Audio Engineering Society

Convention Paper

Presented at the 123rd Convention
2007 October 5–8 New York, NY, USA

Multiple Regression Modeling of the Emotional Content of Film and Music

Rob Parke, Elaine Chew, and Chris Kyriakakis

University of Southern California Viterbi School of Engineering, Los Angeles, CA, 90089, USA
{parke, echew, ckyriak}@usc.edu

ABSTRACT

Our research seeks to model the effect of music on the perceived emotional content of film media. We used participants' ratings of the emotional content of film-alone, music-alone, and film-music pairings for a collection of emotionally neutral film clips and emotionally provocative music segments. Mapping the results onto a three-dimensional emotion space, we observed a strong relationship between the ratings of the film- and music-alone clips, and those of the film-music pairs. Previously, we modeled the ratings in each dimension independently. We now develop models, using stepwise regression, to describe the film-music ratings using quadratic terms, and based on all dimensions simultaneously. We demonstrate that while linear-terms are sufficient for single emotion dimension models, regression models that consider multiple emotion dimensions yield better results.

1. INTRODUCTION

Our research began with the question, "Does music effect the perceived emotion of film?" and our past efforts [8] support the claim that indeed it does. In this paper, we present analyses that validate our previously proposed model for describing emotion response to film with music using music-alone and film-alone ratings as predictors, and explore more complex models for analyzing the emotional impact of music on film. Our work offers a quantitative approach to modeling of the effect of music on emotion response to film.

To investigate the effect of music on perceived emotion of film, we have previously designed an experiment to measure the perceived emotion of music-alone, film-alone, and film-and-music combined. The film clips were chosen to be emotionally neutral, so as to allow for a wide range of interpretations; the music segments were selected to be emotionally provocative, so as to allow for widely polarized perceived emotions in a short duration of time. The results are mapped to a three dimensional emotion space, with axes representing *stress*, *activity*, and *dominance*.

In an earlier study [8], we generated center of gravity plots of the results in two-dimensional subspaces of the original three-dimensional emotion space, and observed that the music-film rating generally falls on a trajectory between the music-alone and the film-alone rating. We then conducted regression analyses of the data within each of the three emotion dimensions to verify that the music-film data can be predicted as a convex combination of the music-alone and film-alone data.

This paper expands upon our earlier analyses. First, we explore whether the added complexity of quadratic terms in the single emotion dimension model improves the regression model results. We use stepwise multiple regression to determine the statistically significant factors in the model. We find that, when treating each emotion dimension independently, our original model with only linear terms fits the data more effectively than one with quadratic and cross terms. This result suggests that linear terms are sufficient when considering single emotion dimension models.

Next, we expand upon the original single emotion dimension linear-order approach by developing multiple emotion dimensions linear-order regression models to describe the resulting film-music ratings in each dimension using film-alone and music-alone ratings from all three dimensions. The improved fit when considering multiple emotion dimensions models, and the statistically significant non-zero coefficients for predictors from other dimensions, suggest that emotion response to film with music in each of the three dimensions is described better as the interaction of perceived emotion ratings of music-alone and film-alone in all three dimensions.

The paper is organized as follows: Section 2 presents some related work on the link between visual and musical stimuli various domains such as film and performance; Section 3 outlines our methods, including the experiment design, and regression models; Section 4 presents the regression analyses results, and their interpretations; and, Section 5 concludes the paper.

2. RELATED WORK

Though the effect of music on the perceived emotion of film has not been studied directly, or in great depth, there has been much research on music and emotion (with the inclusion of video.) In this section, we highlight some related studies. We first discuss research in the area of music and emotion. Next, we provide a brief overview of our initial work, the foundation for the current paper.

Marshall and Cohen's [1] work established that there is indeed a cross-modal interaction between visual and auditory stimuli that affects emotion response. In the study, they paired short animations of geometric shapes, with what they termed "strong" (major key) and "weak" (minor key) classical music excerpts. Participants rated the clips, with and without music, along twelve bipolar adjective pairs. The ratings were then condensed into three separate dimensions: *evaluation*, *potency*, and *dominance*. Their results suggest that the visual and auditory stimuli interact to affect perceived emotion, but that exact relationship is unknown. Given the abstract nature of the visual scenes, these results were not completely generalizable [2].

Lipscomb and Kendall [3] studied perceived congruity between music and film. They asked participants to identify the correct music – that is to say, one intended by the composer or the director for the scene – for a number of film scenes (without audio.) The participants were able to match reliably each scene with its intended music, except in the scenes without humans. Their results suggest that the participants connect the perceived mood of the music with the expressions and mannerisms of the actors on screen.

