Chapter 16
Towards User-Aware Music Information Retrieval: Emotional and Color Perception of Music

Gregor Strle, Matevž Pesek and Matija Marolt

Abstract This chapter presents our findings on emotional and color perception of music. It emphasizes the importance of user-aware music information retrieval (MIR) and the advantages that research on emotional processing and interaction between multiple modalities brings to the understanding of music and its users. Analyses of results show that correlations between emotions, colors and music are largely determined by context. There are differences between emotion-color associations and valence-arousal ratings in non-music and music contexts, with the effects of genre preferences evident for the latter. Participants were able to differentiate between perceived and induced musical emotions. Results also show how associations between individual musical emotions affect their valence-arousal ratings. We believe these findings contribute to the development of user-aware MIR systems and open further possibilities for innovative applications in MIR and affective computing in general.

16.1 Introduction

Research in music information retrieval (MIR) is multidisciplinary, comprising related fields of computer science, machine learning, cognitive science and human-computer interaction, among others. It aims to tackle three fundamental aspects: music information, the user, and the interactions between the two [31, 73]. Music information represents both the inherent musical properties (music content) and the contextual information about music, whereas information about the user includes

G. Strle (✉)
Scientific Research Centre, Institute of Ethnomusicology,
Slovenian Academy of Sciences and Arts, Ljubljana, Slovenia
e-mail: gregor.strle@zrc-sazu.si

M. Pesek · M. Marolt
Faculty of Computer and Information Science, University of Ljubljana,
Ljubljana, Slovenia
e-mail: matevz.pesek@fri.uni-lj.si

M. Tkalčič et al. (eds.), *Emotions and Personality in Personalized Services*,
Human–Computer Interaction Series, DOI 10.1007/978-3-319-31413-6_16

general information (such as personality traits and music preferences), as well as context-related information about the user's use and perception of music (e.g., current mood and emotions). Finally, the interaction involves various communication and interface modalities that integrate general and context-aware information about the user to reflect her music information needs and improve relevancy of results.

All three aspects should be considered for the effective development of MIR. However, the integration of user-context and related interaction modalities is still relatively poor. By ignoring the user, her perception and use of music, system-focused approaches lack grounding in the real world. Furthermore, the design and interaction decisions regarding representation of musical information often come secondary, based on the developer's intuitive notion or a priori assumptions of typical usage scenarios, not on the real data about the user [73, 75, 84, 95, 104]. Recent reviews of the state-of-the-art show that MIR research is still predominantly system-focused [47, 48, 74, 75, 82, 84, 104]. This situation is reflected in the activities of MIREX—Music Information Retrieval Evaluation eXchange,[1] the largest community-based framework that focuses on advances in MIR techniques and algorithms tailored to a variety of music-related tasks. Examples include audio classification, melody extraction, key detection, tempo estimation, music similarity and retrieval, with more recent attempts to integrate user information for purposes of music emotion recognition and mood estimation from audio, lyrics, or collaborative tags and their use in music recommendation and playlist generation.

Research in user-aware MIR is relatively new and efforts towards systematic approach and construction of formal user models are hindered by the scope and multidisciplinary nature of the field [18, 48]. For example, research in music cognition is faced with inherent complexity of music and cognitive issues related to music processing, further intensified by multimodal interactions (e.g., visual and auditory), effects of personality traits, mood and emotions, as well as by the ambiguity of abstract musical concepts and social and cultural differences [13, 20, 41, 49, 65, 66, 69, 89, 90].

This situation is reflected in existing MIR datasets. User information is sparse, typically accounting for basic demographic information and general music preferences (such as genre), while lacking in examination of user's use and perception of music. More complex MIR datasets, aimed at integrating different perceptual modalities to get additional insight into human music processing, are still in initial stage and face a number of problems. For example, in scientific literature much emphasis has been given to emotional aspects of music perception, such as the relationship between musically evoked and perceived emotions [26, 43, 76, 77, 100, 108]. But competing theories and different emotion models add to the overall confusion and make systematic comparison of results difficult, if not impossible [20].

