Measuring music-induced emotion

ARTICLE *in* MUSICAE SCIENTIAE · JULY 2011

Impact Factor: 1.54 · DOI: 10.1177/1029864911403367

2 AUTHORS:

Jonna K. Vuoskoski
University of Oxford

Tuomas Eerola
Durham University

Article

Musicae Scientiae
15(2) 159–173

DOI: 10.1177/1029864911403367
msx.sagepub.com

Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences

Jonna K. Vuoskoski and Tuomas Eerola
University of Jyväskylä, Finland

Abstract

Most previous studies investigating music-induced emotions have applied emotion models developed in other fields to the domain of music. The aim of this study was to compare the applicability of music-specific and general emotion models – namely the Geneva Emotional Music Scale (GEMS), and the discrete and dimensional emotion models – in the assessment of music-induced emotions. A related aim was to explore the role of individual difference variables (such as personality and mood) in music-induced emotions, and to discover whether some emotion models reflect these individual differences more strongly than others. One hundred and forty-eight participants listened to 16 film music excerpts and rated the emotional responses evoked by the music excerpts. Intraclass correlations and Cronbach alphas revealed that the overall consistency of ratings was the highest in the case of the dimensional model. The dimensional model also outperformed the other two models in the discrimination of music excerpts, and principal component analysis revealed that 89.9% of the variance in the mean ratings of all the scales (in all three models) was accounted for by two principal components that could be labelled as valence and arousal. Personality-related differences were the most pronounced in the case of the discrete emotion model. Personality, mood, and the emotion model used were also associated with the intensity of experienced emotions. Implications for future music and emotion studies are raised concerning the selection of an appropriate emotion model when measuring music-induced emotions.

Keywords

emotion models, individual differences, music-induced emotions

Introduction

Music-induced emotions are an inseparable part of music listening, as most listeners claim to experience strong emotions in response to music at least half of the time they spend listening to it (Juslin & Laukka, 2004). Although the focus of music and emotion studies has been quite equally divided between perceived and felt emotions over the past 20 years (Eerola & Vuoskoski,

Corresponding author:
Jonna K. Vuoskoski, Finnish Centre of Excellence in Interdisciplinary Music Research, Department of Music, P.O. Box 35(M), 40014 University of Jyväskylä, Finland
Email: jonna.vuoskoski@jyu.fi

submitted), there is continuing controversy over what kind of emotions music can actually induce in the listener (see, e.g., Juslin & Västfjäll, 2008; Konečni, 2008; Scherer, 2004). Despite this controversy, most previous studies investigating music-induced emotions have applied emotion models developed in fields other than the domain of music. A recent review of music and emotion studies (Eerola & Vuoskoski, submitted) revealed that the two most commonly used emotion models have been the discrete emotion model (i.e., basic emotions such as happiness, sadness, fear, and anger) and the two-dimensional circumplex model (comprising the orthogonal dimensions of valence and arousal; Russell, 1980). The discrete emotion model has often been modified to better describe emotions that are commonly represented by music, by replacing emotions such as disgust or surprise with more suitable emotion concepts such as tenderness or peacefulness (see, e.g., Gabrielsson & Lindström, 1995; Vieillard, Peretz, Gosselin, Khalfa, Gagnon, & Bouchard, 2008). However, it is still unclear whether models and theories designed for 'everyday' emotions can also be applied to an aesthetic context such as music, and there have been doubts whether a few primary basic emotions – or the two dimensions of valence and arousal – are adequate to describe the richness of emotional experiences induced by music (Zentner, Grandjean, & Scherer, 2008).

In 2008, Zentner, Grandjean, and Scherer published a pioneering study in which they introduced a new domain-specific emotion model for the measurement of music-induced emotions: the Geneva Emotional Music Scale (GEMS). They also provided a comparison between the GEMS and the two most commonly used emotion models, the basic emotion model (i.e., discrete emotions) and the two-dimensional circumplex model. However, as they only used western classical music in the comparison, and their formulation of the basic emotion model was somewhat unconventional (in the context of music and emotions), their conclusions about the superiority of the GEMS require further study.

In their comparison study, Zentner et al. (2008) used a version of the basic emotion model that included emotions such as shame, guilt, disgust, and contempt – emotions that are very rarely expressed or induced by music (see, e.g., Juslin & Laukka, 2004) or used in other studies investigating music and emotions (Eerola & Vuoskoski, submitted). The two-dimensional circumplex model was represented in the form of eight adjective groups, which were taken from the extremes of the two dimensions as well as from the quadrants between them, thus covering the entire affective space. However, the stimuli used by Zentner et al. (2008) did not represent the affective space in a comprehensive or balanced manner, as the stimuli were selected based on their potential effectiveness in inducing anger, fear, happiness, sadness, and "some of the novel emotion categories" (p. 508) included in the GEMS. In addition, the novel stimuli (i.e., the stimuli not used in previous experiments) used by Zentner et al. (2008) were not validated in terms of the emotions represented by them. We speculate that the reason for the use of the aforementioned emotion model formulations was to have a similar number of scales in the three models. In the present study, however, we wanted to replicate the comparison done by Zentner et al. (2008) while doing equal justice to all three models by using more conventional and simpler formulations (i.e., formulations that have previously been employed in studies of music and emotions) of the discrete and dimensional models, and by using ecologically valid and emotionally diverse stimulus material.