Bullerjahn and Güldenring [4] presented participants with intentionally ambiguous short films coupled with various genre-specific soundtracks, and asked for participants' interpretations about the outcome of the film. They found that participants were more likely to provide predictions linked to the genre of music, for instance, "melodramatic" music elicited responses suggesting a romantic conclusion, whereas "thriller" music yielded violent endings. This shows that music can affect not only the current mood of a scene, but also global perception of the film and its outcome. More recently, Boltz [5] performed a similar study combining film clips with "positive" and "negative" music, and showed that interpretation and recollection of the scenes was skewed by the music, in directions congruent with the music's mood.

Vines et. al. [6] further explored the effect of the cross-modal interaction of visual and auditory sensations on the emotion ratings of perceived expressiveness in a cello performance. Their factor analysis of participants' responses (via Likert-type scales of emotion) to video-only, audio-only, and video-and-audio combined stimuli suggests that visual stimuli had the stronger impact on the perceived emotional expressiveness. They showed, in further studies using continuous-time ratings of tension (used as a substitute for emotion,) that visual input can both raise and dampen the participants' tension ratings [7].

Our studies seek to model quantitatively and precisely the effect of music on the perceived emotion of a film segment. In a previous paper [8], we investigated quantitative and visual methods for analyzing the effect of the music on the perceived emotion of film. We combined emotionally polarized music with ambiguous film scenes, and we asked participants to rate the perceived emotional content across three orthogonal dimensions: *stress*, *activity*, and *dominance*. We generated weighted average (center of mass) plots in each two-dimensional emotion subspace. Visual inspection of these plots yielded the observation that the music-film average rating is typically on a trajectory between the film-alone and the music-alone ratings. We used linear regression to model this relationship for the data in each dimension independently. The results indicate that a linear equation of film-alone and music-alone data models the participants' responses well when treating each dimension in isolation. Our present study considers more complex models for the relationship between perceived emotion of music-alone, film-alone, and music-film combined, and weights their efficacy against the simpler single emotion dimension linear model.

3. METHODS

This section describes our experiment design, and analysis methods. We first present some background information and motivation for, and a review of, the experiment design. We start with a brief discussion on the emotion space model used in our study; next, we describe the selection of music and film media, and the web-based rating questionnaire presented to the participants. Finally, we describe the regression models employed in the present analysis of the experiment data.

3.1. Emotion Models

There have been many proposed models for representing and measuring perceived emotional content. Osgood, Suci, and Tannebaum [9] are credited with using semantic differential scaling by having participants rate stimuli on bipolar pairs of adjectives, for example, good/bad, weak/strong, using a Likert-type scale (for example, on a scale of 1 to 7, where weak = 1 and strong = 7.) These results are then grouped into three factors, *evaluation*, *activity*, and *potency*, represented by orthogonal axes in a three dimensional space. This method and model was used in Marshall and Cohen's [1] early research, and similar variants have remained popular in the literature. Across various disciplines, these dimensions are given different labels; for our study, we consider *stress*, *activity*, and *dominance* to be analogs of *evaluation*, *activity*, and *potency*.

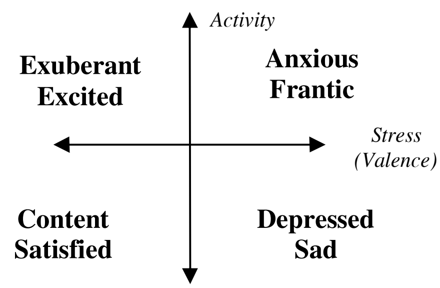

Figure 1: Thayer's Emotion Space (reproduced from [8])

Popular related models, such as Thayer's two-dimensional arousal model [11] and Huron's adapted model [12], are similarly organized and present similar (sub)spaces. *Stress* (or *valence*) and *energy*, the axes of Thayer's and Huron's two-dimensional subspace, can be interpreted as a projection of Osgood's three-dimensional space. As shown in Figure 1, the four resulting quadrants of the space are labeled *content/satisfied*, *depressed/sad*, *anxious/frantic*, and *exuberant/excited*.

3.2. Media

We assembled a collection of short film scenes and two separate music study groups: one with classical music and one with music composed specifically for this project.

