Perceived and Induced Emotion Responses to Popular Music: Categorical and Dimensional Models

Yading Song, Simon Dixon, and Marcus T. Pearce

Centre for Digital Music

School of Electronic Engineering and Computer Science

Queen Mary University of London

London, United Kingdom

Andrea R. Halpern

Department of Psychology

Bucknell University

Abstract

Music both conveys and evokes emotions, and although both phenomena are widely studied, the difference between them is often neglected. The purpose of this study is to examine the difference between *perceived* and *induced* emotion for Western popular music using both *categorical* and *dimensional* models of emotion, and to examine the influence of individual listener differences on their emotion judgment. A total of 80 musical excerpts were randomly selected from an established dataset of 2904 popular songs tagged with one of the four words “happy”, “sad”, “angry”, or “relaxed” on the Last.FM web site. Participants listened to the excerpts and rated *perceived* and *induced* emotion on the *categorical* model and *dimensional* model, and the reliability of emotion tags was evaluated according to participants’ agreement with corresponding labels. In addition, the Goldsmiths Musical Sophistication Index (Gold-MSI) was used to assess participants’ musical expertise and engagement. As expected, regardless of the emotion model used, music evokes emotions similar to the emotional quality perceived in music. Moreover, emotion tags predict music emotion judgments. However, *age*, *gender* and three factors from Gold-MSI, *importance*, *emotion*, and *music training* were found not to predict listeners’ responses, nor the agreement with tags.

Keywords: music and emotion, social tag, induced emotion, perceived emotion, emotion model

Introduction

Music can convey emotion and evoke feelings. Previous studies have shown that listeners' primary motives for listening to music lie in its emotional functions (Juslin and Laukka, 2004). Because of the power that music has to express and induce emotion, much attention has been paid to research on music and emotion in the past few decades (Eerola and Vuoskoski, 2013). The ability to identify emotional content is established at a very early age (Dalla Bella et al., 2001), and people engage with music in different contexts (e.g., travel, everyday routines) and for different purposes (e.g., distraction, mood regulation) (DeNora, 2000; Sloboda et al., 2001; Juslin et al., 2008). However, the ways in which people recognize and experience emotion in music remain unclear. Kreutz et al. (2007) noted that *perceived* emotion refers to intellectual processing, such as the perception of an intended or expressed emotion, whereas *induced* emotions reflect the introspective perception of psychophysiological changes, which are often associated with emotional self-regulation (Thayer and Faith, 2001; Khalfa et al., 2002). However, the distinction between perceived and induced emotion has not been broadly studied. Understanding the similarities and differences between the two will provide us a clearer view of the underlying psychological mechanisms involved in musical emotion (Juslin and Västfjäll, 2008).

In previous studies on induced and perceived emotion, Gabrielsson (2002) stated that induced emotion (also known as felt emotion) is the emotion experienced by the listener whereas the perceived emotion (also known as expressed emotion) is the emotion recognized in the music. Perceiving an expression of sadness in the music without necessarily being affected oneself is mainly a perceptual-cognitive process and it should be distinguished from listeners' emotional response to the music. In general, music seems to evoke emotions similar to the emotional quality perceived in music (Kallinen and Ravaja, 2006). Separating induced emotion from perceived emotion is not straightforward and the distinction is not always observed (Schubert, 2013). Other studies have also suggested that the rating level of induced emotion is higher than the level of perceived emotion in connection with positive valence, but lower in connection with arousal, positive activation and negative activation (Kallinen and Ravaja, 2006). Juslin and Laukka (2004) distinguished the perception and induction of emotion in music from emotions induced in everyday life, but the quantitative relationship between the two has not been examined.

Gabrielsson (2002) proposed and gave examples of four possible relationships between perceived and induced emotion, namely *positive relationship*, *negative relationship*, *no systematic relationship*, and *no relationship*. Those relationships were also discussed by Kallinen and Ravaja (2006), and Evans and Schubert (2008).

- Positive relationship: Occurs when "the listener's emotional response is in agreement with the emotional expression in the music" (p. 131);
- Negative relationship: Occurs when the "listener reacts with an emotion opposite to that expressed in the music: positive emotion in the music elicits negative emotion in the response, or negative emotion in the music elicits positive emotion in the response" (p. 134);
- No systematic relationship: Occurs when the listener stays "emotionally neutral" regardless of the expression of the music, or experiences different emotions on different occasions (p. 136);
- No relationship: Occurs when there is not even a potential relationship, such as when a person feels an emotion that cannot be expressed in music (p. 136).

During the past two decades, *categorical* models (or discrete models), *dimensional* models, *miscellaneous* models (McAdams et al., 2004; Ilie and Thompson, 2006), and *domain-specific* emotion models (e.g., Geneva Emotion Music Scale, which is only used for induced emotion) (Zentner et al., 2008) have been proposed and used in the study of music and emotion. The typical dimensional model (DM) represents emotions in an affective space with two dimensions: one related to valence (a pleasure-displeasure continuum) and the other to arousal (activation-deactivation) (Russell, 1980). Some studies also used the third dimension "dominance", to explain the difference between dominant and submissive emotions (Schubert, 2007). However, participants have found it difficult to rate on the dominance scale (Evans and Schubert, 2008). Only the two core dimensions *valence* and *arousal* were used in our experiment. In the study of music emotion recognition, arousal has been predicted better than valence from musical features (Schubert, 2007; Huq et al., 2010).

In contrast, the categorical model (CM) represents all emotions as being derived from a limited number of universal and innate basic emotions such as happiness, sadness, fear, and anger (Ekman, 1992; Panksepp, 1998). On the one hand, the dimensional model has been criticized for its lack of differentiation when it comes to emotions that are close neighbors in the valence-activation space, one of the consequences of the low dimensionality of the model. On the other hand, the categorical model itself is inadequate to describe the richness of emotional effects of music. However, both the categorical model and dimensional model have received empirical support (Kreutz et al., 2007; Vieillard et al., 2008) and have been commonly used in the study of music and emotion (Juslin et al., 2008; Vieillard et al., 2008; Truong et al., 2009; Vuoskoski and Eerola, 2011b). In addition, Eerola and Vuoskoski (2010) compared the use of these two models and suggested that the categorical emotion model was less reliable than the dimensional model at predicting the rating of excerpts that were ambiguous examples of an emotion category, but that they both produced highly compatible ratings of perceived emotions. Similar to tempo (Repp, 1994; Dixon, 2001), emotion might be instantaneous, which can be recognized and experienced as soon as the music starts, or emotion may evolve and change continuously over time. Due to this dynamic nature of music, we measured the "dominant emotion" from a musical excerpt that participants perceived and felt, which might be considered as a central tendency (not an average, nor the median) of emotion in a certain range of exposure (e.g., 30 or 60 seconds). The goal of this paper is to study both induced and perceived emotion: we compare the perceived and induced responses in both categorical and dimensional models. Although other approaches such as behavioral measurement and physiological reaction are becoming increasingly popular, there exist several issues. For instance, measuring behavioral responses to music including facial expression and body language (Frijda, 1986), and various physiological reactions such as heart rate, respiration, skin conductance, and electroencephalograph (EEG) (Levenson, 2003; Mas-Herrero et al., 2014) can provide us valuable evidence of emotion, but these behaviors do not always associate with certain emotions, and the relationships between emotions and physiological responses are still unclear (Juslin and Laukka, 2004; Eerola and Vuoskoski, 2013). Hence, the most common way to measure subjective emotional responses is still via a format of self-report (Gabrielsson, 2002).

In recent years, researchers have predominantly used excerpts of Western classical or jazz music, film music, and non-western cross-cultural music in the study of music and emotion (Evans and Schubert, 2008; Mohn et al., 2010; Vuoskoski et al., 2011; Vuoskoski and Eerola, 2011a; Kosta et al., 2013). Classical music is the most dominant genre in past studies, but listeners may not be familiar with the genre, and the results may not be applicable to other musical genres such as Western popular music, which is generally well-understood by participants (Juslin and Laukka, 2003). Also, the stimuli are typically annotated, selected, or manipulated by expert musicologists or researchers, which is very expensive in terms of financial cost and human labor. However, with the emergence of web services for music discovery such as Last.FM, we can now easily access rich human-annotated information (e.g., semantic tags) about music, and fetch musical examples tagged with different emotional labels (see Stimuli section below) (Turnbull et al., 2008; Levy and Sandler, 2009). Although previous studies have suggested tag features are more useful than audio content features for certain analyses, and social tags have been widely used in areas of research such as classification, recognition, and identification, the reliability of emotion tags has rarely been evaluated (Lamere, 2008; Wang et al., 2010). Music preference influences perceived and induced emotion, and the emotional reactions may differ across genres due to their inherent musical differences (Kreutz et al., 2007; Zentner et al., 2008; Eerola, 2011). Therefore, our study focuses on Western popular music.

The emotional meaning of music always emerges from the interaction between a listener and a (real or imagined) sound object in some context. The listener plays an important role in the perception and recognition of emotion in music. Different individual factors such as one's personality, musical skills, and music culture, may affect how emotional meaning is elicited (Bigand et al., 2005; Kallinen and Ravaja, 2006). Moreover, several studies have investigated the relationship between personality traits ("Big Five"), musical judgments (Shiota et al., 2006; Vuoskoski and Eerola, 2011a,b), artistic interests (McAdams, 2006), and music preferences (Rentfrow and Gosling, 2003; Chamorro-Premuzic and Furnham, 2007). In addition, previous studies on classical music have suggested that emotional responses and intellectual responses tend not to occur together, and that musically experienced listeners are more likely to focus on intellectual properties, whereas less musically experienced listeners tend to focus on emotional properties (Gabrielsson, 2002). Hence, we also study here the influences of musical education, music engagement, age, and gender on the agreement of human-annotated tags and emotional judgments in two conditions (induced and perceived).

