The Moodo dataset: integrating user context with emotional and color perception of music for affective music information retrieval

Matevž Pesek,

University of Ljubljana

Gregor Strle,

Scientific Research Centre of the Slovenian Academy of Sciences and Arts

Alenka Kavčič,

University of Ljubljana

Matija Marolt,

University of Ljubljana

Abstract

This paper presents a new multimodal dataset Moodo that can aid the development of affective music information retrieval systems. Moodo's main novelties are a multimodal approach that links emotional and color perception to music and the inclusion of user context. Analysis of the dataset reveals notable differences in emotion-color associations and their valence-arousal ratings in non-music and music context. We also show differences in ratings of perceived and induced emotions, especially for those with perceived negative connotation, as well as the influence of genre and user context on perception of emotions. By applying an intermediate data fusion model, we demonstrate the importance of user profiles for predictive modeling in affective music information retrieval scenarios.

I. Introduction

Affective music information retrieval (MIR) is an interdisciplinary effort to understand and model user information needs in the largely uncharted territory of human use and perception of music. Among its many challenges are the systematic analysis and modeling of affect in music (Juslin, 2013; Eerola, 2013), the influences of musical preferences, personality, mood and emotions (Vuoskoski and Eerola, 2011a; Schubert, 2013), the investigations into neural correlates and multisensory integration (Zatorre et al., 2007; Koelsch, 2010; Spence, 2007; Levitin and Tirovolas, 2009), as well as the cognitive, social and cultural influences (Pearce and Rohrmeier, 2012; Stalinski and Schellenberg, 2012; Stevens, 2012).

In recent years, efforts towards personalization in MIR have shifted the emphasis from the system- towards user-centered design. Traditional MIR tasks are predominantly system-centered, e.g. audio classification, melody extraction, music similarity and retrieval, emotion and mood estimation from audio, lyrics or collaborative tags (Song et al., 2012; Kim et al., 2010; Laurier et al., 2009). More often than not they lack grounding in the real world (Weigl and Guastavino, 2013; Schedl et al., 2013). The state-of-the-art is reflected in existing MIR datasets where user information about their perception and use of music is generally sparse (Schedl et al., 2013).

Affective MIR, on the other hand, aims to be 'user-aware'. Its main objective is to integrate context-related user information (such as music preferences) with the perception and use of music to attend to user's music information needs. Specifically, it deals with the affective (emotional and mood related) aspects of music perception and cognition, and their integration in MIR. The main obstacle towards affective MIR is the limited availability of coherent empirical data, as competing theories, different annotation sets, emotion models and methodologies add to the overall complexity and make a systematic comparison of results difficult, if not impossible (Eerola and Vuoskoski, 2013; Lee and Cunningham, 2013).

The enormous amount of digitally available audio recordings has spawned an increased interest in the affective aspects of music. As an important topic within the MIR field, the music emotion recognition (MER) focuses on the development of algorithms for automatic emotion recognition and mood estimation from audio. Although MER is defined as a multiclass-multilabel classification problem (Kim et al., 2010), more recent approaches also aim at integrating subjective aspects, such as user's perception of emotions in music, to improve recognition (Yang and Liu, 2013).

The main problem of existing MER datasets is the incommensurability of emotion modeling approaches used in the studies, as well as a lack of studies on multimodal perception of music. Existing datasets have different research goals, use different emotion models and labels, types of music, number and duration of excerpts, and are often further limited by the number of participants. Many of them have little or no data on user context and participant's affective state, and none, to our knowledge, focus on the integration of multiple modalities when measuring participant's responses to music.

The main motivation for gathering a new dataset is a multimodal approach towards MIR and MER. The Moodo dataset is a dataset of mood-dependent emotional and color responses to music that can be used in research on the multimodal perception of music and in various affective MIR scenarios. By linking emotion-color-music responses to user context (basic demographics and music preferences) we are able to obtain additional insight into emotional and multimodal interactions, and create user profiles for predictive modeling in affective MIR. The color dimension, along with the emotional perception of music and user context, further supports aspects of personalization in the development of affective recommender systems (Sokolova and Fernández-Caballero, 2015). The Moodo dataset is publicly available for download and for online visualization[1].

The rationale for incorporating the visual (color) modality lies in the existing research on human emotions that underline the importance of multimodal approaches, with several experimental studies confirming the link between emotions and colors, e.g. Eysenck (1941); Wexner (1954); Schaie (1961); Adams and Osgood (1973); Valdez and Mehrabian (1994); Sokolova and Fernández-Caballero (2015). As music listening is for most people an emotional experience, and several studies have shown correlations between the emotional perception of colors and music (e.g. (Palmer et al., 2011, 2016)), we argue that color as an additional dimension can give us further insights into relationships between emotions and music, and that its perception is different in the music versus non-music contexts, as well as influenced by the context of the user. The results presented in this paper confirm this hypothesis.

The paper extends the preliminary analysis of emotion-color relations from the Moodo dataset presented in Pesek et al. (2014a). The preliminary analysis gave an overview of the dataset and the data gathering procedure. In this paper, we show how colors and music are mediated by emotions and how emotion-color-music associations are further constrained by the user context of mood, gender, age, and music preferences. These aspects are also included in the experimental section of the paper, where we show how user context can improve prediction of emotional responses in music. The experiment validates the usability of user profiles for predictive modeling in affective MIR and opens new possibilities for innovative applications in affective computing in general.

[1]The Moodo dataset is available at: http://moodo.musiclab.si

The paper is structured as follows. Section II presents the current state-of-the-art in affective MIR with an overview of available datasets for MER. Section III describes the design of the survey prepared for gathering the Moodo dataset. Section IV presents results of the survey and analysis of the gathered data, focusing on emotional and color responses to music given the user context. Section IV-F presents an application of the results using multimodal data fusion, which shows the advantages of integrating user context (demographic data and personal preferences) in predictions of discrete emotions (emotion labels) and their valence-arousal ratings. We conclude the paper with a discussion of the findings and their possible applications to music recommendation.

II. Related work

The following overview addresses several areas: emotion models as the key instrument for gathering user input, the available music datasets used for MER-related tasks, and research on colors in both general MIR visualizations and as a separate research topic, dealing with the relationship between auditory and visual perception.

A. Emotion models in MER

Most MIR researches on the emotional perception of music use traditional discrete or dimensional emotion models for gathering user input, such as variations of the Likert intensity scale (Likert, 1932) or Russell's circumplex model of affect (Russell, 1980). As the names imply, discrete emotion models represent emotions as discrete categories, while dimensional models represent emotions as dimensions, typically in the two-dimensional coordinate space of *valence* and *arousal*.

Variations of Russell's dimensional model have been frequently used in studies on musical emotions (e.g. (Laurier et al., 2009; Mcvicar et al., 2011; Yang and Chen, 2012; Barthet et al., 2013)), with several researchers suggesting additional dimensions to better accommodate the structure of musical emotions. For example, (Schimmack and Reisenzein, 2002) proposed alternative interpretation of the arousal dimension by two separate dimensions of *energetic* (awake-tired) and *tense* (tense-calm), (Bigand et al., 2005) and (Canazza et al., 2001) proposed the additional third dimension of *kinetics* to link perceived emotions with body posture and gestures, and (Eerola et al., 2009) proposed *tension* as an additional third dimension to valence and arousal.

The main criticism of dimensional models refers to the limited number of dimensions used for modeling emotions, with the low-dimensional affective space typically reduced to valence and arousal (e.g. (Fontaine et al., 2007)). Discrete models aim to overcome this limitation by multiple-category ratings of emotions (Kim et al., 2010; Barthet et al., 2012). For example, the MIREX initiative[2] proposes a five-cluster model derived from the All Music Guide mood repository (Hu and Downie, 2007), with label sets consisting of 5 to 7 labels per cluster, resulting in a total of 29 labels. In order to compensate for the limitations of current music emotion models, (Zentner et al., 2008) developed the GEMS - Geneva Emotional Music Scale, as a domain-specific emotion model for musically induced emotions. The GEMS model, based on 45 terms, has been further adapted to shorter variants of 25 and 9 term models (GEMS-25 and GEMS-9) by (Torres-Eliard et al., 2011) and (Aljanaki et al., 2014a).

However, recent surveys and comparative analyses of discrete and dimensional approaches in modeling musical emotions showed little advantages of discrete models – in fact, discrete models exhibited lower discriminative accuracy for emotionally ambiguous musical examples, compared to dimensional models (Eerola and Vuoskoski, 2010; Vuoskoski and Eerola, 2011b). Thus, further and more extensive research is needed to confirm the benefits of models like GEMS and MIREX over the more traditional approaches.

B. *Datasets on emotional perception of music*

To obtain the ground truth for evaluation of MER models, human annotators and online services are typically used. While the annotators can provide more reliable and thorough annotations, the online services offer a greater amount of annotations, which on the other hand are noisier. Moreover, the selection of emotion modeling approaches and sets of emotion labels, which are used to obtain participants' ratings, vary and further influence the scope and nature of different datasets.

The MoodSwings Turk Dataset provides on average 17 valence-arousal ratings of perceived emotions for 240 clips of popular music (Schmidt and Kim, 2011). Authors used a collaborative game MoodSwings (Speck et al., 2011) with marked emoticons to express positive and negative emotions and their intensity in the valence-arousal space, and Amazon Mechanical Turk (paid participation) for building the dataset. The Cal500 provides a set of mood labels for 500 popular

[2] http://www.music-ir.org/mirex/

songs (Turnbull et al., 2008), averaging three annotations per song. The extended dataset CAL10k provides 34 acoustic features, 475 acoustic tags and 153 genre tags for 10.870 songs (Tingle et al., 2010). The MTV Music Dataset (Schuller et al., 2010) contains a set of valence-arousal ratings for 192 popular songs annotated by 5 annotators with different musicological backgrounds. Using the GEMS-9 model, (Aljanaki et al., 2014b) gathered over 8000 responses by 1778 participants on a set of 400 excerpts of classical, rock, pop and electronic music. The Emotify game has been used in a data gathering procedure, with basic demographics of gender, age and language also recorded. (Lykartsis et al., 2013) tested the German version of the GEMS-25 model for electroacoustic and popular music, expanding the original model with three additional categories (GEMS-28-G). The study involved 245 participants annotating 20 music pieces of classical and popular instrumental music and electroacoustic music. Demographic data include age, language, the level of education, music knowledge and the amount of listening to music per day. An interesting application of the GEMS model is a study on the emotional perception of music conducted by (Jaimovich et al., 2013). Here, the shorter GEMS-9 model was used on 4000 participants and 12000 music excerpts from 53 songs in order to measure participants' emotional responses, their electrodermal activity (EDA) and heart rate.