3.2.1. Film Scenes

We selected five film excerpts from an initial pool of 24 scenes, each approximately 20 seconds in length, choosing lesser-known films to limit participants' prior exposure and possible emotional associations with the scenes. The final selection was based on the highest level of ambiguity in terms of content and meaning, as determined independently by the authors. Our intent was to allow for flexible interpretations by choosing neutral scenes and by eliminating, for example, particularly vivid and remarkable scenes. We extracted the scenes from DVDs, removed the original audio, and encoded the video as high quality MPEG-format files. One scene was excerpted from each of the following films: *Amélie*, *Little Man Tate*, *Maria Full of Grace*, *Memento*, and *Three Kings.* Descriptions of the scene contents are provided in Table 1.

Film	Description
Amélie	A woman finds a box hidden in the wall and opens it.
Memento	A truck pulls up to an old dilapidated shack; a man gets out, and walks to the shack (black and white).
Maria	A teenage girl watches a sonogram monitor.
Tate	A young boy embraces a woman and crosses the street to meet another woman.
3 Kings	A group of people and soldiers walk across the desert.

Table 1: Film Scene Descriptions (adapted from [8])

3.2.2. Study 1: Classical Music

For the classical music study, we chose pieces that feature solo piano and/or string instruments. Our criterion for selection was to find brief, emotionally potent clips of music, a decision that was made to allow the music to have non-trivial impact on the perceived emotion, given the short duration of the film scenes. For diversity, we ensured that the four segments of music to be paired with a film scene were selected from each of Thayer's emotion space quadrants. We selected a total of twenty music segments, four for each of the five film clips. We extracted the audio examples from CDs as WAV files, and edited them to the length of the video clips. The music pieces, their determined mood, and the film with which each was paired are listed in Table 2. Each set of four, distinct-mood music clips (along with one audio track of silence) was combined with their respective films, resulting in twenty-five clips in total for the classical music study.

3.2.3. Study 2: Composed Music

Our second study employs music written by composer James Post specifically for this project. Post composed four film-score style soundtracks for each candidate scene, with only one restriction: that all the music clips for a given scene should be distinct one from another in their emotional content.

Composer	Work	Mood	Film
Mahler	*Allegro Assai Und Sehr Trotzig*	Anx.	Amélie
Debussy	Clair de Lune	Cont.	Amélie
Verdi	La Traviata: Act I: *Prelude*	Depr.	Amélie
Grieg	Holberg Suite, Op.40: *Prelude*	Exub.	Amélie
Radiohead	(nice dream)	Anx.	Memento
Chopin	Prelude in Db, Op.28, No.15, "Raindrop"	Cont.	Memento
Chopin	Prelude in e, Op.28, No.4	Depr.	Memento
Beethoven	Rage Over a Lost Penny	Exub.	Memento
Radiohead	Airbag	Anx.	Maria
Schumann	Traumerei, Op.15, No.7	Cont.	Maria
Chopin	Piano Sonata No.2 in bb, Op.35: *Funeral March*	Depr.	Maria
Kreisler	Leibesfreud	Exub.	Maria
Rachmaninov	Piano Concerto No.2: *Allegro scherzando*	Anx.	Tate
Massenet	Thäis: *Meditation*	Cont.	Tate
Liszt	Grandes Etudes de Paganini: No.3 in g#: La Campanella	Depr.	Tate
Mozart	Piano Concerto No.1 in D: *Allegro*	Exub.	Tate
Chopin	Etude in c, Op.10, No.12, "Revolution"	Anx.	3 Kings
Brahms	Intermezzo in Eb, Op.117, No.1	Cont.	3 Kings
Rachmaninov	Prelude in c#	Depr.	3 Kings
Mozart	Violin Concerto No.4 D: *Allegro*	Exub.	3 Kings

Table 2: Classical Music Selections (adapted from [8])

Our intent was to create a similar set of non-overlapping moods as in study 1, while allowing the composer creative license to mimic real-world film scoring. All but one of the twenty clips was instrumental, and the composer recorded the clips and provided them as WAV audio files. Table 3 lists the composer-provided descriptions and intent of each music clip. Each set of four music clips, along with one audio track of silence, was combined with their respective films, resulting in twenty-five clips in total for the composed music study.