Therefore, the aims of this paper are:

1. To systematically examine the relationship between perceived and induced emotional responses to music, using both categorical and dimensional models of emotion;
2. To measure the influence of personal factors (i.e., age, gender, music skills, and engagement) on listeners' musical judgment with emotion tags;
3. To evaluate the reliability of emotion tags, and to establish a set of 80 Western popular music excerpts for future research on music and emotion.

Experiment 1

Method

Participants

Forty English-speaking students (20 male) participated in Experiment 1 using the categorical model. All the participants were recruited through university email lists, and had ages ranging from 18 to 44 years, with various educational backgrounds (from undergraduate to postgraduate) and levels of music training. Among the participants, 50% attentively listened to music more than one hour per day, and 88% of the participants can play at least one instrument. Moreover, 68% of the participants preferred pop/rock music, 10% of them preferred jazz, and the rest preferred classical music. Full details of participants' information (age, gender, and nationality) can be found in Appendix A.

Stimuli

The stimuli were selected from a collection of 2904 musical excerpts retrieved from Last.FM[1] and 7Digital[2] which have been used previously in music and emotion studies (Song et al., 2012b, 2013a,b). Each excerpt had been tagged on Last.FM with one of the four words "happy", "sad", "angry", or "relaxed"[3]. We randomly chose a total of 80 excerpts from these four categories (20 tracks from each category). The excerpts ranged from recent releases back to 1960s, and covered a range of Western popular music styles such as pop, rock, country, metal, and instrumental. It is worth noting that emotion induction usually requires longer excerpts than emotion perception (e.g., 45-60s, Vuoskoski and Eerola, 2011a), but we wanted the length to be consistent for measuring both induced and perceived emotion. Each excerpt was either 30 seconds or 60 seconds long (as provided by 7Digital), and it was played from a standard mp3 format file (bit rate: 128 *kbps* or 64 *kbps*; sample rate: 22050 *Hz* or 44100 *Hz*). Although emotion can be perceived and felt from shorter excerpts, such emotions are liable to vary over time within a piece. We assume that the emotion tags correspond to the overall affect or most frequent emotion over the whole piece, so we took the full length of the excerpts that were available from the database, in order to smooth

1 http://www.last.fm/home
2 http://www.7digital.com/
3 The term "relax" was used in the data collection, which represented the emotion "relaxed".

out short-term affects. This 80-excerpt dataset is described in Appendix C and is available online[4], to enable further studies with this data and comparisons with the current work.

In order to minimize the effect of song order and condition order (perceived and induced emotion), four different list conditions were constructed. The order of presentation of the two rating conditions and two song lists (*n*=40, 10 for each emotion category) was counterbalanced across participants. The songs in each song list were presented in a different random order for each participant (Welch and Krantz, 1996). Therefore, the participants were divided into four groups as shown in Table 1.

TABLE 1. *Group Allocation among Participants*.

Group	**Song list 1**	**Song list 2**
Group 1	Induced emotion	Perceived emotion
Group 2	Perceived emotion	Induced emotion
	Song list 2	**Song list 1**
Group 3	Induced emotion	Perceived emotion
Group 4	Perceived emotion	Induced emotion

Procedure

In the study of music and emotion, the categorical model has been predominantly used, and over 75% of the studies used happiness, sadness, and anger. In order to cover all four quadrants of the 2-dimensional model, four basic emotion classes: "happy", "angry", "sad", and "relaxed" were used. These four basic emotions are widely accepted across different cultures and they are often expressed in music (Laurier et al., 2008; Eerola and Vuoskoski, 2013). Experiment 1 was conducted using the categorical model in a laboratory environment[5] in November 2012. The study was approved by the Queen Mary Research Ethics Committee (REF: QMREC1019). The only way to assess subjective emotional experience is via a format of listening and self-report (Gabrielsson, 2002). First, the participants were asked to read the instruction page:

1. Listen to the songs (they will last either 30 or 60 seconds);

4 https://code.soundsoftware.ac.uk/projects/emotion-recognition
5 http://isophonics.net/emotion/test/

2. After listening, for each piece please choose one of the following: *happy*, *sad*, *relaxed*, *angry*, or *cannot tell/none of above*;

   Note: participants were asked to answer one of the two questions (condition "induced" and condition "perceived").
3. For each track, you may click the "stop" button of the audio player if required;
4. Be careful, do not press too quickly, since you can only listen to each song once;
5. Please answer all the questions; the test will take about 40 minutes to complete.

Participants were then asked to fill in a demographic form including name, age, gender, "type of music they are most familiar with", nationality, and "music culture they grew up with" as well as a selected subscale from the Goldsmiths Musical Sophistication Index (Gold-MSI v0.9[6]) questionnaire to measure participants' level of musical expertise and engagement (Müllensiefen et al., 2012). For Experiment 1, 32 questions regarding the importance of music in everyday life (*importance*), importance of music for psychological functions (*emotion*), and life history of formal music training (*music skills*) were measured by the Gold-MSI. The listeners responded to each excerpt (10 excerpts per page) and rated them with a categorical model. Participants were reminded of the two different rating conditions (induced and perceived emotion). During the listening test, they were asked "*How would you describe the emotional content of the music itself?"* for the emotion perceived, and "*What emotion do you feel in response to the music?*" for the emotion induced. In order not to constrain people in the experiment to four basic emotion classes, we added the fifth option *cannot tell/none of above*. The whole test lasted about 40 minutes without any planned breaks. However, the participants were able to stop whenever they wanted and adjust volume to a comfortable level. At the end of the experiment, their opinions and feedback were collected. Participants were also asked to provide examples of musical pieces for each perceived and induced emotion, as well as the activities involved with music listening and its purposes (see Appendix B).

---

6 http://www.gold.ac.uk/music-mind-brain/gold-msi/

The statistical analyses were all conducted using the Matlab 2012 Statistics Toolbox. Responses were aggregated across participants for song-level analysis, or aggregated across items for individual-level analysis.

Results

Comparison of Responses for Perceived and Induced Emotion

The normality of participants' responses was checked via a Kolmogorov-Smirnov test (K-S test), and the results rejected the assumption of normality with $KS = 0.5$ ($p < .001$). To compare listeners' reports of induced and perceived emotion of music, a nonparametric Wilcoxon signed-rank test was used to analyze the results of the two conditions. Participants' responses were aggregated by their corresponding labels, and the agreement between responses and emotion tags for perceived emotion ($Mdn = 0.54$) was significantly higher than for induced emotion ($Mdn = 0.47$, $Z = -2.09$, and $p < .05$).

Participants' consistency of perceived and induced responses.

Regardless of the emotion tags, to examine the differences in the consistency of participants' perceived and induced responses, one-way analysis of variance (ANOVA) was conducted on the category with the highest number of responses for each music excerpt. The analysis revealed that the consistency of participants' perceived emotional responses ($M = 0.61$) was significantly higher than of their induced emotional responses ($M = 0.52$, $p < .01$, and $F(1) = 10.27$). For each stimulus, we computed an uncertainty score, corresponding to the proportion of participants who indicated *cannot tell/none of the above*. Using this as a dependent variable, a Wilcoxon signed-rank test revealed a significant difference between the uncertainty for induced ($Mdn = 3$) and perceived emotion ($Mdn = 2$, $Z = 5.27$, and $p < .001$). The result is consistent with the literature in suggesting that a higher consistency can be found in the responses of perceived emotion (Gabrielsson and Juslin, 1996). One of the possible reasons for the lower uncertainty level in perceived responses could be explained by the previous findings that the basic emotions such as *happiness*, *sadness*, and *anger* are often expressed in music because of their distinct expressive characteristics (Juslin and Laukka, 2004). A different explanation may come from the fact that emotions induced by music are more sophisticated and context-dependent, and the categorical model is inadequate to describe the richness of subjective emotions (Zentner et al., 2008; Eerola and Vuoskoski, 2013).

The relationship between listening duration and emotional responses.

Given that participants were free to stop whenever they wanted, the listening duration for each emotional response was studied. The duration of each musical stimulus is either 30 or 60 seconds, and the percentage of the length of each excerpt that a participant listened to was computed as the "listening duration". Therefore, the listening durations across four emotion categories were compared using the Kruskal-Wallis test. The results showed that the listening duration for emotional responses of "sad" and "relaxed" was significantly longer than responses of "happy" and "angry", for both induced ($\chi^2(3) = 47.96$, $p < .001$) and perceived emotion ($\chi^2(3) = 57.64$, $p < .001$). However, no significant differences in listening duration were found between induced and perceived emotional responses. This agrees with previous studies that happiness can be easily identified in music, whereas sadness and relaxedness are often confounded (Vieillard et al., 2008).

The relationship between the responses for perceived and induced emotion.

Since participants' perceived and induced responses were different, the relationships between these two emotional responses were investigated. Correlation analyses between perceived and induced emotional responses were performed on the number of responses for each emotion category with the results shown in Table 2. Positive correlations ($p < .001$) were found on corresponding emotions in perceived and induced emotion, and several negative correlations were shown such as induced happiness with both perceived and induced sadness and relaxedness, as well as perceived anger with perceived happiness, sadness, and relaxedness. Interestingly, a positive correlation was found between perceived sadness and induced relaxedness (Pearson's $r(78) = .29$, $p < .01$).

TABLE 2. *Correlations Between Induced and Perceived Emotional Responses*.

		Induced responses (IR)				Perceived responses (PR)		
		Happiness	Sadness	Relaxedness	Anger	Happiness	Sadness	Relaxedness
IR	Sadness	-.57***	-	-	-	-	-	-
	Relaxedness	-.40***	.17	-	-	-	-	-
	Anger	-.09	-.36***	-.59***	-	-	-	-
PR	Happiness	**.85*****	-.52***	-.20	-.26*	-	-	-
	Sadness	-.56***	**.90*****	.29**	-	-.54***	-	-
	Relaxedness	-.26*	.01	**.75*****	-.51***	-.16	-.04	-
	Anger	-.08	-.37***	-.56***	**.96*****	-.28*	-.40***	-.49***

Note. Pearson's *r* values between corresponding emotions in perceived and induced emotion are bold. *$p < .05$; **$p < .01$; ***$p < .001$.