In the present experiment we used Finnish translations of the 26 adjectives representing the nine rating scales of GEMS-9 (Zentner et al., 2008). As Zentner and et al. originally conducted their series of studies in French, there are potential language issues related to the published version of GEMS (which is an English translation of the original). Since the present Finnish translation is based on the English version, it is possible that some nuances of the original GEMS may have been lost in translation. In our formulation of the discrete emotion model we included the emotions *happiness, sadness, anger, fear,* and *tenderness*, which are commonly expressed (and also induced) by music, and easily recognized by listeners (see, e.g., Juslin & Laukka, 2004). As for the formulation of the dimensional emotion model, we decided to use the three-dimensional emotion model (Schimmack & Grob, 2000), which comprises the dimensions of *valence* (pleasant–unpleasant), *energy* (awake–tired), and *tension* (tense–relaxed). Each dimension was represented by two unipolar scales. The three-dimensional model has previously been used in the measurement of perceived emotions in music, where it has provided highly consistent and reliable ratings (Eerola & Vuoskoski, 2011; Ilie & Thompson, 2006). These selections resulted in the three models having a differing number of scales, which has to be taken into consideration in the comparison and interpretation of the results. However, the differing number of scales will also allow us to explore the issue of scale redundancy in relation to discriminative power. In other words, although models with more scales are able to give more detailed and nuanced information, they might not be the most economical self-report instruments for felt emotions.

The consistency of ratings (i.e., agreement among people) is often used as a measure of how well an instrument performs, and it was also one of the measures used by Zentner et al. (2008) in their comparison of the three emotion models. However, in emotional processing, individual differences seem to be the rule rather than the exception (for a review, see Rusting, 1998). Both temporary mood states and stable personality traits have been associated with individual differences in emotional processing: extraverted people and people in positive moods have been shown to be particularly susceptible to positive affect, and neurotic people (and people in negative moods) to negative affect (for a review, see Rusting, 1998). Personality traits and mood states have also been associated with trait- and mood-congruent biases in emotional judgments in general (Rusting, 1998), as well as in the perception of emotions represented by music (Vuoskoski & Eerola, in press). Furthermore, certain personality traits (e.g., absorption and behavioural inhibition/activation) have been shown to play a role in emotional responses to music (Kallinen & Ravaja, 2006; Kreutz, Ott, Teichmann, Osawa, & Vaitl, 2008), and empathy and emotional contagion have been mentioned as possible mechanisms through which music can induce emotions (Juslin & Västfjäll, 2008; Scherer & Zentner, 2001).

Taking the individual differences between listeners into account can enable a more comprehensive picture of the emotional responses induced by music to emerge. Interestingly, Barrett (1998, 2006) has discovered that people with different personalities differ in emotional granularity (i.e., the level of detail people use to describe emotions experienced in everyday contexts). More specifically, people tend to vary in the extent to which they focus on the valence and arousal of their emotional experiences. "Valence focus" is related to valuation sensitivity (i.e., the sensitivity to reward and punishment cues), which is an integral part of the personality traits of extraversion and neuroticism (Barrett, 2006). Individuals high in valence focus tend to emphasize the hedonic contents of their experience, and use different negative (or positive) emotion words interchangeably to describe the same unpleasant (or pleasant) experience. In contrast, individuals low in valence focus describe their emotional experiences in a more differentiated manner, and distinguish between different emotion concepts with the same valence (Barrett, 1998, 2006). Such descriptions may also be used to characterize the distinctions between the nuanced, multi-dimensional GEMS model (highly specific verbal labels, such as nostalgia and wonder), and simpler models such as the two-dimensional circumplex model (broad labels, such as pleasant and unpleasant). Thus, it could be speculated that individuals differing in emotional granularity or valence focus also differ in the way they use different emotion models to describe their emotional responses to music. In the present study, this issue is addressed by exploring the possible contribution of individual difference variables such as personality (the "Big Five" personality traits and empathy) and mood. Perhaps some emotion models might reflect individual differences more strongly than others, and perhaps the applicability of different models depends – to some extent – on individual difference variables.

The first aim of the present study was to compare the applicability of three different emotion models – the discrete emotion model, the three-dimensional model, and the GEMS – in the measurement of music-induced emotions. The second aim was to explore the role of personality and mood in emotional responses to music, with regard to the musical excerpts and the emotion model used. Finally, the third aim was to investigate how the descriptive labels provided by the different emotion models relate to the intensity of experienced emotions. In provocative terms, it could be speculated that if one is required to describe one's emotional responses using inapplicable emotion concepts, the discrepancy between one's experienced emotions and the scales used to describe them may detract from the actual emotional experience, which could be reflected in the ratings of emotional intensity. Thus, ratings of "intensity of experienced emotion" were collected from all participants, including a control group that did not use any other verbal labels to describe their responses.

Method

Participants

The participants were 148 Finnish university students aged 18–49 years ($M = 23.5$, $SD = 4.84$; 77.1% females).

Stimuli and measures

Film music was used as stimulus material, since it is composed for the purpose of mediating powerful emotional cues, and it could serve as an ecologically valid and diverse stimulus material. The stimuli were 16 film music excerpts, ranging from 45 to 77 seconds in length (M = 57.13 s, SD = 9.74). These excerpts were selected on the following basis: using previously obtained valence and arousal ratings (Eerola & Vuoskoski, 2011), the excerpts were selected evenly from the 4 quadrants of the two-dimensional valence-arousal space: 4 excerpts from each quadrant. This resulted in 4 scary, 4 happy, 4 sad, and 4 tender excerpts. Since the original excerpts were too short (15 s) for emotion induction (for a review of music and emotion studies, see Eerola & Vuoskoski, submitted), longer excerpts (approx. 60 s) from the same pieces were taken, ensuring that the emotional characteristics of the original excerpts did not change during the longer versions. In the cases where the expression of a certain emotion did not last long enough (or did not continue in a satisfactory manner), the original excerpts were looped using cross-fading (see Appendix A for the list of excerpts). These new, longer excerpts were then tested in a pilot experiment, where 18 participants (mean age 27.17 years, SD = 6.58; 61.1% females) rated the emotions expressed by the excerpts. The correlations between the mean ratings for the original and the new, longer excerpts were .97 for valence and .97 for arousal, confirming that the new excerpts represented the intended emotion quadrants. Details relating to the stimuli (mean ratings and audio examples) can be found online at www.jyu.fi/music/coe/materials/emotion/soundtracks-1min/.

