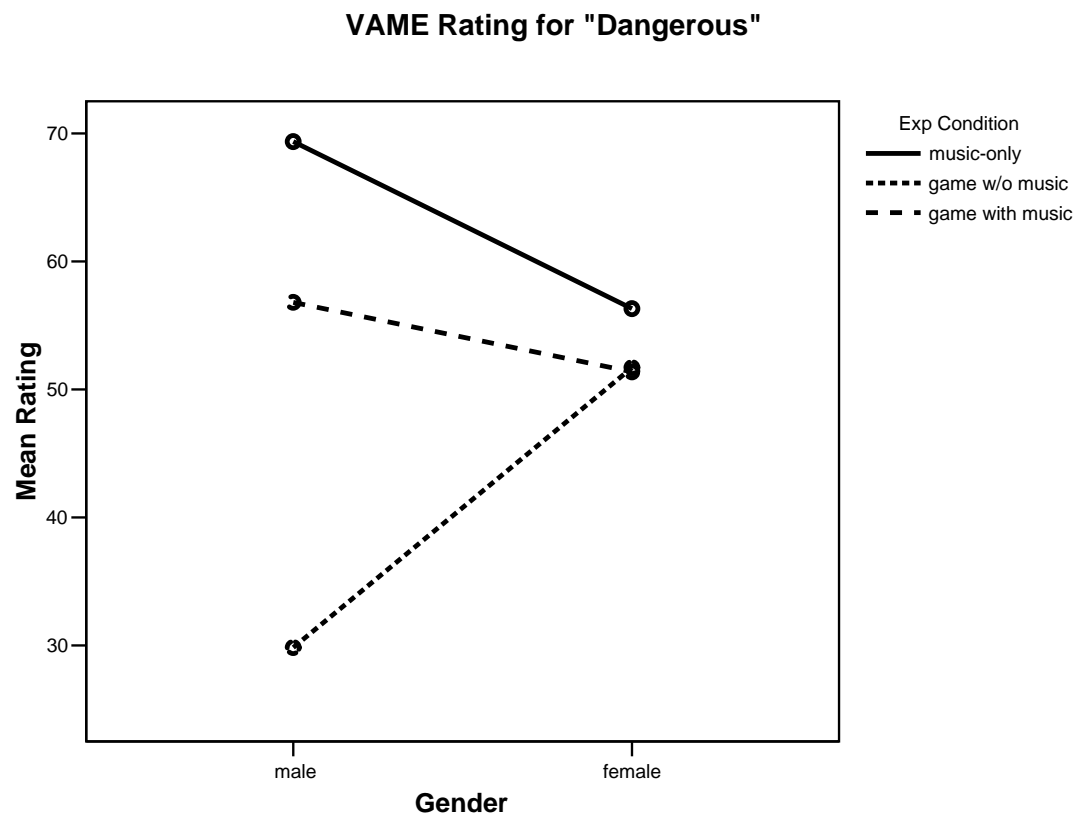


Figure 3.