In order to categorize the relationships between responses for perceived and induced emotion, we used and adapted Gabrielsson's (2002) model (see Table 3). To distinguish between various types of "negative" relationships, the negative category was further divided into three cases: *valence and arousal* (negative valence and arousal in perceived emotion, but positive valence and arousal in induced emotion, etc.), *valence* (negative valence and arousal in perceived emotion, but positive valence and negative arousal in induced emotion, etc.), and *arousal* (negative valence and arousal in perceived emotion, but negative valence and positive arousal in induced emotion, etc.). To categorize the relationship between perceived and induced emotion for each excerpt, we assessed which emotion is dominant among participants' responses. We took the label with the greatest number of votes to be the dominant emotion, as long as it received more than 8 out of 20 votes; otherwise "undecided" was marked. We chose 8 votes, since this gave a p value $< .05$, given the null hypothesis that $p(\text{happy}) = p(\text{sad}) = p(\text{relaxed}) = p(\text{angry}) = 25\%$ and $p(\text{none}) = 0\%$. Therefore, eighty relationships were collected. Seventy-one excerpts had one of the following relationships: *positive relationship*, *negative relationship*, *no systematic relationship*, and *no relationship*. Among these relationships, forty-three songs had a "positive relationship" (the listener's emotional response is in accordance with the emotion expressed in the music), and seven songs had a "negative relationship" (listener reacts with an emotion 'opposite' to that expressed in the music in at least one domain), of which four had a negative valence relationship (only with relaxedness and sadness), and negative arousal relationship took three (only with happiness and relaxedness). No cases of negative relationships were found for both valence and arousal simultaneously.

Meanwhile, fourteen songs had "no systematic relationship" between perceived and induced emotion (*case 1*: the music evokes various[7] emotional responses in different listeners, or *case 2*: listener stays "emotionally neutral" regardless of the expression of the music), of which twelve of them belonged to case 1. The remaining six cases had "no relationship". However, there was one special case in which an equally perceived happy and angry song induced only happiness (song title: Motown Junk, and artist's name: Manic Street Preachers). The relationships for these 80 excerpts are shown in Table 3.

TABLE 3. *Possible Relationships Between Perceived and Induced Emotions in the Categorical Model.*

Relationship	Level	Perceived emotion	Induced emotion	Percentage
Undecided relationship	Valence & Arousal	undecided	undecided	11%
Positive relationship	Valence & Arousal	happy sad relaxed angry	happy sad relaxed angry	54%
Negative relationship	Valence & Arousal	happy sad relaxed angry	sad happy angry relaxed	0%
	Valence	happy sad relaxed angry	angry relaxed sad happy	5%
	Arousal	happy sad relaxed	relaxed angry happy	4%

[7] Various for different listeners and occasions.

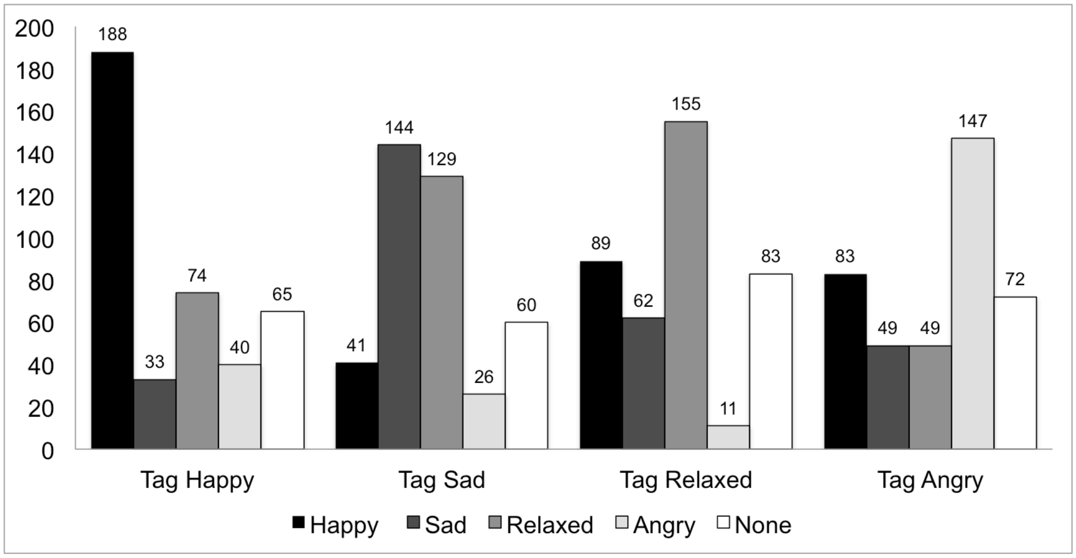
		angry	sad	
No systematic relationship	Case 1	happy sad relaxed angry	various	15%
	Case 2	happy sad relaxed angry	neutral	3%
No relationship	Valence & Arousal	not perceived	happy sad relaxed angry	7%

Note. The special case in which an equally perceived happy and angry song induced only happiness is not shown (covers 1%).

Correspondence between Tags and Emotional Responses

Figures 1 and 2 show the response distributions for excerpts for each emotion tag which was retrieved from the Last.FM web site. As expected, the tags predicted both perceived and induced emotional responses. In both perceived and induced cases, the excerpts labeled with *happy*, *angry*, and *relaxed* were very distinct. The induced responses for the tag “sad” were somewhat blurred between “sad” and “relaxed”. However, for the songs labeled with “anger” and “sadness”, induced emotional responses as seen in Figure 1 received more positive emotion responses (relaxedness and happiness) than perceived emotional responses as seen in Figure 2.

In order to evaluate the reliability of the emotion tags, Wilcoxon signed-rank tests were carried out for each emotion category. The analyses revealed that agreement between emotion tags and participant ratings was well above chance (for perceived emotion $Z = 4.67$, $p < .001$; for induced emotion $Z = 5.34$, $p < .001$). Though the overall agreement of the perceived emotional responses with emotion tags was ranked significantly higher than that for induced responses, in the case of happiness and relaxedness, the agreement of perceived and induced emotional responses showed no significant difference, as shown in Table 4. It suggests that when musical excerpts are labeled as "sad" or "angry", there is a greater chance that listeners supply these labels for perceived emotion than for induced emotion. In contrast, listeners who label excerpts "happy" or "relaxed" are about as likely to base these labels on perceived emotion as induced



emotion.

FIGURE 1. Induced emotional response distribution for each tag. The horizontal axis shows the five responses for each of the four emotion tags *happy*, *sad*, *relaxed*, and *angry*. The vertical axis shows the number of responses.

TABLE 4. *Proportion of Responses Agreeing with Last.FM Tag Data for the Corresponding Song*.

	Happy	Sad	Relaxed	Angry
Induced responses	0.47	0.36	0.39	0.37
Perceived responses	0.45	0.56***	0.37	0.51***

Note. Columns represent the four emotion tags, rows the listener responses for induced and perceived emotion. *$p < .05$; **$p < .01$; ***$p < .001$ using Wilcoxon signed-rank test, for comparing the overall agreement of perceived and induced emotional responses with emotion tags.

Analysis of Individual Factors

In this experiment, Gold-MSI v0.9 was used to assess participants' music expertise and engagement. The three factors measured were *importance* (importance of music in everyday life), *music training* (life history of formal music training), and *emotion* (importance of music for psychological, especially emotional functions). A summary of the responses can be found in Table 5, as well as statistics for a large BBC Internet study (n = 137,633) using Gold-MSI v1.0 (Müllensiefen et al., 2014). Comparing these three musical attributes in our study, a positive correlation was found for *importance* with both *music training* (Pearson's $r(38) = .32, p < .05$) and *emotion* ($r(38) = .57, p < .001$).

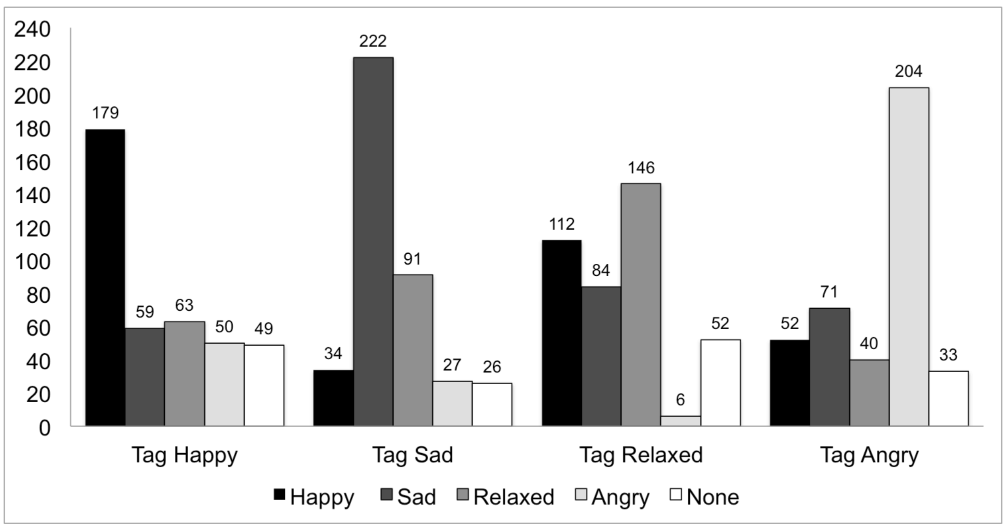


FIGURE 2. Perceived emotional response distribution for each tag. The horizontal axis shows the five responses for each of the four emotion tags *happy*, *sad*, *relaxed*, and *angry*. The vertical axis shows the number of responses.