Presented research aims to contribute to the advances of user-aware MIR by offering the Moodo dataset—a large-scale dataset of mood-dependent, emotional and color responses to music [68]. In total, over 7000 user annotations had been gathered, taking into account demographic information, user's mood and emotions,

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

and her ratings of emotional and color perception of music for a variety of genres. In the process, two novel interfaces for emotion annotation had been developed, the *MoodStripe* and the *MoodGraph*, with the aim to alleviate some of the shortcomings of traditional emotion models.

In what follows, we discuss general aspects of gathering the Moodo dataset and present the analysis of results on emotional and color perception of music. Section 16.2 provides an overview of related work in MIR and music visualization, as well as some background on emotional processing and interactions between visual and auditory modalities, which serves as an introduction to the analysis of the Moodo dataset presented later in the chapter. Section 16.3 describes the design of the survey and the evaluation of two novel interfaces for gathering users' emotional responses to music, the *MoodStripe* and the *MoodGraph*. Section 16.4 provides the analysis of emotional and color responses to music. We show how emotions influence our perception of color and music, as well as interactions between both modalities. We conclude with the discussion on our findings.

16.2 Related Work

The following sections shortly present MIR research on emotional perception of music, issues in music visualization approaches, and give a general overview of emotional processing and interactions between auditory and visual modalities, setting the stage for the analysis in Sect. 16.4.

16.2.1 MIR Datasets and Music Visualization Approaches

16.2.1.1 Emotion Modeling in MIR

Most existing MIR studies on emotions in music use some variation of the discrete or dimensional emotion modeling approach for gathering user input, based on the Likert intensity scale or Russell's Circumplex model of affect [71]. Major difference between the two is the discrete emotion model represents individual emotions as discrete categories, whereas the dimensional model represents emotions as dimensions, typically in the two dimensional coordinate space of *valence* and *arousal*.

Variations of Russell's dimensional model have been used in several music-related studies [5, 46, 55, 106], with several researchers suggesting additional dimensions to better reflect the structure of musical emotions. For example, Schimmack et al. [78] propose two different interpretations of the arousal dimension through *energetic* (awake-tired) and *tense* (tense-calm) dimensions, Bigand et al. [7] and Canazza et al. [10] suggest that the additional third dimension *kinetics* may link perceived emotions with body posture and gestures, whereas Eerola et al. [21] propose *tension* as the additional third dimension to valence and arousal.

The strongest criticism of dimensional modeling is related to the limited number of dimensions used in modeling emotions, with the low-dimensional affective space typically reduced to valence and arousal. Discrete models aim to overcome this limitation by multiple-category rating of emotions [4, 40]. However, recent survey and comparative analysis of discrete and dimensional approaches to modeling emotions in music showed little advantages of the former over the latter—in fact, discrete models exhibit lower discriminative accuracy for musically induced emotions [19, 99].

In order to compensate for the limitations of current music emotion models, Zentner et al. [108] developed the GEMS—Geneva Emotional Music Scale, as a domain-specific emotion model for musically induced emotions. GEMS model, based on 45 terms, has been further adapted to shorter variants of 25 and 9 term models by Torres-Eliard et al. [93] and Aljanaki et al. [3]. The MIREX initiative, on the other hand, proposes a five-cluster model derived from the AllMusicGuide mood repository [32], with label sets consisting of 5–7 labels per cluster, resulting in total of 29 labels. In all, more extensive research is needed to confirm the advantages of initiatives like GEMS over more traditional approaches.

Furthermore, Saari and Eerola presented the affective circumplex transformation which possibly provides the connection between the discrete mood tags and the affective circumplex [72]. Wang et al. [101] proposed a Gaussian mixture representation model to enable the translation between discrete tags and continuous VA space. However, it seems there is currently no available dataset comprising both tag and VA point representations, as the aforementioned researches use multiple datasets to generate links between the two.

In Sect. 16.3 we present our contribution to the user-aware MIR, the *Mood-Graph*—a hybrid emotion model for gathering participants ratings by integrating multiple-category emotion labels with the dimensionality of valence-arousal space.

16.2.1.2 MIR Datasets

There is a growing number of MIR datasets that focus on modeling emotions in music.

