A comparison of the discrete and dimensional models of emotion in music

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

Abstract

The primary aim of the present study was to systematically compare perceived emotions in music using two different theoretical frameworks: the discrete emotion model, and the dimensional model of affect. A secondary aim was to introduce a new, improved set of stimuli for the study of music-mediated emotions. A large pilot study established a set of 110 film music excerpts, half were moderately and highly representative examples of five discrete emotions (anger, fear, sadness, happiness and tenderness), and the other half moderate and high examples of the six extremes of three bipolar dimensions (valence, energy arousal and tension arousal). These excerpts were rated in a listening experiment by 116 non-musicians. All target emotions of highly representative examples in both conceptual sets were discriminated by self-ratings. Linear mapping techniques between the discrete and dimensional models revealed a high correspondence along two central dimensions that can be labelled as valence and arousal, and the three dimensions could be reduced to two without significantly reducing the goodness of fit. The major difference between the discrete and categorical models concerned the poorer resolution of the discrete model in characterizing emotionally ambiguous examples. The study offers systematically structured and rich stimulus material for exploring emotional processing.

Key words

battery, dimensional, music, discrete, emotion, three-dimensional

Music has the ability to convey powerful emotions. This ability has fascinated researchers as well as the general public throughout the ages, and although great strides forward have been made in the field of music and emotion research, much remains unclear. One issue that has been holding back advances in understanding the complex phenomena of music-mediated emotions has been the abundance of different emotion theories and concepts – discrete, dimensional, music-specific or something else altogether. Before focussing on novel, music-specific

Corresponding author:
Tuomas Eerola, Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Seminaarinkatu 15, PO Box 35, Finland.
[email: tuomas.eerola@jyu.fi]

models (e.g., Zentner, Grandjean, & Scherer, 2008), there is a need to first critically compare the discrete and dimensional models of emotions in music because these are the two dominant models used in music and emotion research (Juslin & Sloboda, 2010; Zentner & Eerola, 2009), and they are often implied to be highly convergent although this has not actually been explicitly studied. Secondly, neurological studies (Gosselin et al., 2005; Khalfa et al., 2008b) have indicated that different processes may be involved in the discrete and dimensional assessments of emotions. Thirdly, recent hybrid models of emotion (Barrett, 2006; Christie & Friedman, 2004; Russell, 2003) depend on finding the ways in which core affects (taken to be dimensional) interact with the conscious interpretation of what people know about emotions (best described in discrete terms). And finally the understanding of musical and acoustic features that contribute to emotions would greatly benefit from knowing which model – dimensional or discrete – maps the feature space in the most ecological fashion.

Another hindrance in music and emotion research has been the choice, quality, and amount of musical examples used as stimuli. Previous studies have predominantly used well-known excerpts of Western classical music, which have been chosen arbitrarily by the researchers. Moreover, the stimuli have mostly been highly typical exemplars of the chosen emotions even if the underlying emotion model does not imply that emotions are structured around specific categories. We will address these issues in detail later.

In this article, we focus on perceived emotions (in other words, emotions that are represented by music and perceived as such by the listener). An overview of the literature implies that the border between the two alternatives – emotion recognition and emotion experience – may be somewhat blurred in reality, and it has even been suggested that the two alternatives could be seen as opposite extremes of a continuum (Gabrielsson, 2001). In addition recent empirical studies have found more similarities than differences between the two (Evans & Schubert, 2008; Kallinen & Ravaja, 2006; Vieillard et al., 2008).

To address the theoretical diversity in depth, we will first briefly summarize the prominent kinds of emotion models and their relevance for music. During the past decade, discrete models, different dimensional models, and domain-specific emotion models have all received support in studies of music and emotion (Ilie & Thompson, 2006; Krumhansl, 1997; Schubert, 1999; Zentner et al., 2008). According to the well-known discrete emotion model – the basic emotion model – all emotions can be derived from a limited number of universal and innate basic emotions such as fear, anger, disgust, sadness and happiness (Ekman, 1992, 1999). The basic emotion model builds on the assumption that an independent neural system subserves every discrete basic emotion. However, neuro-imaging and physiological studies have failed to establish reliable, consistent evidence to support this theory (for a review, see Barrett & Wager, 2006; Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000), and the matter remains under debate. In studies investigating music and emotion, the basic emotion model has often been modified to better describe the emotions that are commonly represented by music. For example, basic emotions rarely expressed by music, such as disgust, are often changed to more suitable emotion concepts like tenderness or peacefulness (Balkwill & Thompson, 1999; Gabrielsson & Juslin, 1996; Vieillard et al., 2008). It still remains to be clarified whether models and theories designed for utilitarian emotions (Scherer, 2004) – such as the basic emotion model – can also be applied in an aesthetic context such as music. It has been argued and empirically demonstrated that a few primary basic emotions seem inadequate to describe the richness of the emotional effects of music (Zentner et al., 2008). In their study, Zentner and colleagues (2008) proposed a new model for music-induced emotions by first compiling music-related emotion terms and uncovering the underlying emotion structure with exploratory factor analysis, and then corroborating

their findings by means of confirmatory factor analysis. The resulting nine-factor Geneva Emotion Music Scale (GEMS) model consists of wonder, transcendence, tenderness, nostalgia, peacefulness, power, joyful activation, tension and sadness. For music and emotions studies, this model provides much needed domain-specificity and emphasizes the positive and reflective nature of music-induced emotions. Although it was shown that the GEMS model outperformed discrete and dimensional emotion models in accounting for felt emotions in music, in our opinion, these results can be disputed as they pitted musically non-relevant formulations of the basic emotions and dimensions against GEMS and relied on a handful of familiar classical music examples. The focus of the present study is to compare the traditional models of emotion in music and also to focus on perceived emotions. This is a different emphasis to that of GEMS, although the scale has implications which will be discussed further.

In recent years, two-dimensional models of emotion have gained support among music and emotion researchers (e.g., Gomez & Danuser, 2004; Schubert, 1999; Withvliet & Vrana, 2006). Instead of an independent neural system for every basic emotion, the two-dimensional circumplex model (Posner, Russell, & Peterson, 2005; Russell, 1980) proposes that all affective states arise from two independent neurophysiological systems: one related to valence (a pleasure–displeasure continuum) and the other to arousal (activation–deactivation). In other words, all emotions can be understood as varying degrees of both valence and arousal. In contrast, Thayer (1989) suggested that the two underlying dimensions of affect were two separate arousal dimensions: energetic arousal and tense arousal. According to Thayer's multidimensional model of activation, valence could be explained as varying combinations of energetic arousal and tense arousal. A visual summary of the two-dimensional models of Russell and Thayer is given in Figure 1. In the music domain, Vieillard and colleagues explored emotional excerpts of music by means of similarity ratings, and found that the excerpts could be mapped onto a two-dimensional plane in which the salient dimensions could be best explained in terms of energy and tension (Vieillard et al., 2008). However, the two-dimensional models have been criticized for their lack of differentiation when it comes to emotions that are close neighbours in the valence-activation space, such as anger and fear (see e.g., Tellegen, Watson, & Clark, 1999). It has also been discovered, that the two-dimensional model is not able to account for all the variance in music-mediated emotions (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Collier, 2007; Ilie & Thompson, 2006).

Wilhelm Wundt suggested a distinction be made between three dimensions of emotions already as early as 1896. These three dimensions were pleasure–displeasure, arousal–calmness, and tension–relaxation. Although the two-dimensional models have a dominant position in affect literature, there is some evidence of the model's incompatibility with affect data (Schimmack & Grob, 2000; Schimmack & Reisenzein, 2002). Previous studies have shown that arousal–calmness and tension–relaxation dimensions cannot be reduced to one arousal dimension (Schimmack & Grob, 2000; Schimmack & Reisenzein, 2002). The underlying reason is that the two activation dimensions are related to different causes: unlike tense arousal, energetic arousal is affected by a circadian rhythm (Thayer, 1989; Watson, Wiese, Vaidya, & Tellegen, 1999), and the two arousal dimensions have been shown to change in opposite directions when specifically manipulated (Gold, MacLeod, Deary, & Frier, 1995). In sum, both main theoretical models – discrete and dimensional – will be investigated simultaneously to clarify their dependencies and applicability to music and emotions. The three-dimensional model, shown visually in Figure 1, will be used to collect data regarding the dimensional approach as it still allows us to examine post facto whether a lower dimensional solution (valence and arousal or arousal and tension) could be used instead.
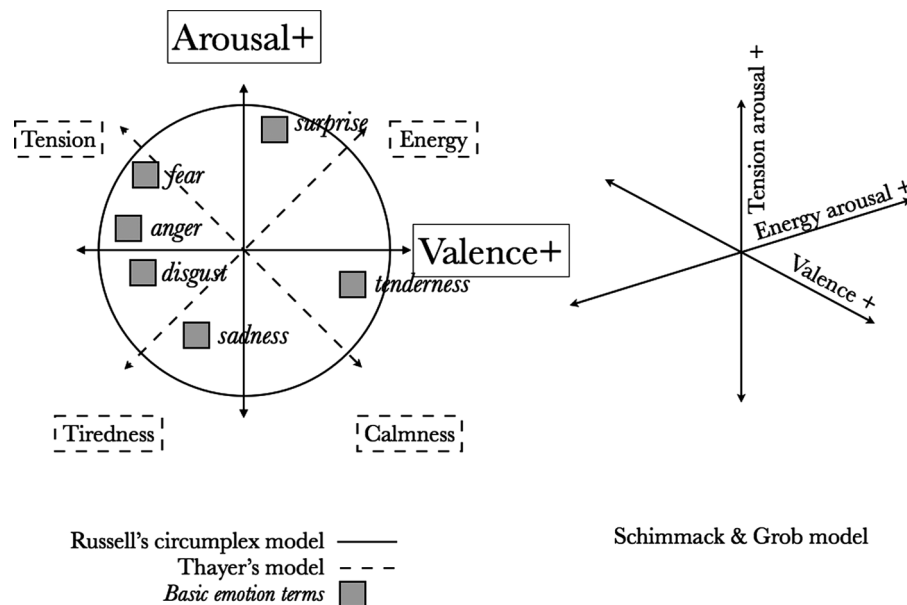
Figure 1. Schematic diagram of the dimensional models of emotions with common basic emotion categories overlaid. Note that the axes of the three-dimensional model are not necessarily orthogonal in actual affect data as depicted here (see Schimmack & Grob, 2000).

Two recent and puzzling findings render the comparison of these conceptual frameworks of emotion even more pressing an issue. Studies with brain damaged patients have documented a dissociation between discrete and dimensional evaluations of emotion in music (Dellacherie, Ehrlé, & Samson, 2008; Gosselin et al., 2005; Khalfa et al., 2008b), implying that separate neural processes are responsible for each of these types of evaluation. Moreover, it has been shown that listeners may experience both sad and happy feelings at the same time when exposed to a stimulus with mixed emotional cues (Hunter, Schellenberg, & Schimmack, 2008). Both of these findings pose challenges for emotion research in music and require a better understanding of the essential similarities and differences between these two conceptual frameworks.

