Gathering a dataset of multi-modal mood-dependent perceptual responses to music

Matevž Pesek[1], Primož Godec[1], Mojca Poredoš[1], Gregor Strle[2], Jože Guna[3], Emilija Stojmenova[3], Matevž Pogačnik[3] and Matija Marolt[1]

[1]University of Ljubljana,
Faculty of Computer and Information Science
{matevz.pesek, matija.marolt}@fri.uni-lj.si
{primoz.godec, mojca.poredos}@lgm.fri.uni-lj.si

[2]Scientfic Research Centre of the Slovenian Academy of Sciences and Arts
Institute of Ethnomusicology gregor.strle@zrc-sazu.si

[3]University of Ljubljana,
Faculty of Electrotechnics
{joze.guna, emilija.stojmenova, matevz.pogacnik}@fe.uni-lj.si

Abstract. The paper presents a new dataset that captures the effect of mood on visual and auditory perception of music. With an online survey, we have collected a dataset of over 6600 responses capturing users' mood, emotions evoked and expressed by music and the perception of color with regard to emotions and music. We describe the methodology of gathering the responses and present two new models for capturing users' emotional states: the *MoodGraph* and *MoodStripe*. Also, general research questions and goals, as well as possible future applications of the collected dataset, are being discussed.

Keywords: color perception, human computer interaction, mood estimation, music information retrieval

1 Introduction

The complexity of human cognitive processes forces researchers to focus on specific narrow problems within their field of expertise. By selectively partitioning the multi-modality of human perception on the respective field-specific features, many other aspects are excluded or presumed to have no impact on the results. For example, much of music information retrieval (MIR) research to date has been *systems-based*, conducted in laboratory environment by assuming some objective "ground truth" (such as genre classification) and ignoring user context and individual preferences [12]. Such models or systems are "one-dimensional", restricted to some simple notion of perception, and fail to capture "real-world" scenarios.

1.1 Aim and scope

In order to propose a system capable of multi-dimensional modeling of human perception, we have considered various perceptual modalities and built a dataset for multi-dimensional model. Our work connects approaches in the fields of human-computer interaction (HCI) and MIR, with an intention to incorporate feature types from both fields into the model. Our aim is to evaluate two aspects of human perception: the color and music perception and concurrently bind these to the emotions induced and perceived from those modalities. By extracting the information about the user's emotional state, we attempt to link the variable components of one's personality which presumably impact the decision-making processes of a human. Following paragraphs provide a brief overview of closely related problem tasks in the fields of HCI, MIR and music visualization.

An impressive amount of attention has been given to user modeling and personalization of user experience (UX) in different application domains. Originating from product-oriented problems, the UX has become one of the most important factors in scientific fields including computer graphics and HCI [1]. In MIR research, on the other hand, there have been only a handful of studies focusing on user satisfaction by including her personal preferences into the decision making model [12]. These typically include music features representing the music genre, artists, instrumentation and mood [16].

Mood has been extensively analyzed within several scientific fields, not only in the fields of psychology and cognitive science [2,10], but also in the fields of HCI [15] and in recent years in MIR [8,12]. Some of the approaches that use mood for personalization capture mood by eye-tracking and visual recognition (with cameras), while others rely on manual input of relevant information by interacting with the interface. In MIR, mood estimation from music is a relatively new task [7], generally evaluated independently from other user-dependent tasks. However, we argue that for mood estimation to be successful, it needs to be integrated with other user-dependent tasks, such as music recommendation task [14]. The evaluation metrics for music recommendation often rely on play-list comparison. The task is formalized as a production of an ordered list of songs, given a query and a set of source data. "The results are evaluated against a ground truth derived from a second source or human evaluation"[1]. While this may seem as a valid evaluation procedure, it discards any personal information about the user, which could, in our opinion, have great impact both on the results and on the functionality and personalization of the recommender system in practical use.