The All Music Guide (AMG) 5 mood clusters were proposed by Hu and Downie in 2007, to "reduce the diverse mood space into a tangible set of categories" (Hu and Downie, 2007). Several studies used the proposed approach in a variety of task-specific applications. For example, (Laurier et al., 2009) compared the AMG mood clusters and *Last.fm* service social tags to analyze possible correlations between the two. (Yang and Chen, 2012) gathered a dataset of 2453 responses to 500 Chinese music pieces (with one expert annotation per song and 5 labels per song on average) to evaluate the acoustic features and compare them to responses on English music pieces.

For comparison of dimensional and discrete emotion models (Eerola and Vuoskoski, 2010) gathered a dataset of 110 film music excerpts annotated by 116 non-musicians. The experiment was divided into two stages. During the first stage, a dimensional model was tested by participants rating individual music excerpts in a three-dimensional valence-arousal-tension space (using bipolar scales), whereas during the second stage a discrete emotion model based on a 9-degree intensity scale was used for rating of individual emotions. The dataset contains mean ratings of perceived emotions gathered by both models. Demographic data about each participant includes gender, age, music education and skills (the ability to play an instrument).

C. Visualization in MIR

User-centered research in music visualization has been largely neglected in the past (Schedl et al., 2013; Serra et al., 2013), but recent attempts (e.g. the Grand challenge at MIREX evaluation exchange initiative[3]) indicate a growing interest in this domain. There are numerous attempts at providing visualizations for a variety of music and audio features, from low-level spectral features to high-level music patterns and music metadata (Donaldson and Lamere, 2009). Most can be separated into two categories: visualizations of music parameters (e.g. harmonies, temporal patterns and other music entities) (Isaacson, 2005; Bergstrom et al., 2007; Mardirossian and Chew, 2007; Grohganz et al., 2013; Jiang and Mueller, 2013), and visualizations of spaces representing relationships among different music pieces (Pampalk, 2003; Kurabayashi and Imai, 2014). The latter are more suitable for music recommendation systems and for data exploration systems in general.

Furthermore, future music visualization approaches should build upon findings from research on multimodal aspects of human perception and cognition. As several studies have shown, the interactions between emotions and visual (color) and auditory (music) modalities significantly depend on user's personality traits, temporary affective state, and music context (Meyer, 1956; Ou et al., 2004; Juslin and Västfjäll, 2008; Spence, 2011). The experiments conducted by (Palmer et al., 2016) further show that cross-modal associations between music and color use emotional mediation as an underlying mechanism. Their findings "associate specific dimensions of color (saturation, lightness, and yellowness-blueness) with specific high-level musical dimensions (tempo and mode), [and] ... show clear evidence of mediation by emotional dimensions (happy-sad and angry-calm) ..." (Palmer et al., 2013, p. 8836). Further evidence for the direct cross-modal emotional mediation of both modalities is presented in this paper, based on analysis of participants' valence-arousal ratings of emotions and their color associations for the individual music excerpts.

As the overview of the state-of-the-art shows, most of the existing MIR and MER research datasets focus exclusively on the auditory modality and gather very limited user context. The Moodo dataset aims to alleviate some of these shortcomings by focusing on emotional mediation of visual and auditory modalities (color and music), and the extensive user data. In what follows, we first present the methodology for gathering the Moodo dataset.

[3]http://www.music-ir.org/mirex/wiki/2016:GC16UX

III. Methodology: survey design and data gathering procedure

The methodology for gathering annotations on emotions, colors, and music, consisted of two parts: the preliminary survey and the main survey. The goal of the preliminary survey was to define the emotion labels and the emotion annotation tools for gathering participants' ratings, and to evaluate different aspects of user experience for the main survey. The main survey was conducted online to collect a maximum number of annotations. Both surveys were conducted in Slovenian language.

A. Preliminary survey

Our main survey used two separate sets of emotion labels: one for ratings of participants' mood and emotions in a non-music context (Section III-B2), and one for annotation of music excerpts (Section III-B3). The decision to use two separate emotion sets was based on findings of several studies on musical emotions arguing that certain emotions are more relevant in the music context than others (e.g. Juslin and Sloboda (2001); Juslin and Västfjäll (2008); Juslin (2013); Schubert (2013); Eerola (2012)).

To define the emotion set for participants' self reports on their mood and emotions in non-music context, we conducted a survey on 64 participants. In total, 48 emotion labels were gathered from various datasets used in scientific studies (Russell, 1980; Remmington et al., 2000; Watson et al., 1988; Ekman, 1992). Participants were then asked to use these emotion labels to describe their current affective state on a 7-degree intensity scale. Principal component analysis (PCA) was used to reveal the semantic salience of the individual emotion labels across the participants and to select a set of emotion labels relevant to our study. The final set represents 64% of the variance along the first three principal components and consists of the following 17 emotion labels: *Active*, *Angry*, *Calm*, *Cheerful*, *Disappointed*, *Discontent*, *Drowsy*, *Happy*, *Inactive*, *Joyous*, *Miserable*, *Relaxed*, *Sad*, *Satisfied*, *Sleepy*, *Tired*, *Wide-Awake*. This emotion set was later used in Part 2 of the main survey for the participants' self-reports on mood and emotions (see Section III-B2).

A separate emotion set was defined for the annotation of musical emotions. This set comprises mainly of the musical emotion labels used across several experimental studies and scientific literature on musical emotions, e.g. Juslin and Sloboda (2001); Juslin and Laukka (2004); Eerola (2012). It represents the following 15 musical emotion labels: *Anger*, *Anticipation*, *Calmness*, *Dreaminess*, *Fear*, *Gloominess*, *Happiness*, *Inspiring*, *Joy*, *Liveliness*, *Longing*, *Relaxation*, *Sad-*

ness, *Surprise*, and *Tension*. This set was later used in Part 3 of the main survey (see Section III-B3).

Additionally, two emotion annotation tools (MoodStripe and MoodGraph) and a discrete color wheel were developed to optimize the data gathering procedure. These are presented in the following section.

B. Main survey

The main survey was conducted online in three consecutive parts. Part 1 gathered the basic demographics about the participants, Part 2 focused on the participants' current mood and their perception of the emotions and colors, and Part 3 gathered the participants' emotional and color responses to music.

1) Part 1: demographic information: The first part of the survey gathered general demographic information about the participants, together with their musical experience and taste. These data included: age, gender, native language and area of living (city/rural area), music education, the experience of playing an instrument, the amount of time listening to music per day, and genre preferences.

2) Part 2: self-reports on mood, emotions and associated colors in non-music context: Part 2 of the survey gathered the participants' self-reports on their mood, emotions, and associated colors. The participants were first required to rate their current mood and emotions in the valence-arousal space and associate their mood with a color. Next, the participants rated the set of 17 emotions (as defined in the preliminary survey). This step required the participant to self-report the presence of the individual emotion labels on a continuous scale, using the MoodStripe emotion annotation tool (Figure 1). The MoodStripe was used instead of the more traditional *n*-point intensity scales (such as the Likert scale), because it enables concurrent annotation of multiple emotions on a one dimensional continuous activation scale, with the horizontal position of each emotion label indicating its activation level. For a detailed presentation of the MoodStripe interface see (Pesek et al., 2016).

Finally, the participants were required to rate emotions in the valence-arousal space and associate them with colors. For each of the 17 emotion labels the participants were first asked to select an associated color from a discrete color wheel, and then place the individual emotion label onto a two-dimensional valence-arousal space using the MoodGraph interface shown in Figure

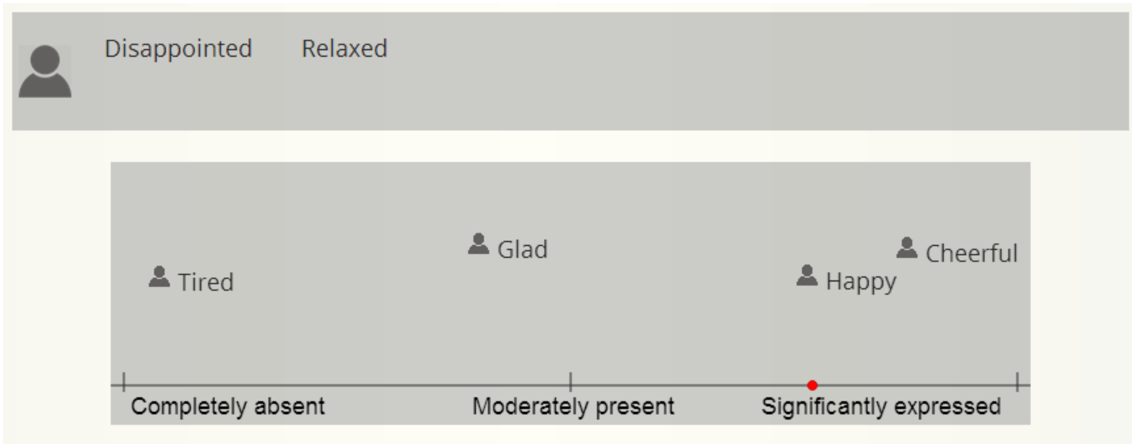

Fig. 1. The MoodStripe tool. The participant drags emotion labels (shown on top) onto a canvas (the grey box below). The canvas represents a one dimensional continuous activation scale: from unexpressed (left) to highly expressed (right). This interface was used by the participants to self-report the presence of individual emotions.

2.[4] The MoodGraph is a hybrid emotion annotation model that combines several properties of discrete and dimensional models by integrating the multiple-category annotations with the dimensionality of the valence-arousal space. The model is based on the drag-and-drop paradigm similar to the MoodStripe and enables comparison of multiple ratings and integration of several categories. The advantages of the MoodGraph are further discussed in Section III-C.