The MoodSwings Turk Dataset contains on average 17 valence-arousal ratings for 240 clips of popular music [79]. The authors used a collaborative game MoodSwings [85] and Amazon Mechanical Turk (paid participation) for gathering perceived emotions in music. The game uses marked emoticons to express positive and negative emotions and their intensity in the valence-arousal space.

The Cal500 contains a set of mood labels for 500 popular songs [94], at approximately three annotations per song. The extended dataset CAL10k is also available, providing 10.870 songs from several thousand artists. It contains 475 acoustic tags and 153 genre tags [91] and 34 acoustic features.

The MTV Music Dataset [81] contains a set of five bipolar valence-arousal ratings, annotated by five different annotators with different musicological backgrounds, for 192 popular songs.

The *Emotion in Music Task* dataset, a part of MediaEval Benchmarking Initiative for Multimedia Evaluation [83], addresses the challenges of music emotion characterization and recognition. For example, the task for 2013 was dynamic emotion characterization, based on the continuous estimation of valence-arousal scores for each musical piece in the dataset. For this purpose, the annual music dataset of 45 s long musical pieces, annotated by a minimum of 10 workers, is gathered (1000 musical pieces for 2013 and 1744 for 2014). The organizers employ crowdsourcing approach, with annotations collected through Amazon Mechanical Turk and partially from publicly available data on Last.fm.

Using the aforementioned GEMS model, Aljanaki et al. [2] collected over 8000 responses by 1778 participants on a set of 400 music excerpts of classical, rock, pop and electronic music, equally represented by 100 excerpts each. The 9 GEMS model was used in the data gathering procedure, based on the Emotify game, developed for this purpose. Demographic data, comprising of gender, age and language was also collected.

Lykartsis et al. tested the GEMS model for electroacoustic and popular music [50], using a German version (GEMS-28-G), based on the GEMS-25 with additional three categories. There were 245 participants included in the study. The study included 20 music pieces of classical and popular instrumental music and electroacoustic music. Some demographic data was collected: age, language, level of education, music knowledge, and amount of listening to music per day.

An interesting application of the GEMS model is the study on emotional reactions to music, conducted by Jaimovich et al. [34], where the GEMS-9 model was used on 4000 participants and 12000 music excerpts based on 53 songs. Here, participants' electrodermal activity (EDA) and heart rate were also recorded.

The All Music Guide (AMG) 5 mood clusters were proposed by Hu and Downie [32], to "reduce the diverse mood space into a tangible set of categories" Several datasets used the proposed approach in a variety of task-specific applications. For example, Yi-Hsuan and Hu [106] collected a dataset of 2453 responses to a set of 500 Chinese music pieces—five labels per song on average and one expert annotation per song—to evaluate the acoustic features and compare them to responses on English music pieces. Laurier et al. [46] used the AMG mood clusters and *Last.fm* service social tags to observe the possible correlations between both. Panda et al. [63] created a multimodal MIREX-like emotion dataset collected from AllMusic database, organized by five emotion clusters from the AMG mood depository. The dataset is based on three sources containing 903 audio clips, 764 lyrics, and 193 midis.

To compare dimensional and discrete emotion models, Eerola et al. [16] gathered an annotated dataset of 360 film music clips, rated by 116 non-musician participants. Additional data about each participant include gender, age, years of musical training, and experience of playing an instrument. The experiment was divided into two stages: during the first stage, participants labeled individual musical excerpts in a three-dimensional valence-arousal-tension space (using bipolar scales), whereas during the second stage, 9-degree scales for each discrete emotion were used. The experiment was relatively time demanding, averaging between 50–60 min for each participant. The soundtracks dataset for music and emotion contains single mean ratings of perceived emotions (labels and values in a three-dimensional model are given).

Most of the existing MIR datasets contain a reasonable amount of demographic information. Yet none focus on interactions between visual and auditory modalities, connecting emotional and color perception of music—this has been the main motivation behind the Moodo dataset, presented in Sect. 16.3. Uncovering the relationship between emotions, colors and music is also relevant for more innovative approach to music visualization.