A notable effort has been put into visualizing the music data on multiple levels: audio signal, symbolic representations and meta-data [9]. Often, these representations contain interpretable psychological dimensions, built upon relationships between multiple qualitative dimensions of represented entities and additional data in form of shape, size or color [3]. Color tone mappings are also applied onto the frequency, pitch or other spectral components [11], in order to

[1] http://www.music-ir.org/mirex/wiki/2009:Music_Recommendation

describe the audio features of the music [18]. However, such representations significantly differ, also due to the choice of the color set, which is typically picked randomly by the creator. Furthermore, there is no standard problem task evaluating the "correctness" of music visualization, since the evaluation procedure would be unclear due to the amount of factors possibly influencing the ground truth produced by a person. Thus, the proposed visualization techniques often serve merely as the visually appealing aspect of the representation. Nevertheless, we believe that certain uniformity exist in the general perception of connections between colors and music.

1.2 Research questions

Based on the proposed model and subsequent studies, our goals are set along the following research questions:

- Does users' mood impact their perception of color and music?
- Is the relationship between color and emotions uniform or individual? To what extent can such relationship be influenced by mood and personal characteristics?
- Does the correlation between sets of perceived (expressed by music) and induced (evoked by music) emotions depend on the personal musical preferences?
- Are there *emotionally ambiguous* music excerpts?

Our general hypothesis is that perceived emotions of a music excerpt are expected to be similar across listeners, whereas the induced emotions are expected to be correlated across groups of songs with similar features (such as genre) and users with similar personal characteristics. Furthermore, if there is a significant variance between sets of perceived and induced emotions, the latter could be used to implicate user's satisfaction of the given selection. We assume such relationship can be integrated into music recommender system, significantly improving current state-of-the-art emotion-based music recommenders. Another aspect, largely ignored by existing recommenders, is the effect of music on user's current mood. If results prove such effect to be significant, there is a real motivation for constructing progressive recommender system, dynamically adjusting music recommendations according to emotionally charged music changes selected by the user. We plan to implement these aspects for playlist generation and evaluate the usefulness of such approach with regard to current state-of-the-art recommenders that do not consider such variability.

From HCI perspective, there are a number of possibilities, where user's mood can be usefully taken into account. One of the most obvious ones is improvement of recommender algorithms for personalized selection of multimedia content (music, movies, books, etc....). Standard personalization techniques have reached the stage where user context is taken into account in addition to user's general preferences. The user's context is usually comprised of a number of parameters such as user's mood, location, time of day, activity, etc. While the number of available movies for example, has reached many thousands, the recommended ones can be consequently counted in hundreds. Taking user's mood into account can further reduce the number of recommended movies to ten or twenty, which makes the recommendation system much more user friendly.

The paper is structured as follows: section 2 provides a detailed methodology of the survey design and the proposed *MoodGraph* emotion model, while section 3 elaborates on the gathered dataset and discusses possible future applications.

2 Methodology

We started our survey design with a preliminary questionnaire (see subsection 2.1), which provided some basic guidelines for the overall design. The following subsections describe the structure of the survey, divided into three parts. These are intended to capture user's emotional state, the perception of colors and corresponding emotions, and emotions perceived and induced from music, along with the corresponding color.

2.1 Preliminary study for bias exclusion

As there are no generally accepted sets of textual descriptors (labels) for various emotional states, beyond the set of basic emotions proposed by [5], many studies choose labeled sets intuitively, with no further explanation, e.g. [17]. In contrast, we performed an initial study of existing emotion sets used within psychology and music research in order to establish a relevant set of emotion labels. In order to eliminate the cultural and lingual bias on the labeling, we performed our survey in Slovenian language for Slovenian-speaking participants. The preliminary questionnaire asked the user to describe their current emotional state through a set of 48 emotion labels, each with an intensity-scale from 1 (inactive) to 7 (active). The questionnaire was solved by 63 participants. Principal component analysis of the data revealed that first three components explained 64% of the variance in the dataset. The majority of the 17 emotion labels for our survey were then chosen from the first three components. To capture the relationship between colors and emotions, we evaluated the effectiveness of the continuous color wheel for choosing colors. Responses indicated the continuous color scale to be too complex for some users and a modified discrete-scale version was chosen instead. The discrete color wheel provides the user with a choice of 49 colors displayed on large tiles, enabling the user to pick the best match. All 49 colors have been chosen to represent a common color spectrum of basic colors and provide a good balance - a trade-off between the complexity of the full continuous color wheel and the limitations of choosing a smaller subset of colors.