TABLE 5. *Summary of Responses to 32 Questions Adapted from the Gold-MSI.*

		Scale Min	Scale Max	Mean	SD
	Importance	15	105	72.5	18.1
Experiment 1	**Music training**	9	63	37.4	15.0
	Emotions	8	56	45.5	6.7
BBC Study	**Active engagement**	9	63	41.5	10.4

Music training	7	49	26.5	11.4
Emotions	6	42	34.7	5.04

Note. For example, a music training score of 26 using Gold-MSI v0.9 could mean that the participant can play one musical instrument, had one year of formal training, and practices the instrument for one hour daily.

To discover whether individual differences such as age, gender, music training, and engagement influence emotion judgments, correlation analysis and Kruskal-Wallis one-way analysis of variance were carried out. However, no significant difference was found between responses of male and female participants, nor for age[8], on their agreement with emotion tags for perceived and induced emotion. We found that older participants were more likely to listen to music for its emotional functions ($r(38) = .57, p < .001$), and there was a tendency of participants to feel less "happy" from music with increasing age. Interestingly, the distribution of individual participant responses showed that some people perceived anger in the music but never became angry listening to the excerpts. Likewise, some people could be easily moved by music. Therefore, we examined the response distribution over the 80 musical excerpts for each participant, but no significant result was found. The reason for the random distribution among people might depend on other factors such as the participant's personality, taste, or current mood.

Discussion

The aim of Experiment 1 was to examine the difference between perceived and induced responses using a categorical model, and to explore associations between participants' emotional judgments and emotion tags (happy, sad, relaxed, and angry), as well as the influence of individual differences. First, participants' responses for perceived emotion in popular music showed a higher level of consistency than induced emotion ratings, thus agreeing with previous studies (Gabrielsson and Juslin, 1996). In addition, we found a

8 Only age categories of 18-24 and 25-34 were compared, as there was not enough data for participants' age over 35.

higher uncertainty level in induced compared to perceived responses, indicating that the basic emotions are often expressed in music and the categorical model may be inadequate to describe the richness of emotion induced by music. Though we found the listening duration for emotional responses of "sad" and "relaxed" were significantly longer than for the responses of "happy" and "angry" for both induced and perceived emotion, no significant differences in listening duration were found between perceived and induced emotional responses.

Secondly, the potential relationships between perceived and induced emotion were studied. The analyses showed the ratings of perceived and induced emotion were positively correlated. Nonetheless, a small but significant difference was found. It suggests most of the times a person feels the emotion that the music expresses. Some interesting correlations were also found such as a perceived sad song could induce relaxedness in listeners, but not the other way round. If a song is perceived as sad or relaxed, the song would be unlikely to induce happiness. The majority of the music excerpts used in the experiment were sung, and lyrics may add some semantic meaning for emotional responses, therefore it is also possible that emotional judgments could be related not to the elements of music structure itself but with the meaning expressed by the lyrics. The analyses showed the same results after removing the instrumental tracks[9]. However, the relationship between emotion perceived and induced is complex, as it may take various forms (Gabrielsson, 2002). Therefore, quantitative analyses were conducted between the two conditions. As reported in the literature, a positive relationship is the most frequent one (Gabrielsson, 2002; Evans and Schubert, 2008). A similar result was found in the experiment that a positive relationship occurred 54% of the time. Moreover, the only negative relationship on the valence level was between the emotions relaxedness and sadness, and the only negative relationship on the arousal level occurred between relaxedness and happiness. Other relationships such as "no relationship" and "no systematic relationship" were found in 7% and 18% of the cases respectively.

---

[9] There were not enough data points (4 in song list 1, and 6 in song list 2) to analyse instrumental tracks separately.

Even though social tags have been widely used in recent literature, emotion tags are rarely explored. Therefore, it is important to evaluate the reliability of the emotion tags for predicting listeners' induced or perceived responses. The results showed that the participant agreement with emotion tags was well above chance for both perceived and induced emotion. For the cases of sadness and anger, perceived emotion showed a significantly higher agreement with tags than induced emotion showed. One of the explanations is that the user tagging behavior is based on their perceived feelings, but other explanations could be the subjectivity of induced emotion and between-subject differences in music experience, culture, personality, situation, current mood, and preference. However, further studies need to be conducted with controlled environment and stimuli. Interestingly, the induced emotion distribution showed responses that were more positive than the emotion tags would have indicated. This agrees with findings that emotional functions of music are generally positive (Juslin and Laukka, 2004; Zentner et al., 2008).

Finally, the influence of individual differences in *age*, *gender*, *importance* (importance of music in everyday life), *music training* (life history of formal music training), and *emotion* (importance of music for psychological, especially emotional functions), on participants' emotion judgments was assessed. No significant differences were found between male and female participants' responses for either perceived or induced emotion, nor their agreement with emotion tags. Studies have suggested that as people get older, they focus more on self-control of their emotions and rate their emotion-regulation skills as better (Gross et al., 1997). Similar results could be found in this experiment that participants' ratings of importance of music for emotional functions increased with age. Previously, different findings on age and emotional experience have been reported (Mroczek and Kolarz, 1998; Charles et al., 2001; Mather and Carstensen, 2005). No significant relationships were found between age, participants' emotional responses and the agreement with emotion tags, but a tendency of older participants to choose "cannot tell/none of above" rather than "happy" of induced emotional responses was shown. In addition, the three musical attributes *importance*, *musical skills,* and *emotion* had no correlation with participants' emotion judgments. Positive correlations were found between *importance* and both *music training* and *emotion*. Individual differences such as preference, personality traits, and listener's current mood are also known to be relevant to emotional judgment (Shiota et al., 2006; Vuoskoski and Eerola, 2011a,b), but they are beyond the scope of this study.

General comments from participants raised other issues which are worthy of further investigation, for example: participants' music preferences may influence their emotion judgments ("*I really like heavy metal, so I think many of the metal songs, normally people would've felt angry, but I just felt happy and energized.*"); the responses for perceived and induced emotion in music may depend on lyrics and cultural factors ("*If I feel sad I will usually listen to songs in my mother tongue*", "*Sometimes, emotional content of the music itself is closely related to the lyrics. Considering that English is not my mother language, it is more difficult for me to get the insight*"). Feedback from participants reinforces the issue of the inadequacy of our categorical model, with comments such as: "*four emotional classes are not enough*", "*more options should be added*", "*many times I was feeling limited because of the small amount of feelings options I had to choose from*", and "*I could feel more than one emotion, or another emotion which was not included in options (like energetic, romantic, etc.)*". To address the limitations in the categorical model, Experiment 2 was conducted using a 2-dimensional continuous model of emotion.

Experiment 2

Method

Participants

Fifty-four English-speaking participants (25 male) took part. They were recruited through professional and academic email lists and social media, and had ages ranging from 15 to 54 years, as well as various educational, cultural, and music training backgrounds. Full details of participants' information (age, gender, and nationality) can be found in Appendix A.

Stimuli

The same 80 musical excerpts used in Experiment 1 were used in Experiment 2. Similar to the design in Experiment 1, four different list conditions were constructed to minimize the effect of song order and conditions (perceived and induced emotion). The order of presentation of the two rating conditions and two song lists (n=40, 10 for each emotion category) was counterbalanced across participants. The songs in each song list were also presented in a different random order for each participant.

Procedure

In the past two decades, about a third of music and emotion studies have used a dimensional model (Eerola and Vuoskoski, 2013). Two dimensions, *valence* (happy versus sad) and *arousal* (calm versus excited), which were proposed by Russell (1980), are the most typical ones. Later, a third dimension *dominance* was also utilized (Killgore, 1999). However, to keep a simple mapping of emotions, only two dimensions, *valence* and *arousal*, on an 11-point scale, were used in Experiment 2. In addition, to map the dimensional model of emotion with emotion tags, the same four basic emotions (*happy*, *sad*, *relaxed*, and *angry*) used in

Experiment 1 are chosen such that each occupies a unique quadrant of the valence-arousal plane as shown in Figure 3.

Experiment 2 was conducted using the dimensional model via an online platform[10] in May 2013. The study was approved by the Queen Mary Research Ethics Committee (REF: QMREC1019). The participants were asked to read a similar instruction page as shown in Experiment 1 except that they rated each piece on the dimensional model: *valence* (happy-sad continuum) and *arousal* (excited-relaxed continuum).

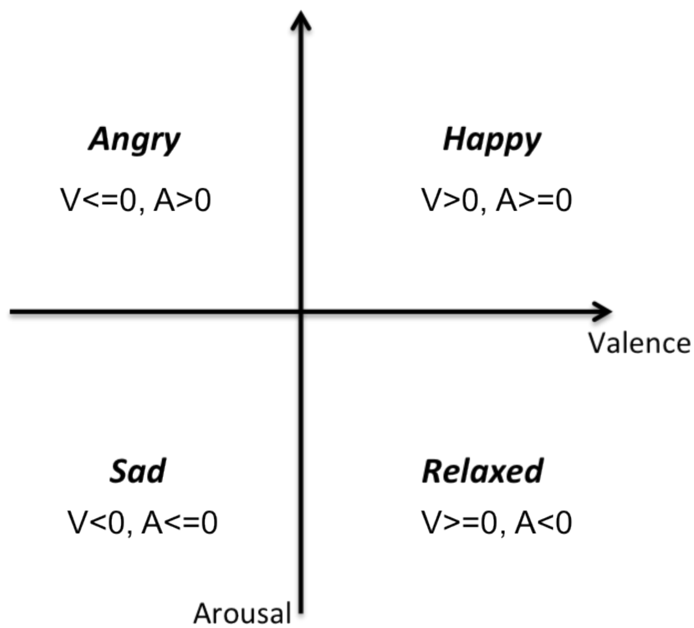


FIGURE 3. Valence-Arousal model showing the quadrants of the four emotion tags used in this experiment.