3) Part 3: emotional and color responses to music: Part 3 of the main survey gathered the participants' emotional and color responses to music. A set of 200 music excerpts, each 15 seconds long and tagged with a genre label gathered from the respective collection, was selected for the task: 80 excerpts from the free online music service Jamendo, distributed across a variety of music genres, another 80 excerpts from the film music dataset (Eerola and Vuoskoski, 2010), 20 excerpts from a collection of Slovenian folk songs, and 20 excerpts from a contemporary electro-acoustic music collection.

Each participant was presented with 10 music excerpts and instructed to:

- choose a color best associated with the music excerpt;

[4]Note that Figure 2 shows the MoodGraph interface as used for the annotation of musical emotions (Part 3 of the main survey). It integrates the two categories of perceived and induced musical emotions, represented by the two boxes with the *note* icon for the perceived and the *person* icon for the induced musical emotions. The MoodGraph used in the non-music context (Part 2 of the main survey) only comprises of one category representing the 17 emotion labels.

- choose one or more emotion labels expressing the perceived emotions of the music excerpt and place them onto the valence-arousal space of the MoodGraph;
- choose one or more emotion labels expressing the induced emotions (emotions evoked in the participant) of the music excerpt and place them onto the valence-arousal space of the MoodGraph.

C. The hybrid emotion annotation model

As discussed in Section II-A, two main types of emotion annotation models exist: dimensional (typically represented by a two-dimensional valence-arousal space) and discrete models. There is no consensus on which of the two approaches is better suited for capturing musical emotions (Eerola and Vuoskoski, 2010). The aim of the MoodGraph is to develop a hybrid model which would capture some of the advantages of both traditional approaches, for two reasons: a) to retain the connection between the specific emotion and its varying position in a two-dimensional affective space (by using the drag-and-drop functionality), and b) to put lesser task-load on the participant by enabling concurrent annotation, thus avoiding the use of several n-degree scales. We believe that the load put on the participant should not be neglected—Part 3 of the survey alone requires that the participants rate 10 music excerpts with at least one emotion label per excerpt and from each category (perceived and induced) of musical emotions. Using the traditional approach based on n-degree intensity scales would require a substantial number of scales to provide all possible annotations for a single excerpt. The MoodGraph is a novel hybrid annotation model that integrates annotation of discrete (induced and perceived) emotion labels with a continuous two-dimensional valence-arousal space and also reduces the cognitive load. By selecting emotion labels from either category and placing them onto the valence-arousal space, we can simultaneously obtain both the dimensional and discrete data (as shown in Figure 2). Such an approach also yields insights into participants' perception of emotional labels and their position in the valence-arousal space, which cannot be fully captured by either of the traditional models.

A subsequent usability evaluation of the MoodStripe and MoodGraph annotation models was conducted. Using a subset of questions from the NASA load task index evaluation questionnaire (Hart, 2006), mental, physical, temporal demands, and frustration were evaluated in comparison to the standard n-degree intensity scale. Both MoodStripe and MoodGraph scored higher in intuitiveness (ease of use), functionality (reduced task-load), and lower-level of user frustration,

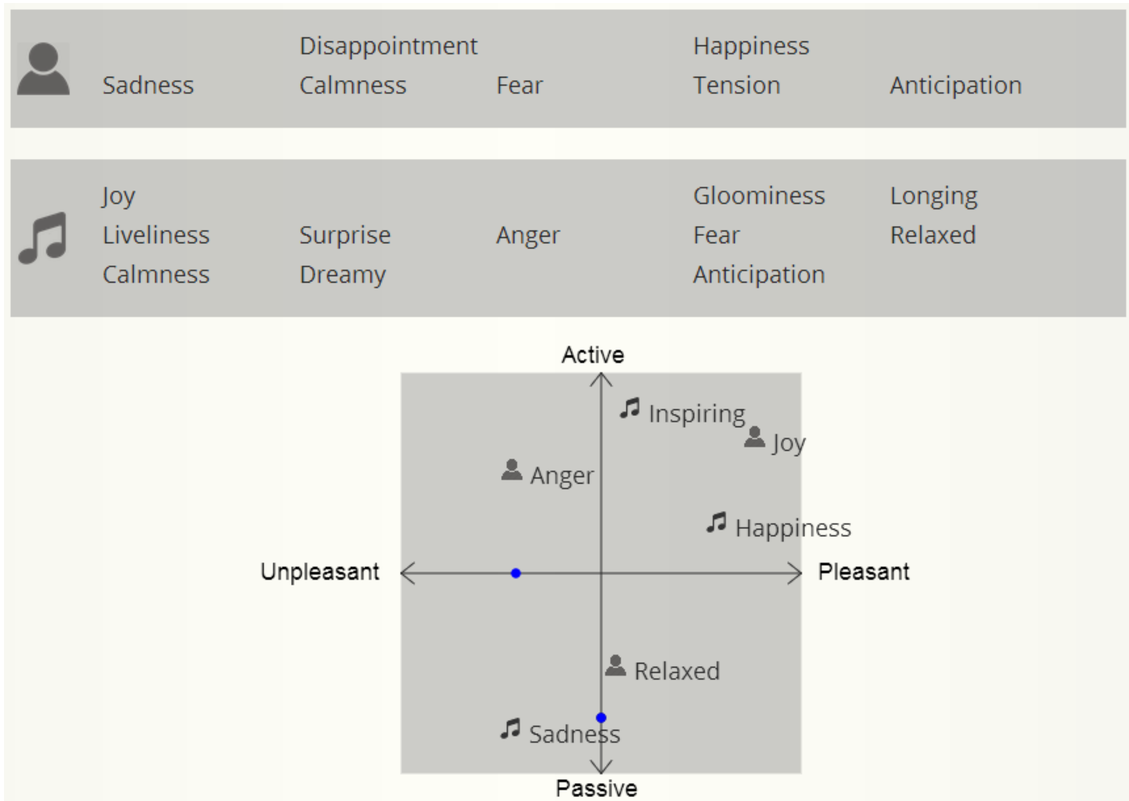

Fig. 2. The two-category MoodGraph interface provides two separate categories for annotation of induced (the *person* icon) and perceived (the *note* icon) musical emotions. Individual emotion labels can be dragged from each category container onto the valence-arousal space. The dots indicate the position of the selected emotion label in the valence-arousal space.

compared to the standard intensity scale. For further details, see (Pesek et al., 2014b) and (Pesek et al., 2016).

IV. Analysis of music, colors, and emotions

A. Results of the survey

The entire survey was completed by 741 participants. The average age of the participants was 26.5 years, the youngest participant was 15 years old and the oldest 64 years old. 66% of the participants were from an urban area, 67% of the participants were women. 50% of the participants, among them 60% men and 44% women, had no music education. Of the participants with music education, 12% of women and 6% of men finished primary music education, a standardized 6-year music curriculum in Slovenia. 47% of the participants did not play an instrument or sing. The amount of music listening per day was evenly spread from less than 1 hour to over 4 hours per day.

In the survey, the participants were asked to rank three favorite genres. The overall favorite music genre was Rock (chosen by 31% of the participants), followed by Pop (chosen by 17% of the participants), whereas Alternative and Classical were chosen by 5% of the participants. Other genres received less than 5% of the votes. As their second favorite genre, 20% of the participants chose Rock, whereas Pop received 14% of the votes. Classical music was the first choice for the third favorite genre (13%), followed by Rock (12%) and Pop (10%).

Altogether, the participants provided 6999 annotations describing their perception of emotions, colors, and music. The amount of the data gathered in the Moodo dataset warrants an informed insight into the emotional processing and the interactions between auditory and visual modalities, and validates an investigation of the following hypotheses:

- a) colors are mediated by emotions;
- b) listeners can differentiate between perceived and induced musical emotions;
- c) emotional perception of color and music varies depending on context;
- d) mood, age, gender, and music preferences can influence the perception of emotions, colors, and music.

These are further explored in the following sections.

B. Emotional mediation of color and music

Several experimental studies have shown the cross-modal emotional mediation of color and music (Palmer et al., 2011, 2013, 2016). Their results are further validated here. The analysis of the dataset shows some remarkable differences between valence-arousal ratings for the individual emotions and their color associations in the non-music and music contexts. Note that the 'non-music context' refers to the emotion-color ratings gathered from the particpants' self-reports on the set of 17 emotion labels, with no music present (see Section III-B2). The 'music context' refers to the ratings of music excerpts based on the set of 15 musical emotions and associated colors (see Section III-B3). The mean and standard deviation of the valence and arousal ratings for several emotions are presented in Table I[5]. Figures 3 and 4 show the valence-arousal ratings and the colors for the selected perceived emotions in non-music and music contexts, respectively. Note that not all of the emotions shown correspond across the two figures: *Energetic* and *Liveliness*

[5]Unless stated otherwise, the musical emotion ratings as presented in the table and the figures are given for the perceived musical emotions. The valence-arousal space is represented by valence on x-axis and arousal on y-axis.

TABLE I

THE MEAN AND THE STANDARD DEVIATION OF VALENCE AND AROUSAL RATINGS FOR EMOTIONS IN THE NON-MUSIC AND MUSIC CONTEXT.