2.2 Structure of the survey

We divided our online survey into three parts. The first part contains basic demographic questions, including questions regarding musical experience. The second part investigates user's current mood and her/his perception of relationships between color and emotions. Part three consists of ten 15-second long music excerpts and questions related to perceived and induced emotions, as well user's perception of relationship between color and individual music excerpt.

2.3 Part one - Demographic data

In order to evaluate how personal characteristics of users impact their music and color perception, we designed a set of questions, shown in Table 1. The set contains three demographic-oriented questions, two questions related to the musical background, including user's preferences and interests in music, and two questions inquiring about medications and drug use (we consider them necessary in order to detect possible discrepancies).

Table 1. Questions from the first part of our survey. Each question is provided with a set/range of possible answers (2nd column) and with comments where needed (third column). The responses provide some background information about the user's music experience and preferences.

Question	Range	Comments
Age	[5, 99]	*in years*
Gender	{Male, Female}	
Area of living	{city, rural area}	
Music school attendance	[0, 20]	*in years,* *0 - meaning not attending*
Instrument playing or singing	[0, 20]	*in years,* *0 - meaning not attending*
Usage of drugs	{yes, no}	*Is participant using drugs*
Influence of drugs	{yes, no}	*Is participant under the influence of drugs when filling the survey*
Genre preference	{Classical, Opera, Country, Folk, Latin, Dance / Disco, Electronic, RnB/Soul, Hip Hop/Rap, Reggae, Pop, Rock, Alternative, Metal, Blues, Jazz, Vocal, Easy Listening, New Age, Punk}	*up to three preferred genres (at least one) can be selected starting with the most preferred genre.*
Time listening to the music	{less than 1, 1-2, 2-3, more than 3}	*in hours per day*

The introduction of a larger set of demographic questions has also been considered. However, as the focus of our research is investigation of the interplay between colors, music and emotions, a very demanding task in itself, the decision has been taken not to put additional stress on the user by conducting lengthy demographic investigation. The amount of time required to finish the survey was averaged under 10 minutes.

2.4 Part two - mood, emotions and colors

This part of the survey is designed to capture information about user's current mood, her/his perception of relationship between color and emotions, and evaluation of the latter in terms of pleasantness and activeness. Here, mood is defined as a longer lasting, but less intense emotional state (generally unaffected by current situation or particular stimulus), compared to emotion (which is usually directly affected by a particular event or stimulus), though such definition is still arguable as both terms are often used interchangeably (see e.g. [6,13]). We use term 'emotional state' in places where both mood and emotion are being referred to. The structure of the second part of the survey is outlined in Table 2.

Table 2. The second part of the survey inquiring about the perception of emotions and colors. The fourth question is visually divided into three blocks, with six, six and five emotions in each block. The fifth question is presented to the user in a combination of a word describing the mood and a color wheel for user input (see Figure 2).