Participants were asked to fill in a demographic form including name, age, gender, "type of music they are most familiar with", nationality, and "music culture they grew up with" as well as the music training background from the Gold-MSI v0.9. The participants then responded to each excerpt (10 excerpts per

[10] http://isophonics.net/dimensional/test/

page) and rated valence and arousal on an 11-point scale. Valence was rated from very positive to very negative; arousal from very calm to very excited. Participants were also reminded of two different rating conditions ("perceived" and "induced") by an alert box when turning each page. During the listening test, they were asked "*How would you describe the emotional content of the music itself?*" for the perceived emotion, and "*What emotion do you feel in response to the music?*" for the induced emotion. The whole test lasted about 40 minutes without any planned breaks. At the end of the experiment, participants' feedback was collected via email.

The statistical analyses were all conducted using the Matlab 2012 Statistics Toolbox. Responses were aggregated across participants for song-level analysis, or aggregated across items for individual-level analysis. Table 6 shows a summary of the two experiments' differences.

TABLE 6. *Summary of Design Differences Between the Two Experiments.*

	Experiment 1	**Experiment 2**
Model	Categorical model	Dimensional model
Condition	Laboratory environment	Online environment
Equipment	Studio quality headphones	Any (internet required)
Participants	40 (students)	54 (unknown)
Reminder of conditions	Yes (oral)	Yes (alert)
Practice page	Yes	No
Ratings	5 options	Two 11-point scales (-5 to 5)
Feedback	Paper	Email

Results

Comparison of Responses for Perceived and Induced Emotion

To check the normality of participants' ratings of valence and arousal, the Jarque-Bera (JB) test was carried out. The results showed that the ratings of valence were normally distributed, whereas the analysis rejected the assumption of normality for ratings of arousal. In the dimensional model of emotion, to assess the effects of rating conditions (perceived and induced emotion) and emotion tags (happy, sad, relaxed, and angry), two-way ANOVA was conducted on the ratings of valence, and Friedman's test and a Wilcoxon signed-rank test were conducted on the ratings of arousal. We found that the four emotion labels differentially predicted the ratings of *valence* ($F(3) = 34.02$, $p < .001$) and *arousal* ($\chi^2(3) = 51.52$, $p < .001$). However, no significant difference was found between the two conditions for the ratings of valence and arousal.

Participants' consistency of perceived and induced responses.

The tags associated with a song are not an absolute ground truth, but are also generated by users, under unknown conditions; we also looked at the consistency of rating quadrants among the participants. The level of participant agreement is defined as the proportion of participants whose ratings are in the quadrant with the highest number of participant ratings. This value has as a lower bound the agreement with the tag quadrant, but can be higher if a greater number of participants agree on a quadrant other than that of the tag. The consistency for the individual dimensions of valence and arousal was also computed. A Wilcoxon signed-rank test was used to compare the consistency between perceived and induced emotion. Results showed that consistency of perceived responses (*Mdn* = 0.60) was significantly higher than of induced responses (*Mdn* = 0.52, $Z = -3.06$, and $p < .01$), in both the ratings of valence ($Z = -2.54$, $p < .05$) and arousal ($Z = -3.83$, $p < .01$), which is also consistent with the literature. In addition, in comparing levels of consistency for the two dimensions *valence* and *arousal*, a higher consistency of responses for arousal can be found. The results are shown in Table 7.

TABLE 7. *Consistency of Participants' Responses for Valence and Arousal.*

	Mean	Standard deviation
Perceived emotion	0.59	0.14
Induced emotion	0.55	0.15
Perceived valence	0.68	0.14
Induced valence	0.65	0.15
Perceived arousal	0.77	0.15
Induced arousal	0.72	0.16

Note. The consistency was calculated by the highest number of responses divided by the overall responses for each excerpt. Rows represent the results in three dimensions: *emotion* (both valence and arousal simultaneously), *valence* and *arousal* respectively, columns the mean and standard deviation across songs.

The relationship between listening duration and emotional responses.

The listening time for each excerpt was recorded, and next we studied the listening duration for four emotional responses using the dimensional model. First of all, participants' ratings of valence and arousal were mapped to the four emotion categories (see Figure 3), and then the listening durations for each emotion were aggregated. To explore the differences between listening durations for four basic emotions (happy, sad, relaxed, and angry), a nonparametric analysis (Kruskal-Wallis test) was conducted. The analysis showed the listening durations of the four emotional responses were significantly different, for both induced ($\chi^2(3)$ = 50.54, p < .001) and perceived emotion ($\chi^2(3)$ = 42.6, p < .001). Additionally, significant shorter listening durations could be found for emotional responses of "happy" in perceived emotion, and both "happy" and "angry" in induced emotion, than the responses of "sad" and "relaxed". This result may suggest that it is easier to feel and recognize emotions in the region of high arousal (happy and angry), whereas it takes longer time to distinguish in the lower region of arousal such as relaxedness and sadness.

The relationship between the responses for perceived and induced emotion.

The comparison of ratings for perceived and induced emotion showed that there was no significant difference between the sets of ratings for the two conditions. Furthermore, Pearson's correlation analyses were performed to study the relationship between perceived and induced emotion ratings of valence and arousal. The results showed that regardless of the emotion tag, the listeners' valence and arousal ratings were highly positively correlated between perceived and induced emotion (*valence*: $r(78) = .94$, $p < .001$; *arousal*: $r(78) = .97$, $p < .001$).

TABLE 8. *Possible Relationships between Perceived and Induced Emotions in the Dimensional Model*.

Relationship	Level	Perceived emotion		Induced emotion		Percentage
		Valence	**Arousal**	**Valence**	**Arousal**	
Undecided relationship		undecided	undecided	undecided	undecided	14%
Positive relationship	Valence & Arousal	high	high	high	high	16%
		high	low	high	low	
		low	low	low	low	
		low	high	low	high	
	Valence	high		high		12.5%
		low		low		
		zero		zero		
	Arousal		high		high	45%
			low		low	
			zero		zero	
Negative relationship	Valence & Arousal	high	high	low	low	0%
		high	low	low	high	
		low	high	high	low	
		low	low	high	high	
	Valence	high		low		0%
		low		high		
	Arousal		high		low	0%
			low		high	
No systematic relationship	Case 1	decided	undecided	undecided	decided	2.5%
		undecided	decided	decided	undecided	
	Case 2	decided	undecided	undecided	undecided	6%
		undecided	decided	undecided	undecided	
No relationship		undecided	undecided	decided	undecided	4%
		undecided	undecided	undecided	decided	

We therefore quantitatively categorize the responses of perceived and induced emotion. Similar to the relationship we defined for the categorical model, we also added the relationship *undecided* in the dimensional model, for the case where both valence and arousal ratings are undefined among listeners. Both negative and positive relationships were further divided into three levels: *valence and arousal* (valence and arousal simultaneously), *valence* (only the response of valence), and *arousal* (only the response of arousal). Table 8 gives the possible relationships between perceived and induced emotion in the dimensional model. The terms "high" (greater than zero), "low" (less than zero) and "zero" were used to represent the values for valence and arousal to avoid confusion with the relationship names. The term (high, low, and zero) with the highest number of responses was selected, but only if the highest number was more than 18 out of 25 participants' responses, or 20 out of 29 responses; otherwise "undecided" was used. These numbers of responses were selected to give a small p value ($p < .05$), given the null hypothesis that p(valence or arousal rating greater than zero) = p(valence or arousal rating less than zero) = 50% and p(zero) = 0. The relationships were calculated for each of the 80 songs, of which sixty-nine had the following relationships: *positive relationship*, *negative relationship*, *no systematic relationship*, and *no relationship*. Among these relationships, fifty-nine of them had a "positive relationship", of which thirteen were at the level of valence and arousal simultaneously, and thirty-six cases were only at the level of arousal. Three cases had "no relationship" and seven had "no systematic relationship". However, no "negative relationship" was found. The distribution of relationships is shown in Table 8.

Correspondence between Tags and Emotional Responses

Considering that these basic emotions are widely accepted across different cultures, we are able to assess the agreement between tags and participant ratings according to the extent that participants' ratings correspond with the quadrant belonging to the song's tag as shown in Figure 3. For each song, the averages of participants' valence and arousal ratings were calculated for both perceived and induced emotion, to give a centroid for each song. The quadrant of this song centroid was then compared with the expected quadrant based on the emotion tag associated with the song. The proportion of songs for which the centroid quadrant corresponded with that of the tag as well as the standard deviations (*SD*) of the valence and arousal ratings are shown in Table 9. The highest values are shown in bold. Apart from the excerpts tagged with "relaxed", more than 60% of the average valence and arousal ratings lay in the song's corresponding tag quadrant.

Fewer than 20% of ratings for songs labeled “relaxed” were located in the correct quadrant. Moreover, the standard deviation of valence and arousal ratings for both perceived and induced emotion was high, indicating that the ratings of “relaxed” excerpts were not consistent across songs.

TABLE 9. *Agreement of Valence-Arousal Ratings with Tag Quadrants, and Spread of Per-song Ratings (Averaged over Participants).*

	Happy	**Sad**	**Relaxed**	**Angry**
Perceived emotion				
Rating = Tag	0.65	0.70	0.15	0.60
Valence Mean	1.56	-1.16	0.49	-0.73
Valence SD	1.05	1.33	**1.76**	1.21
Arousal Mean	1.74	-0.91	-0.14	2.11
Arousal SD	1.57	1.43	**2.42**	1.86
Induced emotion				
Rating = Tag	0.75	0.70	0.20	0.60
Valence Mean	1.66	-0.90	0.54	-0.50
Valence SD	1.10	1.26	**1.42**	1.03
Arousal Mean	1.58	-0.87	-0.15	1.90
Arousal SD	1.38	1.35	**2.03**	1.53

Note. The highest standard deviation values in valence and arousal are bold.