	Non-music context				Music context			
	$\mu(V)$	$\mu(A)$	$\sigma(V)$	$\sigma(A)$	$\mu(V)$	$\mu(A)$	$\sigma(V)$	$\sigma(A)$
Anger	-0.68	0.57	0.26	0.47	-0.41	0.34	0.42	0.49
Joy	0.69	0.51	0.26	0.33	0.49	0.52	0.30	0.34
Happiness	0.75	0.31	0.26	0.46	0.53	0.48	0.31	0.37
Fear	-0.75	0.22	0.24	0.54	-0.45	0.31	0.41	0.45
Sadness	-0.76	-0.33	0.22	0.48	-0.37	0.04	0.45	0.48
Relaxation	0.67	-0.16	0.32	0.50	0.44	0.24	0.34	0.48

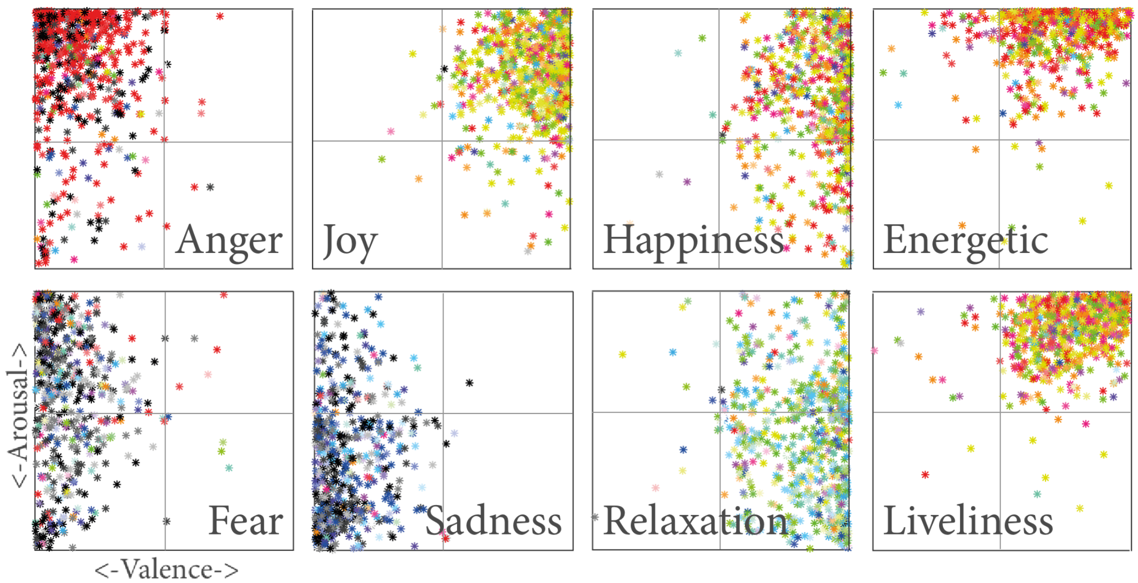

Fig. 3. Emotions and their color associations in the valence-arousal space (non-music context). The participants assigned a valence-arousal point and a color for each of the emotions in the second part of the main survey (see Section III-B2).[7]

in Figure 3 are replaced with *Anticipation* and *Calmness* in Figure 4. This is due to the different emotion labels used for annotation of emotions in non-music and music context (see Section III-A).

The color associations for the individual emotions in the non-music context, as presented in Figure 3, are in line with previous research (Adams and Osgood, 1973; Ou et al., 2004; Palmer

[7]Colors appear in grayscale in the printed version of this paper. For a colored Figure, please refer to the online version.

et al., 2013). The dark blue-violet and black hues are associated with more negative emotions such as *Fear* and *Sadness*. On the other hand, the more vibrant yellow-green-red hues are associated with more positive emotions, such as *Joy*, *Happiness*, *Energetic*, and *Liveliness*, with the prevalent red hues for the latter two. Two noticeable exceptions are *Relaxation* and *Anger*. *Relaxation* occupies the positive part of the valence dimension, but with almost exclusively light green hues, whereas *Anger* is dominated by the bright red and black hues. All emotions exhibiting high activity (high arousal) contain more red hues, regardless of their valence (e.g. compare *Anger*, *Energetic* and *Liveliness*). In the music context (Figure 4), the red-purple hues prevail over the dark blue-violet and the grey-black hues for the perceived negative emotions *Anger* and *Fear*, whereas the green hues prevail over the yellow hues for the perceived positive emotions of *Joy* and *Happiness*, and dominate for *Relaxation* and *Calmness*. *Sadness* differentiates itself from the rest of the emotions with the prevalent blueness. There is an interesting relation between the color associations and the valence-arousal ratings for *Anticipation*, where the color hues associated with the negative emotions dominate the negative pole of the valence space, whereas the green-blue hues associated with the more passive but pleasant emotions (such as *Relaxation* and *Calmness*) dominate in the lower right quadrant of positive valence and negative arousal.

The comparison of Figures 3 and 4 and the associated data in Table I shows that the emotions in the music context occupy a more central position of the valence-arousal space and with higher variance on the valence dimension, in comparison to the emotions rated in the non-music context. The mean and standard deviation of the valence and arousal ratings for the emotions in the non-music and music context are presented in Table I. There is a notable difference in the valence-arousal ratings and the variance for the following emotions, where their valence is compared for the non-music vs. music context, respectively: *Anger*: $\mu(V) = -0.68$ vs. $\mu(V) = -0.41$, $\sigma(V) = 0.26$ vs. $\sigma(V) = 0.42$; *Fear*: $\mu(V) = -0.75$ vs. $\mu(V) = -0.45$, $\sigma(V) = 0.24$ vs. $\sigma(V) = 0.41$; and *Sadness* $\mu(V) = -0.76$ vs. $\mu(V) = -0.37$, $\sigma(V) = 0.22$ vs. $\sigma(V) = 0.45$. Also notable is the positive arousal for the typically more passive musical emotions, such as *Sadness* ($\mu(A) = 0.04$) and *Relaxation* ($\mu(A) = 0.24$). This shows that musical emotions are being perceived differently from their non-music counterparts and can at times occupy semantically opposite positions (Zentner et al., 2008), especially along the valence dimension, where for example sad music can sometimes be perceived as pleasant (Vuoskoski

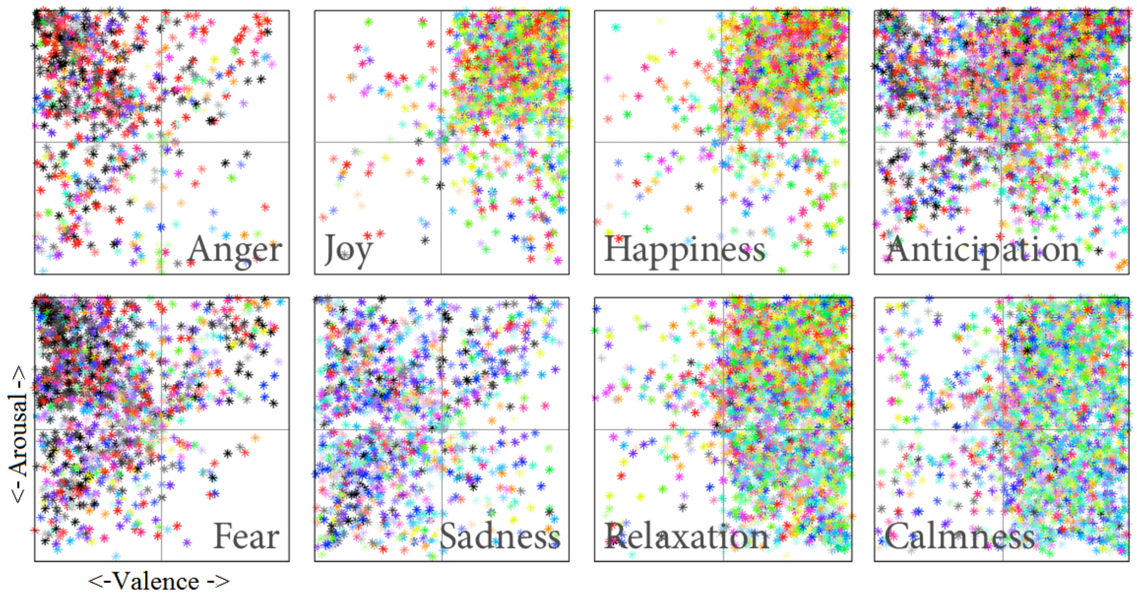

Fig. 4. Emotions and their color associations in the valence-arousal space (music context). The participants assigned a valence-arousal point and a color for perceived emotions for each music excerpt in the third part of the main survey (described in Section III-B3). Compared to Figure 3 there is a larger variety in positioning of emotions in the valence-arousal space as well as the chosen colors.[9]

and Thompson, 2012).

As already discussed, the rationale for incorporating the visual (color) modality lies in the relations between emotions, colors and music, as demonstrated in several studies (e.g. (Wexner, 1954; Schaie, 1961; Adams and Osgood, 1973; Valdez and Mehrabian, 1994; Palmer et al., 2011, 2016)). As Figures 3 and 4 show, the dimension of color uncovers some qualitative differences among emotions that are otherwise constrained by the two-dimensional affective space of valence and arousal. One of the limitations of such space is the reduction of emotions onto two dimensions where perceptually different emotions (e.g. compare *Sadness* and *Fear* in Figure 4) often occupy the same regions of the affective space. In this context, color can give us additional insight and show how emotions, even those sharing the same regions of the affective space, are visually perceived as different. The role of color in the emotional perception of music is further discussed in Sections IV-D and IV-E.

[9]Colors appear in grayscale in the printed version of this paper. For a colored Figure, please refer to the online version.

C. Perceived and induced musical emotions

Emotional processing of music generally involves two types of emotions (Evans and Schubert, 2008): the emotions that are conveyed by music (perceived) and those evoked in a listener (induced). The analysis of the Moodo dataset shows that the participants do differentiate between the perceived and induced musical emotions. The Kruskal-Wallis test shows significant differences in the ratings of: *Sadness*, where for induced *Sadness* valence is less negative ($p < 0.001$, $r = 0.11$), *Anticipation*, where for induced *Anticipation* valence is more positive ($p < 0.001$, $r = 0.13$) and arousal is higher ($p = 0.001$, $r = 0.05$), and *Anger*, where for induced *Anger* valence is more negative ($p = 0.015$, $r = -0.08$).

Figure 5 shows that in certain music contexts, especially those with perceived negative connotation, music can produce a variety of, sometimes polar, perceived-induced emotion responses in the listener. This is in line with previous research on perceived and induced emotions in music (Schubert, 2013), and while the differentiation between the two categories is not always clear (Gabrielsson, 2002), these results show that both aspects should be accounted for when modeling emotions in music.