Description	Film
Mysterious, foreboding, darkness	Amélie
Curious, playful, sneaky, "I wonder what is in the box"	Amélie
Sad, mourning, longing, "this is my last hope"	Amélie
Anxious, happy, anticipating	Amélie
Sneaky, investigative, film noir	Memento
Thought-provoking, passing time, building	Memento
Chaos, angry, insanity	Memento
Moving on, mellow, "everything will be alright in time"	Memento
Sweetness, caring, motherly	Maria
Evil, omen, "now I have them!"	Maria
Sadness, "a let down"	Maria
Craziness, out-of-control, nonsense	Maria
Sadness, loss, confusion	Tate
Moving on for the better, inevitable, reserved emotions	Tate
Cheesy reunion, warm feelings, 80s nostalgia	Tate
Anxiety, predator lurking near, scary	Tate
Courage, with hope, "saving the people"	3 Kings
Dreamy, ethereal, a nice walk	3 Kings
Sneaky, risky journey, "danger may lie ahead"	3 Kings
Scary, agitated	3 Kings

Table 3: Composed Music Descriptions (adapted from [8])

3.3. Web Interface

We adapted a web-based interface, shown in Figure 2, to present the clips in random order in order to minimize error, and to allow participants to rate the perceived emotion in the dimensions of stress, activity, and dominance. To facilitate this rating process, each dimension was accompanied by a five-point Likert-type scale and self-assessment mannequins (SAMs), which are simple images that provide meaning and visual feedback about each rating. This interface and its illustrations (SAMs) have been used in previous research, and shown to be an effective method for capturing perceived emotional content [13].

Forty-seven participants – 19 females, 28 males, mean 32.6 years – completed both studies, classical and composed, by rating the perceived emotion of each music and film combination, as well as the film-alone baselines. Participants were asked to rate the perceived emotion of the clip, and not the mood it induced in them. There was no time limit; clips could be replay as many times as necessary; and, each study had to be completed in a single sitting.

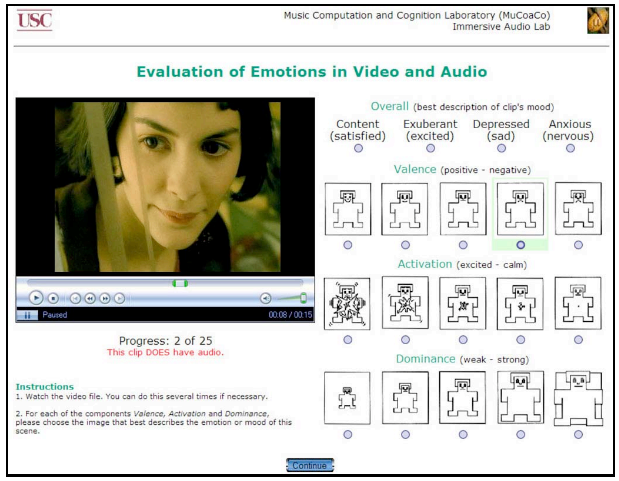

Figure 2: Web Interface

The order of presentation of clips was also an important aspect of the interface. All five clips for a film – four with music and one silent – were shown as a group. We mitigated potential biases by randomizing not only the intra-scene playback, that is to say, the order of each different soundtrack for the same film, but also the inter-scene selection, the order of each film scene group. We arranged for the silent version each film scene to play before the four versions with music, in order to provide a control-group baseline for the film-alone response. A separate study with seven participants – 4 females and 3 males, mean age of 28.9 years – was conducted for the music-alone control-group.

3.4. Linear Regression Models

This section describes linear regression models employed in the analysis of the data collected using the web-based interface described in the previous section. Unlike our previous analyses, which allowed only linear terms, the present analyses utilize both linear and quadratic terms in the model to capture possible higher degree relationships amongst the different factors.

The analysis in this paper can be divided into two parts. Section 3.4.1 presents the model for the first analysis, which considers each dimension of the emotion space independently. This analysis models the music-film rating as a function of the music-alone and film-alone ratings in one emotion dimension, allowing for quadratic terms. Section 3.4.2 describes the second analysis, which considers the mixture of different dimensions of the emotion space in the modeling of the music-film rating.

As in our previous paper [8], we consider the aggregate ratings from all participant responses. We compute a weighted average (center of mass) representing a triplet in the there-dimensional emotion space with axes representing stress, activity, and dominance, for each film-alone, music-alone, and film-and-music combined clip. Please refer to [8] for a complete list of data and figures.