Question	Range	Comments
Current mood	Valence/Arousal space	*User selects her/his mood in a 2 dimensional space according to the pleasantness and activeness of the mood - see Figure 1*
Color of the mood	Color wheel	*User chooses the color tone most reflecting her/his current mood - see Figure 2*
Perception of emotions	{fear, energetic, angry, relaxed, happiness, sadness, liveliness, joy, disappointment, discontent}	*User places the set of emotions onto valence/arousal space - see Figure 3*
Emotional state	{active, wide awake, drowsy, inactive, miserable, discontent, disappointment, relaxed, happiness, cheerful, joyous, satisfied, sleepy, sad, calm, angry}	*User places all emotions onto a stripe with a continuous scale ranging from unexpressed to highly expressed - see figure 4*
Colors of emotions	{energetic, discontent, sad, disappointed, relaxed, angry, fearful, happy, joyous, lively}	*For each word in a set, user selects a color most resembling the described emotion.*

First, the participants are asked to describe their current emotional state. Instead of having them choose a set of emotional labels, they performed this task by placing a point in the valence-arousal space (Figure 1). This is a standard mood estimation approach, also frequently used for the estimation of perceived emotions in music. The same approach was used for self-estimation in the *MoodStripe*. We believe, that the valence-arousal space is a fast and intuitive way for the user to describe her/his current emotional state.

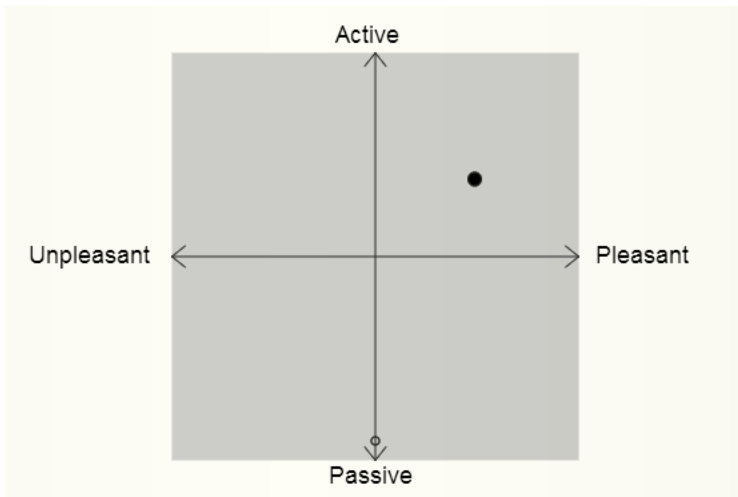

Fig. 1. Valence-arousal space for estimation of user's emotional state. The graph axes are marked *Uncomfortable* and *Comfortable* for the abscissa, and *Passive* and *Active* for the ordinate values.

Questions two and five refer to a color wheel, shown in Figure 2. Participants were asked to choose the best-matching color for particular emotion. The color wheel contains a total of 49 possible color selections (see 2.1).

Question three was designed to assess how users perceive the pleasantness and activeness of emotions in the valence/arousal space, called *MoodGraph* (Figure 3). The users were asked to place a set of ten emotion labels in the *MoodGraph*, according to their perception of pleasantness and activeness of individual emotion. Although we assume that emotions will form distinctive clusters in the valence/arousal space across users (thus, gathering the *stereotypical* representation of the emotions), we nevertheless wish to evaluate the variability of the emotion labels placements in terms of their activeness and pleasantness, and compare with the results in part three, where users described musical excerpts in a similar manner.

The fourth question was designed to capture users' current emotional state, by annotating 17 emotions on a scale from unexpressed to highly expressed. To make the task more intuitive and keep in line with the overall UX (user experience) of the survey, we designed a new input modality, *MoodStripe* (Figure 4). As in *MoodGraph*, the user drags emotion labels onto space representing

Fig. 2. A color wheel with total of 49 colors presented. The black dot indicates the selected color.