However, this analysis and the results were based on ratings for each excerpt averaged across participants. To analyze the relationship between individual ratings and emotion tags, we computed the proportion of ratings that were in the same quadrant as the emotion tag for the song, and compared this with the baseline of 25% for random choice of quadrants. Wilcoxon signed-rank tests were used to test whether the agreement with the emotion tag was significantly above chance level. The results are shown in Table 10. It was found that the songs labeled with “happy” for perceived emotion had the highest agreement at 58% ($Z = 4.64$ and $p < .001$). Significant results were also found for tags “sad” ($Z = 4.20$ for perceived emotion, $Z = 3.93$ for induced emotion and $p < .001$) and “angry” ($Z = 4.57$ for perceived emotion, $Z = 4.24$ for induced emotion and $p < .001$). However, the agreement of participant ratings and the expected quadrant for songs labeled with "relaxed" was at the level of chance. In addition, as can be seen in Figures 4 and 5, showing the response distribution in each emotion category, excerpts labeled with "happy" are the most distinct, but other tags "sad" and "angry" also predict well for both perceived and induced emotion responses.

TABLE 10. *Agreement of Participant Ratings with the Quadrant of the Emotion Tag for Each Category*.

	Happy	**Sad**	**Relaxed**	**Angry**
Perceived emotion	0.58***	0.48***	0.24	0.50***
Induced emotion	0.56***	0.43***	0.26	0.47***

Note. Values above chance level according to Wilcoxon signed-rank tests are shown for the following significance levels: *$p < .05$, **$p < .01$, ***$p < .001$.

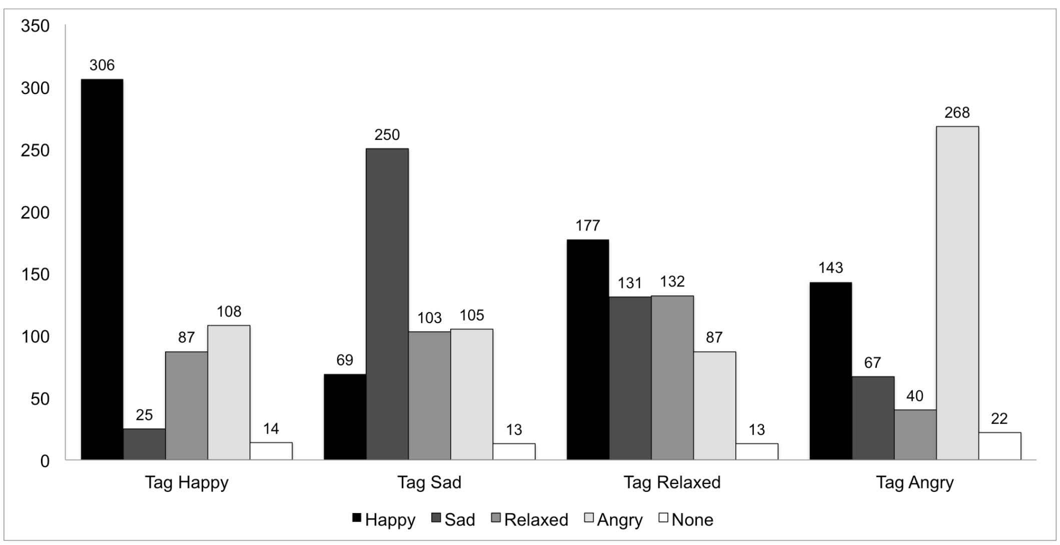


FIGURE 4. Perceived emotion response distribution for each tag. The horizontal axis shows the five responses for each of the four emotion tags *happy*, *sad*, *relaxed*, and *angry*. The vertical axis shows the number of responses.

Analysis of Individual Factors

Unlike Experiment 1, we only collected data on participants' music training via Gold-MSI v0.9. A summary of the responses can be found in Table 11.

In the study using the dimensional model, we explored the influence of musical factors such as age, gender, and music training, on the emotion judgments. Correlation analysis and Kruskal-Wallis one-way analysis of variance were used. However, the results showed no difference between male and female participants' ratings in perceived and induced responses, nor in the agreement with emotion tags. Also, no significant relationships were found between participants' judgments, age, and music training[11].

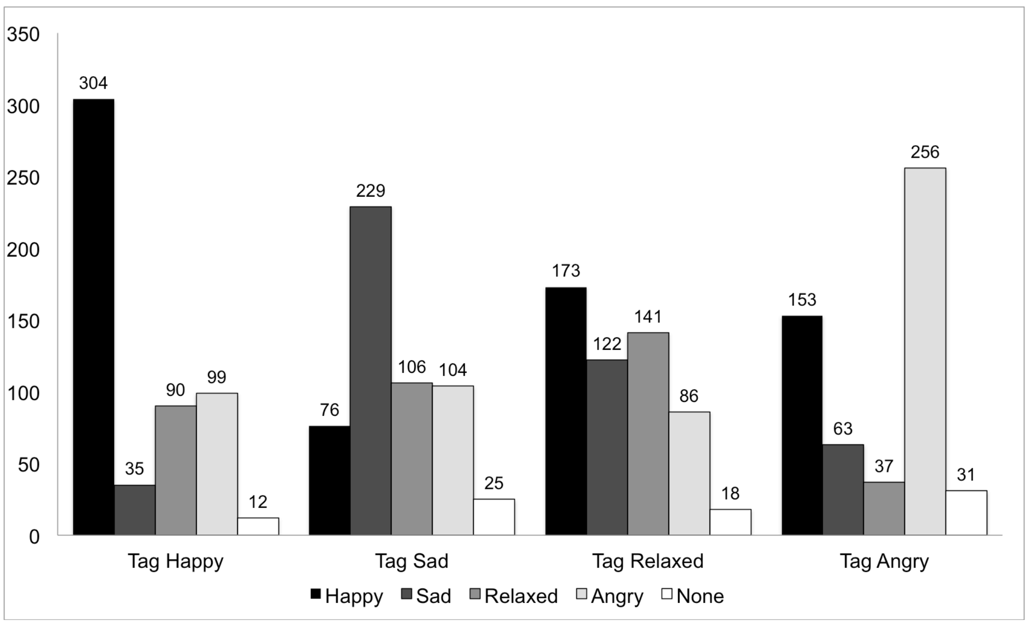


FIGURE 5. Induced emotion response distribution for each tag. The horizontal axis shows the five responses for each of the four emotion tags *happy*, *sad*, *relaxed*, and *angry*. The vertical axis shows the number of responses.

TABLE 11. *Summary Scores for Music Training to 9 Questions Adapted from the Gold-MSI.*

	Scale Min	Scale Max	Mean	SD
Music training	9	63	34.8	13.98

Note. For example, a music training score of 26 could mean that the participant can play one musical instrument, had one year of formal training, and practices the instrument for one hour daily.

[11] Only age categories of "18-24" and "25-34" were considered, as the number of the participants aged over 35 was very small.

Discussion

Experiment 2 investigated the associations between social tags and judgments of perceived emotion and induced emotion using the *valence-arousal* two-dimensional model of emotion. No significant difference was found between the rating conditions, perceived and induced emotion. However, the four emotion tags differentially predicted the ratings of valence and arousal. The between-participant agreement on perceived and induced emotion was then measured by a Wilcoxon signed-rank test. The result showed a higher consistency among participants for perceived emotion. This agrees with previous studies on classical music using the dimensional model, which also suggested that consistency of arousal is higher than valence consistency. The same results were found in this study of Western popular music whether instrumental tracks were included or excluded from consideration[12]. Additionally, the analyses between emotional responses and listening duration showed that feeling and recognizing emotion "happy" in music took significantly less time than experiencing emotions such as "relaxed" and "sad". It also indicates that in the two dimensional model of emotion, emotions with high arousal are easier to identify and distinguish.

Furthermore, we studied the relationship between perceived and induced ratings. As expected, strong positive correlations between the two conditions for both ratings of valence and arousal were found. These results again suggest that listeners will typically feel the emotions expressed by the music. Due to the fact that perceived and induced emotion are often difficult to clearly delineate, a quantitative analysis was used to measure the possible relationships. The results showed that "positive relationship" dominates, covering 64% of the cases, and no cases of "negative relationship" were found. Due to the strict threshold, only the strongly consistent responses across participants were retained. Comparing Tables 7 and 10, the levels of participant agreement among themselves are higher than the agreement with the emotion tags, suggesting

[12] There were not enough data points (4 in song list 1, and 6 in song list 2) to analyse instrumental tracks separately.

that at least some of the tags do not correspond either with participants' perceived or induced emotion. It might also suggest that the controlled collection of responses creates more systematic results than data created for other purposes without specific instructions.

Thirdly, the reliability of emotion tags in the dimensional model was evaluated. A mapping between *valence*, *arousal*, and four emotion tags was created, so that each emotion occupies a unique quadrant of the two-dimensional plane. The analyses indicated that songs labeled with "happy", "sad", and "angry" had ratings in the corresponding quadrants of the valence-arousal plane at a level that was significantly above chance. For songs tagged "relaxed", however, the agreement of ratings with the positive-valence, negative-arousal quadrant was at chance for both perceived and induced emotion. Comparing these four tags, regardless of song or person, the excerpts tagged with "happy" are most likely to produce responses in the corresponding quadrant of the valence-arousal plane.

Finally, we explored the influence of individual factors on the responses of perceived and induced emotion. Gender, age, and participants' music training did not significantly mediate any relationships.

General Discussion and Future Work

This paper examined the relationship between perceived and induced emotion, and evaluated the reliability of emotion tags from music discovery websites as well as individual differences in emotion judgments. The majority of the previous studies on music and emotion deal with classical and film music, but as previous results may not be applicable to other genres, this work extends the study to Western popular music. In addition, two models of emotions, the *categorical* and *dimensional* models, were used in our experiments to structure the investigations and interpret our findings. Though the participants of the two experiments were recruited in different ways, the distribution of key variables (age and music training score) did not differ significantly between experiments[13].