3.4.1. Single Emotion Dimension Analysis

In this section, we consider each of the emotion dimensions independently. In our previous experiments, we observed that the music-film combined rating occurred on an approximately linear trajectory between the music-alone and the film-alone ratings. Thus, we originally proposed a linear model [8] between the film-alone and the music-alone data (the predictors) and the film and music combined data (the response.) We now evaluate the validity of the simple, linear model by comparing it with a more complex one containing second order and cross-terms of the film-alone and the music-alone data. Our beginning regression equation is as follows:

$$FM_{i,j} = \alpha_{i,j} \cdot F^2{}_{i,j} + \beta_{i,j} \cdot F_{i,j} + \gamma_{i,j} \cdot M^2{}_{i,j} + \zeta_{i,j} \cdot M_{i,j} + \eta_{i,j} \cdot F_{i,j} \cdot M_{i,j} + \varepsilon \quad , \quad (1)$$

where $i \in \{1,2\}$ is the experiment set (1 indicates the classical study, and 2 indicates the composed study) and $j \in$ {stress, activity, dominance} is the emotion dimension. MF is the combined film-and-music center of mass, F is the film-alone center of mass, M is the music-alone center of mass, the coefficients $\{\alpha, \beta, \gamma, \zeta, \eta\}$ represent the weights of each predictor, and ε represents noise.

Using stepwise linear regression techniques (as implemented in MATLAB), we alternately add the most statistically significant predictors and remove the least statistically significant predictors from the model, which we evaluate against our previous linear-order term model,

$$FM_{i,j} = \alpha_{i,j} \cdot F_{i,j} + \zeta_{i,j} \cdot M_{i,j} + \varepsilon \quad , \quad (2)$$

where $i \in \{1,2\}$ is the experiment set, and $j \in$ {stress, activity, dominance} is the emotion dimension. MF is the combined film-and-music center of mass, F is the film-alone center of mass, M is the music-alone center of mass, the coefficients $\{\alpha, \beta\}$ represent the weights of the film-alone and music-alone predictors, respectively, and ε represents noise. If the backward and forward stepwise regression procedures result in different model predictors, we select the optimal model by maximizing R^2 and minimizing root mean square error (RMSE).

3.4.2. Multiple Emotion Dimensions Analysis

In this section, we expand our single emotion dimension model to include all three of the emotion dimensions simultaneously. Only linear terms are used. The expanded model now allows, for example, film-alone ratings in the stress dimension to be used in the model for film-music response in the activity dimension. The resulting regression equation is:

$$FM_{i,j} = \alpha_{i,str} \cdot F_{i,str} + \beta_{i,str} \cdot M_{i,str} + \gamma_{i,act} \cdot F_{i,act} + \zeta_{i,act} \cdot M_{i,act} + \eta_{i,dom} \cdot F_{i,dom} + \iota_{i,dom} \cdot M_{i,dom} + \varepsilon \quad , \quad (3)$$

where $i \in \{1,2\}$ is the experiment set, and $j \in$ {stress, activity, dominance} is the emotion dimension. MF is the combined film-and-music-combined center of mass, F is the film-alone center of mass, M is the music-alone center of mass, the coefficients $\{\alpha, \beta, \gamma, \zeta, \eta, \iota\}$ represent the weights of each predictor, and ε represents noise.

As with the single-dimensional model, we use stepwise regression techniques to iteratively add the most statistically significant predictors, and remove the least statistically significant predictors from the model. Again, if the backward and forward stepwise procedures result in different model predictors, we select the optimal model by maximizing R^2 and minimizing RMSE

4. RESULTS

This section discusses the results of the single emotion dimension and the multiple emotion dimensions regression analysis with quadratic terms. We present the results of the two analyses for the classical study and the composed study separately. Section 4.1 contains the results for the classical music study, with the single emotion dimension analysis in Section 4.1.1, and the multiple emotion dimensions analysis in Section 4.1.2. Similarly, Section 4.2 contains the results for the composed music study, with the single emotion dimension analysis in Section 4.2.1, and the multiple emotion dimensions analysis in Section 4.2.2.

4.1. Classical Music

4.1.1. Single Emotion Dimension Analysis

This section presents the stepwise procedure results for the regression analysis of the classical study data, starting with Equation 1. In separate procedures, using both forward and backward stepwise regression, our analysis for each of the three dimensions show only the linear-order terms of film-alone and music-alone to be statistically significant. Table 4a shows the summary statistics for the single emotion dimension model, and Table 4b lists the predictor coefficients (PCs) for the optimized models.

i	j	R^2	RMSE	F-Val	p
1	str	0.8056	0.3589	35.2170	9.0060e-07
	act	0.9045	0.2597	80.4798	2.1433e-09
	dom	0.8790	0.1845	61.7734	1.5934e-08

Table 4a: Forward and Backward Stepwise Regression Results Summary – Single Emotion Dimension Model, Classical Study

i	j	PC 1			PC 2		
		var	val	t-stat	var	val	t-stat
1	str	F_{str}	0.3443	2.6955	M_{str}	0.6038	8.0141
	act	F_{act}	0.7908	2.6897	M_{act}	0.5819	12.3218
	dom	F_{dom}	0.7116	3.0219	M_{dom}	0.5565	10.6620

Table 4b: Predictor Coefficients for Optimized Single Emotion Dimension Model, Classical Study

All three models yield high R^2 values with small RMSE, especially the model for the dominance dimension. We also observe more variability in the stress dimension than in the activity and dominance dimensions.