The majority of studies on music and emotions have used excerpts of relatively well-known Western classical music pieces (e.g., Kreutz, Ott, Teichmann, Osawa, & Vaitl, 2008; Krumhansl, 1997; Nawrot, 2003; Schmidt & Trainor, 2001), which have been arbitrarily chosen by the researchers. Other types of musical stimuli used – occasionally together with classical music – include popular music and jazz (Altenmüller, Schuermann, Lim, & Parlitz, 2002; Gomez & Danuser, 2004), film music (Etzel, Johnsen, Dickerson, Tranel, & Adolphs, 2006), music from other cultures (Balkwill & Thompson, 1999; Balkwill, Thompson, & Matsunaga, 2004), and synthetic music that has been composed especially for the research task at hand (e.g., Khalfa, Peretz, Blondin, & Manon, 2002; Vieillard et al., 2008). Well-known music examples are potentially problematic as stimuli because participants may already be familiar with the excerpts. Emotions elicited by this type of stimuli can be closely entwined with extra-musical associations. Synthetic stimuli are free of such problems and have provided opportunities to study and manipulate musical features. However, such stimuli often sound artificial and lack some of real

music's intricate features, such as expressive performance and timbre, which might be essential in evoking an emotional response (Juslin, Friberg, & Bresin, 2002; Leman, Vermeulen, De Voogdt, Moelants, & Lesaffre, 2005). Moreover, the stimuli used to investigate dimensional emotion models have typically represented the extremes of the dimensions (e.g., low arousal, high valence) as opposed to points along the continuum of each dimension. This may have led to the neglect of potentially relevant affect data.

It is worth of noting that, with the notable exception of Bigand et al. (2005), musical examples have been chosen solely in terms of wholly discrete emotions (e.g., happy and sad) in the existing studies that use both discrete and dimensional models of emotion in music (Gosselin et al., 2006; Gosselin, Peretz, Johnsen, & Adolphs, 2007; Khalfa et al., 2002, 2008a, 2008b; Kreutz et al., 2008; Krumhansl, 1997; Nyklìcek, Thayer, & Van Doornen, 1997; Terwogt & Van Grinsven, 1991; Vieillard et al., 2008). Hence, the ratings of dimensional concepts such as valence and arousal describe the known points of discrete emotions in the dimensional affective space (e.g., Nyklìcek et al., 1997). However, it is currently unclear what happens in the reverse situation where the musical examples are selected systematically from various points within the affective space. In such cases the discrete emotions may not be easily assigned to examples that are distant from the prescribed point in the affective space for that particular discrete emotion. Lastly, the number of music examples used in previous studies has been relatively low compared with the stimulus sets used by emotion researchers in other fields, such as the visual domain. The International Affective Picture System (IAPS) uses 12 series of 60 pictures each (Lang, Bradley, & Cuthbert, 2005), out of which typical emotion studies tend to use 50–60 images (this was the median in a sample of 20 studies using IAPS). However, in the music domain, much smaller sets are more commonly used.

Aims of the study

The primary aim of the present study is to contribute to the theoretical debate currently occupying music and emotion research by systematically comparing evaluations of perceived emotions using two different theoretical frameworks: the discrete emotion model, and dimensional model of affect. The importance of the comparison lies not only in the prevalence of these models in music and emotion studies, but also in the suggested neurological differences involved in emotion categorization and the evaluation of emotion dimensions (Khalfa et al., 2008b), as well as in the categorically constrained affect space the excerpts have represented to date. Moreover, the various alternative formulations of the dimensional model have not been investigated in music and emotion studies before.

A secondary aim is to introduce a new, improved set of stimuli – consisting of unfamiliar, thoroughly tested and validated non-synthetic music excerpts – for the study of music-mediated emotions. Moreover, this set of stimuli should not only include the best examples of target emotions but also moderate examples that permit the study of more subtle variations in emotion.

Expert selection of the stimulus materials

In order to obtain a large sample of unknown yet emotionally stimulating musical examples, a large expert panel was organized for choosing the material. The primary goal of this panel was to choose emotionally representative musical material from a large selection of film soundtracks according to predefined criteria.

It was decided to use film music because it is composed for the purpose of mediating powerful emotional cues, and could serve as a relatively 'neutral' stimulus material in terms of musical preferences and familiarity. Unfamiliar excerpts were chosen to avoid *episodic memories* from particular films influencing perceived emotions in the music. And yet, since it was film music, and listeners are generally accustomed to this genre from media exposure, the excerpts were nevertheless expected to conjure up *schematic memories*. But this also meant we could not prevent excerpts from triggering memories by simple virtue of them resembling others from a listener's previous experience. With the exception of Vertigo (from 1958), the selection of 60 soundtracks was limited to those published within the last three decades (1976–2006). This was to keep the sound quality of the corpus relatively homogeneous. It should also be noted that the soundtracks came from a wide range of films that included romantic, sci-fi, horror, action, comedy, and drama.

The panel consisted of 12 expert musicologists (staff members and third to fifth year university students) who had all studied a musical instrument for 10 years or more. Each panel member was given five different soundtracks and asked to find five examples of the six target emotions. Half the experts focused on discrete emotions (six targets), and half on the extremes of the three-dimensional model (six targets). For the discrete emotions we chose happiness, sadness, fear, anger, surprise and tenderness, as these have been favoured in previous studies of music and emotion (Juslin, 2000; Kallinen, 2005; Krumhansl, 1997). For the dimensional model, we chose the six extremes of the three-dimensional model of emotion by Schimmack and Grob (2000). This meant the panel should find examples of both positive and negative valence, high and low tension arousal, as well as high and low energy arousal. Each extreme was characterized using three adjectives taken from Schimmack and Grob (2000). For valence, these were pleasant–unpleasant, good–bad, and positive–negative. For the energy dimension the adjectives were awake–sleepy, wakeful–tired, and alert–drowsy. The adjectives used to represent the extremes of the tension dimension were tense–relaxed, clutched up–calm, and jittery–at rest.

To ensure a degree of uniformity in the choice of sound examples, the following criteria were established: each excerpt should be between 10 and 30 seconds (depending on the natural phrasing of the excerpt): it should not contain lyrics, dialogue, or sound effects (car sounds, etc.), and, though familiar with the schematic memory, it should not be familiar in the episodic sense (see earlier). It was also stressed that the goal was to choose examples that could convey the target emotion to the general listener in an optimal way. The experts also made a note of the musical features and devices which informed their choice. This resulted in 360 audio clips (12 × 5 × 6), equally representative of the discrete emotion and three-dimensional models. Details related to the stimuli (names, ratings, and audio examples) may be found online.[1]

Pilot experiment

The aim of the pilot experiment was to rate all the examples previously selected by experts in terms of both the models (discrete and dimensional). This was done to understand how the emotions were conveyed by the examples. The aim was also to reduce the number of excerpts and homogenize the selection for further investigation.

Method

Participants, stimuli, apparatus and procedure. The participants of the pilot experiment were the same group of experts who originally chose the examples (mean age 24.1 years, SD = 3.9 years, 7 females). The stimuli consisted of all 360 excerpts that had been chosen. The panel then

Table 1. Consistencies, means, standard deviations for all emotion ratings and repeated measures ANOVA results for excerpts grouped by target emotion (η^2 for effect sizes)

Type of excerpt/ concept	Cons. (α) (concept)	Target (concept)	Non-target (concept)	ANOVA (η^2) (excerpts grouped by target)
Happy	.93	5.49 (1.53)	2.02 (0.69)	0.63***
Sad	.89	5.46 (1.60)	2.22 (0.98)	0.71***
Tender	.92	5.69(1.89)	2.38 (0.88)	0.72***
Fearful	.92	5.29 (1.80)	2.63 (0.91)	0.63***
Angry	.92	5.38 (1.61)	2.03 (0.74)	0.69***
Surprising	.66	3.36 (2.05)	1.72 (1.02)	0.23***
Pos. valence	.92	5.68 (0.94)	3.96 (1.00)	0.64***
Neg. valence	–	1.81 (0.68)	4.31 (1.02)	0.88***
Pos. energy	.90	5.67 (0.88)	3.92 (0.96)	0.30***
Neg. energy	–	2.30 (0.77)	4.23 (0.97)	0.58***
Pos. tension	.93	5.98 (0.75)	4.31 (0.86)	0.69***
Neg. tension	–	2.48 (0.90)	4.36 (0.89)	0.68***

Note: *p < .05; **p < .01; ***p < .001; *df* = 5,179 for basic emotion concepts, *df* = 2,89 for dimensional concepts.

received instructions to rate the *perceived* emotion in each audio clip. Half of them were instructed to rate the discrete emotions on a scale of 1–7 and the rest gave ratings for the three dimensions using bipolar scales. Familiarity with the excerpts was also rated (0 = unfamiliar, 1 = somewhat familiar, 2 = very familiar). Note that each of the panel's own selections constituted a mere 8.3% of the whole material (30 items out of 360). The rating task was also divided into four sections, and between each of them, the emotion models were switched between the two halves of the experts. The sections each consisted of 90 excerpts and lasted about 50 minutes. All the excerpts were played in a random order, but this order was the same for the whole group since it was done as a classroom exercise using high-quality audio in laboratory conditions. In total, the task lasted approximately four hours and was carried out on two separate days. The student participants received course credits for their efforts.

Results

To rule out the order effects, a linear trend analysis was conducted against the rating order for all scales. All the trends yielded non-significant *F*-ratios (*F* values between 0.2 and 2.82, *p* = ns, df = 358). Next the consensus between the raters was investigated. Cronbach's alpha was employed to measure the consistency between raters, rather than between items, as there was a large proportion of excerpts given similar (low) ratings, which meant negligible item variance. In other words the lowest ratings were given by all raters for those discrete emotions that manifestly did not seem applicable to a particular excerpt, e.g., ratings of sadness for happy excerpts. Most emotion concepts scored relatively high consistency using this procedure, as indicated in Table 1. The notable exception in inter-rater consistencies was surprise (α = .66), which was actually unsurprising, as a number of previous studies have also observed this concept to be problematic for music-mediated emotions (Gabrielsson & Juslin, 1996; Laukka & Juslin, 2007). Surprise was therefore eliminated from further analyses, as the alpha for it was considerably lower than for the other emotion terms.

The pilot showed that the raters were not particularly familiar with the excerpts as 89.9% indicated 'unfamiliar', 6.4% 'somewhat familiar' and only 3.8% 'very familiar'. To assess the role familiarity for the emotion categories and dimensions, a variance analysis was conducted using a non-parametric method of calculation (Kruskal-Wallis) as the familiarity ratings contained mainly zeros and were thus not normally distributed (Lilliefors test, $p < .001$). This yielded no significant differences across the discrete emotions, χ^2 (5,179) = 10.18, p = ns, and yet significant differences across the dimensional extremes, χ^2 (5,179) = 30.02, $p < 0.001$. A closer look at the familiarity ratings showed a small number of highly familiar excerpts (above three SDs above the mean familiarity: three in total) in two categories: high valence and low tension. We kept these familiar excerpts in the data set since they represented a minor portion of the total number (0.83%).

Next we explored whether the emotion targets were clearly evident in the ratings. This was done by separately comparing the mean ratings for each type of excerpt using an analysis of variance. This repeated measures ANOVA yielded significant main effects for all of the discrete emotion targets, and with large effect sizes (mostly above 0.60, shown in Table 1). The exception was a surprise, as mentioned previously, which, although statistically significant, exhibited a fairly low effect size. The other weak effect size (0.30) was seen in the ratings of high energy, the reason for which is probably due to its collinearity with the other dimensions. Excerpts representing high energy were also rated high in tension, as well as high in negative valence, and thus the ANOVA results display weaker discrimination between these categories. A post-hoc analysis between the discrete emotion categories (using Holm-Sidak adjusted t-tests at $p < .05$ level) revealed this statistically, namely how anger ratings could not be discriminated from fear ratings when anger was the target category. Similarly, ratings of surprise were statistically indistinguishable from ratings of fear and anger. Post-hoc analyses were also used to characterize the differences between dimensional extremes, resulting in statistically significant differences at a level of $p < .05$ in all comparisons.

Discussion

The primary aim of the pilot experiment was to evaluate the chosen set of musical stimuli in terms of two conceptual frameworks in order to establish a systematic basis for selecting the stimuli for the actual experiment. Such selection rationale has been noted to be largely absent in the previous studies of music and emotions (Juslin & Västfjäll, 2008). The chosen target emotions were clearly evident in the ratings, with the exception of surprise as previously noted, and so this was removed. As the set of stimuli contains a large variety of music excerpts rated in terms of both emotion models, it is now possible to compare these models in a theory-based and systematic way.