16.2.1.3 Music Visualization

In the past, user-oriented research in music visualization has been largely neglected [75, 82], but recent attempts (e.g., the Grand challenge at MIREX evaluation exchange initiative[2]) indicate a growing interest in the domain. There are numerous attempts of providing visualizations for a variety of music and audio features, from the low-level spectral features to the high-level music patterns and music metadata [15]. Most can be separated into two categories: visualizations of music parameters (e.g., harmonies, temporal patterns, and other music entities) [6, 29, 33, 35, 51] and visualizations of spaces representing relationships among different music pieces [44, 62]. The latter are more suitable for music recommendation systems and for data exploration in general. Examples include visualization of musical pieces as thumbnail images by Yoshii and Goto [107], visualization of personal music library (Torrens et al. [92]), and visualizations designed for the exploration of music collections on small-screen devices (Van Gulik et al. [96, 97]). Julia and Jorda [36] developed visualization for exploring large music collections in tabletop applications, thus extending user interaction beyond the standard keyboard/mouse interfaces, whereas Lamere and Eck [45] developed three-dimensional space visualization for music.

While there are significant advances in music visualization, most approaches still rely on the intuitive interpretation of color in music and lack real world data gathered with the analysis of various user scenarios [82]. For example, topographic visualizations of similarities in music, based on Self-Organizing Maps [42] or Islands of Music [61, 62]—a very popular approach in music visualization that efficiently reduces the dimensionality of data—use arbitrary sets of colors to differentiate between individual clusters. We believe that user-context is essential for improving music visualizations, as well as the overall design of MIR systems. The following section briefly discusses the importance of multimodal integration, both in MIR and affective computing in general.

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

16.2.2 Multimodal Interactions

Integrating sensory information from various modalities is essential for a coherent perceptual experience. This integration is ongoing—the brain continually collects and estimates multiple sources of sensory information and, based on our prior knowledge, personality traits, as well as our affective state and understanding of the momentary context, attempts to make coherent representations of reality [9, 12, 88]. As the flux of information signals is noisy and incomplete, "the brain reduces the variance in the integrated estimate and increases the robustness of the percept by combining and integrating sources of sensory information from within and across modalities" [23]. The brain's effort to provide a coherent percept from concurrent stimulation of multiple modalities is not always successful. In one of the most famous examples of audio-visual integration, the McGurk effect [54], the auditory experience of speech perception is modulated by visual information, combining incongruent audio-visual stimuli, and as a result, altering the phonetic processing into an illusory perception of a sound. Thus, for multimodal integration to be successful, the key constraints of semantic and spatio-temporal congruency should be met [86, 87]. Overall, research on auditory and visual perception [8, 14, 25, 53, 64, 86, 98, 105] has shown significant benefits of multimodal interactions in terms of "filling in" the missing information, enhancing individual modalities and increasing the accuracy and robustness of resulting percept.

16.2.2.1 Relations to Music

Emotional processing of music is affected by many factors, most notably by individual's personality and age, music preferences (e.g. genre), musical features (e.g., rhythm, tempo, and mode), and mood. For example, Vuoskoski et al. [100] identified personality factors involved in the emotional processing of music and found mood- and trait-congruent biases in the perception of musical emotions, while Zentner, Grandjean, and Scherer [108] found significant variations in the emotional responses to various musical genres. Strong correlations have been found along valence and arousal, the two primary dimensions of emotion, and the individual musical features. Overall, rhythm and tempo are the two most prominent musical parameters associated with emotional processing, with rhythm having significant correlations along both valence (together with major and minor mode) and arousal dimensions, and tempo typically correlated with the arousal dimension [58]. Moreover, existing research has shown there are positive associations between the overall sound intensity and arousal [27], cases of cross-modal transfer of arousal to vision [52], and the effects of individual emotions on color-music associations [60]. In general, research on audio-visual integration shows flexible integration of both modalities [86], with the multisensory integration reducing ambiguity and providing the "faster and more accurate categorization" [11], and thus the richer and more coherent percept.

The underlying mechanisms governing emotion induction are not unique to music [39], and most of the presented findings can be extended to other domains of affective computing. Besides cognitive aspects, the benefits of multimodal integration are particularly relevant for the user-aware MIR systems, both in terms of the overall user experience, with audio-visual modalities being the dominant aspects of interface design, and the development of intelligent recommendation algorithms.