Hence, results for our single emotion dimension model suggest that our originally proposed linear relationship (Equation 2) fits the data more effectively than a model that includes quadratic and cross terms (Equation 1).

4.1.2. Multiple Emotion Dimensions Analysis

This section presents the stepwise procedure results for the regression analysis of the classical study data, starting with Equation 3. For this multiple emotion dimensions model, we include the film-alone and music-alone ratings for all three of the emotion dimensions as predictors for the film-music rating in each separate emotion dimension. The backward and forward multiple emotion dimensions models for modeling the music-film rating in each dimension yield higher R^2 values and lower error statistics than the single emotion dimension models. Table 5a summarizes both all the models (optimal models are italicized); the table indicates the variables in each model using an 'x,' and Table 5b lists the predictor coefficients for the optimized models. We show the coefficients for only the optimal model, derived from both the forward and backward stepwise procedures, where optimal is defined by having maximum R^2 and minimum RMSE.

	F_s	M_s	F_a	M_a	F_d	M_d	R^2	RMSE	F-Val	p
str F	×	×				×	0.8764	0.2950	37.8055	1.7191e-07
str B		×	×		×	×	0.8904	0.2870	30.4725	4.8212e-07
act F			×			×	0.9174	0.2415	94.4240	6.2183e-10
act B				×	×	×	0.9260	0.2356	66.7292	2.9046e-09
dom F			×	×			0.8790	0.1845	61.7734	1.5934e-08
dom B			×	×		×	0.8997	0.1732	47.8603	3.2538e-08

Table 5a: Forward and Backward Stepwise Regression Results Summary – Multiple Emotion Dimensions Model, Classical Study

i	j	PC 1			PC 2		
		var	val	t-stat	var	val	t-stat
1	str B	M_{str}	0.5578	8.8767	F_{act}	-2.1725	-3.8773
	act B	M_{act}	0.4012	2.1568	F_{dom}	0.7502	2.8052
	dom B	F_{act}	0.8209	3.5842	M_{act}	0.3166	2.2477

i	j	PC 3			4th PC		
		var	val	t-stat	var	val	t-stat
1	str B	F_{dom}	1.8315	3.6949	M_{dom}	-0.1722	-3.1320
	act B	M_{dom}	0.3414	2.8590			
	dom B	M_{dom}	0.1651	1.8172			

Table 5b: Predictor Coefficients for Optimized Multiple Emotion Dimensions Model, Classical Study

The optimal music-film-combined stress model includes the music-alone stress predictor, film-alone activity predictor, and music-alone stress and dominance predictors. The optimal music-film-combined activity model includes film-alone dominance predictors, and music-alone activity and dominance predictors. The optimal music-film-combined dominance includes the film-alone activity predictor, and the music-alone activity and dominance predictors. These results suggest that perceived dominance in music alone is an important factor in the modeling of perceived stress, activity, and dominance in music-film pairs.

4.2. Composed Music

4.2.1. Single Emotion Dimension Analysis

This section presents the stepwise procedure results for the regression analysis of the composed study data, starting with Equation 1. The single emotion dimension data from our second study of composed music displays more variability than that of the classical study. The optimal models for the music-film rating in each emotion dimension were different from the original linear-order term model, and the forward and backward stepwise regressions for a given dimension result in identical, or almost identical, models in all three cases. Table 6a gives a summary of the forward and backward stepwise regression results (optimal models are italicized), and Table 6b lists the predictor coefficients for the optimized models.

	F^2	F	M^2	M	FM	R^2	RMSE	F-Val	p
str F			×	×		0.8819	0.2646	63.4965	1.2969e-08
str B	×		×	×		0.9019	0.2485	49.0547	2.7271e-08
act F				×		0.6539	0.3300	34.0042	1.5950e-05
act B				×		0.6539	0.3300	34.0042	1.5950e-05
dom F		×				0.2067	0.3093	4.6893	4.4027e-02
dom B	×	×			×	0.4785	0.2660	4.8928	1.3369e-02

Table 6a: Forward and Backward Stepwise Regression Results Summary – Single Emotion Dimension Model, Composed Study

The optimal stress model includes the linear and quadratic music-alone and film-alone terms. For activity, the optimal models determined by both backward and forward stepwise regression are the same, and include only the quadratic film-alone predictor. The optimized model for dominance contains the linear film-alone predictor, the quadratic film-alone predictor, and the film-alone, music-alone cross-term.