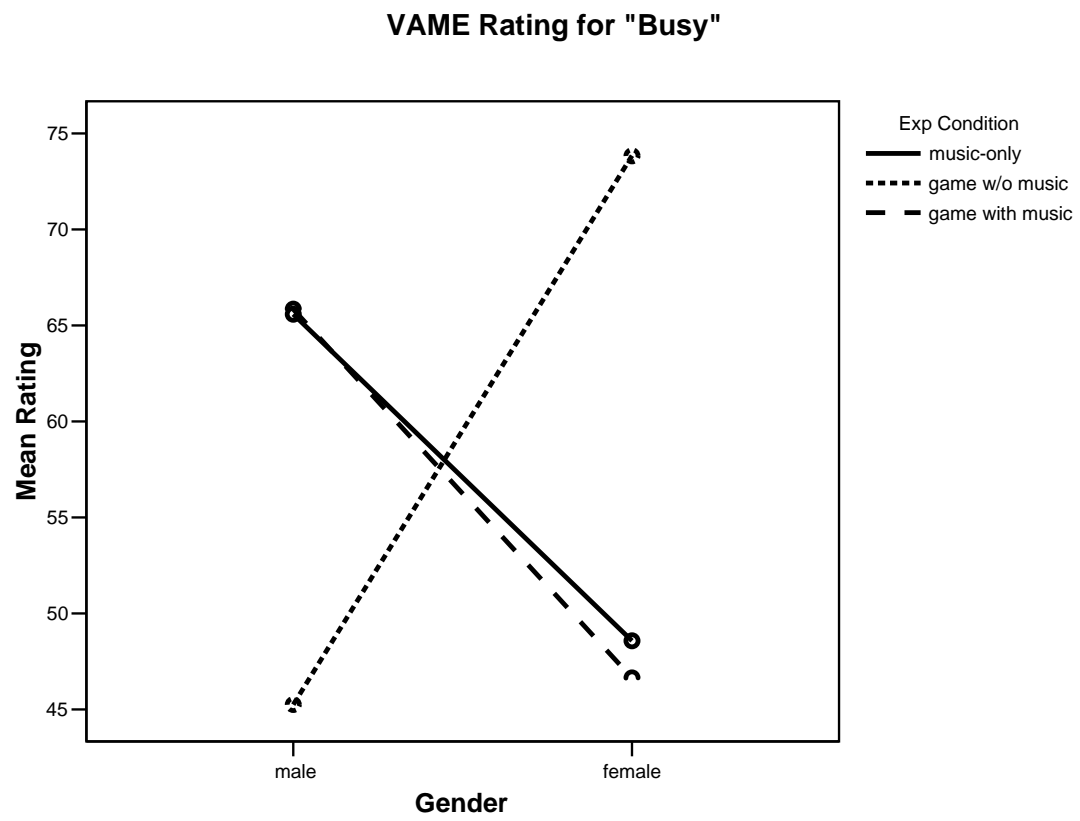


Figure 4.

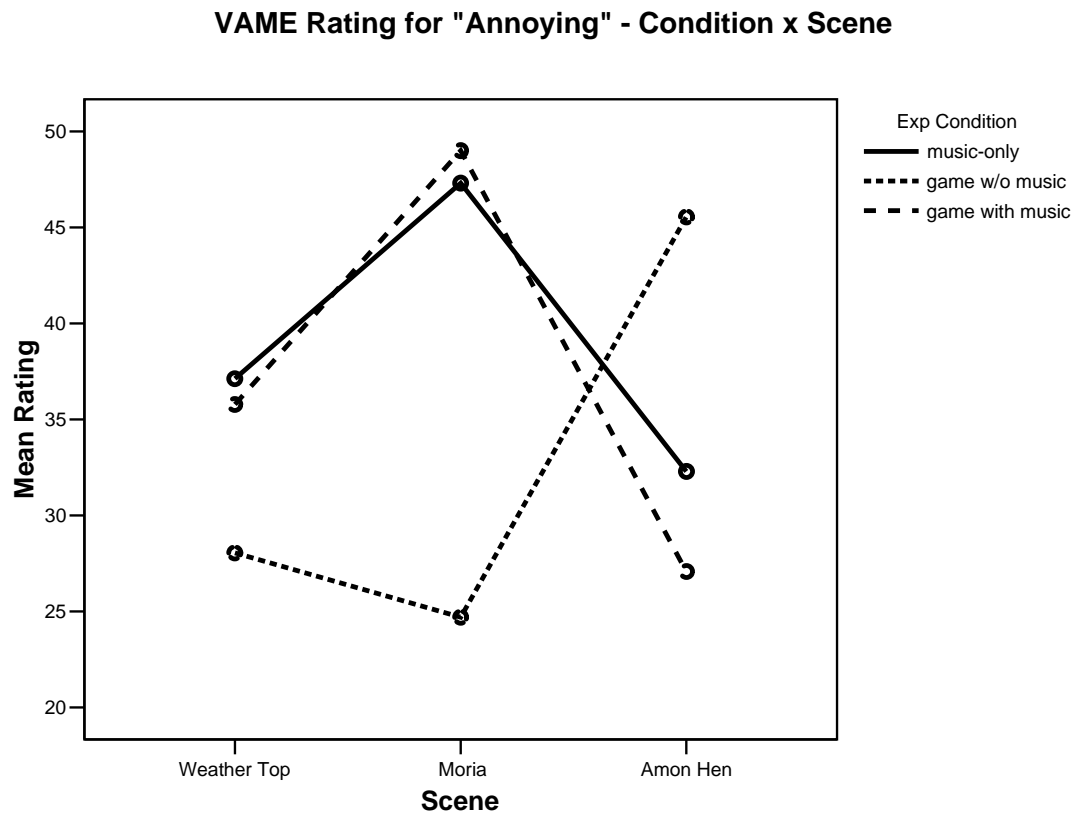


Figure 5.