D. Influence of genre on the perception of musical emotions

Emotional mediation of color and music is to some extent further influenced by music genre. Beyond various musical characteristics and styles that differentiate one genre from another, the music genres can also convey a plethora of (sometimes very distinctive) musical emotions. These emotions can be perceived differently and sometimes even as divergent, depending on the underlying musical characteristics and style of the particular genre (Eerola, 2011).

Figure 6 shows the top three music genres for the perceived musical emotions: *Anger*, *Joy*, *Anticipation*, *Fear*, *Sadness*, and *Calmness*. Each point represents the average valence-arousal rating for the music excerpt belonging to one of the top three genres (different genres are shown with different icons). The prevalent (most frequently used) color associated with each music excerpt is shown as a color of each point, whereas the prevalent genre color is shown in the icon next to the genre label in the graph's legend. The genre categorization of the music excerpts is based on their categorization in the music collections used in the study (see Section III-B3). The results show that different genres convey different sets of musical emotions and that even the same emotion might be perceived differently across musically and stylistically heterogeneous genres. For example the Kruskal-Wallis test shows that *Sadness* in Rock is perceived as significantly

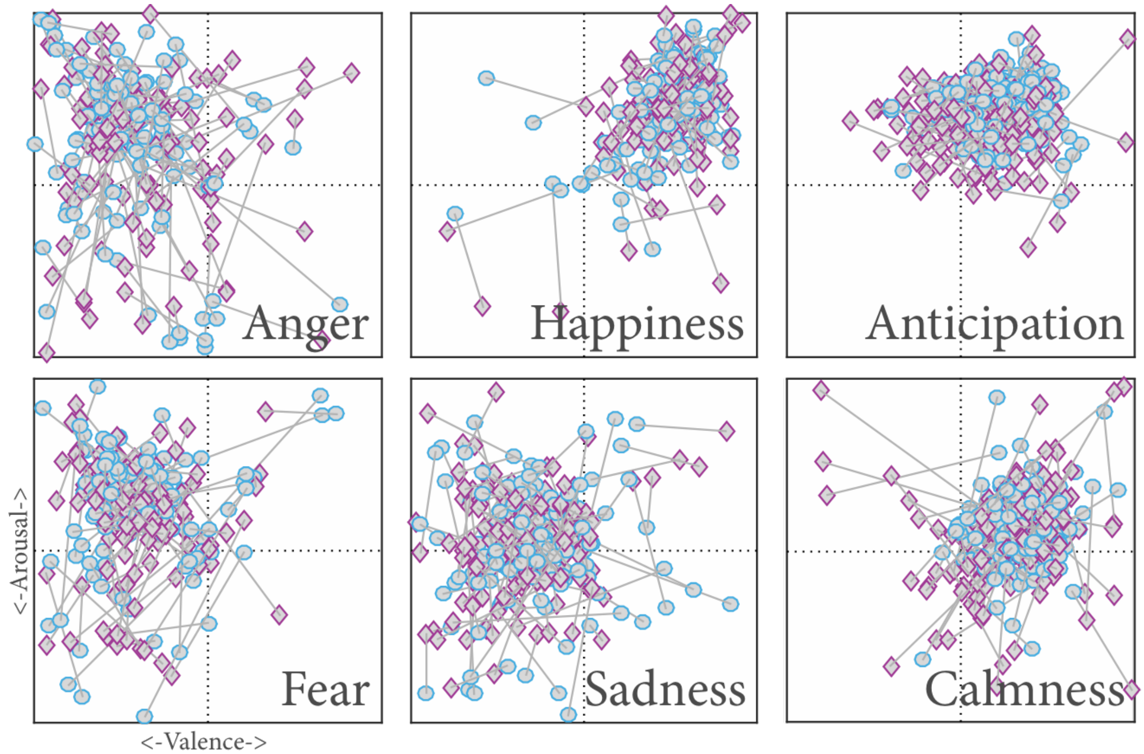

Fig. 5. Differences between the induced (diamond) and perceived (circle) musical emotions in the valence-arousal space. The average participants' ratings (centroids) for individual music excerpts are shown as induced-perceived emotion pairs.

more negative than in Pop (p = 0.02, r = -0.28) or Country (p = 0.001, r = -0.45), *Anticipation* is perceived as more neutral in Electronic genre than in New-Age (p = 0.01, r = -0.2) or Pop (p = 0.01, r = -0.22), and *Calmness* is perceived as less positive in Pop than in New-Age (p = 0.02, r = -0.22) or Easy Listening (p <0.001, r = -0.33).

E. Additional user context: mood, gender and age

The analyses presented above show the role of emotions in the perception of color and music. Additional user context can give further insight. In what follows, the influences of mood, gender, and age, as well as the effects of various emotion combinations are shortly presented. These findings further emphasize the role of user context in MIR and MER research.

The analysis of variance (Kruskal-Wallis test) for the emotional perception of color and music, and with regard to user context, revealed several interesting differences. With increased age, the active emotions (e.g. *Energetic*) are perceived to be less positive ($p < 0.001$, Bonferroni

[11]Colors appear in grayscale in the printed version of this paper. For a colored Figure, please refer to the online version.

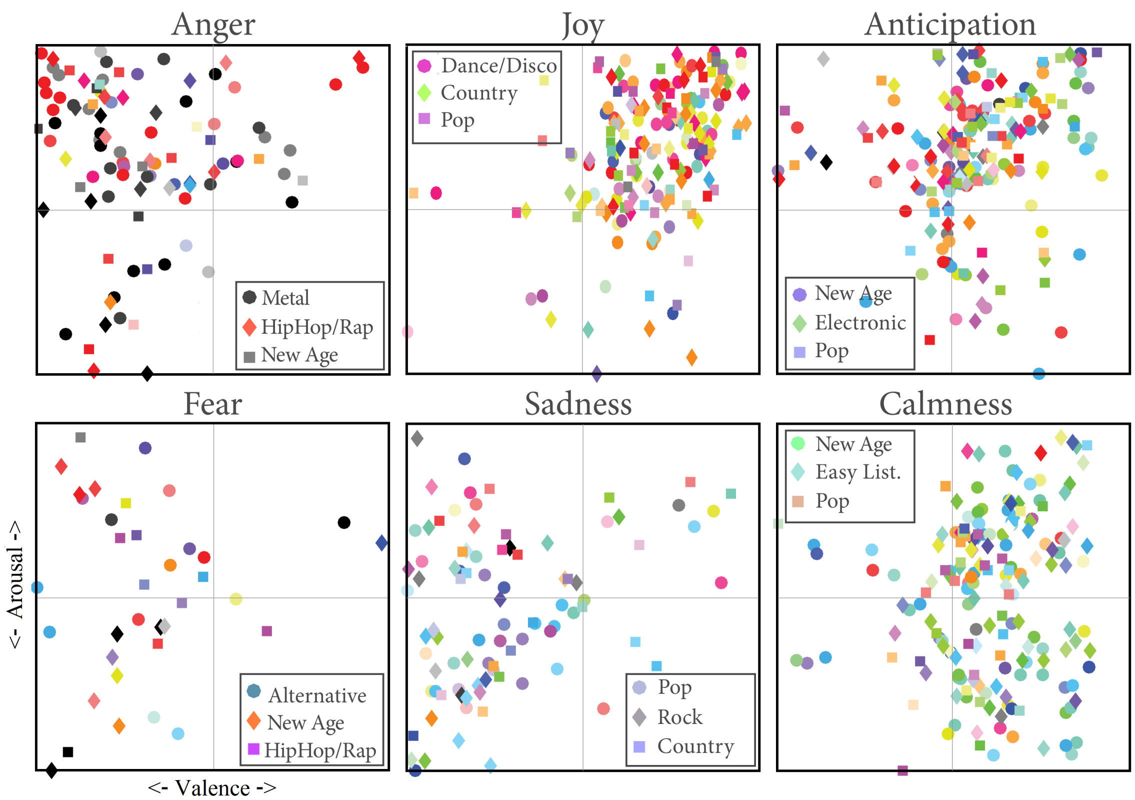

Fig. 6. Top three genres for the individual emotions: *Anger*, *Joy*, *Anticipation*, *Fear*, *Sadness*, and *Calmness*. Points represent the average valence-arousal ratings and prevalent colors for the individual music excerpts from the respective genres (refer to the genre icons in the graph legend), whereas the colors of the icons in the legend represent the prevalent color for each genre.[11]

corrected, df = 4, largest effect r = -0.21), whereas emotions with typically low arousal (e.g. *Relaxed*) are perceived as more active (p = 0.047, Bonferroni corrected, df = 4, largest effect r = 0.11). Women prefer more extreme ranges in the valence-arousal ratings than men (e.g. see Figure 7), as well as more intensive colors for the positive and negative emotions, as expressed in higher color saturation for *Energetic* ($p < 0.001$, r = 0.18) and also for *Happiness*, *Liveliness* and *Joy*. Also, women perceive *Dreaminess* as more positive ($p < 0.001$, r = 0.11) and *Tension* as more negative ($p < 0.001$, r = -0.14).

To further illustrate these differences, the quantized valence-arousal space visualization (Ju-Chiang et al., 2015) was used in Figures 7 and 9. Figure 7 shows the gender-based differences in perception of *Surprise* and *Tension*. These emotions represent the dynamic properties of music (Juslin and Sloboda, 2001; Juslin and Västfjäll, 2008) and are perceived as such on the arousal dimension, with both gender groups rating them as active. However, some interesting differences

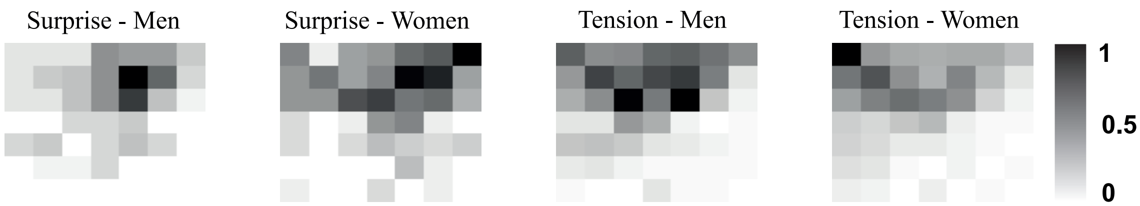

Fig. 7. Gender differences in perception of musical emotions for *Surprise* and *Tension*. The heatmap shows the frequency of valence and arousal ratings for both genders.

can be found on the valence dimension, with the ratings for women more distributed for *Surprise* and the ratings for men more distributed for *Tension*, as shown in Figure 7.