For stress and activity, the results suggest that the music-alone results have a stronger influence on the overall perceived emotion. The dominance dimension, however, is more affected by the film content and possibly the interaction of the film and music.

i	j	PC 1			PC 2		
		var	val	t-stat	var	val	t-stat
2	str B	$\mathbf{F}^2_{str}$	-0.4300	-1.8065	$\mathbf{M}^2_{str}$	-0.3500	-3.6604
	act F&B	$\mathbf{M}_{act}$	0.6408	5.8313			
	dom B	$\mathbf{F}^2_{dom}$	9.7690	2.3695	$\mathbf{F}_{dom}$	10.4001	2.5486

i	j	PC 3		
		var	val	t-stat
2	str B	$\mathbf{M}_{str}$	0.9689	11.5842
	act F&B			
	dom B	$\mathbf{F}_{dom} \cdot \mathbf{M}_{dom}$	-0.7376	-2.1677

Table 6b: Predictor Coefficients for Optimized Single Emotion Dimension Model, Composed Study

4.2.2. Multiple Emotion Dimensions Analysis

This section presents the stepwise procedure results for the regression analysis of the composed study data, starting with Equation 3. As with the single emotion dimension analysis for the composed study, the forward and backward stepwise regression results for the composed music result in similar models. Comparing the R^2 and RMSE values, each multiple emotion dimensions model for the composed study is slightly less statistically significant, and has higher error values, than its classical study counterpart. The forward and backward stepwise regression results are shown in Table 7a (optimal models are italicized), and Table 7b lists the predictor coefficients for the optimized models. Again, only the optimal models are shown.

The optimal stress includes the music-alone stress predictor and the film-alone activity predictor. The optimal activity model only includes the music-alone dominance predictor. The forward and backward dominance models were equal, each including the film-alone dominance predictor, and the music-alone dominance predictor.

	F_s	M_s	F_a	M_a	F_d	M_d	R^2	RMSE	F-Val	p
str F		×					0.8190	0.3184	81.4404	4.2300e-08
str B		×	×				0.8494	0.2988	47.9407	1.0269e-07
act F						×	0.6539	0.3300	34.0042	1.5950e-05
act B		×		×		×	0.7507	0.2970	16.0595	4.3949e-05
dom F					×	×	0.6095	0.2233	13.2665	3.3795e-04
dom B					×	×	0.6095	0.2233	13.2665	3.3795e-04

Table 7a: Forward and Backward Stepwise Regression Results Summary – Multiple Emotion Dimensions Model, Composed Study

i	j	PC 1			PC 2		
		var	val	t-stat	var	val	t-stat
2	str B	M_{str}	0.7974	9.6229	F_{act}	-0.7815	-1.8528
	act B	M_{str}	0.1648	1.9527	M_{act}	-0.3103	-1.9144
	dom F&B	F_{dom}	0.8647	2.6503	M_{dom}	0.2256	2.6887

i	j	PC 3		
		var	val	t-stat
2	str B act B dom F&B	M_{dom}	0.7228	6.5116

Table 7b: Predictor Coefficients for Optimized Multiple Emotion Dimensions Model, Composed Study

5. CONCLUSIONS

As we previously noted, when comparing weighted average (center of mass) plots of two-dimensional emotion subspaces, we observed that the music-film combined clip average rating appears on to be a trajectory between the film-alone and the music-alone rating. Our present research sought to validate the use of a simple one-emotion-dimension, linear-order terms only model, and also to consider the inter-emotion-dimension interaction.

In the single emotion dimension analysis with classical music, forward and backward stepwise regression on models with quadratic terms all confirm that the linear order model is adequate.

Including all emotion dimensions in the stepwise regression, we see much more variability in the optimal models for the classical study. All multiple emotion dimensions models perform either the same, or slightly more accurately, than the single emotion dimension models, as evaluated by higher R^2 and lower RMSE values. In addition, the music-alone dominance rating is a statistically significant factor in five of the six regression models. It should also be noted that we chose the music clips based on their distribution in Thayer's quadrants, which does not take into account the dominance dimension.