Experiment

The first aim of this experiment was to systematically compare the effectiveness of discrete and dimensional models in the study of perceived emotions in music. A related aim was to explore whether some of the ways of representing emotions within the discrete and dimensional models could be merged or eliminated (e.g., collapsing the three-dimensional model into two dimensions, or removing some of the overlapping concepts from the discrete emotion model).

The second aim was to form a refined set of musical examples, which would not only include the clearest exemplars of the discrete emotion categories, but also ones that are less easily

attributable to a category. Such moderate examples would provide more realistic and interesting material for empirical work on recognition and induction of emotions in music than the standard paradigm, where only a few highly characteristic examples of emotion categories are used. An experiment was designed to address these questions using a subset of the stimuli from the pilot experiment.

Method

Stimuli.   To validate the stimulus material and to compare the conceptual frameworks, stimuli were needed to represent emotion concepts from both discrete and three-dimensional models in order to do justice to both (see Mikels et al., 2005 for a selection method aimed at discrete emotion categories). Therefore, a sampling of the 360 excerpts from the pilot experiment using both models was carried out. In the first stage, all excerpts that were rated as moderately or very familiar in the pilot experiment were eliminated. To obtain both highly and moderately typical examples of discrete emotions, it was necessary to calculate the *typicality* (T) of the target emotion for each excerpt (i). This was done by subtracting the mean of the excerpt's non-target emotion rating ($N\bar{E}_i$) and the standard deviation of its target emotion rating (SE_i) from the mean of the target emotion rating ($\bar{E}_i$)

$$T_i = \bar{E}_i - SE_i - N\bar{E}_i$$

Highest typicality values for excerpts occurred when they were highly and consistently rated on the target emotion, and *not* attributed to other emotion categories. For example, two excerpts scoring 6 on sadness had a different typicality value if they differed in consistency and their scores for other emotion categories. If a sadness rating had a mean of 6 with a SD of 2, and a mean of 1 in other emotion categories, it would result in a typicality of 3 (6–2-1). However a mean sadness of 6 with more deviation, and therefore less consistency (SD 3), together with higher attributions to other categories (1 in fear, anger, surprise, and happiness but 5 in tenderness, which is a $N\bar{E}_i$ of 1.8) would have a typicality of only 1.2 (6–3-1.8). The excerpts were thus ranked according to their typicality values for each emotion. From these ranked lists, the top five examples were chosen as best examples of each discrete emotion (happiness, sadness, tenderness, anger and fear), called high examples hereafter. Five moderate examples were taken from the ranked positions of 51 to 55 of each similarly ranked list. This yielded a total of 50 examples for all the discrete emotions ([5 high + 5 moderate] × 5 categories). Surprise was not incorporated into this experiment due to the low consistency and recognition it received in the pilot experiment.

For the dimensional model, another scheme for obtaining representative examples from the pilot experiment was adopted. Each dimension was sampled at points along its axis whilst the other two dimensions were kept constant. The purpose of this was to maximize the variance according to the dimension in question, although this was not always entirely possible due to the collinearity of the dimensions. The axis of each dimension was then split into four percentiles as follows: extreme low (< 10%), moderate low (20%–40%), moderate high (60%–80%) and extreme high (> 90%). From each of these percentiles five excerpts were taken, which meant a total of 20 for each dimension, exhibiting a similar range of typicality. During this process, the other two dimensions were kept in control by minimizing the error distance so that the excerpts were chosen as close to the target dimensional axis as possible. In this way 60 audio clips (4 × 5 × 3) were picked to cover the essential variance of the three-dimensional affect space. Again, the purpose of choosing examples that were only moderate examples of the

target emotion concept was to increase the coverage of the stimulus space, and to compare the effectiveness of each model at rating ambiguous excerpts. In the case of the dimensional model, we also wanted to provide stimulus material which could be used to test whether the conceptual framework does actually operate in a dimensional fashion which is otherwise hard to do if only best examples of the bipolar extremes are used. A list of the final stimuli may be found in the Appendix, and the details are also documented online[2] (50 discrete + 60 dimensional = 110 examples in total).

Participants. The participants for the main experiment were 116 university students aged 18–42 years (mean 24.7, SD 3.75, 68% females and 32% males). Forty-eight percent of the participants did not play any instrument and were not musically trained, 41% had experience of playing an instrument or some level of musical training, and 11% were in between with music as a hobby for less than three years. The participants received cinema tickets in return for their participation, and a number of individual variables were collected from each of them. First, a survey of their musical taste was made using a localized version of the STOMP questionnaire (Rentfrow & Gosling, 2003). Second, their current mood at the time of taking the test was evaluated using the POMS-A questionnaire (Terry, Lane, Lane, & Keohane, 1999). Third, an assessment was made of the participants' personality traits using a 44-item personality measure known as 'The Big Five Inventory' (John & Srivastava, 1999). And finally a short questionnaire was given out to gather information about the participants' film genre preferences, musical training, and any hearing problems.

Procedure. The experiment was divided into two blocks: block A with the 50 discrete emotion excerpts and block B with the 60 dimensional model excerpts. The order of the blocks was counterbalanced across the participants (67 participants did block A first followed by block B), and the order of the examples within blocks was individually randomized. The participants were instructed to rate the emotions represented by the music excerpts (perceived emotions), and the difference between perceived and induced emotions was explained to them. While one group of participants were rating excerpts in block A on a scale of 1–9 for each discrete emotion, the other group was rating excerpts on bipolar scales of 1–9 for each of the three axes of the dimensional model. Then the blocks were switched for each group. This meant that the experiment used both models to rate the emotions in all 110 excerpts without taking an unreasonably long time. The total duration of the experiment was between 50 and 60 minutes, depending on the participant's rating speed.

Participants were also asked to mark how much they liked each example (with a preference rating) and how beautiful they considered each example to be (with a beauty rating). In both cases this was on a scale from 1 to 9. These additional measures were added to clarify the role of valence because valence and preference have been shown to be separate constructs (Schubert, 2007). For example, one can be fond of harsh and rough sounding music despite the fact that most people associate those qualities with negative valence. Ratings of preference and beauty would also briefly allow the exploration of the relation between sadness and valence in this context.

Before the actual experiment, a short practice session was carried out by each participant to become familiar with the interface, likert scales, and type of music used. The participants also had the possibility to ask questions about the task before the start of the experiment.

Apparatus. The listening experiments were conducted in a soundproof room. To gather the emotion ratings, a special patch was designed in *Pure Data* graphical programming

environment (Puckette, 1996), running on Mac OS X. The patch enabled the participants to move from one excerpt to the next at their own pace and to repeat an excerpt if needed. Participants listened to the excerpts through studio quality headphones (AKG K141 Studio), and were able to adjust the sound volume according to their own preferences.

Results

To rule out extreme mood states that might affect the participants' emotion ratings, POMS-A ratings were aggregated and the distance from the mean rating (1.87, SD = 0.47 on a scale of 1–5) was calculated for each participant. One participant whose score was more than three SDs above the mean was removed from the analysis as that person's current mood appeared considerably pessimistic, tired and negative. The intersubject correlations were used to identify possible outliers and anyone who scored more than three SDs from the mean intersubject correlation was removed from the dataset. This resulted in the removal of five participants, leaving the total of 110 that was eventually used. Cronbach's alpha was used to assess the inter-rater reliability across both experimental blocks. All emotion ratings received alphas above .99 and only the preference ratings were slightly lower (.94), probably since personal opinions are prone to vary across individuals. Subsequently the data was pooled together for further analyses.

We reviewed the contribution of individual variables (personality, musical preferences, film genre preferences, and musical training) to the emotion ratings by correlating the individual ratings with the background variables. Only a few statistically significant correlations emerged: the personality trait 'openness to experience' (John & Srivastava, 1999) appeared related to increased ratings of anger ($r = 0.25$, $p < .05$) and valence ($r = 0.33$, $p < .05$), and 'extroversion' seemed related to decreased ratings of tension ($r = -0.38$, $p < .01$). These traits may therefore indicate different rating strategies. 'Negative mood' (POMS-A; Terry et al., 1999) was, perhaps unsurprisingly, related to increased ratings of sadness ($r = 0.29$, $p < .05$). These initial observations are potentially interesting and are known to influence emotional evaluations (Kreutz et al., 2008; Rusting, 1998), but the figures indicate that their role is only moderate at best, and thus these factors will be left unexplored at this stage.

Discrimination of emotion categories and levels. An examination of the discrete emotion rating was carried out next using a mixed design repeated measures ANOVA for groups of excerpts representing each target emotion. The five emotion concepts (anger, fear, sadness, happiness, and tenderness) as the within subjects variable. The two levels (high and moderate) at which excerpts conveyed an emotion (i.e., high and moderate examples) provided the within group variable. Taken together this gave significant main effects for all target emotion concepts, and a significant main effect for levels in one target emotion. There was also significant interaction effects between these two factors in four of the target emotions (see Table 2 for effect sizes and *p* values). In other words, the effect sizes were robust for concept (between 0.79 and 0.83; see Table 2), but negligible for emotion levels (0.000–0.006). And most of the interactions between concept and level were within each target emotion were significant. These analyses were later followed by post hoc tests, in which *p* values were adjusted using the Holm-Sidak procedure to avoid the effects of multiple comparison tests (Ludbrook, 1998). These analyses revealed that in the high examples the target emotion was never confused with other emotion concepts, but the moderate examples exhibited confusions between one or two other concepts of emotion (see Table 2). Figure 2 clearly shows this pattern. For example, moderate examples of anger are indistinguishable from fear and sadness, and moderate examples of tenderness could easily be
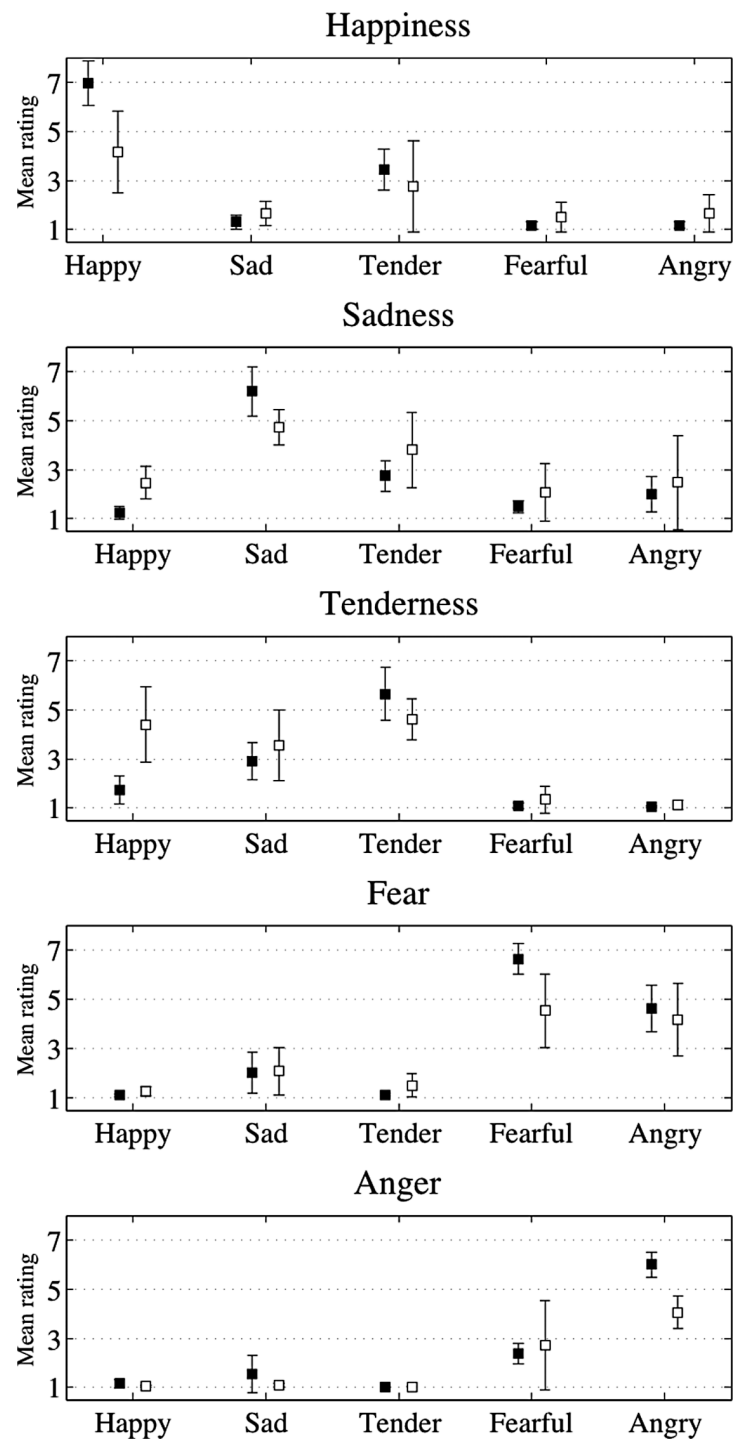
Figure 2. Mean ratings and 95% confidence intervals for five discrete emotions and 50 excerpts representing these target emotions (black markers for high examples and white for moderate examples).