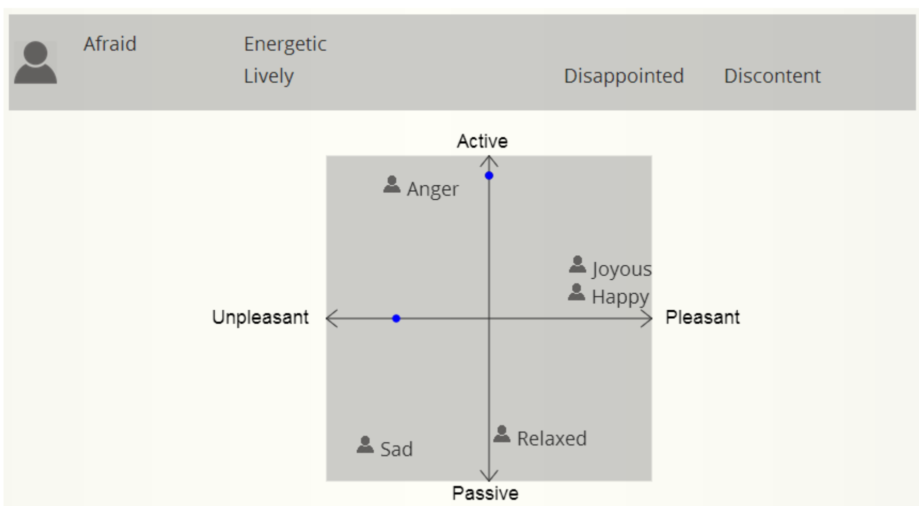

Fig. 3. The *MoodGraph*: emotions are dragged from the container above the graph onto the valence/arousal space. Blue dots indicate the position of the selected emotion in the valence/arousal space.

expressivity of emotions. To ensure the ease the positioning, the emotions are divided into three separate *MoodStripes*.

Both novel user input modalities, the *MoodGraph* and *MoodStripe*, could easily be replaced by a set of ordinal scales, implemented as radio buttons for each emotion. However, we believe that our approach is more efficient, intuitive

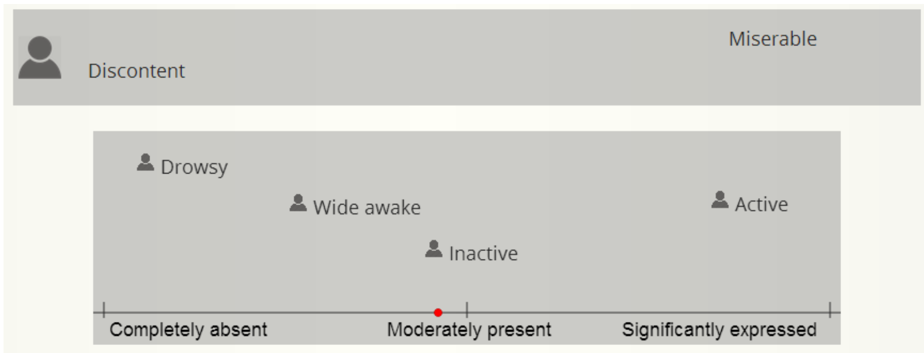

Fig. 4. The *MoodStripe*: to express their emotional state, users position a set of emotions on a scale from unexpressed on the left to highly expressed on the right. The scale is marked from *absent* over *present* in the middle, to *expressed* the right.

and puts less stress on the user. We provided an optional feedback form after users finished the survey and preliminary analysis reveals very positive remarks about the proposed innovative input modalities.

2.5 Part three - music in relation to colors and emotions

In part three of the survey, users are asked to complete two tasks on a set of ten randomly selected 15-second long music excerpts. After listening to an excerpt, the user is first asked to choose the color (from the color wheel) that best represents current music excerpt (Figure 2). Next, the user is asked to place a set of emotion labels on the *MoodGraph*, according to two categories: a category of induced emotions and a category for perceived emotions. The category of induced emotion labels is marked with a person icon, while the perceived emotion labels are represented with a note icon. The user may place any number (but at least one from each category) of emotions on the *MoodGraph*. An example is shown in Figure 5.

3 Contributions and future work on the collected dataset

This survey was published in March. With the help of social networks we have gathered responses of 952 users to date. This number is much higher than expected, particularly as the survey was only distributed among Slovenian speaking population. Users provided 6609 mood/color-perception responses for the 200 music excerpts used in the survey.