The first step toward user-aware MIR research is the creation of more comprehensive MIR datasets and integration of user-related information. In what follows, we present the methodology for gathering emotional and color responses to music in the *Moodo* dataset.

16.3 Online Survey: Gathering Emotional and Color Responses to Music

The specification for online survey was based on the preliminary study, with the aim to select relevant emotion labels, create interfaces for gathering participants' responses, evaluate different aspects of user experience and set guidelines for the overall design of the survey.

16.3.1 Preliminary Survey: Selection of Emotion Labels and Colors

To establish a relevant set of emotion labels for the main survey, we had performed a preliminary survey of emotion labels gathered from the music research literature [18, 22, 32, 37, 38, 108]. The preliminary survey was conducted in Slovenian language and asked 64 participants to describe their current affective state on a 7-degree Likert intensity scale, based on a set of 48 emotion labels selected from selected studies [70, 71, 103]. Principal component analysis of the gathered data revealed 64 % variance in the first three components, covering most of the 17 emotion labels chosen for the main survey. The final set of emotions used in the survey: *Anger, Anticipation, Calmness, Disappointment, Dreamy, Energetic, Fear, Gloominess, Happiness, Inspiring, Joy, Liveliness, Longing, Relaxed, Sadness, Surprise, Tension*. Additionally, mood labels were derived from the above set for gathering participants' self-reports on mood in the second part of the survey: *Active, Angry, Calm, Cheerful, Disappointed, Discontent, Drowsy, Happy, Inactive, Joyous, Miserable, Relaxed, Sad, Satisfied, Sleepy, Tired, Wide Awake*.

Next, we evaluated the effectiveness of the continuous color wheel for gathering color annotations for individual emotions. Most participants found the continuous color scale too complex. Consequently, a modified discrete-scale version with 49 colors displayed on large tiles was created. The set of 49 colors has been rated by most participants as providing a good balance between the complexity of the full continuous color wheel and limitations of choosing a smaller subset of colors.

Another important finding of the preliminary survey was participants' feedback on the interfaces for gathering emotion and color responses. Traditional Likert scale variations present a high task load, as each of the 48 emotions used in the main survey had to be rated on a separate scale. Thus, two novel graphical user interfaces for gathering mood and emotion ratings had been developed, the *MoodStripe* and *MoodGraph*. These are presented in the following sections together with other aspects of the main survey design.

16.3.2 Main Survey

The main survey was conducted online in three stages. Part one contains basic demographic questions, including questions regarding participant's musical experience. Part two focuses on participant's current mood, emotions and associated colors, and part three focuses on emotional and color responses to music.

16.3.2.1 Part One: Demographic Information

The first part of the survey captures basic demographic information (age, gender, area of living, native language) and participant's music-related information, including music education and skills (e.g., ability to play an instrument or sing), the amount of time listening to music, and genre preferences. Additional, more detailed questions were omitted, to reduce the overall duration of the survey (to estimated 15 min) and allow participants to focus on emotional, visual, and musical aspects in the second and third part of the survey.

16.3.2.2 Part Two: Gathering Participants' Self-reports on Mood, Emotions, and Colors

The second part of the survey focuses on participant's self-report on currently felt mood, emotions and associated colors.

Participant's affective state was captured in several ways. To estimate their current mood, participants were first asked to place a point in the valence-arousal space (Fig. 16.1, left). This is a standard mood estimation procedure in dimensional modeling. Self-reports on mood were only gathered at this stage of the survey, under the assumption that participant's mood will not change considerably throughout the remaining parts of the survey (average duration of the survey is 15 min).

Participants were then asked to choose colors best associated with currently felt emotions by selecting a color in the discrete color wheel (Fig. 16.1, right).[3] Next, the

[3]for colored figures (Fig. 16.1 (left) and Figs. 16.5, 16.6, 16.7, 16.8, 16.9, 16.10, 16.11, 16.12 and 16.13) refer to the electronic version of this paper.

