Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th - 15th June 2013. Geoff Luck & Olivier Brabant (Eds.)

USING TAGS TO SELECT STIMULI IN THE STUDY OF MUSIC AND EMOTION

Yading Song, Simon Dixon, Marcus Pearce, György Fazekas

Centre for Digital Music
Department of Electronic Engineering and Computer Science
Queen Mary University of London, United Kingdom
yading.song@eecs.qmul.ac.uk

Abstract

A wealth of literature on musical emotion exists, including investigation of the use of tags to classify musical emotions. However, the relationship between musical emotions and human annotated information is still unclear. Likewise, the understanding of the differences between induced emotion (also known as felt emotion) and perceived emotion (also known as expressed emotion) is at an early stage. In previous work, lists of songs labelled with one of the four basic emotion tags "happy", "sad", "angry" and "relaxed" were retrieved from Last.FM, and audio excerpts were fetched from 7Digital.com. In this study, we asked listeners to rate musical excerpts with the perceived or induced emotion fitting the excerpt. 80 excerpts (20 for each of the four emotions considered) were rated by 40 participants from various backgrounds and levels of musical expertise. The results show that in majority of the selected songs the tags agreed more closely with the ratings of perceived emotion than induced emotion. In addition, each induced emotion was highly correlated with its corresponding perceived emotion and induced anger can also be very distinct from its perceived ratings. However, the participants' emotional judgements were not related to measured cultural or musical factors.

Keywords: perceived emotion, induced emotion, social tags

1. Introduction

Music provides powerful means of communication and self-expression and has attracted increasingly significant research interest in the past decades (Eerola and Vuoskoski, 2013; Barthet et al., 2013). People report that their primary motivation for listening to music is its emotional effect (Juslin and Laukka, 2004). Additionally, research indicates that people are listening to music more often than any of the other activities (Rentfrow and Gosling, 2003) (i.e. watching television, reading books, and watching movies).

With the emergence of music discovery websites such as Last.FM[1], social tags have received increasing interest for the study of music and emotion in the past ten years (Eck et al., 2007; Levy and Sandler, 2009; Wang et al., 2010). Social tags are words or groups of words supplied by a community of internet users. They are more and more commonly used to aid navigation through large media collections (Wu et al., 2006), allowing users to get a sense of what qualities characterise a song at a glance (Hoffman et al., 2009). Compared with traditional human annotation, semantic tags provide large-scale, cost-efficient, rich and easily accessible source of metadata (Turnbull et al., 2008). In addition, the information they provide is highly relevant to music

1. http://www.last.fm

information retrieval, including genre, mood and instrument, which account for 70% of the tags (Lamere, 2008).

Though the use of social tags is a powerful tool that can assist searching and the exploration of music (Levy and Sandler, 2007), several problems with tags have been identified, such as the "cold start" problem (new or unknown music has no tags), noise, malicious tagging, and bias towards popular artists or genres (Lamere, 2008). There are a number of incentives and motivations for tagging, such as to aid memory, provide context for task organisation, social signalling, social contribution, play and competition, and opinion expression (Ames and Naaman, 2007). However, we know very little about the criteria on which tagging is based.

Music can both induce and express emotion, and it is a fundamental difference that induced emotion (also known as felt emotion) is the emotion experienced by the listener, while perceived emotion (also known as expressed emotion) is the emotion recognised in music (Gabrielsson, 2002). However, separating induced emotion from perceived emotion is not always straightforward. Previous studies have suggested that music seems to induce emotions similar to the emotional quality perceived in music. Felt emotions are often stronger than perceived emotion in connection with pleasure, but weaker in connection with arousal, positive activation, and negative activation (Kallinen and Ravaja, 2006; Evans and Schubert, 2008). To our knowledge, these two facets of emotion communication in music were rarely studied in comparison with semantic tags.

In the study of music and emotion, two popular theoretical frameworks, the categorical model and dimensional model, have both received empirical support (Vieillard et al., 2008; Eerola and Vuoskoski, 2010). The dimensional approach considers all affective terms arising from independent neurophysiological systems: one related to valence (a pleasure-displeasure continuum) and the other to arousal (activation-deactivation) (Russell, 1980). In contrast, the discrete or categorical model describes all emotions as being derived from a limited number of universal and innate basic emotions such as anger, happiness, sadness and fear. The experimental work in this paper is based on the discrete model, since the data was collected through human-annotated social tags, which are categorical in nature (Ekman, 1992; Panksepp, 1998). We used four basic emotion classes: *happy*, *angry*, *sad* and *relaxed*, considering these four emotions were widely accepted across different cultures and covered the four quadrants of the two dimensional model of emotion (Laurier and Grivolla, 2008).