The Big Five Inventory (BFI; John & Srivastava, 1999) was used to assess the participants' personality traits. The BFI has 44 items that measure five broad personality domains labelled extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience.

The items were rated on a 5-point Likert scale (1 = disagree strongly, 5 = agree strongly). Empathy was assessed using the Interpersonal Reactivity Index (IRI; Davis, 1980), which taps into four separate aspects of the global concept of empathy: fantasy, perspective-taking, empathic concern, and personal distress. The participants had filled in the Big Five Inventory prior to the listening experiment, and the Interpersonal Reactivity Index was sent to the participants after the experiment. In all, 131 participants (88.5% of the total number of participants) returned the IRI. In addition, the UWIST Mood Adjective Checklist (UMACL; Matthews, Jones, & Chamberlain, 1990) was used to measure the participants' current mood prior to the experiment.

Procedure

The listening experiments were conducted individually for each participant in a soundproof room. To gather the emotion ratings, a special patch was designed in MAX/MSP graphical programming environment (version 5.1), running on Mac OS X. The patch presented the music excerpts in a different random order to each participant, and enabled the participants to move from one excerpt to the next at their own pace. Participants listened to the excerpts through studio quality headphones and were able to adjust the sound volume according to their own preferences.

The participants were instructed to rate the emotions that the music evoked in them, and the difference between perceived and felt emotions was explained to them. They were also told that the music might not evoke any emotions in them, and they were able to indicate whether the music evoked an emotional response or not by rating the intensity of their response. In the experiment patch, the scale measuring the intensity of emotional responses was labelled as "the strength of experienced emotion", where 1 signified "no emotion experienced" and 7 signified "very strong". The participants were divided into 4 groups; Group 1 (n = 46) only evaluated the intensity of their emotional responses (i.e., strength of experienced emotion) and how much they liked each excerpt, thus acting as a control group, whereas the other groups could also describe their potential emotional responses using different scales. Group 2 (n = 34) used discrete emotion scales (sadness, happiness, tenderness, fear, and anger), Group 3 (n = 35) used 6 unipolar scales (each represented by two adjectives) derived from the three-dimensional emotion model (positive and negative valence, high and low energy, and high and low tension; Schimmack & Grob, 2000), and Group 4 (n = 33) used the 9 scales (each represented by two or three adjectives) of the music-specific GEMS-9 emotion model (wonder, transcendence, power, tenderness, nostalgia, peacefulness, joyful activation, sadness, and tension; Zentner et al., 2008). The participants were instructed to describe their felt emotions using as many adjective scales as they found appropriate. All emotion ratings were done on 7-point Likert scales.

Results

Model comparison

Scale consistencies. In order to compare the consistency of ratings obtained with the different emotion models, Cronbach alphas and intra-class correlations were calculated for each scale in every model. Although emotional responses to music can potentially be highly subjective, the comparison of the inter-rater agreements can point us towards those models or scales where individual differences might play a more substantial role. The Cronbach alphas and intra-class

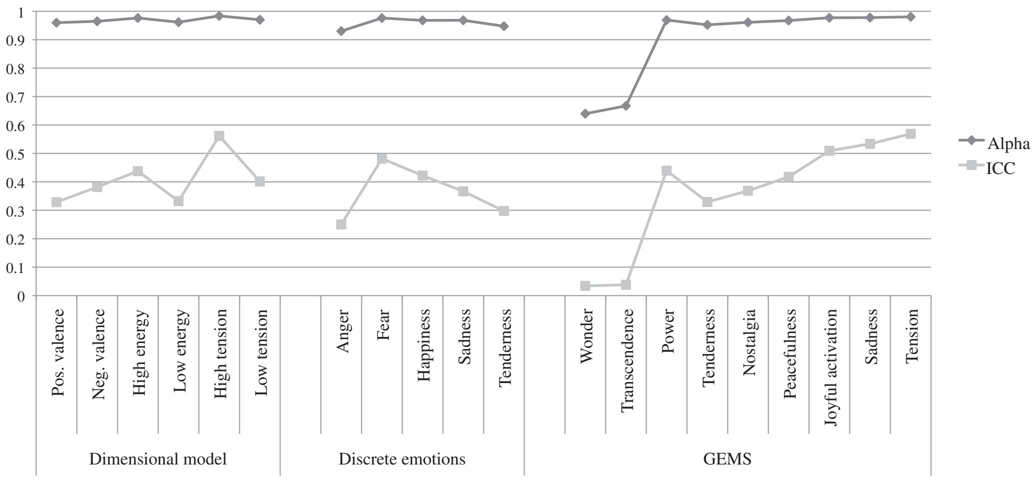

Figure 1. Intraclass correlations (ICC) and Cronbach alphas for all scales in every model.

correlations for every scale are displayed in Figure 1. As the figure illustrates, there were weak and strong scales in every model, and the GEMS-model possessed both the weakest and some of the strongest scales among all three models. The scales with the lowest consistencies were the wonder and transcendence scales of the GEMS-model (α .64, .67; ICC .03, .04), whereas the ratings of high tension (dimensional model) and tension (GEMS-model) had the highest consistencies (α .98, .98; ICC .56, .57). The low consistencies of the wonder and transcendence scales might either reflect the subjectivity of these types of responses, or that the concepts were inapplicable or unclear for the participants. Another explanation might be that the musical material used as stimuli was not able to evoke these types of responses in the participants. However, this seems unlikely, as the mean ratings of wonder and transcendence for all 16 excerpts were 2.46 and 2.14, whereas the overall mean of all GEMS scales was 2.26.