Table 2. Mixed-design repeated measures ANOVA results for groups of excerpts representing each target emotion (η^2 for effect sizes). Post-hoc tests display target emotions that do not reliably differ in means (using Holm-Sidak adjusted values for $p < 0.05$)

Type of excerpt	Category	Level	Interaction	Post-hoc (moderate)
Happy	0.70***	0.003**	0.21***	S, T
Sad	0.83***	0.003	0.05*	T
Tender	0.78***	0.000	0.05	H, S
Fearful	0.79***	0.001	0.08**	A
Angry	0.79***	0.006	0.06**	F, S
Pos. valence	0.52***	0.01	0.08*	En, Te
Neg. valence	0.70***	0.000	0.08*	En, Te
Pos. energy	0.44***	0.11*	0.06	Va, Te
Neg. energy	0.54***	0.01	0.11*	Te
Pos. tension	0.81***	0.02**	0.07*	En
Neg. tension	0.58***	0.09*	0.03	En

*Notes: $p < .05$; **$p < .01$; ***$p < .001$; df = 4,49 for basic emotion concepts, df = 2,29 for dimensional concepts.
S = Sadness, T = Tenderness, H = Happiness, A = Anger, F = Fear, Va = Valence, En = Energy, Te = Tension.*

confused with happiness and sadness (the precise confusion pattern is given in the last column of Table 2). The effect sizes and patterns of confusion are similar to the ones observed in the pilot experiment, and confirm prior research (e.g., Gabrielsson & Juslin, 1996).

Similarly for dimensional examples, mixed design repeated measures ANOVAs were carried out to investigate how ratings for each group of excerpts representing different target emotions varied between the dimensions and levels within that dimension. This was done in the same way as with the discrete emotion examples. Table 2 and Figure 3 display the results of this analysis, which yielded significant main effects for the concept and significant effect for level in three of the dimension extremes. There was also a significant interaction effect between concept and level for four of the dimension extremes. The effect sizes were comparable to those obtained with discrete emotions (between 0.44 and 0.81 in the concept effect sizes and between 0.00 and 0.11 for the level effect sizes). As with the discrete emotion examples, Holm-Sidak adjusted post hoc analyses were also performed (results shown in the last column of Table 2). However, the results have to be interpreted in a different manner since the ratings given with the three-dimensional model only indicate the excerpt's location in the dimensional space.

Reliability of discrete and dimensional model ratings. Although the overall reliability of the dimensional and basic emotion model scales was earlier found to be high, a closer scrutiny may reveal interesting differences, especially related to numerous ambiguous (moderate) examples. To compare the applicability of the discrete and dimensional models in the rating of emotionally ambiguous as well as highly characteristic excerpts, the raw ratings for the high examples of target discrete emotions were compared with the raw ratings for moderate examples of target emotions. This was done using inter-rater agreements across the excerpts (Cronbach's alpha). The agreements within the high and moderate examples of target emotions were different, $\alpha = .74$ for high and $\alpha = .49$ for moderate examples. To evaluate the statistical significance of this difference, we used bootstrapping, in which the confidence intervals (CI) of the statistics in question were obtained for 1,000 bootstrapped calculations (Davison & Hinkley, 1997). On the basis of these confidence intervals, the means were found to be different at $p < .001$ level (mean
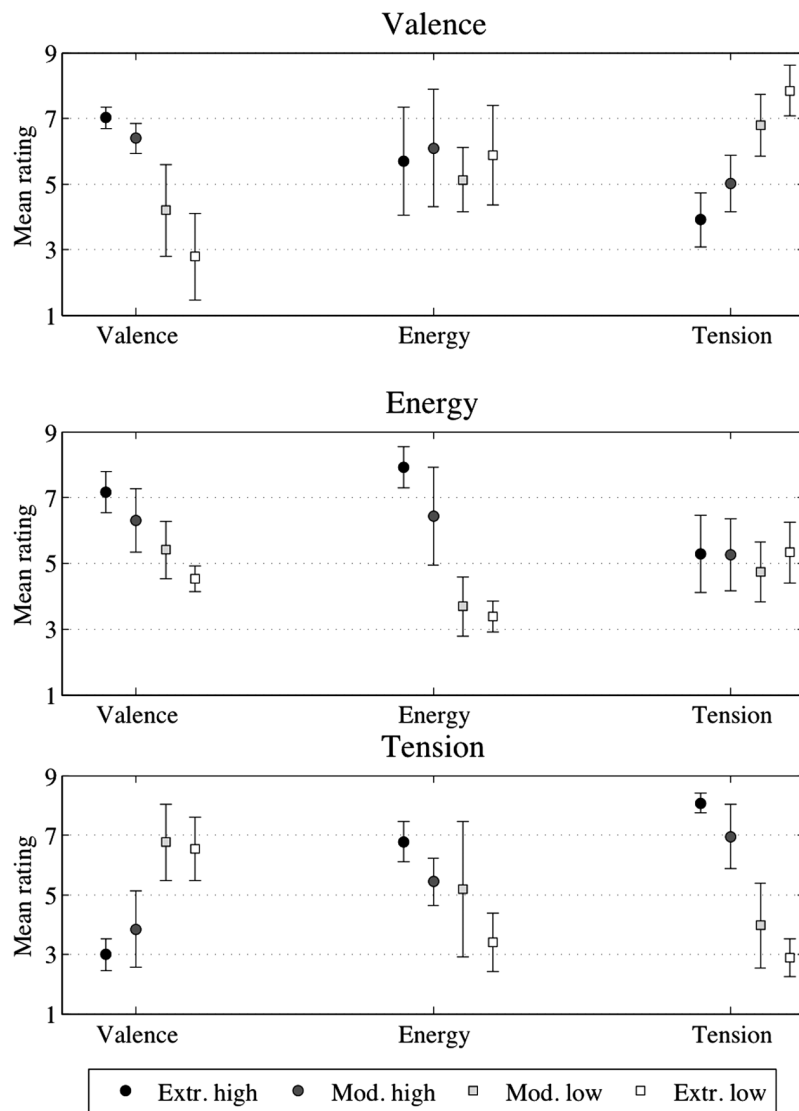
Figure 3. Mean ratings and 95% confidence intervals for 60 examples representing 6 target emotions and 2 levels of the three dimensions.

CI 99.9% for high examples was .74 .69 .78, and for moderate examples it was .49 .29 .61). Similar analysis was performed for the raw ratings of the three dimensions, and the resulting reliability estimates were also significantly different at $p < .001$ level (mean CI 99.9% for high examples was .95 .94 .96 and for moderate examples it was .77 .72 .82). Although the reliabilities for both models show significantly lower overall reliabilities for moderate excerpts, presumable as they may be a mixture of several emotions, those for moderate excerpts from the dimensional model are at the same level as the reliabilities for the high excerpts of the basic emotion model, implying a higher overall reliability in the dimensional model ratings. However, perhaps the selection of the basic

emotion excerpts is such that they have lower overall reliability in general compared with the dimensional model excerpts and the difference is not in the measurement model. To evaluate this argument, we replicated the same analysis but reversed the concepts (high examples of target emotions selected using the basic emotion model were rated with the dimensional model). The reliabilities and their confidence intervals showed again a similar pattern. For basic emotion model ratings of the dimensional targets, α was .87 $_{.85}$ $_{.89}$ for characteristic examples, and .70 $_{.66}$ $_{.74}$ for ambiguous examples. For dimensional model ratings of the basic emotion model targets, the alphas were higher (.94 $_{.93}$ $_{.95}$ for high examples, and .85 $_{.81}$ $_{.87}$ for moderate examples). In both cases, significant differences in the reliabilities between the high and moderate excerpts were found but the overall reliability of the dimensional ratings was again higher. So the dimensional model provides somewhat higher inter-rater consistency no matter which way the musical excerpts have been chosen. Whether consistency is a crucial detail in assessing the adequacy of these models is another question. High consistency could for example also indicate that the measurement scale is trivial and thus offers little insight into the actual emotion process.

The applicability of the discrete emotions to the appropriate emotion prototype areas (Russell & Feldman Barrett, 1999) is visualized in Figure 4, where both the strength and the variation of all discrete target emotions are shown as densities in the valence-energy space. Marked in the plot are the target centroids of the chosen high and moderate examples of the five discrete emotions, which indeed lie in the approximate (attractor) areas of high ratings for these emotions. Moreover, the variation demonstrates that the high examples are mostly located within areas of high agreement (sadness and tenderness are the notable exceptions), and the most moderate examples are located in less well-defined areas. In contrast, the deviations in the ratings of the dimensional model are lower and spatially more uniformly distributed, and do not have such ill-defined areas around the attractors of the emotion categories. This is also demonstrated by the means of the ANOVA above. The practical implication of this is that, whereas both models may be used to adequately describe emotional excerpts representing clear examples of discrete emotions, the strength of the dimensional model lies in its ability to describe such emotional examples that lie outside these discrete attractor areas. The utilization of this asymmetry between the models would allow us to explore how a hybrid model of emotions might manifest itself in music (Russell, 2003), and is of consequence to clinically oriented studies that are interested in the patients' processing of emotionally ambiguous examples (e.g., Bouhuys, Bloem & Groothuis, 1995; Cavanagh, & Geisler, 2006).

In the end participants were able to recognize the target emotions represented by the high examples consistently, and in moderate examples the target emotion was confused with at least one other emotion. This has also been observed in previous research, and could be interpreted more generally as fuzziness in the definition of the emotion categories (for examples, see Dailey, Cottrell, Padgett, & Adolphs, 2002; Russell & Fehr, 1994). In the case of the emotion dimensions, the results have to be viewed from a different angle, because here the 'emotion targets' were actually the six bipolar extremes of three dimensions. Instead of concentrating on the confusion between different dimensions, we should focus our attention towards any possible confusion between the bipolar extremes of a given dimension. For example, valence and tension ratings for excerpts representing moderate positive energy did not differ (see Table 2). To set these observations into a wider context, it is necessary to examine the whole pattern of correlations between the emotion concepts.