The participants' current mood also influences their perception of emotions and colors. The more positive and active the mood, the more are the positive emotions perceived as positive and active—e.g. for *Joy* (p = 0.011, Bonferroni corrected, df = 4, largest effect r = 0.13) and *Happiness* (p = 0.002, Bonferroni corrected, df = 4, largest effect r = 0.19)—whereas the negative emotions are perceived as less negative (and often positive in valence) and more active, e.g. *Tension* (p = 0.001, Bonferroni corrected, df = 4, largest effect r = 0.16). Mood also influences participants' perception of musical emotions, as shown in Figure 8. The figure illustrates the effects of mood on perception of color for *Sadness* and *Happiness* and their valence-arousal ratings. The participants in a *satisfied* mood tend to rate *Sadness* as pleasant (more positive in valence) more frequently, compared to the participants feeling *discontent*. The color of musical emotions is also perceived differently, depending on the mood. In Figure 8, the effect of negative mood on the color of *Sadness* is obvious, with the black hues only present for the participants feeling *discontent*. However, as the figure shows, the participant's negative mood only influences their color perception of *Sadness*, but not *Happiness*.

Also widespread are the effects of emotion combinations used by the participants, as presented in Figure 9. The figure shows the quantized valence-arousal space for *Anticipation*, *Liveliness*, and *Tension*, and the effects of their combinations with *Anger*, *Fear*, *Happiness*, and *Joy*, when they concurrently appeared in the ratings. Common to all three is the positive position on the arousal dimension, with *Tension* leaning towards the negative and *Liveliness* towards the positive valence. Figure 9 shows how their association with the negative emotions (*Anger* and *Fear*) affects

[13]Colors appear in grayscale in the printed version of this paper. For a colored Figure, please refer to the online version.

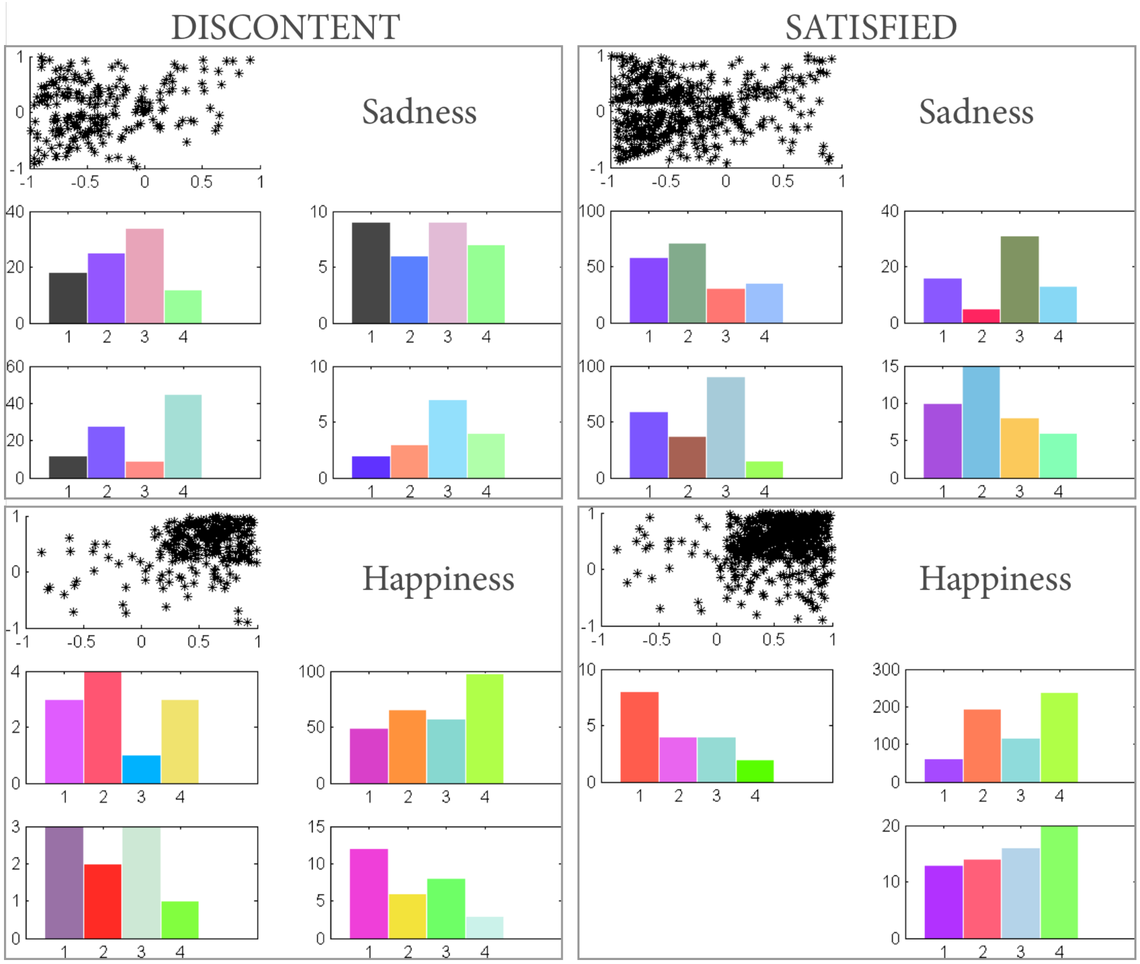

Fig. 8. The effects of the participants' negative (Discontent) and positive mood (Satisfied) on the color perception of *Sadness* and *Happiness*. The figure shows the valence-arousal ratings for *Sadness* and *Happiness* as perceived by the discontent and satisfied participants, with the frequencies and colors of the most prevalent hues (four per quadrant of the valence-arousal space) shown in the bar charts.[13]

their valence-arousal ratings towards the negative valence, whereas their association with the positive emotions (*Happiness*, *Joy*) gravitates the ratings towards the positive valence (Kruskall-Wallis, $p < 0.001$, Bonferroni corrected, df = 3, largest effect r = 0.51 for Anticipation-Fear vs. Anticipation-Joy). All three, *Anticipation*, *Liveliness*, and *Tension*, are often considered as candidates for the alternative third dimension, used in modeling the dynamic aspects of music, such as tempo and rhythm (Juslin and Sloboda, 2001; van der Zwaag et al., 2011).

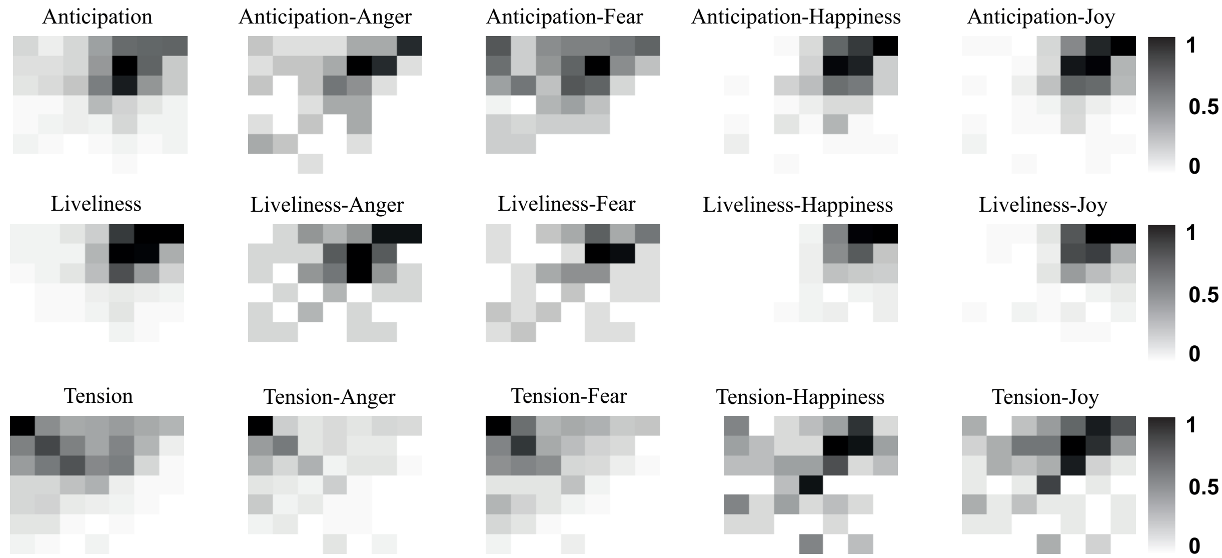

Fig. 9. The effects of different emotion combinations on the perception of musical emotions for *Anticipation*, *Liveliness* and *Tension*. The heatmap shows the differences in the valence-arousal ratings as influenced by the associated negative (*Anger*, *Fear*) and positive (*Happiness*, *Joy*) emotions.

F. Predicting emotional responses: applying machine learning to the Moodo dataset

The size of the Moodo dataset provides a suitable foundation for application of machine learning algorithms to predict users' responses, given the user context. To demonstrate its usefulness, an experiment was conducted with the goal of predicting perceived emotion labels and their valence-arousal ratings. An intermediate data fusion approach called Data Fusion by Matrix Factorization (DFMF) (Zitnik and Zupan, 2015) was used for modeling the dataset. The DFMF is based on the constrained matrix factorization and has shown success in analyzing heterogeneous data sources, overcoming several difficulties that many machine learning methods face when processing sparse and high-dimensional data. The DFMF is flexible, requires minimal to no preprocessing of input data (Zitnik and Zupan, 2015), and can model multi-dimensional and multi-entity data. Overall, the DFMF is capable of modeling the heterogeneity of the Moodo dataset.