The overall consistency of the three-dimensional model (mean alpha .97; mean ICC .41) was higher than the consistencies of the discrete (.96; .36) and GEMS-9 (.90; .36) models. The pattern of results resembles those obtained by Zentner et al. (2008), except that in the present study the scale consistencies were more uniform across models. The higher overall consistencies of the discrete and dimensional models probably stem from the choice of more relevant items for the discrete and dimensional models, as well as the selection of emotionally varied (and previously validated) stimuli.

Discrimination of music excerpts. Cluster analysis was used to compare the discriminative power of the three models. A model that most reliably captures the variety and range of emotions evoked by music should have a higher power of discrimination and should provide coherent clusters of excerpts. In order to do justice to the dimensional model, the six unipolar scales (positive and negative valence, high and low energy, and high and low tension) were transformed into three bipolar dimensions (valence, energy, and tension) by deducting the mean ratings of negative valence from the mean ratings of positive valence and so forth. This is a standard procedure when collecting data using the three-dimensional model of affect (Schimmack & Grob, 2000; Schimmack & Reisenzein, 2002). The correlations between the two unipolar scales representing each dimension were between −.86 and −.95, supporting the notion that the three dimensions are indeed bipolar.

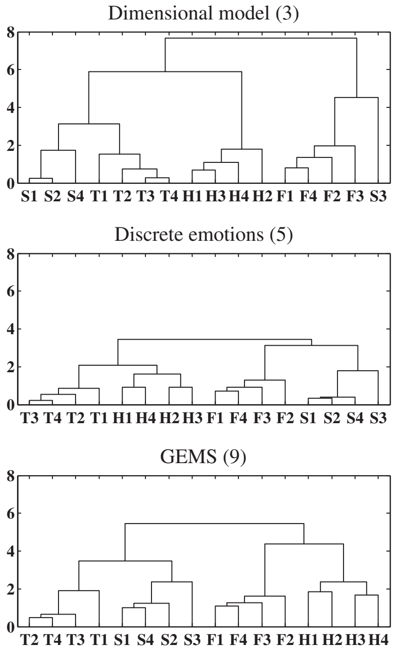

Figure 2. The cluster solutions for the three emotion models showing the clustering of the 16 music excerpts (scary excerpts F1–4, happy excerpts H1–4, sad excerpts S1–4, and tender excerpts T1–4; see Appendix A for the list of stimuli).

Clustering of the 16 music excerpts was performed using average linkage method and Euclidean distances. The mean emotion ratings given in response to each excerpt (on all scales) were used in the analysis. The resulting cluster solutions for the three emotion models are shown in Figure 2. The discriminative power of the cluster solutions was evaluated using aggregated distances (within a cluster solution), which were also used by Zentner et al. (2008) to compare the discriminative accuracy of the three models. The aggregated distances for the dimensional, discrete, and GEMS models were 33.4, 19.2, and 30.9 (respectively), the dimensional model having the largest distances between excerpts and thus the highest discriminative accuracy. These differences are also illustrated in Figure 2.

The qualitative aspects of the cluster structures are also revealing (see Figure 2). The discrete and GEMS models both cluster all 4 examples of each emotion into distinct low-level branches, but the main difference is that in the cluster solution of the GEMS model these branches are organized in terms of arousal (one mid-level branch for all tender [T1–4] and sad excerpts [S1–4], and a separate branch for scary [F1–4] and happy excerpts [H1–4]). In the cluster solution of the discrete emotion model, however, the mid-level branches seem to be organized according to valence. Interestingly, in the cluster solution of the dimensional model, the branch containing all scary excerpts also has a single sad excerpt, which was not evident in the cluster solutions of the other models. In addition, the three remaining sad excerpts were clustered close to the tender excerpts, clearly separated from the cluster of happy excerpts, and even more so from the scary (plus one sad) excerpts. This pattern appears to be reflecting both arousal and valence, and it brings forth the interesting fact that the participants reported experiencing positive (rather than negative) valence in response to three of the sad excerpts, but negative valence in response to one of the sad excerpts (and all of the scary excerpts). Thus, the cluster containing the scary excerpts and one sad excerpt represents the only group of excerpts that evoked negative valence in the participants.

If we consider the economy of the models by taking into account the number of parameters (dimensions or emotion categories) within each model, this provides another perspective on the discriminative power of the models (Lee, 2001). When Bayesian Information Criterion (Schwarz, 1978) is applied to the predictive power of the cluster solutions (using cophenetic correlation, which compares the cluster structure with the original distance matrix), the information theoretic values clearly favor the simplest model, the three-dimensional model. The Bayesian Information Criterion values for the dimensional, discrete, and GEMS models were −35.1 (with 3 predictors), −30.5 (5 predictors), and −18.5 (9 predictors) respectively, illustrating that the simplest models are in this case the most efficient ones, since the lower values indicate better overall model performance.

In sum, the three-dimensional emotion model seems to have superior discriminative power, especially when both the accuracy and simplicity of the models are considered. However, these results are naturally dependent on the musical material used, which in this case may favour the dimensional model. Furthermore, one may ask whether the discrete and GEMS models could be simplified in order to perform better in terms of economy and predictive accuracy. This issue is explored next.