Patterns of correlations between the emotion concepts.   Our intention was to explore how the two dominant conceptual frameworks for emotions in music (dimensional and discrete) can be used
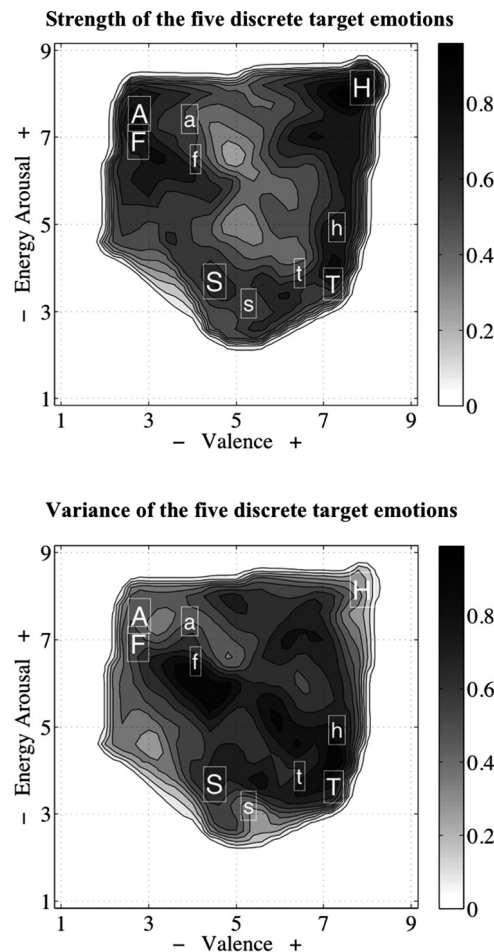
Strength of the five discrete target emotions

Variance of the five discrete target emotions

Figure 4. The intensity of each discrete target emotion is shown via accumulated density distribution of the emotion ratings for each category (upper panel). To define each of the five discrete target emotions, the examples that received ratings above the upper 50% percentile of the appropriate target ratings have been selected. The labels refer to the centroids in the valence–arousal space defined by the five discrete target emotion examples (A = anger, F = fear, H = Happiness, T = tenderness, and S = sadness, capitals refer to high examples and small letters moderate examples). The lower panel displays a similar projection of discrete emotion areas within the valence – arousal space, but the gradient indicates the amount of variation (normalized standard distribution of the ratings) for each emotion category. The centroids of the high examples are located clearly on well-defined areas (high intensity and low deviation) whereas the moderate examples are on less clearly defined areas (lower intensity and higher variation).

to describe perceived emotions in music. We also set out to clarify what type of dimensional model would be the most appropriate for such studies. To make this comparison we used a variety of correlational techniques. To begin with, mean ratings for both sets of stimuli were visualized across the three dimensions (Figure 5) and the ratings of the discrete emotion categories were indicated with appropriately sized markers (the greater the marker size, the greater the
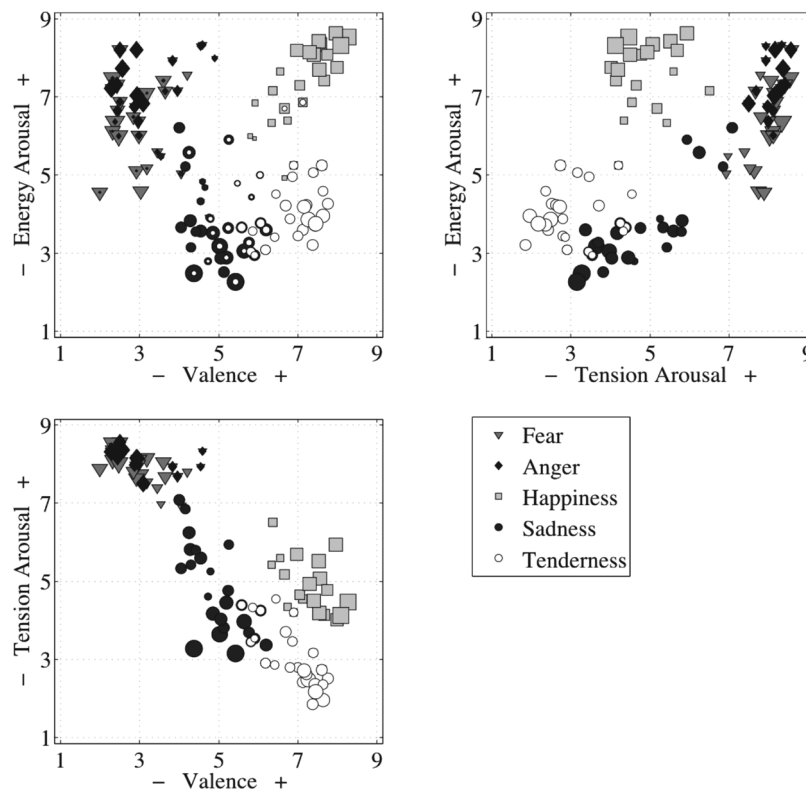
Figure 5. Mean ratings of three dimensions and discrete emotions for all excerpts (N = 110). The marker types represent the target emotion categories and the sizes indicate the mean target emotion rating for each excerpt.

mean rating for the discrete emotion indicated by the marker type). The scatterplots in Figure 5 show a highly collinear structure between valence and tension. However, the remaining dimensions display less evident correlational structures. The discrete emotion categories – as represented by marker types and sizes – suggest a clear separation of happy excerpts from the rest of the discrete emotions. Also tenderness and sadness stand out as distinct areas within the dimensional space. None of these three categories overlap with anger and fear, even if these two do so between themselves in the dimensional structure (see Kreutz et al., 2008; Schubert, 1999; Vieillard et al., 2008). As we remember from the analysis of discrete emotion ratings, this overlap was also evident in the ratings of anger and fear, and thus is not a feature particular to the dimensional emotion model alone.

Correlations between the different emotion concepts, together with preference and beauty, are shown in Table 3. Fear and anger can be observed to correlate highly with each other ($r = .69$, $p < .001$), which suggests that these two emotion concepts might not be easily distinguishable in the context of music (Juslin, 2000; Kallinen & Ravaja, 2006). Interestingly, tenderness received higher correlations with sadness ($r = .36$, $p < .001$) than with happiness ($r = .15$), as traditionally tenderness has been associated with positive emotions in general (Juslin, 2001, p. 315). Another noteworthy observation is that valence and sadness did not

correlate with each other. This is in line with results obtained by Bigand et al. (2005) and Kreutz et al. (2008), who both discovered that sad music was not systematically associated with negative valence. Although sadness is generally considered to be an unpleasant emotion, the classification is not as straightforward in the context of music. For instance Schellenberg and colleagues (2008) found that, in some instances, sad music was liked as much as happy music. It seems that in music-mediated emotions, happiness and sadness do not represent the opposite extremes of valence: although happiness had a strong positive correlation with valence ($r = .80$, $p < .001$), sadness and valence did not correlate ($r = −.03$). Sad music is often considered beautiful, and therefore it may be difficult to perceive sadness in music as unpleasant. Schubert (1996) has offered a theoretical solution to this dilemma using a neural inspired associative network model, in which negative emotions in an aesthetic context may activate enjoyment. In fact, it has even been reported that sad music activates neural networks involved in biological reward (Blood & Zatorre, 2001). For example, sadness correlated with preference ($r = .38$, $p < .001$) and beauty ($r = .59$, $p < .001$) significantly more highly than happiness ($r = .22$, $p < .05$; $r = .16$). However, tenderness and valence correlated with preference ($r = .58$, $p < .001$; $r = .56$, $p < .001$) and beauty ($r = .77$, $p < .001$; $r = .61$, $p < .001$) even more highly than sadness. Lastly, the ratings given by the participants in the main experiment were to a great extent similar to the ratings given by the small group of experts in the pilot (the final row in Table 3).

When comparing our results to earlier studies of the three-dimensional model (conducted in non-music contexts) we find there are certain differences, particularly when the correlations between separate dimensions of the models are taken into account. For example, in a study based on current mood ratings, Schimmack and Grob (2000) found a strong positive correlation between energy and valence ($r = .49$), a strong negative correlation between valence and tension ($r = −.70$), and a moderate negative correlation between tension and energy ($r = −.33$). In our study, energy and valence did not correlate with each other ($r = −.08$), valence and tension had a very strong negative correlation ($r = −.83$), and tension and energy had a strong positive correlation ($r = .57$). This might be due to the different qualities of music-mediated emotions compared to mood or everyday emotions, but there are no other points of comparison from the field of music-mediated emotions. Another possible cause for the difference is that, despite our efforts of sampling the three-dimensional space in a systematic manner, the chosen sound examples may have represented the geometric space in a different way to how it was

Table 3. Correlations between the concepts ($N = 110$).

	Happiness	Sadness	Tenderness	Fear	Anger	Valence	Energy	Tension	Pref.
Sadness	−.48***								
Tenderness	.15	.36***							
Fear	−.61***	−.28**	−.67***						
Anger	−.41***	−.31**	−.58***	.69***					
Valence	.80***	−.03	.63***	−.91***	−.71***				
Energy	.44***	−.79***	−.64***	.28***	.47***	−.08			
Tension	−.42***	−.38***	−.87***	.87***	.75***	−.83***	.57***		
Pref.	.22*	.38***	.58***	−.63***	−.37***	.56***	−.31***	−.63***	
Beauty	.16	.59***	.77***	−.73***	−.56***	.61***	−.58***	−.81***	.87***
Experts†	.94***	.88***	.90***	.93***	.92***	.86***	.90***	.94***	

*Notes:*** $p < .001$; ** $p < .01$; * $p < .05$; †Correlation with the ratings of the experts from the pilot experiment for the same 110 excerpts.

represented in the previous studies. For instance, we have relatively few examples of excerpts representing low tension and negative valence, as the correlations suggest that these variables mostly co-varied. At this point it is difficult to conclude whether this phenomenon is specific to music or the particular set of stimuli we used, as we do not have alternative samples of stimuli at our disposal. Nevertheless these correlations, as well as the figures, hint at a possible reduction of the three-dimensional model. We will address this issue after first making a direct comparison of the discrete and dimensional approaches.

Correspondence between discrete and dimensional models of emotions.   To assess the compatibility of the two main conceptual frameworks for emotions in music, we adopted two correlational techniques, canonical correlation and regression. In the canonical correlation, the interdependency of the two frameworks could be measured in a single analysis. This analysis provided three canonical variates. The first canonical correlation was .99, the second .94 and the third was .57, and the model with the three canonical correlations included was highly significant, $\chi^2 (15) = 634.05$, $p < .0001$. The first three pairs of canonical variates accounted for a significant relationship between the two sets of variables. Data on the canonical variates are displayed in Table 4. Indicated in the table are correlations between the variables and canonical variates, within-set variance accounted for by the canonical correlations (percent of variance), redundancies and canonical correlations. Total percentage of variance indicates that the third canonical variate was minimally related to the two sets of variables and therefore the interpretation of the third pair is questionable, even though this variate was also statistically significant ($\chi^2 = 41.80$, $p < .001$). The interpretation of the first canonical variate can be drawn from the correlations, indicating that the variate may be labelled as valence (inverted). This is because tension (.88), fear (.94) and anger (.75) as well as valence (−.98) are projected with high loadings onto the first variate. The second canonical variate could then be labelled as activity (inverted). This is because energy (−.92), happiness (−.64), and sadness (.85) receive the highest correlations with the second canonical variate. The interpretation of the third canonical variate is precarious due to the low percent of variance that can be explained (0–5%). This analysis as a whole puts forward the notion that the two conceptual frameworks are largely similar and the minimal description of this mapping might reasonably have a two-dimensional structure.