Gabrielsson (Gabrielsson, 2002) stated that people generally agree on the basic emotion that a particular piece of music is expressing. In this paper, we studied whether the same agreement exists between human-annotated tags and emotional judgements. We examine a number of questions: (1) Do semantic tags agree show a greater agreement with perceived emotion or with induced emotion? (2) How can we quantify the reliability of tags, for example, for selecting stimuli in the study of music and emotion? (3) What are the influences of musical education, music engagement and culture on the attribution of emotion to music?

2. Method

2.1. Participants

Forty English-speaking students (20 male and 20 female) from undergraduate to postgraduate participated in this study. They were recruited through university email lists, and had ages ranging from 18 to 44 years (age 18-24: 55%; age 25- 34: 35%; age 35-44: 10%) with various educational and musical training backgrounds. Among these participants, 32.5% of them were Chinese and 50% of the participants attentively listened to music more than one hour per day. Also 87.5% of the participants can play at least one instrument. Moreover, 67.5% of the participants prefer pop/rock music, 10% of them prefer jazz, and the rest prefer classical music.

To assess the participants' musical expertise and engagement the Goldsmiths Musical Sophistication Index questionnaire (GOLD-

MSI) was given (see subsection 2.3). The three factors measured were *importance* (importance of the music in everyday life), *musical training* (life history of formal musical training) and *emotion* (importance of music for psychological functions) (Müllensiefen et al., 2012). The values were calculated in a provided template[2] (importance (min: 15 - max: 105), musical training (min: 9 - max: 63), emotion (min: 8 - max: 56)). A summary of the responses can be found in **Table 1**.

Skills	Min	Max	Mean	SD
Importance	26	99	72.5	18.1
Musical Training	9	59	37.4	15
Emotion	30	56	45.5	6.7

Note: 32 questions adapted from GOLD-MSI

Table 1. Summary of Musical Attributes

In order to minimise the effect of song sequence and conditions, the study was conducted in four equal groups (n=10 for each group). The order of presentation of the two rating conditions (perceived and induced emotion) and two song blocks (m=40, 10 for each emotion category) was counterbalanced across subjects. The songs in each block were randomly distributed across participants (see **Table 2**).

Group	Block 1	Block 2
Group 1	Induced Emotion	Perceived Emotion
Group 2	Perceived Emotion	Induced Emotion
	Block 2	**Block 1**
Group 3	Induced Emotion	Perceived Emotion
Group 4	Perceived Emotion	Induced Emotion

Table 2. Group allocation

2.2. Stimuli

The stimuli were selected from a collection of 2904 excerpts retrieved from Last.FM[3] and 7Digital[4], which were previously used in music and emotion studies (Song et al., 2012). Each excerpt had been tagged on Last.FM with one of the four words "*happy*", "*sad*", "*angry*" and "*relaxed*". We randomly chose a total of 80 excerpts from these four categories (n=20 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. Each excerpt was either 30 seconds or 60 seconds long which was randomly selected by 7Digital, and it was played from a standard mp3 format file (bitrate: 128 kbps or 64 kbps, sample rate: 22,050 kHz or 44,100 kHz).

2.3. Procedure

The study was approved by Queen Mary Research Ethics Committee (REF: QMREC1019). Participants were asked to sit in a laboratory environment. They were given studio quality headphones, and they could adjust volume to a comfortable level. First of all, participants were asked to familiarise themselves with the interface and read an instruction page, for example:

1. Listen to the songs one after one (30 or 60 seconds).
2. From page 1-4, you will be asked to answer "What emotion do you feel in response to the music" (induced ratings) and from page 5-8, you will be asked to answer "How would you describe the emotional content of the music itself" (perceived ratings).
3. For each single track, once you are sure about the question, click the "stop" button on the audio player.
4. Choose from one of the options: happy, sad, relaxed, angry, or cannot tell/none of above.
5. Note: you can only listen to each song once.

The participants filled in a demographic form including name, age, gender, "type of music they are most familiar with", nationality, and "music culture they grew up with". Then they responded to each excerpt (n=10 per

2. http://www.gold.ac.uk/music-mind-brain/gold-msi/
3. http://www.last.fm/home
4. http://www.7digital.com/

page) and chose one of corresponding emotions: *happy*, *sad*, *relaxed*, *angry* and *none/cannot tell*. They also indicated whether they knew the song by selecting *yes* or *no*.

At the end of the experiment, their opinions and feedback were collected. A selected Goldsmiths Musical Sophistication Index (GOLD-MSI) questionnaire was given to measure participants' musical expertise and engagement (see section 2.1. The whole experiment lasted about an hour without any breaks. However, the participants were able to stop whenever they wanted.