Model reduction. In order to examine the possible collinearity and correspondence of the scales in all three models, a principal component analysis was conducted using the mean ratings for each excerpt. The analysis yielded 3 principal components: the first component (with Eigenvalue 13.8) accounted for 55.4% of the variance, the second component (with Eigenvalue 8.6) accounted for 34.5%, and the third component (with Eigenvalue 1.4) accounted for 6%. Since the contribution of the third component was minor and the component itself was difficult to interpret, components 1 and 2 were rotated using varimax rotation. Figure 3 displays the rotated two-dimensional solution, which accounted for 89.9% of the variance. Based on the loadings of the different emotion scales, the two components could be labelled as valence (inverted) and energy – or tension and energy – respectively.

As Figure 3 illustrates, some of the scales in the three models displayed very similar loadings. For example, high tension (dimensional) and tension (GEMS) were mapped next to each other in the principal component space, as were low tension (dimensional) and peacefulness (GEMS), high energy (dimensional) and power (GEMS), tenderness (discrete) and tenderness (GEMS), and sadness (discrete) and sadness (GEMS). This suggests that the three models have scales that measure very similar emotional qualities, and that the GEMS-model appears to combine aspects of the discrete and dimensional models – with the exception of nostalgia, transcendence, and wonder. Another noteworthy observation is that the scales of wonder and transcendence had very low loadings on all the three principal components, which is in line with the finding that these two scales had the lowest consistency of ratings among all three models.

Intensity of experienced emotion. In order to investigate whether the type of scales (GEMS, discrete, or dimensional) the participants used to describe their emotional responses had an effect on the reported intensity of their experienced emotions, we compared the mean intensity ratings of the participants in the four groups. Group 1 served as a control group, as they only rated the intensity of their experienced emotions and how much they liked each excerpt. A 1-way ANOVA followed by multiple comparisons of means (adjusted by Tukey's method) revealed that the participants using discrete emotion scales (Group 2) reported significantly

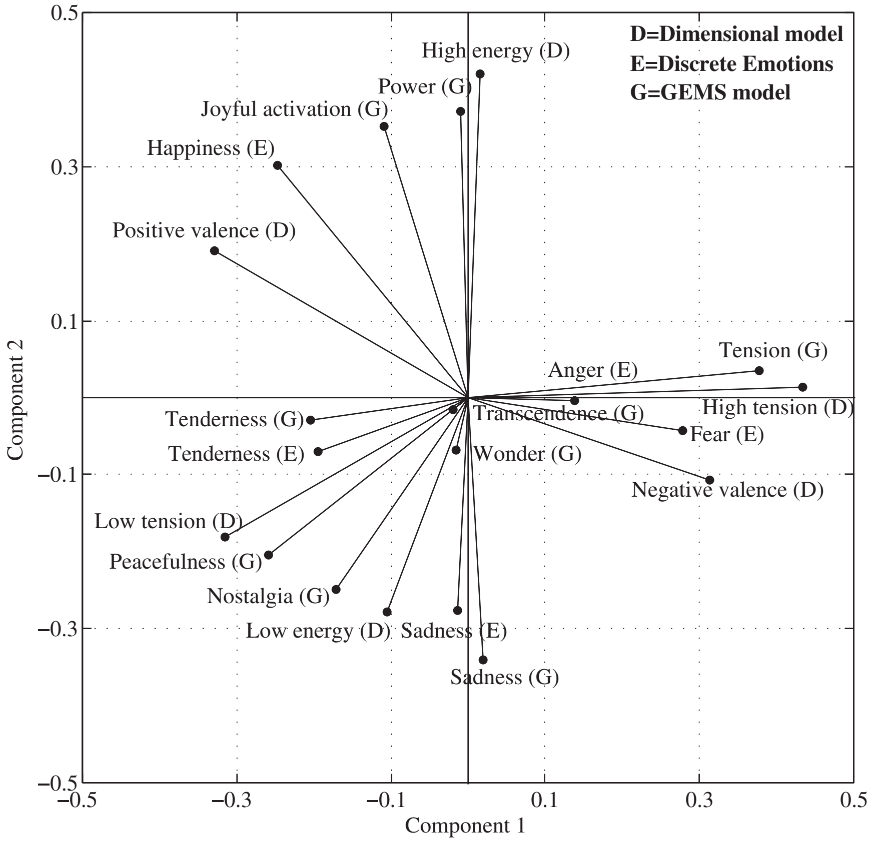

Figure 3. The scales of all three emotion models mapped according to their loadings on the two rotated principal components.

lower intensities of experienced emotion than the other groups: $F(3,144) = 4.83$, $p < 0.01$, $\eta_p^2 = .09$. This suggests that some participants may not have found the discrete emotion scales (anger, fear, happiness, sadness, and tenderness) applicable to their emotional responses, which affected their reported level of intensity. In comparison, the mean intensity ratings of the participants who used dimensional and GEMS scales (groups 3 and 4) did not differ from the intensity ratings of the control group (Group 1).

Individual differences in emotional responses

The effect of rating scales used. Next, we proceeded to investigate whether some emotion models reflect individual differences in emotional responses more strongly than others, and whether the low consistency of ratings is (at least in some cases) the result of these individual differences. Correlation analysis was used to explore whether certain personality traits were related to the experience of certain kinds of music-induced emotions. In the analyses we used mean emotion ratings of each participant, which were calculated from the ratings given in response to all 16 excerpts. There were no statistically significant correlations between personality traits and the mean emotion ratings of the participants using dimensional or GEMS scales, which may partly be due to the small sample size. More specifically, the GEMS scales of wonder and transcendence – which had the lowest consistency of ratings – did not correlate with any of the personality traits or moods under investigation. One might have expected that the trait of openness to experience could play a significant role in these types of responses, as it has previously been associated with the experience of aesthetic emotions such as awe (Shiota, Keltner, & John, 2006). However, in the group using discrete emotion scales, several statistically significant correlations emerged: extraversion correlated positively with experienced happiness ($r = .42$, $p < .05$), sadness ($r = .44$, $p < .01$), and tenderness ($r = .38$, $p < .05$). Extraversion also correlated with the intensity of experienced emotions ($r = .37$, $p < .05$), but only in the group using discrete emotion scales.