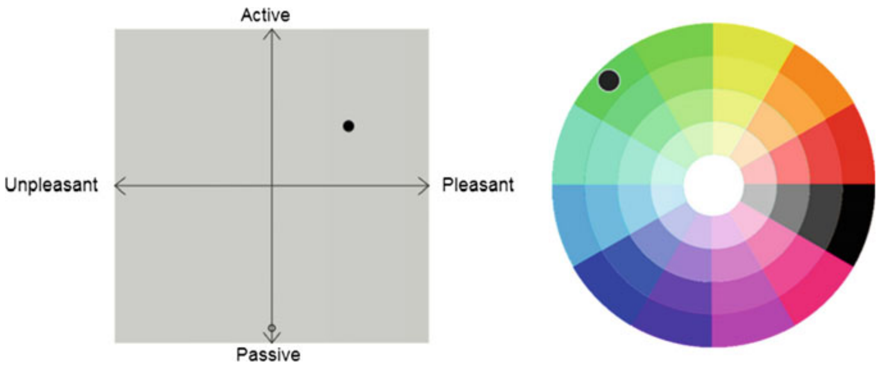

Fig. 16.1 *Left*: the *valence-arousal space*. The graph axes are marked *Unpleasant* and *Pleasant* for the abscissa, and *Passive* and *Active* for the ordinate values (the *black dot* indicates participant's selection in valence-arousal space). *Right*: the *discrete color wheel* with a set of 49 colors (the *black dot* indicates the selected color)

Fig. 16.2 The *MoodStripe*: participant drags emotions on the canvas according to their level of activation: from unexpressed to highly expressed (areas of the scale are marked from absent (*left*), moderately present (*middle*), to highly expressed (*right*)). Here, only a selection of emotion labels is presented. This interface is a substitute for a set of n-degree scales typically used in gathering user ratings

level of activity for 17 emotions was evaluated by positioning individual emotion labels in the *MoodStripe* interface (see Fig. 16.2). To make the task as easy and intuitive as possible, participants were able to drag and drop individual emotion labels onto the continuous activation space. This significantly reduced the overall task load, compared to the more traditional approach of using n-degree scales, where annotations for individual emotions need to be conducted on separate scales (see Sect. 16.3.3 for the evaluation of proposed interfaces).

Finally, participants assessed pleasantness and activity of individual emotions by positioning emotion labels onto the valence-arousal space of the *MoodGraph* interface (Fig. 16.3). The decision to use the *MoodGraph* interface instead of classical Russell's Circumplex model of affect [71] was to avoid the assumptions made in the latter, where the placement of individual emotions is designated to the specific areas (sections in the four quadrants) of the valence-arousal space. In music, this is not always the case (e.g., *sadness* is sometimes perceived as pleasant; see analysis

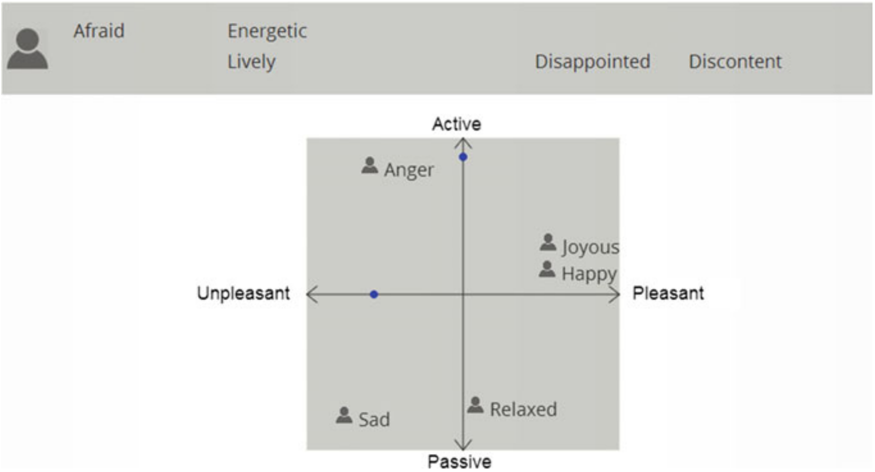

Fig. 16.3 The *MoodGraph*: emotions are dragged from the category container onto the valence-arousal space. *Blue dots* indicate the positions of selected emotion on both axes

in Sect. 16.4.4) and without imposing discrete areas of valence-arousal space the *MoodGraph* can account for the variability in perception and annotation of musical emotions.