3. Results

Depending on the group (refer to **Table 2**), each participant rated the perceived emotion and induced emotion for different blocks. Therefore, we analysed perceived ratings and induced ratings separately. The data analysis was conducted using the Matlab 2012 Statistics Toolbox. We considered the results with and without the none/cannot tell option. However, as similar conclusions were reached in the analyses, we only present the output with none/cannot tell option included in this section.

3.1. Comparison of agreement of perceived and induced emotion ratings with social tags

In order to compare the agreement of perceived emotion and induced emotion results with social tags, a Wilcoxon Signed Rank test was used for the entire 80 excerpts (see **Table 3**). The analyses revealed that though both are well above chance, the agreement of the perceived emotion ratings with tags was ranked significantly higher than that of the induced data (*p* values <0.0001).

	Perceived	Induced
Median	0.5	0.35
Mean	0.4694	0.3962
Standard Deviation	0.2919	0.2385
P-value	6.78E-04	
Zvalue	-3.3983	

Table 3. Wilcoxon Signed Rank Test

3.2. Differences across emotion categories

Though section 3.1 showed perceived emotion is closer to human tags, for each emotion a Wilcoxon Signed Rank test was carried out to analyse the individual differences (see **Table 4**). In the case of happiness (induced mean = 0.47 and perceived mean = 0.45) and relaxedness (induced mean = 0.39 and perceived mean = 0.37), interestingly, the agreement for induced emotion and perceived emotion showed no significant difference (p = 0.51 and p = 0.52 respectively). However, the agreement of tags for anger and sadness in perceived emotion were significantly higher that the agreement for induced emotion (p = 1.6e-04 and p = 6.78e-04).

	Happy	Sad	Relaxed	Angry
Induced	**0.47**	0.36	**0.39**	0.37
Perceived	0.45	**0.56**	0.37	**0.51**
P-value	0.51	*1.6E-04*	0.52	*6.78E-04*
Zvalue	0.66	-3.77	0.64	-3.4

Table 4. Agreement with tags for each emotion

In order to study the relationships between perceived and induced emotions, correlation analyses were performed, with the results shown in **Table 5**. Positive correlations between perceived and induced emotions for each corresponding emotion were found ($p < 0.001$), and several other significant correlations were revealed such as negative correlations between induced happiness and perceived sad- ness ($r = -0.59$ and $p < 0.05$), and induced anger with perceived happiness ($r = -0.57$ and $p < 0.01$), sadness ($r = -0.77$ and $p < 0.001$) and relaxedness ($r = -0.69$ and $p < 0.001$).

	I Happy	I Sad	I Relax	I Angry
PHappy	***0.76	*-0.56	*-0.52	**-0.57
PSad	*-0.59	***0.89	0.13	***-0.77
PRelaxed	-0.27	*-0.48	***0.73	***-0.69
PAngry	-0.23	-0.44	-0.26	***0.96

*p <0.05, **p <0.01, ***p <0.001

Table 5. Correlations between perceived (P) and induced (I) emotions, for each emotion

3.3. Cultural and musical factors analysis

To discover whether there was any association between gender, cultural background, or musical training and engagement with the musical emotion judgements, a Wilcoxon Signed-Rank test, Mann Whitney U test and correlation analysis were carried out respectively. No significant difference was found between male and female in their agreement with tags (p = 0.9038). In addition, there was no significant impact on musical judgement for Asian (n=15) and Western (m=25) participants (p = 0.7474)[5]. As shown in **Table 6**, no correlation was found between perceived and induced agreement with the three musical attributes: importance, musical training and emotion. However, there is a weak negative correlation with musical training in induced emotion for group 2 and group 3 (r = -0.5137 and p < 0.05).

	Group 1 and Group 4	
	Induced	**Perceived**
Importance	-0.0877	-0.3107
Musical Training	0.0134	0.1062
Emotion	-0.1784	-0.2594
	Group 2 and Group 3	
	Perceived	**Induced**
Importance	0.2399	-0.3154
Musical Training	-0.0772	*-0.5137
Emotion	0.1115	-0.1897

*p <0.05, **p <0.01, ***p <0.001

Table 6. Correlation between levels of tag-rating agreement and the GOLD-MSI scale, where tag-rating agreement can be divided into induced and perceived emotion

4. Discussion

The experiment reported in this paper aimed at finding associations between social emotion tags (happy, sad, angry and relaxed) and participants' emotional judgements (perceived and induced emotion). First, we found that participants' ratings of perceived emotion showed a significantly higher level of agreement with tags than that shown by their induced emotion ratings. Furthermore, we tested the difference for each individual emotion. The results revealed that there is no significant difference for the happy and relaxed tags, but for the excerpts labelled with angry and sad, these labels are more likely to correspond to perceived emotion (p < 0.001).