In the second conceptual comparison, we employed regression to predict the dimensional ratings from the discrete ratings and vice-versa. This technique has the advantage of providing a well-known measure of fit (R^2). As all 110 music examples were rated using both conceptual frameworks, such a comparison was possible. Knowing that relatively high correlations exist between the emotion ratings, collinearity of predictors was evaluated using a variance inflation factor (VIF) for each set of predictors (basic emotion and dimensional models). For basic emotion ratings, all VIF values remained lower than the suggested threshold value for collinearity (10; see Cohen, Cohen, West, & Aiken, 2003, p. 423) but for valence, energy arousal and tension arousal, VIF values indicated high collinearity (11.4, 5.3, 16.9). For this reason, the regression estimates for each basic emotion concept with the dimensional model ratings as predictors was performed using ridge regression. This technique is less influenced by collinearity due to the inclusion of constant variance parameter (λ), which attenuates the influence of collinearity in the calculation of the least squares optimization in regression (Cohen et al., 2003). In this case, optimal λ was set at 50 (three predictors) and 100 (two predictors), which was established by 10-fold-over cross-validation with this data. The results – displayed in Table 5 – demonstrate that the discrete emotion model can more accurately explain the results obtained with the three-dimensional model than vice-versa. This may partly be due to the higher amount of explanatory

Table 4. Correlations, canonical correlations percents of variance and redundancies between dimensional and categorical ratings and their corresponding canonical variates

	1st Can. Var.	2nd Can. Var.	3rd Can. Var.
Dimensional			
Valence	−.98	−.11	.04
Energy	.20	−.92	−.03
Tension	.88	−.37	−.13
Percent of variance	.59	.33	.00
Redundancy	.57	.29	.00
Categorical			
Happiness	−.75	.64	.05
Sadness	−.07	.85	.05
Tenderness	−.71	.53	.33
Fear	.94	−.11	.94
Anger	.75	−.37	.42
Percent of variance	.51	.31	.06
Redundancy	.49	.30	.05
Canonical correlation	.99	.94	.57

variables (five) but also the fact that discrete emotions are an easier concept for the general public to understand than emotion *dimensions*. Nevertheless, the difference between the mean prediction rates of the models (displayed in Table 5) is not large (17%) and this considerable degree of overlap between the conceptual frameworks is noteworthy considering that neurological evidence has suggested that separate processes might be involved (Dellacherie et al., 2008), and that it is only relatively recently that both models have started to occur within the same study.

To examine the validity of the three-dimensional model, the coefficients of determination for it were also compared with the circumplex model (Russell, 1980) and the multidimensional model of activation (Thayer, 1989). The results suggest that these two-dimensional models can explain the results obtained with the discrete emotion model virtually as accurately as the three-dimensional model, with the exception of anger (see Table 5). The differences between the prediction rates of the three alternative dimensional models were evaluated using a comparison of the difference between two multiple correlations (Steiger, 1980), which involves transforming the multiple correlations of the predicted models into Z scores and adjusting for mutual correlation and sample size (Tabachnick & Fidell, 2001, p. 146). This analysis yielded significant differences between all the different prediction rates of the models (see Table 5). It is worth pointing out that, in comparison to other emotion categories, sadness was explained equally modestly (R^2 = .63) by all dimensional models. This may reflect the participants' difficulty with rating the valence of sad music, as previously mentioned. Despite this irregularity, these analyses suggest fairly high mutual correspondence between the two conceptual frameworks and stimulus sets, and further suggest that the common denominator between these frameworks might be two-dimensional.

A requisite number of dimensions. The previous summary of the correlations between the emotion ratings suggested high collinearity within the three-dimensional model (the overlap between valence and tension). To examine whether the three-dimensional model could be reduced to two dimensions, the independence of the three dimensions needed to be scrutinized.

Table 5. Regression summary of dimensions and discrete emotions explaining discrete and dimensional emotion ratings ($N = 110$)

	R^2 (β)		
	3D	2D (Russell)	2D (Thayer)
Dimensions as predictors (valence, energy, tension)			
Happiness	.89 ($V_{0.93}$, $E_{0.79}$, $T_{-0.35}$)*	.89 ($V_{0.85}$, $E_{0.49}$)*	.86 ($E_{0.64}$, $T_{-0.62}$)
Sadness	.63 ($V_{-0.20}$, $E_{-0.84}$, $T_{-0.22}$)	.63 ($V_{-0.05}$, $E_{-0.69}$)	.60 ($E_{-0.65}$, $T_{-0.13}$)
Tenderness	.77 ($V_{0.33}$, $E_{-0.45}$, $T_{-0.58}$)**	.74 ($V_{0.50}$, $E_{-0.51}$)	.77 ($E_{-0.34}$, $T_{-0.61}$)**
Fear	.87 ($V_{-0.83}$, $E_{0.07}$, $T_{0.63}$)**	.87 ($V_{-0.90}$, $E_{0.24}$)**	.74 ($E_{0.03}$, $T_{0.85}$)
Anger	.64 ($V_{-0.52}$, $E_{0.32}$, $T_{0.35}$)**	.68 ($V_{-0.55}$, $E_{0.35}$)**	.54 ($E_{0.22}$, $T_{0.52}$)
Mean R^2	.76	.76	.70
Discrete emotions as predictors (happiness, sadness, tenderness, fear, anger)			
Valence	.97 ($H_{0.35}$, $S_{-.11}$, $T_{0.20}$, $F_{-0.50}$, $A_{-0.14}$)		
Energy	.88 ($H_{0.47}$, $S_{-0.32}$, $T_{-0.42}$, $F_{-0.05}$, $A_{0.36}$)		
Tension	.93 ($H_{-0.29}$, $S_{-0.23}$, $T_{-0.55}$, $F_{0.18}$, $A_{0.12}$)		
Mean R^2	.93		

Notes: β = Standardized beta coefficients. For predicting basic emotions with emotion dimensions, ridge regression was used ($\lambda = 50$ for 3-dimensions, and $\lambda = 100$ for 2-dimensions). For all models, F tests were significant at $p < .001$. Asterisks denote significant difference between the regression models within each categorical emotion (** at $p < 0.01$ level; * at $p < 0.05$ level).

For instance, Roberts and Wedell (1994) have aptly demonstrated that the amount of dimensions needed to explain common mood terms is influenced by stimulus density. In their study, the common two-dimensional solution (valence and arousal) was not sufficient when a core set of mood terms were supplemented with terms representing variants of anger and fear. In the present study, two different types of reductions could be attempted: energy and tension dimensions could be collapsed into one arousal dimension (Russell & Feldman Barrett, 1999) or energy and valence, and tension and valence could both be collapsed into separate dimensions, forming another two-dimensional construct (Thayer, 1989). The plausibility of these more parsimonious models could be investigated by looking at the partial correlations between the dimensions or by testing these explicitly with separate structural equation models. However, when the intercorrelation between the model predictors is high (> .85), it is known to pose severe problems for such iterative models that are based on a covariance matrix (Kline, 2004). Therefore we resorted to a simpler strategy and looked at the partial correlations.

To investigate reducing three dimensions into the traditional two dimensions of valence and arousal (Russell, 1980), we checked whether tension and energy, which correlate positively (.57), could be collapsed into a single arousal dimension by partialling out the contribution of the valence dimension. This analysis yielded a partial correlation of $r_{te.v} = .90$ ($p < .001$), indicating a considerable overlap between the concepts. As valence and energy did not correlate significantly (−.08), we consequently received support for the traditional two-dimensional model. Note that this is in contrast to the results obtained by Schimmack and Reizenstein (2002), who used structural equation modelling to test the independence of the energy and tension dimensions by controlling for valence, and they found no correlation between the residuals of the energy and tension dimensions.

The other two-dimensional model (Thayer, 1989) casts valence into two separate dimensions that consist of energy and tension. As we know from the first order correlations, valence

and tension correlate highly ($r = -.83$), but if we control for the contribution of energy then this high correlation might be revealed to be spurious. Nevertheless, valence and tension correlate even more highly when the energy ratings are partialed out ($r_{vt.e} = -.95$), suggesting that at least the high collinearity is not affected by the energy dimension. Partialling out the tension from the correlation between energy and valence also makes them correlate even more highly ($r_{ev.t} = .85$), lending further support for the possibility of dimensional reduction. In sum, both theoretically derived ways of reducing the three dimensions into two are supported by these analyses although the difference between the two reductions is not discernible.

A requisite number of emotion categories.   We may also ask whether the ratings of the five discrete emotions contain significant overlap. In comparing conceptual frameworks using regression, we already observed that three to five discrete emotions were necessary to explain over 90% of the ratings in three dimensions. We also noticed how fear and anger seemed to overlap ($r = .69$) when looking at the first order correlations. Here we looked again at the partial correlations by controlling the contribution of the other three discrete emotions while examining the correlation of each pair of discrete emotions. Table 6 displays the results of such analysis. The most striking partial correlations can be seen between happiness, sadness, and fear, all negative and highly significant. This implies that if we removed the ratings of happiness from the data, we would still be able to deduce that happy examples are those which are rated low on sadness, anger, and fear. The case of fear is more complicated, however, as it correlates with sadness and tenderness and so its removal could not be entirely constructed from the three other remaining discrete emotions. Interestingly, fear and anger do not correlate with each other when the other discrete emotion categories are partialled out. It seems that the contribution of happiness ($r_{fa.h} = .60$, $p < .001$) and sadness ($r_{fa.hs} = .18$, $p = ns$) is enough to create this effect. Thus, the overlap between happy and sad examples indicates that the real simplification of the discrete emotion model may be in terms of the valence dimension. From the previous canonical correlation and regression analysis we already know that this is a viable way of reducing the number of variables in question. Nevertheless, the issue needs to ultimately be considered in a larger context where the purpose of the measurement model can be taken into account.

Discussion

The results of the experiment suggest that the three-dimensional model of emotions may be collapsed into a two-dimensional one when applied to music. The support for this interpretation comes from (1) canonical correlations that highlighted two canonical variates which could account for the correspondence between the discrete emotion and dimensional models, (2) regression analysis which demonstrated that the ratings of discrete emotions may be recovered to a large extent by a two dimensional model ($\approx 80\%$) and vice versa ($\approx 90\%$), and (3) analysis of partial correlations, which emphasized the high correlational nature of valence and tension. Nevertheless, the two possible formulations of the two-dimensional model could not be clearly ranked using these analyses, although the canonical correlations indicated that the version by Russell (1980; valence and arousal) was somewhat more appropriate. Also, the regression approach suggested that the version by Russell was slightly better in accounting ratings of discrete emotions (mean $R^2 = .76$), than the alternative version by Thayer (1989; mean $R^2 = .73$).

These results are in contrast to the ones obtained by proponents of the three-dimensional model (Schimmack & Grob, 2000; Schimmack & Reisenzein, 2002), though we must emphasize the main differences between the design of our study and theirs. Schimmack and Reisenzein

Table 6. Partial correlations between the ratings of basic emotions ($N = 110$)

	Happiness	Sadness	Tender	Fear
Sadness	−0.88***			
Tenderness	−0.50***	−0.34***		
Fear	−0.84***	−0.73***	−0.63***	
Anger	−0.39***	−0.39***	−0.34***	−0.08

Note: The contribution of all other discrete emotions has been partialled out except the two (row and column) used in the comparison. *** $p < .001$.

(2002) used a questionnaire study with a large number of questions (18) that covered the affect dimensions with several questions (three for each polar extreme). This allowed them to construct and test latent variables from the separately observed variables using structural equation modelling. Also the ratings of their study were based on current mood, and there was no stimulus or manipulation of mood. In our study the ratings were given to emotions that the participants thought the excerpts conveyed, and the excerpts were selected to portray discrete emotions and polar extremes of the dimensional model. Because of this, the observed emotion structure reflects more directly the stimulus structure. It should also be noted that, despite our careful attempts to control the dimensions when selecting excerpts for the experiment based on the results of the pilot experiment, valence and tension were already correlated. Therefore it is difficult to estimate whether it is even possible to separate these dimensions in musical examples. In other words, the question remains as to whether there is an abundant number of musical pieces that could be highly tense and highly positively valenced at the same time. This is therefore something that needs to be studied further.