To achieve the most relevant results in regard to the emotional influence of music, we've carefully selected the songs with an intention to cover the whole variety of genres, and at the same time avoid popular (or most known songs), that could bias users' ratings. The database includes 80 songs from the online service

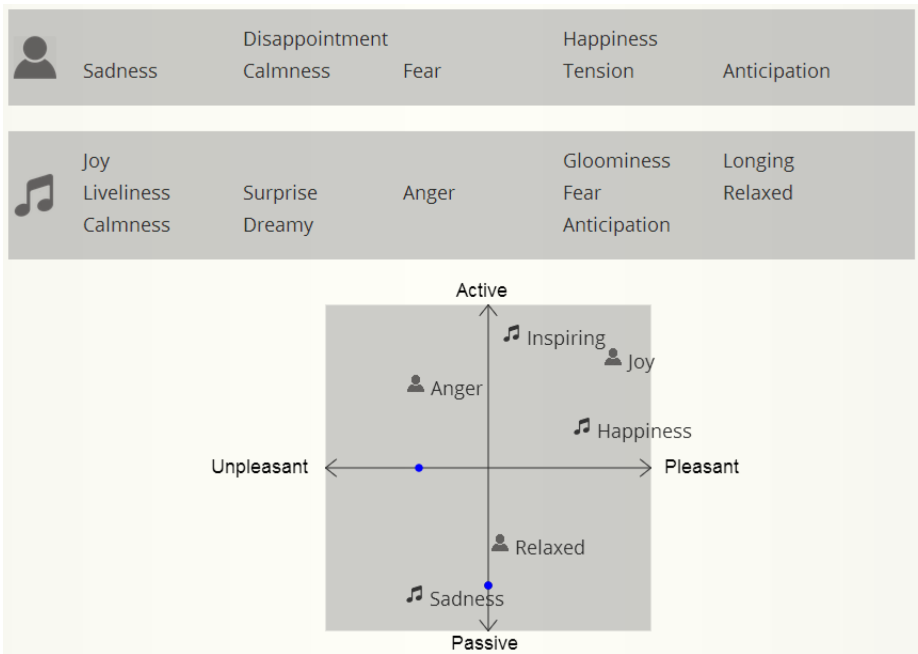

Fig. 5. An example of two-category *MoodGraph*. Induced emotions are marked with a person icon, perceived emotions with a note icon.

Jamendo[2], which is widely used by less-known artists. We have also included a part of a previously collected dataset [4] containing 80 excerpts of film music, a subset of 20 Slovenian folk music excerpts and 20 music excerpts from past International Computer Music Conference proceedings containing contemporary electro-acoustic music. The diversity of the chosen sets will allow us to analyse the differences of responses and draw possible conclusions by extracting the audio liable features. By gathering a vast amount of responses (each excerpt has on average 33 responses), including induced and perceived emotions and relationships between music excerpts and color, we've provided a foundation for future research on person's emotional state, perception of music and color, and complex relationships between perceived and induced emotions in music.

To our knowledge, no currently available music-mood dataset has such a high ratio of user annotations per music excerpt. By analysing the data provided by *Google analytics*, we've established that the average time to complete the survey was 8.27 minutes, thus we reached our goal of a less than 10 minute long survey.

We intend to make the entire dataset available to the public, including musical excerpts, the data on users' emotional states, their placement of emotions within the valence/arousal space, their perceived and induced emotional responses to music and their perception of color in relation to emotions and music.

[2] `http://www.jamendo.com`

This will open new possibilities for evaluating and re-evaluating mood estimation and music recommendation approaches on a well annotated dataset, where the ground truth lies in the statistically significant amount of responses per song, rather than relying on annotations of a small number of users.

Shortly, we will also publish an English version of the survey, where an additional set of responses will be gathered. We also intend to enlarge the number of music excerpts in the music dataset and provide it to the users who have already participated in the study. Thus, we hope to further extend and diversify the music collection.