Second, we studied the correlations between perceived and induced emotions. As expected, there is a highly significant positive correlation between induced emotion and its corresponding perceived emotion. Previous studies mentioned that happy and sad excerpts were identified easily (Terwogt and Grinsven, 1991; Vieillard et al., 2008). However, a highly negative correlation between induced anger and the perceived emotions of happiness, sadness and relaxedness showed that induced anger can also be very distinct from its perceived emotion. Despite the fact that individual differences such as personality traits and listeners' current mood are relevant to music preference and emotional judgement (Vuoskoski and Eerola, 2011b,a; Shiota et al., 2006), we focused on the influence of gender, culture, musical expertise and engagement on emotional judgements. There was an only weak negative association between induced emotion agreement and musical expertise. However, the results remain tentative, since the use of the discrete model, particularly with only four emotions, is inadequate to describe the richness of emotional effects of music.

Feedback from participants reinforces the issue of the inadequacy of our discrete 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 investigate this difference further, it is suggested that the dimensional model and a music-specific model, the Geneva Emotion Music Scale (GEMS (Zentner et al., 2008)), should be considered in future studies.

Comments from participants raised other issues which are worthy of further investiga-

5. Asian: Chinese, Indian, and Pakistani excluding those who were born in the UK

tion, for example: "*If I feel sad I will usually listen to songs in my mother tongue, it is always sad.*"; "*Sometimes I felt angry because I didn't like the song and didn't want to listen*."; "*I feel multiple emotions.*"; "*I really like heavy metal, so I think many of the metal songs, normally people would've felt angry, but I just felt happy + energised.*". Further research is required to understand cross-cultural differences and the influence of participants' musical taste on their ratings.

One interesting outcome of this work is the comparison with our previous studies on automatic emotion recognition, in which we obtained classification results of up to 54% based on analysis of musical features extracted from the audio excerpts (Song et al., 2012). Surprisingly, this is higher than the level of agreement between human listeners, for either perceived or induced emotion. In future work, we will compare further the predictions of classifiers based on high-dimension musical features with the results of human listening tests for ratings of perceived and induced emotion.

5. Acknowledgements

This study was supported by the School of Electronic Engineering and Computer Science at Queen Mary University of London. We would also like to thank the SEMPRE Award, the Centre for Digital Music (C4DM), and the China Scholarship Council (CSC) for their financial support. We are very grateful to all the participants, reviewers and colleagues for their valuable suggestions.

References

Ames, M. and Naaman, M. (2007). Why we tag: motivations for annotation in mobile and online media. *In Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 971–980.

Barthet, M., Fazekas, G., and Sandler, M. (2013). Music emotion recognition: From content to context-based models. In Aramaki, Barthet, K.-M. . Y., editor, *CMMR 2012 Post-proceedings (in press), Lecture Notes in Computer Science*. Springer.

Eck, D., Lam ere, P., Bertin-Mahieux, T., and Green, S. (2007). Automatic generation of social tags for music recommendation. *Advances in neural information processing systems*, 20:385–392.

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

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6:169–200.

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

Gabrielsson, A. (2002). Emotion perceived and emotion felt: same or different? *Musicae Scientiae, Special Issue: Current Trends in the Study of Music and Emotion*, (123- 147).

Hoffman, M., Blei, D., and Cook, P. (2009). Easy as CBA: A simple probabilistic model for tagging music. *10th International Society for Music Information Retrieval*, pages 369–374.

Juslin, P. 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.

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

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

Laurier, C. and Grivolla, J. (2008). Multimodal music mood classification using audio and lyrics. *In Int. Conf. Machine Learning and Applications*, pages 1–6.

Levy, M. and Sandler, M. (2007). A semantic space for music derived from social tags. *In 8th International Conference for Music Information Retrieval*.

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

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

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

Rentfrow, P. and Gosling, S. (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.

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

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.

Song, Y., Dixon, S., and Pearce, M. (2012). Evaluation of musical features for emotion classification. *In 13th International Society for Music Information Retrieval Conference*.

Terwogt, M. and Grinsven, F. V. (1991). Musical expression of moodstates. *Psychology of Music*, pages 90–109.

Turnbull, D., Barrington, L., and Lanckriet, G. (2008). Five approaches to collecting tags for music. *In 9th International Conference for Music Information Retrieval*, pages 225–230.

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. and Eerola, T. (2011a). 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):1099–106.

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

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

Wu, X., Zhang, L., and Yu, Y. (2006). Exploring social annotations for the semantic web. *In Proceedings of the 15th international conference on World Wide Web - WWW '06*, pages 417–426, New York, New York, USA. ACM Press.

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