Results for both the categorical and the dimensional model showed that the four emotion labels "happy", "sad", "relaxed", and "angry" did correlate with the perceived and induced emotional responses. The inter-participant consistency in perceived emotion was, however, significantly higher than the consistency in induced emotion. This supports the argument that perceived emotion refers to intellectual processing (objective), such as the perception of an intended or expressed emotional character, whereas felt emotions reflect the introspective perception of psychophysiological changes (subjective), which are often associated with emotional self-regulation. But also, it may be unreasonable to expect induced emotions to change so rapidly from moment to moment and the results could be different in extended listening of a single piece. The excerpts used in these experiments were 30-60 seconds in duration. Since participants were able to stop whenever they wished, the listening time for each song was recorded. The analysis of listening duration and emotional responses showed that for both induced and perceived conditions, emotions such as "happy" and "angry" were easier to recognize and feel in music, whereas participants needed a significantly longer time to experience emotions such as "sad" and "relaxed", which are located in the lower region of the dimensional model (low arousal). However, no significant differences in listening duration were found between induced and perceived emotion responses.

Similarly, a higher uncertainty level was found for induced emotional responses. To an extent, it suggests that these four basic emotions better capture the musical emotions which are expressed in music than those induced by music. The feedback from participants also implies that the categorical model is inadequate to describe the richness of emotions induced by music. In the analysis of the dimensional model, higher consistency was also found in the ratings of arousal than of valence, which is in accordance with previous results in music emotion prediction (Schubert, 2007; Huq et al., 2010). The majority of music

[13] The same participant could take part in both experiments.

excerpts used in the experiment had lyrics, and the overall results remained the same after removing the instrumental tracks. However, it is possible that emotional judgments could be related not to the elements of music structure but to the meaning expressed by the lyrics.

In addition, positive correlations were found for corresponding emotions between perceived and induced emotion. This result is consistent with the finding that music evokes emotions similar to the emotions perceived in music. However, other correlations showed that a perceived sad song might induce relaxedness. Meanwhile, a perceived angry song is less likely to induce sadness and relaxedness. In addition, we tested for four possible relationships proposed by Gabrielsson: *positive relationship*, *negative relationship*, *no systematic relationship*, and *no relationship*. We found that in both the categorical and dimensional models, a positive relationship is the most frequent one. Negative relationships were found in our analysis of the categorical model, but not in the dimensional model. This may be caused by the fundamental differences between the models, that one allows people to respond in a graded way whereas the other forces them into a small number of discrete categories; or this may be caused by the thresholding, that only the strongest response is retained in the dimensional model. "No systematic relationship" and "no relationship" also existed for a small minority of stimuli. There was one special case in the categorical model, where a song perceived to be equally happy and angry induced only happiness. Our empirical results follow Gabrielsson's (2002) framework; however, the case in which both the perceived and induced emotions are disagreed by listeners was not included in this framework. We therefore expand Gabrielsson's framework with the relationship "*undecided*". There are various underlying mechanisms involved in explaining the relationships between induced and perceived emotion. For example, Juslin and Västfjäll (2008) mentioned that the same perceived and induced emotional response is due to the emotional contagion mechanism, whereas the negative relationship may involve the episodic memory mechanism. Details of underlying mechanisms are beyond the scope of this paper, but for future studies it will be useful to consider all relevant mechanisms to provide an accurate understanding of the relationships between induced and perceived emotion.

The reliability of emotion tags was evaluated via the level of agreement between participants' responses and the tags. The results revealed that the agreement between social tags and participants' ratings was well above chance for both the categorical model and the dimensional model. However, the excerpts labeled with "relaxed" had the lowest agreement with tags in the categorical model and agreed with the tags only at the level of chance in the dimensional model. Moreover, the distribution of listeners' responses for the excerpts corresponding to each tag indicated that the emotion tags predicted both the perceived and induced emotional responses well. It is worth mentioning that with respect to valence, listeners often collectively confused sadness with relaxedness, and happiness with anger. Responses may vary because of individual factors such as current mood, music culture, and preference.

Finally, the analysis of individual differences showed that *age*, *gender*, and *music training* were found not to mediate listeners' emotion judgments, nor their agreement with tags. One interpretation is that the excerpt selections are Western popular music which is generally well-understood by participants who are English speakers. However, a tendency was found that older participants were more likely to feel and experience less "happy", and use music more for its emotional functions. Also, positive correlations between three factors from the Gold-MSI v0.9, *importance* and *music training*, *emotion* were found in Experiment 1.

We also need to acknowledge two potential limitations of collecting both perceived and felt emotion responses in a laboratory setting. First, it is likely that people do not feel as much in this setting as they would in a natural condition during which listeners are actively choosing music pieces for obtaining a given internal feeling (Altenmüller et al., 2002). However, we believe that whereas the setting might affect quantitative responses (how much is felt), it should not alter the qualitative responses (which emotion is felt). Second, it is possible that participants might confuse the two rating conditions. However, they were given very precise instructions and were reminded of the two different rating conditions (perceived and induced emotion) just prior to the ratings. Also, the obtained significant differences between the conditions show that, on the whole, the participants understood the tasks.

In summary, our study of the relationship between perceived and induced emotion showed similar results using both the categorical and dimensional models. This study supports the previous work on classical and film music, suggesting that the robustness of the models does not depend on the genre of music considered. Emotion, like music, is dynamic and it may change and evolve continuously. In our experiments, emotional response was measured by a static value. Future research should consider the dynamic emotional judgments (Egermann et al., 2013). Musical emotional meaning can not only be influenced by subjective factors (taste, musical abilities, and personality), but also social factors (music culture and context). In further studies, we explore music emotional judgment in various musical cultures and contexts. Also, more objective measurements such as behavioral and physiological reactions could be conducted in combination with self-report to provide richer evidence of emotional responses. A greater understanding of these factors would be beneficial in the design of subjective music recommendation systems (Song et al., 2012a).

Acknowledgement

This study was supported by the School of Electronic Engineering and Computer Science at Queen Mary University of London. We would like to thank the China Scholarship Council (CSC) for financial support and the Centre for Digital Music (C4DM). ARH was partly supported by a Leverhulme Visiting Professorship during her sabbatical year at Queen Mary. We are very grateful to all the participants, reviewers, and colleagues for their valuable suggestions.

References

Altenmüller, E., Schürmann, K., Lim, V. K., and 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.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., and 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.

Chamorro-Premuzic, T. and Furnham, A. (2007). Personality and music: Can traits explain how people use music in everyday life? *British Journal of Psychology*, 98(2):175-185.

Charles, S. T., Reynolds, C. A., and Gatz, M. (2001). Age-related differences and change in positive and negative affect over 23 years. *Journal of Personality and Social Psychology*, 80(1):136-151.

Dalla Bella, S., Peretz, I., Rousseau, L., and Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, 80(3):B1-10.

DeNora, T. (2000). *Music in everyday life*. Cambridge University Press, Cambridge, UK.

Dixon, S. (2001). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39-58.

Eerola, T. (2011). Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *Journal of New Music Research*, 40(4):349-366.

Eerola, T. and Vuoskoski, J. K. (2010). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1):18-49.

Eerola, T. and Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception*, 30(3):307-340.

Egermann, H., Pearce, M. T., Wiggins, G. A., and McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective & Behavioral Neuroscience*, 13(3):533-553.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3/4):169-200.

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

Frijda, N. H. (1986). *The emotions*. Cambridge University Press, Cambridge, UK.

Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, 5(1):123-147.

Gabrielsson, A. and Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, 24(1):68-91.

Gross, J. J., Carstensen, L. L., Tsai, J., Skorpen, C. G., and Hsu, A. Y. C. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging*, 12(4):590-599.

Huq, A., Bello, J. P., and Rowe, R. (2010). Automated music emotion recognition: A systematic evaluation. *Journal of New Music Research*, 39(3):227-244.

Ilie, G. and Thompson, W. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23(4):319-330.

Juslin, P. N. and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5):770-814.

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

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., and Silva, A. (2008). An experience sampling study of emotional reactions to music: Listener, music, and situation. *Emotion*, 8(5):668-683.

Juslin, P. N. and Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *The Behavioral and Brain Sciences*, 31(5):559-621.

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

Khalfa, S., Isabelle, P., Jean-Pierre, B., and Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, 328(2):145-149.

Killgore, W. (1999). Affective valence and arousal in self-rated depression and anxiety. *Perceptual and Motor Skills*, 89(1):301-304.

Kosta, K., Song, Y., Fazekas, G., and Sandler, M. B. (2013). A study of cultural dependence of perceived mood in Greek music. In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, pages 317-322, Curitiba, Brazil.

Kreutz, G., Ott, U., Teichmann, D., Osawa, P., and Vaitl, D. (2007). Using music to induce emotions: Influences of musical preference and absorption. *Psychology of Music*, 36(1):101-126.

Lamere, P. (2008). Social tagging and music information retrieval. *Journal of New Music Research*, 37(2):101-114.

Laurier, C., Grivolla, J., and Herrera, P. (2008). Multimodal music mood classification using audio and lyrics. In *Proceedings of the 7th International Conference on Machine Learning and Applications (ICMPLA)*, pages 1-6, San Diego, USA.

Levenson, R. W. (2003). Autonomic specificity and emotion. In *Handbook of Affective Sciences*, pages 212-224. Oxford University Press, Oxford, UK.

Levy, M. and Sandler, M. B. (2009). Music information retrieval using social tags and audio. *IEEE Transactions on Multimedia*, 11(3):383-395.

Mas-Herrero, E., Zatorre, R. J., Rodriguez-Fornells, A., and Marco-Pallarés, J. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Current Biology*, 24(6):699-704.

Mather, M. and Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences*, 9(10):496-502.

McAdams, D. P. (2006). *The person*. John Wiley, New Jersey, USA.

McAdams, S., Vines, B. W., Vieillard, S., Smith, B. K., and Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception*, 22(2):297-350.

Mohn, C., Argstatter, H., and Wilker, F. (2010). Perception of six basic emotions in music. *Psychology of Music*, 39(4):503-517.

Mroczek, D. K. and Kolarz, C. M. (1998). The effect of age on positive and negative affect: A developmental perspective on happiness. *Journal of Personality and Social Psychology*, 75(5):1333-1349.