Intensity of emotional responses evoked by different types of excerpts. Finally, we investigated whether personality and mood variables were related to the intensity of emotional responses evoked by the different types of excerpts (scary, happy, sad, and tender). In the correlation analyses, we used the mean intensity ratings of all 148 participants for emotions evoked by the four different types of excerpts (note that $n = 131$ in the correlations regarding empathy and its subscales). The personality trait of agreeableness correlated positively with the intensity of emotional responses evoked by tender excerpts ($r = .16$, $p < .05$). Similarly, openness to experience and the empathy-subscale fantasy correlated with the intensity of emotional responses evoked by tender excerpts ($r = .18$ and $r = .19$, $p < .05$) as well as sad excerpts ($r = .17$ and $r = .19$, $p < .05$).

The hedonic tone (i.e., valence) and activity of current mood were both positively correlated with the overall intensity of experienced emotions ($r = .19$ and $r = .17$, $p < .05$). That is, participants in positive and active moods tended to experience more intense emotions in response to the music excerpts. In addition, the hedonic tone of mood was positively connected to the intensity of emotional responses evoked by scary excerpts ($r = .23$, $p < .01$), while the activity of mood was related to higher intensity ratings in response to sad excerpts ($r = .23$, $p < .01$). Interestingly, the third dimension of mood – tension – was negatively correlated with the intensity of emotional responses evoked by happy ($r = -.17$, $p < .05$) and tender ($r = -.20$, $p < .05$) excerpts. This suggests that music expressing positive emotions evokes milder emotional responses when the emotions expressed are incongruent with the listener's current mood state. However, this does not appear to be the case with music expressing negative emotions.

Conclusions

Model comparison

The dimensional model outperformed the GEMS and the discrete emotion model in the discrimination of music excerpts, illustrating that the simplest model may often be the most efficient one. In addition, the ratings of music-induced emotions obtained using the dimensional model had the highest overall consistency. The overall consistencies of the GEMS and the discrete emotion model were difficult to rank, but two of the scales in the GEMS model – wonder and transcendence – had substantially lower consistencies than the any of the other scales. In addition, these two scales did not load onto any of the three principal components in the principal component analysis, and they also did not correlate with any of the individual difference variables (such as openness to experience) that might be related to the experience of aesthetic emotions (see, e.g., Shiota et al., 2006). These findings suggest that these two scales – *wonder* and *transcendence* – may be controversial, especially when using types of music other than western classical music. One may ask whether the musical material used as the stimuli in the present experiment was able to evoke these types of responses in participants in the first place. It may be that these two scales tap into those types of strong, aesthetic experiences that only occur under certain circumstances and with very specific types of music. However, the mean ratings of wonder and transcendence (across all excerpts) were comparable to the mean ratings of the other GEMS scales, which suggests that the low inter-rater agreement might result from the subjective nature of these types of responses, or from the subjective definition of the emotion adjectives used to represent the scales. In the present study we used Finnish translations of the English GEMS scales and adjectives (which were originally translated from French adjectives by Zentner et al.), which may have also contributed to the differences in the present findings compared with Zentner et al. (2008).

Although the discrete emotion model had a relatively high consistency of ratings, it was also the weakest model in the discrimination of music excerpts. Most interestingly, those participants who used discrete emotion scales to describe their emotional responses reported significantly lower intensities of experienced emotions than the other participants. This finding suggests that at least some participants did not find the discrete emotion scales applicable to their experienced emotions, which influenced the reported intensity of their emotional responses. In comparison, the reported intensities of the participants using dimensional and GEMS scales were comparable to the intensities reported by the control group that did not use any adjective scales to describe the quality of their responses.

The results of the principal component analysis suggest that there is a great deal of scale redundancy in the GEMS model and the discrete model, as 89.9% of the variance in the mean ratings of all the scales (in all three models) was accounted for by two components that could be labelled as valence (inverted) and energy (or tension and energy). Similarly, Zentner et al. (2008) found that the intercorrelations in the 9 GEMS factors could be accounted for by 3 second-order factors labelled sublimity, vitality, and unease, which bear strong resemblance to the dimensions of valence, energy, and tension (respectively). However, the results of the PCA also indicate that there is some degree of scale redundancy in the three-dimensional model. This finding is supported by a previous study investigating perceived emotions in music (see Eerola & Vuoskoski, 2011), where valence and tension were observed to be highly collinear constructs. Furthermore, judging by the loadings of all the scales on the two principal components, most of the scales in the GEMS model appear to measure emotional attributes that are very similar to scales in the dimensional and discrete models – with the exception of nostalgia, transcendence, and wonder. However, as the two latter scales appear to be rather problematic, the contribution of the music-specific GEMS model to the measurement of music-induced emotions seems to require further investigation.