The DFMF model (Figure 10) was applied to the Moodo dataset for prediction of perceived emotion labels and their valence-arousal ratings based on the user profiles, ratings of musical excerpts, high-level musical features annotated by a musicologist (rhythm, tempo, harmonic complexity, tunefulness), and music similarities in the acoustic feature space calculated by block

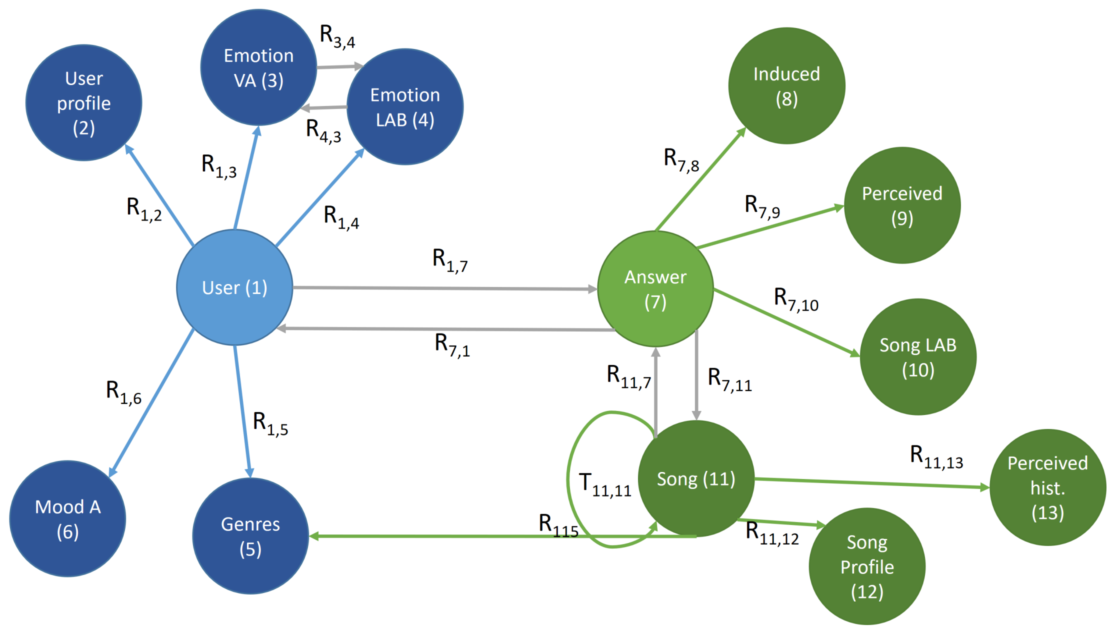

Fig. 10. A graph representation of the fusion model applied to the Moodo dataset. The 'user' entity (blue) represents information about the user: the demographic profile, music preferences, current mood, user's ratings of emotions and colors. The 'answer' entity (green) represents all the ratings for the selected music excerpt: the ratings for the perceived and induced musical emotions and the associated colors. The 'song' entity represents the music excerpt, together with the description of its high and low-level musical features.

level features (Seyerlehner et al., 2010).

Two models were trained and tested: one model used both the user profile data and the musical features for training and prediction, whereas the other model was trained only on the musical features (without user profiles). Leave-one-out cross-validation was used to compare the performance of both models. The results showed that the model that included user profiles significantly outperformed the one without (F-measure 0.3 vs. 0.25 for prediction of labels, and mean absolute difference of 0.32 vs. 0.41 for prediction of valence-arousal ratings). The models used for prediction were simple linear models, so results are not on par with the state-of-the-art models, however, the difference between both models is significant, thus confirming the usability of user profiles for predictive modeling in affective MIR scenarios.

V. Conclusions

The overall goal of this paper was to address some of the current challenges in modeling emotions in music and argue towards a tighter integration of affect and user context in MIR and MER research. The presented Moodo dataset of emotional and color perception of music is publicly accessible online[14] and can serve as a basis for future research.

The presented analyses reveal interactions between emotions, color, and music, and emphasize the role of user context (mood, gender, age, and musical preferences) in MIR and MER research. Analyses of the gathered data show that color is mediated by emotions in both music and non-music contexts, with the former revealing substantially larger variance in the color and valence-arousal ratings. The results also reveal the importance of user context, showing gender and age differences in perception of musical emotions and their color associations. We discuss the differences between ratings of perceived and induced emotions, as well as the influence of music genre on preceived emotions and colors.

The dimension of color was used to uncover some of the qualitative differences in the perception of emotions (both in non-music and in music context) within the two-dimensional affective space of valence and arousal. One obvious limitation of such a space is reduction of emotional information onto two dimensions, where color can give us additional insight by showing how emotions, even those sharing the same region of the valence-arousal space, are visually perceived as different.

Affective MIR research can give important insights into emotional processing and interactions between visual and auditory modalities and presents a number of challenges for the future. From a MIR perspective, the integration of affect and user context can improve automatic emotion recognition in music, as shown by the experiment on the prediction of emotional responses. However, the underlying mechanisms governing multimodal interactions are not unique to music and most of the findings presented here (e.g. context-dependent emotion-color associations) can be extended to other domains of affective computing (Sokolova and Fernández-Caballero (2015)). Furthermore, user-aware systems research can benefit from such studies, both in terms of improving user interfaces and overall user experience—with the audio-visual modalities as the dominant features of user interface design—as well as by improving existing recommendation algorithms.

References

P. N. Juslin, "What does music express? Basic emotions and beyond," *Frontiers in Psychology*, vol. 4, no. September, pp. 1–14, 2013.

[14]The Moodo dataset is available at: http://moodo.musiclab.si

T. Eerola, "Modelling Emotional Effects of Music: key Areas of Improvement," in *Proceedings of the Sound and Music Computing Conference 2013, SMC 2013, Stockholm, Sweden*, Aug. 2013.

J. K. Vuoskoski and T. Eerola, "The role of mood and personality in the perception of emotions represented by music," *Cortex*, vol. 47, no. 9, pp. 1099–1106, 2011.

E. Schubert, "Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives," *Frontiers in Psychology*, vol. 4, no. December, pp. 1–18, 2013.

R. J. Zatorre, J. L. Chen, and V. B. Penhune, "When the brain plays music: auditory-motor interactions in music perception and production." *Nature reviews. Neuroscience*, vol. 8, no. 7, pp. 547–558, 2007.

S. Koelsch, "Towards a neural basis of music-evoked emotions," *Trends in Cognitive Sciences*, vol. 14, no. 3, pp. 131–137, 2010.

C. Spence, "Audiovisual multisensory integration," *Acoustical science and technology*, vol. 28, no. 2, pp. 61–70, 2007.

D. J. Levitin and A. K. Tirovolas, "Current advances in the cognitive neuroscience of music," *Annals of the New York Academy of Sciences*, vol. 1156, pp. 211–231, 2009.

M. Pearce and M. Rohrmeier, "Music Cognition and the Cognitive Sciences," *Topics in Cognitive Science*, vol. 4, no. 4, pp. 468–484, 2012.

S. M. Stalinski and E. G. Schellenberg, "Music Cognition: A Developmental Perspective," *Topics in Cognitive Science*, vol. 4, no. 4, pp. 485–497, 2012.

C. J. Stevens, "Music Perception and Cognition: A Review of Recent Cross-Cultural Research," *Topics in Cognitive Science*, vol. 4, pp. 653–667, 2012.

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)*, London, 2012, pp. 395–410.

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)*, Utrecht, 2010, pp. 255–266.

C. Laurier, M. Sordo, J. Serrà, and P. Herrera, "Music Mood Representations from Social Tags," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2009, pp. 381–386.

D. Weigl and C. Guastavino, "User studies in the Music Information Retrieval Literature." in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2013, pp. 335–340.

M. Schedl, A. Flexer, and J. Urbano, "The neglected user in music information retrieval research," *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 523–539, 2013.

T. Eerola and J. K. Vuoskoski, "A review of Music and Emotion Studies: Approaches, Emotion Models, and Stimuli," *Music Perception*, vol. 30, no. 3, pp. 307–340, 2013.

J. Lee and S. Cunningham, "Toward an Understanding of the History and Impact of User Studies in Music Information Retrieval," *Journal of Intelligent Information Systems*, 2013.

Y.-H. Yang and J.-Y. Liu, "Quantitative Study of Music Listening Behavior in a Social and Affective Context," *IEEE Transactions on Multimedia*, vol. 15, no. 6, pp. 1304–1315, Oct. 2013.

M. Sokolova and A. Fernández-Caballero, "A Review on the Role of Color and Light in Affective Computing," *Applied Sciences*, vol. 5, no. 3, pp. 275–293, 2015. [Online]. Available: http://www.mdpi.com/2076-3417/5/3/275/

H. J. Eysenck, "A critical and experimental study of colour preferences," *The American Journal of Psychology*, vol. 54, no. 3, pp. 385–394, 1941.

L. B. Wexner, "The degree to which colors (hues) are associated with mood-tones." *Journal of applied psychology*, vol. 38, no. 6, p. 432, 1954.

K. W. Schaie, "Scaling the association between colors and mood-tones," *The American Journal of Psychology*, vol. 74, no. 2, pp. 266–273, 1961.

F. M. Adams and C. E. Osgood, "A Cross-Cultural Study of the Affective Meanings of Color," *Journal of Cross-Cultural Psychology*, vol. 4, no. 2, pp. 135–156, 1973.

P. Valdez and A. Mehrabian, "Effects of color on emotions." *Journal of experimental psychology: General*, vol. 123, no. 4, p. 394, 1994.

S. E. Palmer, T. Langlois, T. Tsang, K. B. Schloss, and D. J. Levitin, "Color, music, and emotion," *Journal of Vision*, vol. 11, no. 11, p. 391, 2011.

S. E. Palmer, T. A. Langlois, and K. B. Schloss, "Music-to-color associations of single-line piano melodies in non-synesthetes," *Multisensory Research*, vol. 29, no. 1-3, pp. 157–193, 2016.

M. Pesek, P. Godec, M. Poredoš, G. Strle, J. Guna, E. Stojmenova, M. Pogačnik, and M. Marolt, "Capturing the Mood: Evaluation of the MoodStripe and MoodGraph Interfaces," in *Management Information Systems in Multimedia Art, Education, Entertainment, and Culture*

(MIS-MEDIA), IEEE International Conference on Multimedia & Expo (ICME), 2014, pp. 1–4.