Insights from the second study of composed music are less obvious, which is to be expected since the emotion of these pieces was not constrained to fit a strong, singular mood, but rather, the composer was allowed to freely choose diverse moods for his compositions. In the first study of classical music, we selected music with non-overlapping moods covering the four quadrants of Thayer's emotion space, and we used a similar model to evaluate the perceived emotion. Because the composer of the second study was not constrained to the four quadrants of Thayer's space, the emotion of the composed pieces span the emotion space in a different way, which may represent some transformation of the four quadrants, or another space entirely. For example, one of the composed music segments is described as "curious, playful, sneaky," which could simultaneously suggest excitement (low stress) and danger (high stress). Thus, the less systematic coverage of Thayer's emotion space may have resulted in the higher variability we see in the linear regression models for the second study.

There are, however, a few points to be noted. In the single emotion dimension analysis, the linear-order music-alone data is included as a predictor for stress and activity, while the linear-order film-alone data is a predictor for both dominance regressions, forward and backward. Dominance may have been more difficult to determine from these music clips, causing the film to figure more prominently.

In the multiple emotion dimensions analysis, again the music-alone dominance data figures prominently, appearing in four of the six models, suggesting the greater interaction of this dimension on perceived emotion. The dominance models also include film-alone dominance, similar to what was stated above.

In summary, our present research suggests that, when considering single emotion dimension models, the single order film-alone and music-alone terms only are sufficient for a regression model. Our results also suggest that emotion perception of film and music combined may interact across multiple emotion dimensions, and not only in isolation within a single emotion dimension. Furthermore, music with less strongly defined emotional content will result in more variability in perceived emotion, and less accuracy in the regression models. Since emotion ambiguity occurs frequently in real-world scenarios, this area would benefit from further investigation, and more rigorous statistical modeling.

6. ACKNOWLEDGEMENTS

This work was supported in part by the University of Southern California Annenberg Center, and the Integrated Media Systems Center, an NSF Engineering Research Center, under Cooperative Agreement No. EEC-9529152, and NSF Grants Nos. 0347988 and 0321377. It made use of Integrated Media Systems Center Shared Facilities supported by the NSF Cooperative Agreement No. EEC-9529152. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect those of the Annenberg Center or NSF.

7. REFERENCES

[1] Marshall, Sandra K. and Cohen, A.J., “Effects of Musical Soundtrack on Attitudes toward Animated Geometric Figures,” Music Perception, 6, 95-112 (1988).

[2] Sirius, George and Clarke, Eric F., “The Perception of Audiovisual Relationships: A Preliminary Study,” Psychomusicology, 13, 119-132 (1994).

[3] Lipscomb, Scott D. and Kendall, Roger A., “Perceptual Judgment of the Relationship between Musical and Visual Components in Film,” Psychomusicology, 13, 60-98 (1994).

[4] Bullerjahn, Claudia and Güldenring, Markus, “An Empirical Investigation of Effects of Film Music Using Qualitative Content Analysis,” Psychomusicology, 13, 99-118 (1994).

[5] Boltz, Marilyn G., “Musical Soundtracks as a Schematic Influence on the Cognitive Processing of Filmed Events,” Music Perception, 18, 427–454 (2001).

[6] Vines, Bradley W., Krumhansl, Carol L., Wanderley, Marcelo M., Dalca, Ioana M., and Levitin, Daniel J., "Dimensions of Emotion in Expressive Musical Performance," Annals of New York Acad. of Sci., 1060, 462-466 (2005).

[7] Vines, Bradley W., Krumhansl, Carol L., Wanderley, Marcelo M., and Levitin, Daniel J., "Cross-Modal Interactions in the Perception of Musical Performance," Cognition, 101, 80-113 (2006).

[8] Parke, R., Chew, E., and Kyriakakis, C., “Quantitative and Visual Analysis of the Impact of Music on Perceived Emotion of Film,” ACM Computers in Entertainment, vol. 5, no. 3 (Jul/Sep 2007).

[9] Osgood, C. E., Suci, G. J., and Tannebaum, P.H., The Measurement of Meaning, Urbana: University of Illinois Press (1956).

[10] Juslin, Patrik N., “Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception,” J. Experimental Psychology, 26, 1797-1813 (2000).

[11] Thayer, R, The Biopsychology of Mood and Arousal, New York: Oxford University Press (1989).

[12] Huron, D, "Perceptual and Cognitive Applications in Music Information Retrieval," ISMIR (2000).

[13] Grimm, Michael and Kroschel, Kristian, “Evaluation of Natural Emotions Using Self Assessment Manikins,” Proc. IEEE Automatic Speech Recognition and Understanding Workshop (2005).