The results of the present study are somewhat incongruent with the results obtained by Zentner et al. (2008), and suggest that the performance of GEMS might depend on the type of music used. In the present experiment we used film music excerpts expressing emotions that varied in valence and arousal, while Zentner et al. used excerpts of western classical music masterpieces that were potentially effective in inducing sadness, happiness, anger, fear, and some of the novel emotion categories included in GEMS (for details, see Zentner et al., 2008). It has to be taken into consideration that the results of the present study might at least partly reflect the stimuli, which were originally chosen from the four quadrants of valence-arousal space. However, the stimulus selection was based on ratings of perceived emotions, which can be different from felt emotions (see, e.g., Evans & Schubert, 2008; Gabrielsson, 2002; Kallinen & Ravaja, 2006). Another noteworthy difference between the present study and the study by Zentner et al. is that the formulations of the discrete and dimensional models were quite different, and the version of the GEMS used in the present study (GEMS-9; 9 rating scales each represented by 2 or 3 emotion terms) was simpler than the version used by Zentner et al. (43 separately rated emotion terms). As in the present study we aimed to use more conventional formulations of the discrete and dimensional models, we inevitably had a different number of scales in the three models.

Individual differences in emotional responses

Personality-related differences in the experience of music-induced emotions were the most pronounced in the case of the discrete emotion model: the personality trait of *extraversion* was positively related to the experience of happiness, sadness, and tenderness in response to music excerpts, as well as to the overall intensity of emotional responses. The correlation between extraversion and experienced sadness is not in line with personality theory (according to which extraversion is associated with the tendency to experience positive emotions; see, e.g., John & Srivastava, 1999), which complicates the interpretation of the results. One may argue that the familiar discrete emotion concepts might bring out personality-related differences in emotional processing more clearly than the other two models, which are constructed from more abstract and nuanced emotion concepts. However, this interpretation does not account for the affect-incongruent correlation between extraversion and experienced sadness, nor does it account for the finding that extraversion was related to the reported intensity of experienced emotions only in the discrete emotion group, but not in the three other groups (dimensional, GEMS, and control). Therefore another, perhaps more plausible explanation might be that some people find the discrete emotion concepts more applicable to their emotional responses (to music) than other people do. According to Barrett (1998, 2006), people differ in terms of the level of detail they use to describe and define their emotional experiences. Furthermore, personality traits related to valuation sensitivity, namely extraversion and neuroticism, are associated with high valence focus (i.e., the tendency to use emotion terms such as "angry", "sad", and "nervous" interchangeably to indicate negative feeling; Barrett, 2006). It may be that in comparison to other participants, extraverted participants found the discrete emotion scales more applicable to their experienced emotions, perhaps reflecting high valence focus. This interpretation is further supported by the finding that the mean intensity of emotional responses was significantly lower in the group using discrete emotions to describe their responses, suggesting that some participants may have found the discrete emotion scales inapplicable to their emotional responses.

Certain personality traits and mood states were also related to the intensity of emotional responses evoked by different types of excerpts. Agreeableness was connected with the intensity of emotional responses evoked by tender excerpts, which is in line with the definition of agreeableness as a prosocial trait (see, e.g., John & Srtivastava, 1999). The positive correlation between agreeableness and the intensity of emotional responses evoked by tender music can be interpreted as a form of trait-congruence, reflecting the tendency of agreeable people to be altruistic, tender-minded, and trustful (see, e.g., John & Srtivastava, 1999). Similarly, the connection between the empathy-subscale fantasy and the intensity of emotional responses evoked by tender and sad excerpts can be seen to reflect an empathic individual's tendency to experience feelings of warmth, compassion, and concern for others undergoing negative experiences (see, e.g., Davis, 1980). More specifically, the fantasy scale taps into the tendency to transpose oneself imaginatively into the feelings and actions of fictitious characters in books and films (Davis, 1980), and thus it could be speculated that the trait also facilitates emotional contagion from music – one of the mechanisms through which music is thought to induce emotions (Juslin & Västfjäll, 2008).

The trait of openness to experience was also positively connected with the intensity of emotional responses evoked by tender and sad excerpts. One of the definitions of openness to experience is the sensitivity to art and beauty (McCrae & Sutin, 2009), which may at least partly explain the trait's involvement in the intensity of responses evoked by tender and sad excerpts. Openness to experience has also been found to play an important role in the enjoyment of sad music – together with empathy (Vuoskoski, Thompson, McIlwain, & Eerola, submitted). Furthermore, studies investigating the role of personality in music genre preferences have associated openness to experience with the preference for "reflective and complex" music styles (e.g., Rentfrow & Gosling, 2003), which further supports the interpretation that this trait might be related to the aesthetic appreciation of music.

Finally, positive and active moods were related to higher overall intensity of experienced emotions. A similar finding was reported by Dibben (2004), who found that high arousal (induced physiological arousal) was connected with higher intensity of emotional responses evoked by music. Thus, the connection between active mood and experienced intensity might be at least partly explained by the higher arousal level of active mood states. Positive mood also correlated with the intensity of emotional responses evoked by scary excerpts, while active mood correlated with the intensity evoked by sad excerpts. One explanation for these findings may be that the participants in negative and tired moods tried to repair their mood (see, e.g., Isen, 1985) and did not let themselves be absorbed in music expressing negative emotions. Another explanation involves the intriguing notion that music-induced emotions may be somewhat different from everyday emotions. For example, even if scary excerpts evoke intense emotional responses in listeners, the experienced emotion might not be fear (or any other clearly unpleasant emotion; see, e.g., Gabrielsson, 2002). This interpretation is rendered even more interesting by the finding that music expressing *positive* emotions evoked milder emotional responses when it was incongruent with the listener's current mood state (a finding consistent with the mood-congruency literature), whereas music expressing *negative* emotions functioned in an opposite manner. Perhaps the emotions evoked by music expressing "negative" emotions are indeed fundamentally different from everyday negative emotions, as they appear to function in a manner different from that described in the mood-congruency literature. This conclusion, however, requires further study, as the present study has examined the issue only briefly.