The information collected during this research will possibly provide the basis for the realization of the following research goals and future work:

- Previously introduced mood estimation algorithms can be evaluated by weighting the correctness of their predictions with the distribution of perceived-emotion responses for a music excerpt. A new mood estimation algorithm will be developed, building upon the newly obtained data.
- We will explore the modelling of relations between music and the corresponding colors chosen by users in the survey. Results may be useful for music visualization, provided that correlations between audio and visual perception will be consistent enough.
- A user interface for music recommendation will be developed, presenting recommendations in a visual manner, possibly raising the user satisfaction by reducing the textual information needed to be presented to the user. The interface will include personal characteristics and their variability in the decision model.
- The dataset can also be used in several other domains. Responses that link color and mood perception based on the user's emotional state can be used independently.

References

1. William Albert and Thomas Tullis. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Google eBook).* Newnes, 2013.
2. Tiziano Colibazzi, Jonathan Posner, Zhishun Wang, Daniel Gorman, Andrew Gerber, Shan Yu, Hongtu Zhu, Alayar Kangarlu, Yunsuo Duan, James A Russell, and Bradley S Peterson. Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion (Washington, D.C.)*, 10(3):377–89, June 2010.
3. Justin Donaldson and Paul Lamere. Using Visualizations for Music Discovery. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, page Tutorial, 2009.
4. T. Eerola and J. K. Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1):18–49, August 2010.
5. Paul Ekman. An argument for basic emotions. *Cognition and Emotion*, 169-200(6), 1992.

6. Alf Gabrielsson. Emotion Perceived and Emotion Felt: Same or Different? *Musicae Scientiae*, 5(1_suppl):123–147, September 2002.
7. Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull. Music emotion recognition: A state of the art review. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 255–266, Utrecht, 2010.
8. Cyril Laurier, Owen Meyers, Joan Serrà, Martin Blech, Perfecto Herrera, and Xavier Serra. Indexing music by mood: design and integration of an automatic content-based annotator. *Multimedia Tools and Applications*, 48(1):161–184, October 2009.
9. Nicola Orio. Music Retrieval: A Tutorial and Review. *Foundations and Trends® in Information Retrieval*, 1(1):1–90, 2006.
10. N A Remmington, L R Fabrigar, and P S Visser. Reexamining the circumplex model of affect. *Journal of personality and social psychology*, 79(2):286–300, August 2000.
11. Shigeki Sagayama and Keigo Takahashi. Specmurt anasylis: A piano-roll-visualization of polyphonic music signal by deconvolution of log-frequency spectrum. In *ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing*, Jeju, Korea, 2004.
12. Markus Schedl, Arthur Flexer, and Julián Urbano. The neglected user in music information retrieval research. *Journal of Intelligent Information Systems*, 41(3):523–539, July 2013.
13. K. R. Scherer and M. R. Zentner. Emotional effects of music: production rules. In P. N. Juslin and J. A Sloboda, editors, *Music and emotion.* Oxford University Press, New York, 2001.
14. Y. Song, S. Dixon, and M. Pearce. A survey of music recommendation systems and future perspectives. In *Proc. 9th Int. Symp. Computer Music Modelling and Retrieval (CMMR)*, pages 395–410, London, 2012.
15. Marko Tkalcic, Ante Odic, Andrej Kosir, and Jurij Tasic. Affective Labeling in a Content-Based Recommender System for Images. *IEEE Transactions on Multimedia*, 15(2):391–400, February 2013.
16. Jill Palzkill Woelfer. The role of music in the lives of homeless young people. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 367–372, Porto, 2012.
17. B. Wu, S. Wun, C. Lee, and A. Horner. Spectral correlates in emotion labeling of sustained musical instrument tones. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 415–421, 2013.
18. Ho-Hsiang Wu and Juan P. Bello. Audio-based music visualization for music structure analysis. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, 2010.