General discussion

The work presented in this article aimed to systematically compare distinct models of emotion. Although a small number of previous studies exist where discrete and dimensional data have been collected (Gosselin et al., 2006; Kreutz et al., 2008; Vieillard et al., 2008; Zentner et al., 2008), these have been incomplete with regard to the structure of emotion due to their (1) reliance on discrete emotions only, (2) focus on unambiguous exemplars, or (3) insufficient stimulus quantity. Here the set of musical stimuli was carefully selected in a large pilot study to represent emotion concepts in the dimensional as well as the discrete emotion model. Moreover, both models were represented not only by the clearest examples, but also by more moderate examples. This provided subtle nuances for emotion recognition and linear geometry for comparing the two conceptual sets using linear mapping methods.

The comparison of discrete and dimensional models yielded interesting results. Initially we thought that the discrete emotion model would lead to more consistent ratings of emotions than the dimensional model because the terminology (sad, happy, angry, etc.) is already familiar to participants due to their prevalence in the everyday language of the general public. But this was not observed in the data, as the overall consistencies between the ratings in the dimensional and discrete models did not exhibit any substantial differences. However, the discrete emotion model was clearly less reliable in rating excerpts that were ambiguous examples of an emotion category when compared with the dimensional model. This has direct implications for studies that seek to (1) explore mixed emotions (e.g., Hunter et al., 2008), (2) understand the provocative differences in neural processes between dimensional and discrete emotion ratings

(e.g., Gosselin et al., 2006), (3) examine processing biases exhibited by clinical populations to inherently ambiguous emotion stimuli (e.g., Bouhuys et al., 1995), or (4) attempt to clarify the way conceptual decisions are made within the framework of the hybrid model of emotions. For such studies, it is important to be aware of this asymmetry in the reliability of the ratings between the two models.

Despite the discrepancy in the resolution between the models, a high correspondence between the discrete and dimensional models was observed. Probably a large part of the assumed differences between the models has been caused by methodological differences. In many of the previous studies on discrete emotions (Dellacherie et al., 2008; Kallinen, 2005; Khalfa et al., 2008b), a forced-choice paradigm is used in emotion recognition. In the present study, all discrete emotions were available in the form of Likert scales, allowing more subtlety in definition than in a forced choice. In this way the methodologies used in both emotion models were similar and thus were perhaps more likely to lead to converging results.

Another way of representing the high correspondence between the two conceptual models is to consider a hybrid model of emotions (Christie & Friedman, 2004; Russell, 2003). This model uses the components of a dimensional model (valence and arousal) to explain the underlying affect space, which is mainly physiologically driven. When the changes in these core affects are interpreted consciously, however, discrete emotion terminology is used to label the emotional experiences. In this way common discrete emotions can be regarded as attractors or hot spots in the affect space. This view is entirely compatible with the results of our experiment – due to the selection of moderately and highly representative examples of discrete emotion categories – and these attractors are explicit in the figures portraying the excerpts along the three dimensions (see Figures 4 and 5). This model could also be used to characterize the main difference between the utilitarian and aesthetic emotions (Scherer, 2004). Whereas the utilitarian emotions such as fear or anger have specific connections to underlying physiology due to the adaptive function they have (in order to protect the physical integrity of the individual), aesthetic emotions do not. Indeed most of the domain-specific emotions established by Zentner et al. (2008) concern positive emotional responses and match the established functions of music as a reminder of past events (North, Hargreaves & Hargreaves, 2004), or have a direct correlate in a core affect (e.g., joyful activation).

The comparison between different versions of the dimensional model indicated that two dimensions are probably sufficient to represent perceived emotions in music. Nevertheless, this should still be studied further as the stimuli in our experiment were initially chosen to represent each of the six dimension extremes separately. Therefore the tense examples in the selection were also the ones which were often negatively valenced. A random sample of a large corpus of music that is known to manipulate emotions could be used to test for the validity of the three-dimensional model in music. Intuitively, the additional dimension of tension makes perfect sense and a great deal of the effects of music deal with patterns of tension and release (Lerdahl, 2001). Moreover, the tension dimension did actually vary independently for certain discrete emotion categories (e.g., sad examples in the lower panel of Figure 5). Tension is also one of the nine factors in the GEMS model of music-induced emotions (Zentner et al., 2008). Whether the other factors in GEMS model actually correspond with the traditionally used terms (sadness, tenderness, peacefulness, and joyful activation to name the obvious ones) is an interesting future research question. Therefore in our opinion, more research as to the specific dimensions required to represent emotions in music is warranted.

In light of the results obtained in this study, we should also investigate more thoroughly the influence of individual factors on the processing of music-mediated emotions, such as

personality and musical expertise. It would also be important to evaluate the degree of mismatch between perceived and felt emotions. That would entail changing the rating instruction towards induced emotions but should also incorporate physiological measures, in which significant steps have been recently taken (Baumgartner, Esslen, & Jancke, 2006; Roy, Mailhot, Gosselin, Paquette, & Peretz, 2008; Withvliet & Vrana, 2006). Another important issue is the limitation imposed by the musical style used. Although film music makes use of stereotypical conventions from the romantic era of classical music, as well as more recent ways to use artificial sound schemes and elements from popular music, it has to be acknowledged that the results could well be different with other genres such as pop or jazz. The genre-specificity and stimulus density effects are two related questions that warrant further systematic research in the near future. Also, a thorough dissection of the acoustical and musical features of the stimuli should be carried out in order to address some of the focal issues that have been raised by the study (mixed feelings, the role of timbre, small differences in fear and anger, etc.). Finally, non-verbal methods (such as paired similarity ratings) could give crucial insights into the requisite number of emotion dimensions and categories (Bigand et al., 2005; Vieillard et al., 2008), provided that the initial coverage of the stimuli is reasonably varied (Roberts & Wedell, 2004).

In conclusion, our study demonstrated that discrete and dimensional models of emotion produce highly compatible ratings of perceived emotions, when using a large, systematically chosen set of authentic music from film soundtracks. We also highlighted the noteworthy differences between the models that mostly relate to the constrained resolution of the discrete emotion model. In these respects, our study provides a useful point of reference for exploring the connections between the recognition, experience and physiological manifestations of emotions, as well as the individual variables that moderate all of these.

Acknowledgements

Notes

1. https://www.jyu.fi/music/coe/materials/emotion/soundtracks/
2. https://www.jyu.fi/music/coe/materials/emotion/soundtracks/

References

Altenmüller, E., Schuermann, K., Lim, V.K., & Parlitz, D. (2002). Hits to the left, flops to the right: Different emotions during listening to music are reflected in cortical lateralisation patterns. *Neuropsychologia*, *40*(13): 2242–2256.

Balkwill, L.-L., & Thompson, W. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*(1): 43–64.

Balkwill, L.-L., Thompson, W., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, *46*(4): 337–349.

Barrett, L.F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1): 20–46.

Barrett, L.F., & Wager, T.D. (2006). The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science*, *15*(2): 79–83.

Baumgartner, T., Esslen, M., & Jancke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, *60*(1): 34–43.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, *19*(8): 1113–1139.

Blood, A.J., & Zatorre, R.J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of National Academy of Sciences*, *98*(20): 11,818–11,823.

Bouhuys, A.L., Bloem, G.M., & Groothuis, T.G.G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects. *Journal of Affective Disorders*, *33*(4): 215–226.

Cacioppo, J.T., Berntson, G.G., Larsen, J.T., Poehlmann, K.M., & Ito, T.A. (2000). The psychophysiology of emotion. In R. Lewis & J.M. Haviland-Jones (Eds.), *The handbook of emotion* (2nd ed., pp. 173–191). New York: Guilford Press.

Cavanagh, J., & Geisler, M.W. (2006). Mood effects on the ERP processing of emotional intensity in faces: A P3 investigation with depressed students. *International Journal of Psychophysiology*, 60(1): 27–33.

Christie, I.C., & Friedman, B.H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology*, *51*(2): 143–153.

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Collier, G.L. (2007). Beyond valence and activity in the emotional connotations of music. *Psychology of Music*, *35*(1): 110–131.

Dailey, M., Cottrell, G., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience*, *14*(8): 1158–1173.

Davison, A.C., & Hinkley, D.V. (1997). *Bootstrap methods and their application.* Cambridge: Cambridge University Press.

Dellacherie, D., Ehrlé, N., & Samson, S. (2008). Is the neutral condition relevant to study musical emotion in patients? *Music Perception*, *25*(4): 285–294.

Ekman, P. (1992). Are there basic emotions. *Psychological Review*, *99*(3): 550–553.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M.J. Power (Eds.), *Handbook of cognition and emotion* (p. 45–60). New York: John Wiley & Sons.

Etzel, J.A., Johnsen, E.L., Dickerson, J., Tranel, D., & Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction. *International Journal of Psychophysiology*, *61*(1): 57–69.

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

Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different?. *Musicae Scientiae*, *6*(Special Issue 2001/2002): 123–147.

Gabrielsson, A., & Juslin, P.N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1): 68.

Gold, A.E., MacLeod, K.M., Deary, I.J., & Frier, B.M. (1995). Hypoglycemia-induced cognitive dysfunction in diabetes mellitus: Effect of hypoglycemia unawareness. *Physiology & Behavior*, *58*(3): 501–511.

Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, *53*(2): 91–103.

Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, C., Baulac, M., et al. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain*, *128*(3): 628–640.

Gosselin, N., Samson, S., Adolphs, R., Noulhiane, M., Roy, M., Hasboun, D., et al. (2006). Emotional responses to unpleasant music correlates with damage to the parahippocampal cortex. *Brain*, *129*(10): 2585–2592.

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, *45*, 236–244.

Hunter, P.G., Schellenberg, E.G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition & Emotion*, *22*(2): 327–352.

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.

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. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6): 1797–1813.

Juslin, P.N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P.N. Juslin & J.A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–340). Oxford: Oxford University Press.

Juslin, P.N., Friberg, A., & Bresin, R. (2002). Toward a computational model of expression in music performance: The GERM model. *Musicae Scientiae*, *Special Issue 2001–2002*, 63–122.

Juslin, P.N. & Sloboda, J.A. (Eds.) (2010). *Handbook of music and emotion.* New York: Oxford University Press.

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. (2005). Emotional ratings of music excerpts in the western art music repertoire and their self-organization in the Kohonen neural network. *Psychology of Music*, *33*(4): 373–393.

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

Khalfa, S., Peretz, I., Blondin, J.P., & Manon, R. (2002) Event-related skin conductance responses to musical emotions in humans. *Neuroscience letters*, *328*(2): 145–149.

Khalfa, S., Roy, M., Rainville, P., Dalla Bella, S., & Peretz, I. (2008a). Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *International Journal of Psychophysiology*, *68(1):* 17–26.

Khalfa, S., Delbe, C., Bigand, E., Reynaud, E., Chauvel, P., & Liégeois-Chauvel, C. (2008b). Positive and negative music recognition reveals a specialization of mesio-temporal structures in epileptic patients. *Music Perception*, *25*(4): 295–302.

Kline, R.B. (2004). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

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.

Krumhansl, C.L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, *51*(4): 336–352.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2005). *International affective picture system (IAPS): Digitized photographs, instruction manual and affective ratings*. Gainesville, FL: University of Florida.

Laukka, P., & Juslin, P.N. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion*, *31*(3): 182–191.

Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D., & Lesaffre, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research*, *34*(1): 39–67.