In summary, although the GEMS may provide more nuanced information about musically induced emotions (especially in situations where strong, aesthetic emotions are experienced), the dimensional model appears to be the most efficient and reliable way of collecting and representing musical emotion data. Conversely, the discrete emotion model appears to be the most problematic model in the measurement of music-induced emotions, as it had significantly lower discriminative accuracy than the other two models, and it also appeared to attenuate the reported intensity of experienced emotions. There were also individual differences in the use of the discrete emotion model, as extraverted participants seemed to find the discrete emotion scales more applicable to their experienced emotions than the other participants. Furthermore, certain personality traits (agreeableness, openness to experience, and the empathy-subscale fantasy) and mood states were associated with the intensity of emotional responses evoked by different types (scary, happy, tender, and sad) of excerpts. Perhaps the most intriguing direction for future research would be to explore further how music-specific and general emotion models are connected to individual differences in music-induced emotions, and to identify the range of listener features that contribute to these differences.

Acknowledgements

This research was supported by the Academy of Finland (project number 118616). A part of this study has previously been published in the proceedings of the 11th International Conference on Music Perception and Cognition in Seattle, USA (August 2010, University of Washington).

References

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion*, *12*(4), 579–599.

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality*, *40*(1), 35–55.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.

Dibben, N. (2004). The role of peripheral feedback in emotional experience with music. *Music Perception*, *22*(1), 79–115.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49.

Eerola, T., & Vuoskoski, J. K. (submitted). A review of music and emotion studies: Approaches, emotion models and stimuli. Manuscript submitted for publication.

Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, *12*(1), 75–99.

Gabrielsson, A. (2002). Emotion perceived and emotion felt. *Musicae Scientiae* [Special issue 2001/2], 123–147.

Gabrielsson, A., & Lindström, E. (1995). Emotional expression in synthesizer and sentograph performance. *Psychomusicology*, *14*(1), 94–116.

Ilie, G., & Thompson, W. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*(4), 319–329.

Isen, A. M. (1985). Asymmetry of happiness and sadness in effects on memory in normal college students: Comment on Hasher, Rose, Zacks, Sanft, and Doren. *Journal of Experimental Psychology: General*, *114*(3), 388–391.

John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (eds), *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford Press.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*(3), 217–238.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–575.

Kallinen, K., & Ravaja, N. (2006). Emotion perceived and emotion felt: Same and different. *Musicae Scientiae*, *10*(2), 191–213.

Konečni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *2*, 115–129.

Kreutz, G., Ott, U., Teichmann, D., Osawa, P., & Vaitl, D. (2008). Using music to induce emotions: Influences of musical preference and absorption. *Psychology of Music*, *36*(1), 101–126.

Lee, M. D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology*, *45*, 149–166.

Matthews, G., Jones, D. M., & Chamberlain, A. (1990). Refining the measurement of mood: the UWIST Mood Adjective Checklist. *British Journal of Psychology*, *81*, 17–42.

McCrae, R. R., & Sutin, A. R. (2009). Openness to experience. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 257–273). New York: Guilford.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*, 1236–1256.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.
Rusting, C. L. (1998). Personality, mood, and cognitive processing of emotional information: Three conceptual frameworks. *Psychological Bulletin*, *124*(2), 165–196.
Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, *14*(4), 325–345.
Schimmack, U., & Reisenzein, R. (2002). Experiencing activation: Energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion*, *2*(4), 412–417.
Scherer, K. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research*, *33*(3), 239–251.
Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research*. Oxford: University Press.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
Shiota, M. N., Keltner, D., & John, O. P. (2006). Positive emotion dispositions differentially associated with Big Five personality and attachment style. *The Journal of Positive Psychology*, *1*(2), 61–71.
Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*(4), 720–752.
Vuoskoski, J. K., & Eerola, T. (in press). The role of mood and personality in the perception of emotions represented by music. *Cortex*.
Vuoskoski, J. K., Thompson, W. F., McIlwain, D., & Eerola, T. (submitted). Who enjoys listening to sad music and why? Manuscript submitted for publication.
Zentner, M., Grandjean, D., & Scherer, K. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*(4), 494–521.

Appendix A

List of stimuli used to induce emotions

Excerpt	Emotion	Soundtrack name	Track	Time	Duration	Number in the set of 110**
F1	Scary	The Alien Trilogy	9	00:00-00:56	56 s	003
F2	Scary	Batman Returns	5	00:00-00:46	46 s	011
F3	Scary	The Fifth Element	17	00:00-01:01	61 s	018
F4	Scary	The Alien Trilogy	11	02:04-02:58	54 s	091
H1	Happy	Oliver Twist	8	01:32-02:09 L*	72 s	027
H2	Happy	Dances with Wolves	10	00:00-00:46	46 s	055
H3	Happy	The Untouchables	6	01:26-02:06 L*	67 s	071
H4	Happy	Pride & Prejudice	4	00:10-01:06	56 s	105
S1	Sad	The English Patient	18	00:00-00:59	59 s	031
S2	Sad	The Portrait of a Lady	9	00:00-00:23 L*	45 s	033
S3	Sad	Running Scared	15	01:45-02:40	55 s	086
S4	Sad	Pride & Prejudice	13	00:40-01:30	50 s	109
T1	Tender	The Portrait of a Lady	3	00:23-01:08	45 s	029
T2	Tender	Shine	10	01:01-02:00	59 s	041
T3	Tender	Pride & Prejudice	1	00:10-00:49 L*	77 s	042
T4	Tender	The Godfather III	5	01:13-02:19	66 s	107

*L = looped
**see the list of stimuli for the set of 110 excerpts, Eerola & Vuoskoski, 2011