Both novel interfaces, the *MoodStripe* and the *MoodGraph*, pose an alternative approach to gathering participants' ratings, replacing a set of ordinal n-degree scales. A subsequent evaluation of user experience and proposed interfaces showed that concurrent rating of emotions in the *MoodStripe* and the *MoodGraph* interfaces is intuitive and effective at reducing task load put on the participant (see Sect. 16.3.3).

16.3.2.3 Part Three: Emotional and Color Perception of Music

In part three of the survey, participants are asked to complete two tasks related to emotional processing of music. First, participants are presented with a set of 10 randomly selected 15 s long music excerpts. After listening to the excerpt, participant is first asked to select best matching color for the excerpt (Fig. 16.1, left). Next, participant is asked to place a set of emotion labels in the *MoodGraph* valence-arousal space, differentiating between two different categories of musical emotions: emotions evoked in the listener (induced emotions) and emotions expressed by music (perceived emotions). The category of induced emotion labels is marked with a person icon, whereas perceived emotion labels are represented with a note icon. Participants are instructed the first category (person icon) represents their personal emotions, what they feel when listening to the individual music fragment, and the second category (note icon) represents emotions that the music expresses (that one recognizes in music). Participants may place any number of emotions (but at least one from each category) in the *MoodGraph* (as shown in Fig. 16.4).

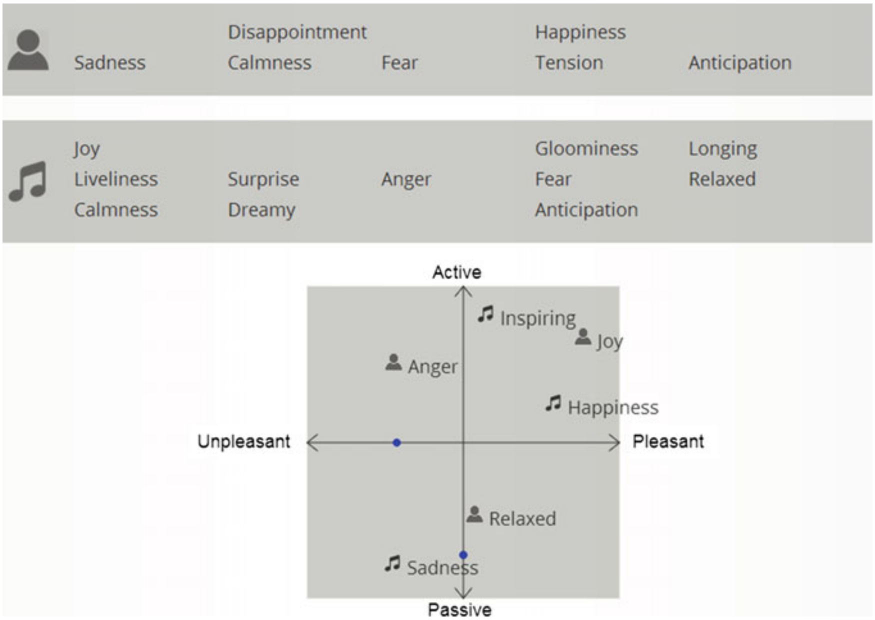

Fig. 16.4 The two-category *MoodGraph*: an extension of the one-category interface provides the participant with two categories, each denoted with an icon: a person icon for induced emotions and a note icon for perceived emotions

16.3.3 Evaluation of the MoodStripe and the MoodGraph Interfaces

We have conducted a subsequent evaluation of the *MoodStripe* and the *MoodGraph* interfaces, based on the feedback from participants of the survey. Participants were asked to evaluate several aspects of the survey: user experience (UX) [1], complexity of the questionnaire, and both interfaces. Our goal was to determine the functionality and user-friendliness of both interfaces compared to the standard approaches of gathering participant ratings. The evaluation of both interfaces contained a subset of the NASA load task index [30] evaluation survey and a set of specific questions. Results are presented in Sect. 16.3.3.1.