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS ONE*, 9(2):e89642.

Müllensiefen, D., Gingras, B., Stewart, L., and Musil, J. (2012). Goldsmiths Musical Sophistication Index (Gold-MSI): Technical report and documentation. Technical report, Goldsmiths, University of London, London, UK.

Panksepp, J. (1998). *Affective neuroscience: The foundation of human and animal emotions*. Oxford University Press, Oxford, UK.

Rentfrow, P. J. and 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.

Repp, B. H. (1994). On determining the basic tempo of an expressive music performance. *Psychology of Music*, 22:157-167.

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

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

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives. *Frontiers in Psychology*, 4(December):1-18.

Shiota, M., Keltner, D., and John, O. (2006). Positive emotion dispositions differentially associated with Big Five personality and attachment style. *The Journal of Positive Psychology*, 1(2):61-71.

Sloboda, J. A., O'Neill, S. A., and Ivaldi, A. (2001). Function of music in everyday life: An exploratory study using the experience sampling method. *Musicae Scientiae*, 5(1):9-32.

Song, Y., Dixon, S., and Pearce, M. T. (2012a). A survey of music recommendation systems and future perspectives. In *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, pages 395-410, London, UK.

Song, Y., Dixon, S., and Pearce, M. T. (2012b). Evaluation of musical features for emotion classification. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, pages 523-528, Porto, Portugal.

Song, Y., Dixon, S., Pearce, M. T., and Fazekas, G. (2013a). Using tags to select stimuli in the study of music and emotion. In *Proceedings of the 3rd International Conference on Music & Emotion (ICME)*, Jyväskylä, Finland.

Song, Y., Dixon, S., Pearce, M. T., and Halpern, A. R. (2013b). Do online social tags predict perceived or induced emotional responses to music? In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, pages 89-94, Curitiba, Brazil.

Thayer, J. and Faith, M. (2001). A dynamic systems model of musically induced emotions. *Annals of the New York Academy of Sciences*, 930:452-456.

Truong, K. P., Leeuwen, D. A. V., Neerincx, M. A., and Jong, F. M. G. D. (2009). Arousal and valence prediction in spontaneous emotional speech: Felt versus perceived emotion. In *Proceedings of the Interspeech Conference*, pages 2027-2030, Brighton, UK.

Turnbull, D., Barrington, L., and Lanckriet, G. (2008). Five approaches to collecting tags for music. In *Proceedings of the 9th International Society for Music Information Retrieval (ISMIR)*, pages 225-230, Philadelphia, USA.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., and Bernard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, 22(4):720-752.

Vuoskoski, J. K. and Eerola, T. (2011a). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, 15(2):159-173.

Vuoskoski, J. K. and Eerola, T. (2011b). The role of mood and personality in the perception of emotions represented by music. *Cortex (Special Section on Music in the Brain): Research Report*, 47(9):1099-1106.

Vuoskoski, J. K., Thompson, W. F., and Eerola, T. (2011). Who enjoys listening to sad music and why? *Music Perception*, 29(3):311-318.

Wang, D., Li, T., and Ogihara, M. (2010). Are tags better than audio features? The effect of joint use of tags and audio content features for artistic style clustering. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, pages 57-62, Miami, USA.

Welch, N. and Krantz, J. H. (1996). The World-Wide Web as a medium for psychoacoustical demonstrations and experiments: Experience and results. *Behavior Research Methods, Instruments, and Computers*, 28(2):192-196.

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

A Appendix - Participants Information

TABLE A1. *Statistics for 40 and 54 Participants in Experiment 1 and 2.*

Participants	Category	Nos. in Experiment 1	Nos. in Experiment 2
Age	<18	0	2
	18-24	22	23
	25-34	14	25
	35-44	4	1
	45-54	0	3
Gender	Male	20	25
	Female	20	29
Nationality	Chinese	13	10
	British	6	4
	Greek	3	7
	Polish	2	2
	Italian	2	4
	French	0	2
	Slovenian	0	2
	Taiwanese	0	2
	American	0	2
	Argentinian	0	2
	Canadian/Romanian	1	1
	English/Australian	0	1
	British-Pakistani	1	0
	British-Indian	1	0
	Iranian	1	1
	Indian	1	1
	Canadian	1	1
	Belgian	1	1
	Pakistani	1	1
	Romanian	1	1
	Portuguese	1	1
	Australian	1	1
	Lithuanian	1	0
	Croatian	1	0
	German	1	0
	Cypriot	0	1
	Moroccan	0	1
	Spanish	0	1
	Jamaican	0	1
	Egyptian	0	1
	Sri Lankan	0	1
	Asian - other	0	1

B Appendix - Activities Involving with Music Listening and its Purposes

TABLE B1. *Participants' Feedback for Activities Involving Music Listening and its Purposes*.

Activity	No. responses	Purposes	No. responses
Commuting/travelling (walking, tube, driving)	22	Relax	12
Exercising (gym)	20	Engage	2
Studying	13	Keep me away	1
Dancing (club)	11	Motivate	1
Party	11	Synonym to music	1
Working	5	Feel excited	1
Sleeping	4	Help focus	1
Cooking	4	Make atmosphere	1
Chatting/talking with friends	4	Keep me company	1
Concerts/Watching band	4	Feel happy	1
Surfing the internet	3	Relax my brain	1
Reading	3		
Housework	3		
Eating	2		
Socializing	1		
Background	1		
Practicing instrument	1		
Making music	1		
Music research	1		
Mechanical work	1		
Sitting/waiting	1		
Dressing up	1		
Copying something	1		

C Appendix - List of Stimuli Used in the Experiments

TABLE C1. *List of Stimuli Used in the Experiments*.

Artist	Song title	Tag	Artist	Song title	Tag
Mary J. Blige	Be Happy	Happy	Sunparlour Players	If The Creeks Don't Rise	Happy
The Busters	Skank Down	Happy	Ezio	Steal Away	Happy
Muse	Starlight	Happy	Dragonette	Get Lucky	Happy
Die Happy	Cry For More	Happy	Die Happy	Supersonic Speed	Happy

Toploader	Dancing In The Moonlight	Happy	Uner	Labaneria	Happy
Louis Armstrong	Hello Dolly	Happy	13th Floor Elevators	I'm Gonna Love You Too	Happy
Kakkmaddafakka	Restless	Happy	Amy Winehouse	Valerie	Happy
Al Green	Nothing Impossible With Love	Happy	MGMT	Kids	Happy
Passion Pit	Make Light	Happy	Hymns	St. Sebastian	Happy
Die Happy	Genuine Venus	Happy	The Go! Team	Ladyflash	Happy
Die Toten Hosen	Böser Wolf	Sad	Fleetwood Mac	Dreams (LP Version)	Sad
James Taylor	Fire And Rain (LP Version)	Sad	Paul McCartney	Here Today	Sad
Iliketrains	Death Is The End	Sad	Tsunami	Fits And Starts	Sad
TLC	Unpretty	Sad	HIM	Love's Requiem	Sad
In This Moment	World In Flames	Sad	Birdy	Skinny Love	Sad
Die Toten Hosen	Was Zählt	Sad	54.4	One Gun (Album Version)	Sad
Santana Feat. Steven Tyler	Just Feel Better	Sad	The Secret Show	Old Blacktop	Sad
Avantasia	Inside	Sad	Richard Youngs	Broke Up By Night	Sad
The Veronicas	Heavily Broken (Live Version)	Sad	Giorgia	Gocce Di Memoria	Sad
Wu-Tang Clan	Josephine	Sad	Shania Twain	It Only Hurts When I'm Breathing	Sad
Tok Tok Tok	Walk On The Wild Side	Relaxed	Scorpions	Destiny	Relaxed
ATB	Mysterious Skies	Relaxed	Seven Foot Wave	In The Ocean	Relaxed
Blur	Sweet Song	Relaxed	Planet Funk	Chase The Sun	Relaxed
Carbon Based Lifeforms	Mos 6581	Relaxed	Free Planet Radio	Dhijaz	Relaxed
Kay Kyser	On A Slow Boat To China	Relaxed	Mia Moi Todd	Digital	Relaxed
The Mahotella Queens	Mbube	Relaxed	Bohren & Der Club of Gore	Karin	Relaxed
Chicane	Halcyon	Relaxed	Tori Amos	Crucify (LP Version)	Relaxed
Nightmares on Wax	The Sweetest	Relaxed	Joe Satriani	Come On Baby	Relaxed
Future Sound of London	Papua New Guinea (12" Version)	Relaxed	Red Hot Chili Peppers	Tell Me Baby	Relaxed
Zwan	Honestly (Album Version)	Relaxed	Stephan Micus	Flowers In Chaos	Relaxed
Manic Street Preachers	Motown Junk	Angry	The Distillers	Drain The Blood	Angry
Wumpscut	Bunkertor 7 (German Texture)	Angry	Skinny Puppy	Anger	Angry
Dido	Stan	Angry	Savage Garden	Gunning Down Romance	Angry
Metric	Hustle Rose	Angry	Skinny Puppy	Scrapyard	Angry
Natalie Imbruglia	Want	Angry	Vanessa Carlton	Paint It Black	Angry
Hole	Violet	Angry	Kittie	Pain (Live Version)	Angry
Fear Factory	Edgecrusher (Urban Assault Mix)	Angry	Rilo Kiley	A Better Son/Daughter	Angry
Incubus	Blood On The Ground	Angry	Texas	Summer Son	Angry
Soulfly	Arise Again (Album Version)	Angry	4LYN	Incomplete	Angry
Three Days Grace	I Hate Everything About You	Angry	Stone Sour	Cold Reader	Angry

Author Note

Correspondence concerning this article should be addressed to Yading Song.

Yading Song, Student, Centre for Digital Music, School of Electronic Engineering and Computer Science, Queen Mary University of London.

Address: Queen Mary University of London Mile End Road London E1 4NS United Kingdom Email: y.song@qmul.ac.uk Tele: +44(0)7936356089