R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22 140, p. 55, 1932.

J. A. Russell, "A circumplex model of affect," *1Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

M. Mcvicar, T. Freeman, and T. De Bie, "Mining the Correlation Between Lyrical and Audio Features and the Emergence of Mood," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Miami, 2011, pp. 783–788.

Y.-H. Yang and H. H. Chen, "Machine Recognition of Music Emotion," pp. 1–30, 2012.

M. Barthet, D. Marston, C. Baume, G. Fazekas, and M. Sandler, "Design and evaluation of semantic mood models for music recommendation using editorial tags," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Curitiba, 2013.

U. Schimmack and R. Reisenzein, "Experiencing activation: energetic arousal and tense arousal are not mixtures of valence and activation." *Emotion (Washington, D.C.)*, vol. 2, no. 4, pp. 412–7, Dec. 2002.

E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, and A. Dacquet, "Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts," *Cognition & Emotion*, vol. 19, no. 8, pp. 1113–1139, Dec. 2005.

S. Canazza, G. D. Poli, A. Rodà, A. Vidolin, and P. Zanon, "Kinematics-energy space for expressive interaction in music performance," in *Proceedings of MOSART. Workshop on current research directions in computer music*, Jan. 2001, pp. 35–40.

T. Eerola, O. Lartillot, and P. Toiviainen, "Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2009, pp. 621–626.

J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. Ellsworth, "The world of emotion is not two-dimensional," *Psychological Science*, vol. 18, pp. 1050–1057, 2007.

M. Barthet, G. Fazekas, and M. Sandler, "Multidisciplinary perspectives on music emotion recognition: Implications for content and context-based models," in *Proc. CMMR*, London, 2012, pp. 492–507.

X. Hu and J. S. Downie, "Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, 2007.

M. Zentner, D. Grandjean, and K. R. Scherer, “Emotions evoked by the sound of music: characterization, classification, and measurement,” *Emotion*, vol. 8, no. 4, p. 494, 2008.

K. Torres-Eliard, C. Labbé, and D. Grandjean, “Towards a Dynamic Approach to the Study of Emotions Expressed by Music,” *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 78, pp. 252–259, 2011.

A. Aljanaki, F. Wiering, and R. C. Veltkamp, “Computational Modeling of Induced Emotion Using GEMS,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Taipei, 2014, pp. 373–378.

T. Eerola and J. K. Vuoskoski, “A comparison of the discrete and dimensional models of emotion in music,” *Psychology of Music*, vol. 39, no. 1, pp. 18–49, Aug. 2010.

J. K. Vuoskoski and T. Eerola, “Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences,” *Musicae Scientiae*, vol. 15, no. 2, pp. 159–173, Jul. 2011.

E. M. Schmidt and Y. E. Kim, “Modeling Musical Emotion Dynamics with Conditional Random Fields.” in *ISMIR*, 2011, pp. 777–782.

J. A. Speck, E. M. Schmidt, B. G. Morton, and Y. E. Kim, “A Comparative Study of Collaborative vs. Traditional Musical Mood Annotation,” in *2Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Miami, 2011, pp. 549–554.

D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, “Semantic annotation and retrieval of music and sound effects,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 2, pp. 467–476, 2008.

D. Tingle, Y. E. Kim, and D. Turnbull, “Exploring automatic music annotation with ”acoustically-objective” tags,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, New York, Mar. 2010, pp. 55–62.

B. Schuller, C. Hage, D. Schuller, and G. Rigoll, “‘Mister DJ, Cheer Me Up!’: Musical and textual features for automatic mood classification,” *Journal of New Music Research*, vol. 39, no. 1, pp. 13–34, 2010.

A. Aljanaki, D. Bountouridis, J. A. Burgoyne, J. van Balen, F. Wiering, H. Honing, and R. C. Veltkamp, “Designing games with a purpose for data collection in music research. Emotify and Hooked: Two case studies,” *Lecture Notes in Computer Science*, p. 2014, 2014.

A. Lykartsis, A. Pysiewicz, H. Coler, and S. Lepa, “The Emotionality of Sonic Events : Testing the Geneva Emotional Music Scale (GEMS) for Popular and Electroacoustic Music,” in

Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Jan. 2013, pp. 1–15.

J. Jaimovich, N. Coghlan, and R. Knapp, "Emotion in Motion: A Study of Music and Affective Response," in *From Sounds to Music and Emotions*. Springer, 2013, pp. 19–43.

X. Serra, M. Magas, E. Benetos, M. Chudy, S. Dixon, A. Flexer, E. Gómez, F. Gouyon, P. Herrera, S. Jordà, O. Paytuvi, G. Peeters, J. Schlüter, H. Vinet, and G. Widmer, *Roadmap for Music Information ReSearch*. London: MIReS Consortium, 2013.

J. Donaldson and P. Lamere, "Using Visualizations for Music Discovery," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2009, p. Tutorial.

E. Isaacson, "What You See Is What You Get: On Visualizing Music," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, 2005, pp. 389–395.

T. Bergstrom, K. Karahalios, and J. Hart, "Isochords: visualizing structure in music," *Proceedings of Graphics Interface ...*, pp. 297–304, 2007.

A. Mardirossian and E. Chew, "Visualizing Music: Tonal Progressions and Distributions," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, 2007, pp. 189–194.

H. Grohganz, M. Clausen, N. Jiang, and M. Mueller, "onverting path structures into block structures using eigenvalue decompositions of self-similarity matrices," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Curitiba, 2013.

N. Jiang and M. Mueller, "Automated methods for analyzing music recordings in sonata form," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Curitiba, 2013.

E. Pampalk, "Islands of music analysis, organization, and visualization of music archives," *OGAI Journal (Oesterreichische Gesellschaft fuer Artificial Intelligence)*, vol. 22, no. 4, pp. 20–23, 2003.

S. Kurabayashi and T. Imai, "Chord-Cube: Music Visualization and Navigation System with an Emotion-Aware Metric Space for Temporal Chord Progression," *International Journal On Advances in Internet Technology*, vol. 7, no. 1, pp. 52–62, 2014.

L. B. Meyer, *Emotion and Meaning in Music*. Chicago: University of Chicago Press, 1956.

L.-C. Ou, M. R. Luo, A. Woodcock, and A. Wright, "A study of colour emotion and colour preference. Part I: Colour emotions for single colours," *Color Research & Application*, vol. 29, no. 3, 2004.

P. N. Juslin and D. Västfjäll, “Emotional responses to music: The need to consider underlying mechanisms,” *Behavioral and brain sciences*, vol. 31, no. 5, pp. 559–575, 2008.

C. Spence, “Crossmodal correspondences: A tutorial review,” *Attention, Perception, & Psychophysics*, vol. 4, no. 1, pp. 971–995, 2011.

S. E. Palmer, K. B. Schloss, Z. Xu, and L. R. Prado-León, “Music–color associations are mediated by emotion,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 22, pp. 8836–8841, 2013.

P. N. Juslin and J. A. Sloboda, *Music and emotion: Theory and research.* Oxford University Press, 2001.

T. Eerola, “Modeling Listeners’ Emotional Response to Music,” *Topics in Cognitive Science*, vol. 4, pp. 607–624, 2012.

N. A. Remmington, L. R. Fabrigar, and P. S. Visser, “Reexamining the circumplex model of affect.” *Journal of personality and social psychology*, vol. 79, no. 2, pp. 286–300, Aug. 2000.

D. Watson, L. A. Clark, and A. Tellegen, “Development and validation of brief measures of positive and negative affect: the PANAS scales.” *Journal of personality and social psychology*, vol. 54, no. 6, pp. 1063–1070, 1988.

P. Ekman, “An argument for basic emotions,” *Cognition and Emotion*, vol. 6, pp. 169–200, 1992.

P. N. Juslin and P. Laukka, “Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening,” *Journal of New Music Research*, vol. 33, no. 3, pp. 217–238, Sep. 2004. [Online]. Available: http://www.ingentaconnect.com/content/routledg/jnmr/2004/00000033/00000003/art00004

M. Pesek, A. Isaković, G. Strle, and M. Marolt, “Improving the usability of online usability surveys with an interactive Stripe scale,” in *Proceedings of the 19th International Multiconference INFORMATION SOCIETY*, Ljubljana, 2016, pp. 21–24.

S. G. Hart, “Nasa-Task Load Index (NASA-TLX); 20 Years Later,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006.

M. Pesek, P. Godec, M. Poredoš, G. Strle, J. Guna, E. Stojmenova, M. Pogačnik, and M. Marolt, “Capturing the Mood: Evaluation of the MoodStripe and MoodGraph Interfaces,” in *Management Information Systems in Multimedia Art, Education, Entertainment, and Culture (MIS-MEDIA), IEEE Internation Conference on Multimedia & Expo (ICME)*, 2014, pp. 1–4.

J. K. Vuoskoski and W. F. Thompson, “Who Enjoys Listening to Sad Music and Why?” pp.

311–317, 2012.

P. Evans and E. Schubert, “Relationships between expressed and felt emotions in music,” *Musicae Scientiae*, vol. 12, pp. 75–99, 2008.

A. Gabrielsson, “Emotion Perceived and Emotion Felt: Same or Different?” *Musicae Scientiae*, vol. 5, no. 1_suppl, pp. 123–147, Sep. 2002.

T. Eerola, “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*, vol. 40, no. 4, pp. 349–366, 2011.

W. Ju-Chiang, W. Hsin-Min, and G. Lanckriet, “A Histogram Density Modeling Approach to Music Emotion Recognition,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2015.

M. D. van der Zwaag, J. H. D. M. Westerink, and E. L. van den Broek, “Emotional and psychophysiological responses to tempo, mode, and percussiveness,” *Musicae Scientiae*, vol. 15, pp. 250–269, 2011.

M. Zitnik and B. Zupan, “Data fusion by matrix factorization,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 1, pp. 41–53, 2015.

K. Seyerlehner, G. Widmer, M. Schedl, and P. Knees, “Automatic Music Tag Classification based on Block-Level Features,” in *Proceedings of the 7th Sound and Music Computing Conference (SMC 2010)*, Barcelona, Spain, July 2010.