Ludbrook, J. (1998). Multiple comparison procedures updated. *Clinical and Experimental Pharmacology and Physiology*, *25*, 1032–1037.

Lerdahl, F. (2001). *Tonal pitch space*. New York: Oxford University Press.

Mikels, J., Fredrickson, B., Larkin, G., Lindberg, C., Maglio, S., & Reuter-Lorenz, P. (2005). Emotional category data on images from the International Affective Picture System. *Behavioral Research Methods*, *37*(4), 626–630.

Nawrot, E. (2003). The perception of emotional expression in music: Evidence from infants, children and adults. *Psychology of Music*, *31*(1): 75–92.

North, A., Hargreaves, D., & Hargreaves, J. (2004). Uses of music in everyday life. *Music Perception*, *22*(1), 41–77.

Nyklìcek, I., Thayer, J.F., & Van Doornen, L.J.P. (1997). Cardiorespiratory differentiation of musically-induced emotions. *Journal of Psychophysiology*, *11*(4): 304–321.

Posner, J., Russell, J.A., & Peterson, B.S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3): 715–734.

Puckette, M. (1996). Pure data. In *Proceedings of the International Computer Music Conference* (pp. 269–272). San Francisco, CA: International Computer Music Association.

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*(6): 1236–1256.

Roberts, J.S., & Wedell, D.H. (1994). Context effects on similarity judgments of multidimensional stimuli: Inferring the structure of the emotion space. *Journal of Experimental Social Psychology*, *30*(1), 1–38.

Roy, M., Mailhot, J.-P., Gosselin, N., Paquette, S., & Peretz, I. (2008). Modulation of the startle reflex by pleasant and unpleasant music. *International Journal of Psychophysiology*, *71*(1): 37–42.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6): 1161–1178.

Russell, J.A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1): 145–172.

Russell, J.A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology*, *67*(2): 186–205.

Russell, J.A., & Feldman Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5): 805–819.

Rusting, C.L. (1998). Personality, mood, and cognitive processing of emotional information: Three conceptual frameworks. *Psychological Bulletin*, *124*(2): 165–196.

Schellenberg, E.G., Peretz, I., & Vieillard, S. (2008). Liking for happy- and sad-sounding music: Effects of exposure. *Cognition & Emotion*, *22*(2): 218–237.

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.

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.

Schmidt, L.A., & Trainor, L.J. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion*, *15*(4): 487–500.

Schubert, E. (1996). Enjoyment of negative emotions in music: An associative network explanation. *Psychology of Music*, *24*(1): 18–28.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology, 51*(3): 154–165.

Schubert, E. (2007). The influence of emotion, locus of emotion and familiarity upon preference in music. *Psychology of Music, 35*(3), 499–515.

Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251.

Tabachnick, B.G., & Fidell, L.S. (Eds.) (2001). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.

Tellegen, A., Watson, D., & Clark, L.A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*(4): 297–303.

Terry, P., Lane, A., Lane, H., & Keohane, L. (1999). Development and validation of a mood measure for adolescents. *Journal of Sports Sciences, 17*(11): 861–872.

Terwogt, M., & Van Grinsven, F. (1991). Musical Expression of Moodstates. *Psychology of Music, 19(2):* 99–109.

Thayer, R.E. (1989). *The biopsychology of mood and arousal*. New York: Oxford University Press.

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.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence: The structure of emotion. *Journal of Personality and Social Psychology, 76*(5): 820–838.

Withvliet, C.V.O., & Vrana, S.R. (2006). Play it again Sam: Repeated exposure to emotionally evocative music polarises liking and smiling responses, and influences other affective reports, facial EMG, and heart rate. *Cognition & Emotion, 21*(1): 1–23.

Wundt, W. (1896). *Grundrisse der psychologie* [Outlines of psychology]. Leipzig, Germany: Engelmann.

Zentner, M., & Eerola, T. (2010). Self-report based measures and models. In P.N. Juslin & J.A. Sloboda (Eds.), *Handbook of music and emotion* (pp. 187–221). Oxford: Oxford University Press.

Zentner, M., Grandjean, D., & Scherer, K. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion, 8*(4): 494–521.

Appendix. List of audio tracks used in Experiment

No.	Emotion & level	Album name	Track	Min:Sec
001	Anger high	Lethal Weapon 3	8	04:15–04:29
002	Anger high	The Rainmaker	7	01:45–02:00
003	Anger high	The Alien Trilogy	9	00:03–00:18
004	Anger high	Cape Fear	1	02:15–02:30
005	Anger high	The Fifth Element	19	00:00–00:20
006	Anger mod.	Crouching Tiger, Hidden Dragon	8	01:12–01:25
007	Anger mod.	Batman Returns	2	00:18–00:33
008	Anger mod.	Man of Galilee CD1	6	00:40–01:07
009	Anger mod.	The Untouchables	8	01:38–01:53
010	Anger mod.	Oliver Twist	15	02:05–02:25
011	Fear high	Batman Returns	5	00:09–00:25
012	Fear high	JFK	8	01:26–01:40
013	Fear high	JFK	8	00:08–00:25
014	Fear high	The Alien Trilogy	5	00:26–00:41
015	Fear high	Hannibal	1	00:40–00:54
016	Fear mod.	Running Scared	6	02:53–03:07
017	Fear mod.	The Untouchables	8	01:38–01:53
018	Fear mod.	The Fifth Element	17	00:00–00:19
019	Fear mod.	Lethal Weapon 3	7	00:00–00:16
020	Fear mod.	Man of Galilee CD1	2	03:45–04:02
021	Happy high	The Rainmaker	3	02:55–03:13
022	Happy high	Batman	18	00:55–01:15
023	Happy high	Shallow Grave	6	02:02–02:17
024	Happy high	Man of Galilee CD1	2	03:02–03:18
025	Happy high	Oliver Twist	1	00:17–00:34
026	Happy mod.	The Omen	9	00:00–00:24
027	Happy mod.	Oliver Twist	8	01:40–02:04
028	Happy mod.	Grizzly Man	1	00:00–00:27
029	Happy mod.	The Portrait of a Lady	3	00:23–00:45
030	Happy mod.	Nostradamus	2	01:09–01:28
031	Sad high	The English Patient	18	00:07–00:32
032	Sad high	Running Scared	15	02:06–02:27
033	Sad high	The Portrait of a Lady	9	00:00–00:22
034	Sad high	Big Fish	15	00:55–01:11
035	Sad high	Man of Galilee CD1	8	01:20–01:37
036	Sad mod.	Angel Heart	4	00:08–00:28
037	Sad mod.	Batman	5	01:08–01:22
038	Sad mod.	Dracula	7	00:00–00:12
039	Sad mod.	Shakespeare in Love	3	00:59–01:17
040	Sad mod.	The English Patient	7	00:00–00:31
041	Tender high	Shine	10	01:28–01:48
042	Tender high	Pride & Prejudice	1	00:10–00:26
043	Tender high	Dances with Wolves	4	01:31–01:48
044	Tender high	Pride & Prejudice	12	00:01–00:15
045	Tender high	Oliver Twist	8	00:14–00:30
046	Tender mod.	Batman	9	00:00–00:19
047	Tender mod.	Oliver Twist	8	01:15–01:32
048	Tender mod.	Dracula	4	00:55–01:09
049	Tender mod.	Juha	2	02:11–02:26
050	Tender mod.	Oliver Twist	2	00:00–00:29

(Continued)

Appendix. (Continued)

No.	Emotion & level	Album name	Track	Min:Sec
051	Valence pos. high	Juha	10	00:20–00:38
052	Valence pos. high	Blanc	12	00:51–01:06
053	Valence pos. high	Gladiator	17	00:14–00:27
054	Valence pos. high	Pride & Prejudice	9	00:01–00:21
055	Valence pos. high	Dances with Wolves	10	00:28–00:46
056	Valence pos. mod.	Man of Galilee CD1	2	00:19–00:42
057	Valence pos. mod.	Shakespeare in Love	21	00:03–00:21
058	Valence pos. mod.	Vertigo OST	6	02:02–02:17
059	Valence pos. mod.	Vertigo OST	6	04:42–04:57
060	Valence pos. mod.	Outbreak	6	00:16–00:31
061	Valence neg. mod.	Juha	18	02:30–02:46
062	Valence neg. mod.	Shakespeare in Love	11	00:21–00:36
063	Valence neg. mod.	Batman	9	00:57–01:16
064	Valence neg. mod.	The Fifth Element	9	00:00–00:18
065	Valence neg. mod.	Big Fish	15	00:15–00:30
066	Valence neg. high	The English Patient	8	01:35–01:57
067	Valence neg. high	Lethal Weapon 3	7	00:00–00:16
068	Valence neg. high	Road to Perdition	6	00:34–00:49
069	Valence neg. high	Hellraiser	5	00:00–00:15
070	Valence neg. high	Grizzly Man	16	01:05–01:32
071	Energy pos. high	The Untouchables	6	01:50–02:05
072	Energy pos. high	Man of Galilee CD1	2	03:02–03:18
073	Energy pos. high	Shine	5	02:00–02:16
074	Energy pos. high	Shine	15	01:00–01:19
075	Energy pos. high	Batman	18	00:55–01:15
076	Energy pos. mod.	Juha	2	00:07–00:18
077	Energy pos. mod.	Lethal Weapon 3	4	01:40–02:00
078	Energy pos. mod.	Crouching Tiger, Hidden Dragon	13	01:52–02:10
079	Energy pos. mod.	Batman	4	02:31–02:51
080	Energy pos. mod.	Oliver Twist	7	01:30–01:46
081	Energy neg. mod.	Juha	16	00:00–00:15
082	Energy neg. mod.	Big Fish	15	00:55–01:11
083	Energy neg. mod.	Big Fish	11	01:26–01:40
084	Energy neg. mod.	Blanc	18	00:00–00:16
085	Energy neg. mod.	Oliver Twist	6	00:51–01:07
086	Energy neg. high	Running Scared	15	02:06–02:27
087	Energy neg. high	Road to Perdition	16	00:17–00:32
088	Energy neg. high	Blanc	10	00:13–00:31
089	Energy neg. high	Blanc	16	00:00–00:15
090	Energy neg. high	Batman Returns	12	00:57–01:14
091	Tension pos. high	The Alien Trilogy	11	02:12–02:27
092	Tension pos. high	The Fifth Element	13	00:17–00:31
093	Tension pos. high	Babylon 5	3	02:47–03:00
094	Tension pos. high	Hellraiser	10	02:44–03:00
095	Tension pos. high	Oliver Twist	15	02:05–02:25
096	Tension pos. mod.	The Missing	3	02:45–03:06
097	Tension pos. mod.	Shallow Grave	4	01:04–01:19
098	Tension pos. mod.	Naked Lunch	7	01:01–01:20
099	Tension pos. mod.	Dracula	5	00:11–00:27
100	Tension pos. mod.	Cape Fear	2	01:25–01:40

(Continued)

Appendix. (Continued)

No.	Emotion & level	Album name	Track	Min:Sec
101	Tension neg. mod.	Juha	2	02:11–02:26
102	Tension neg. mod.	Shakespeare in Love	6	00:00–00:19
103	Tension neg. mod.	The Fifth Element	12	00:00–00:17
104	Tension neg. mod.	Crouching Tiger, Hidden Dragon	11	00:28–00:46
105	Tension neg. mod.	Pride & Prejudice	4	00:10–00:29
106	Tension neg. high	Lethal Weapon 3	10	01:59–02:17
107	Tension neg. high	The Godfather	5	01:12–01:28
108	Tension neg. high	Gladiator	4	00:48–01:06
109	Tension neg. high	Pride & Prejudice	13	01:02–01:20
110	Tension neg. high	Big Fish	8	00:12–00:34