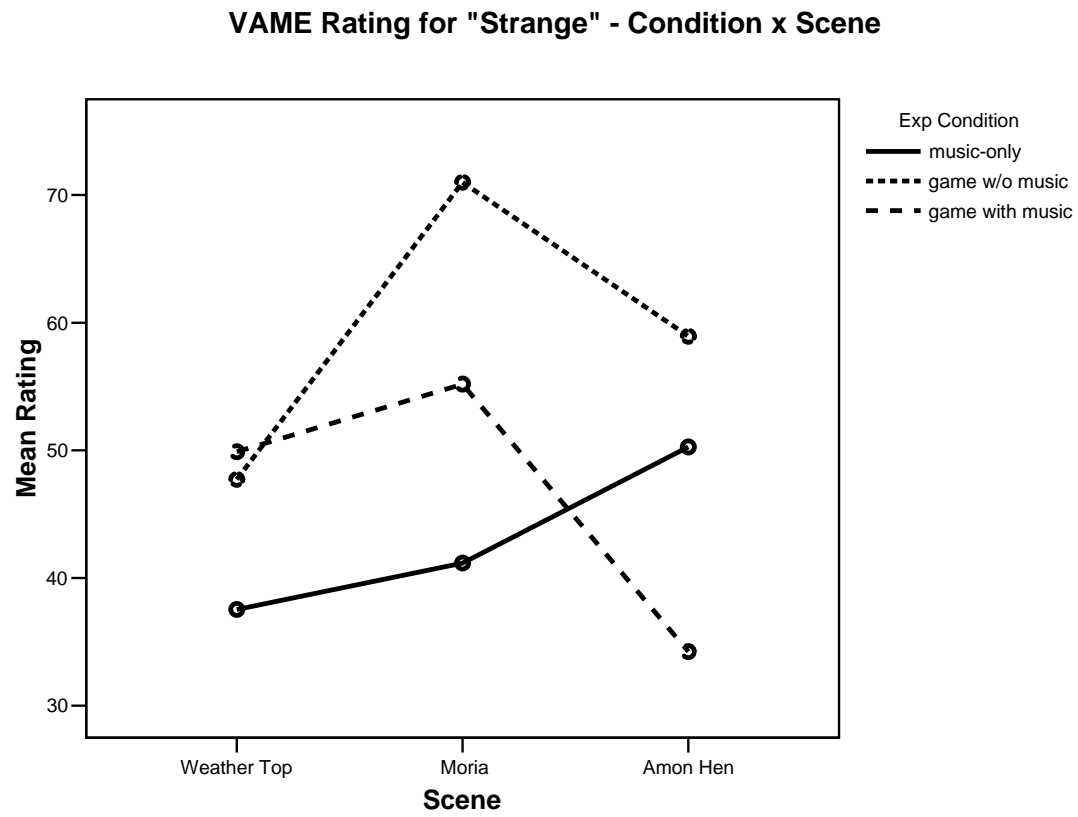
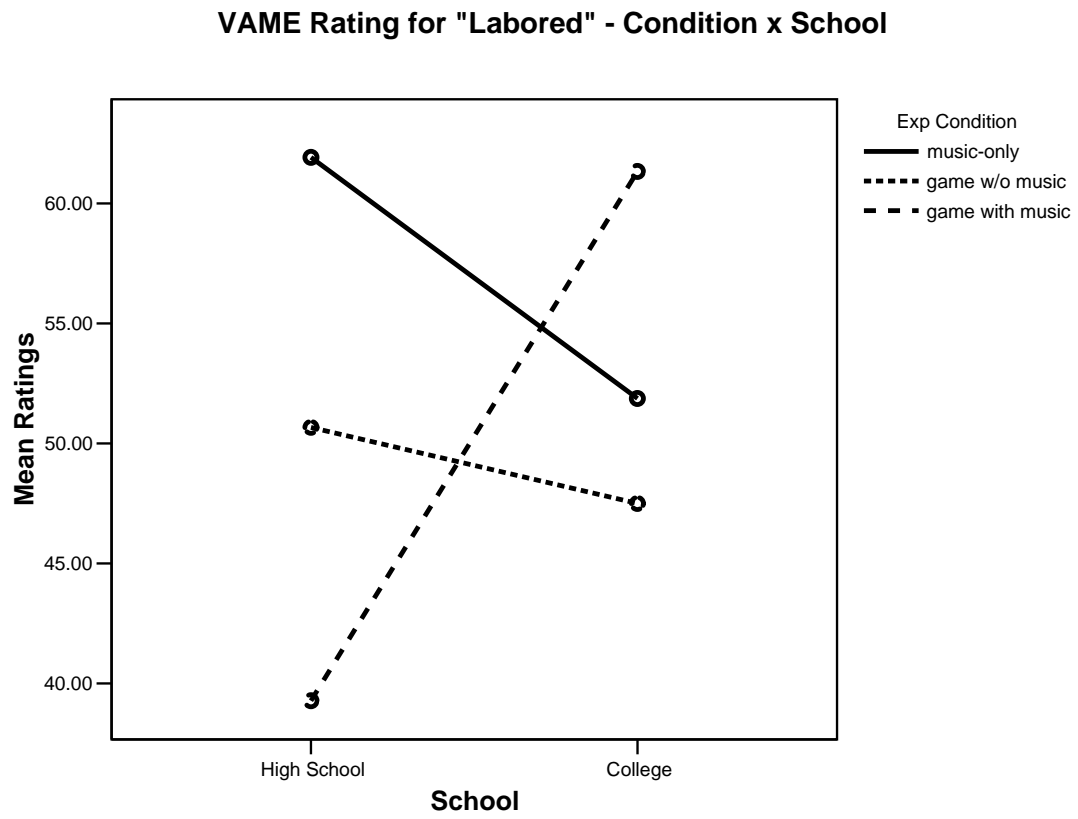


Figure 6



Legends to Figures

Figure 1. Graphic representation of the mean scores for all VAME scales that exhibited a significant main effect of group assignment.

Figure 2. VAME rating for “dangerous,” revealing an interaction between group assignment and gender.

Figure 3. VAME rating for “busy,” revealing an interaction between group assignment and gender.

Figure 4. VAME rating for “annoying,” revealing an interaction between group assignment and video game segment.

Figure 5. VAME rating for “strange,” revealing an interaction between group assignment and video game segment.

Figure 6. VAME rating for “labored,” revealing an interaction between group assignment and age level.