The online evaluation questionnaire was completed by 125 participants that previously participated in the main survey (detailed presentation of the results can be found in [67]). Results were generally positive and indicate overall balance of the survey and user-friendliness of both interfaces. They are summarized in Fig. 16.5. The majority of participants spent 11–15 min solving the survey (11). Although responses show balanced mental difficulty of the survey (1), the physical difficulty seems to be more uniformly distributed across participants (2). Thus, it can be speculated that the listening part of the survey presented a challenge for many participants.

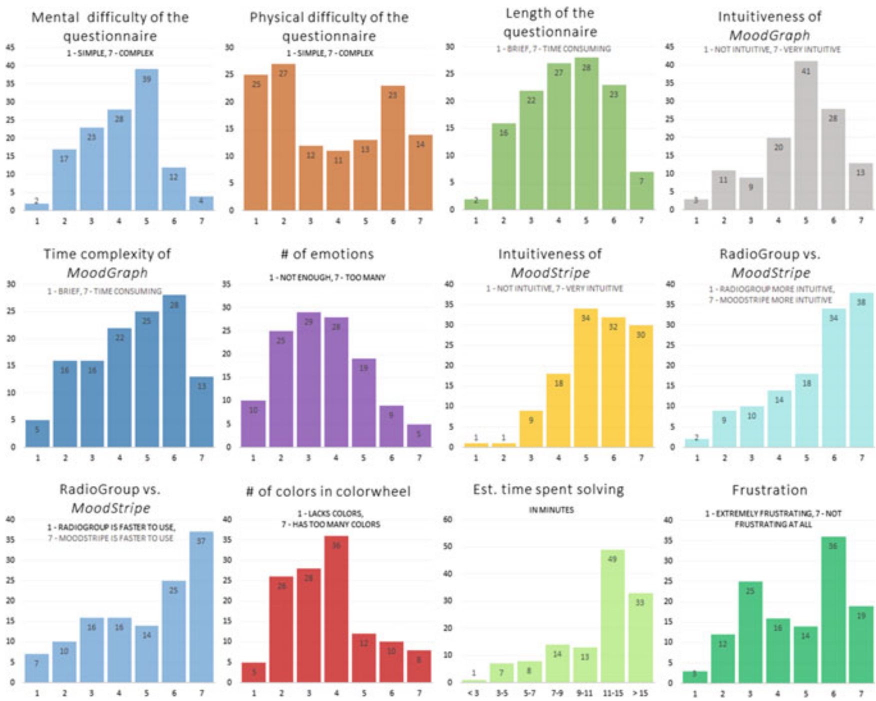

Fig. 16.5 The cumulative histograms for twelve evaluation questions

16.3.3.1 Results

The *MoodGraph* interface was evaluated as quite intuitive (4), however, it was also time consuming (5). Considering the complexity of required tasks (differentiating and rating emotions from two conceptually separate emotion categories, induced vs. perceived) and task load (e.g., the number of music excerpts, colors and emotions) put on participants, such feedback is realistic. Participants suggested the number of emotions in *MoodGraph* categories is slightly unbalanced (6), and we plan to review and extend both categories in the future. The *MoodStripe* interface represents a significant improvement over a variation of the Likert scale (a group of radio buttons), both in intuitiveness (7, 8) and time complexity (9). Participants also indicated that the set of 49 colors may not be large enough, so we will consider expanding the existing set.

Results of the evaluation demonstrate the usefulness of the proposed interfaces. The key advantages, compared to the standard input types, are reduced time complexity and task load, and an increased intuitiveness of the two novel interfaces, resulting in a lower mental difficulty and frustration of participants. At the same time, participants' comments give some useful future directions for improving the design and methodology of gathering participants' ratings on emotions, colors and music.

16.4 Analysis: Music, Colors, and Emotions

In what follows, we first present the demographic information on participants of the survey. Then we continue the discussion on emotional processing and multimodal interactions in music. We present correlations between emotions, colors and music in non-music and music contexts, the effects of genre preferences, differences between perceived and induced musical emotions, as well as differences in valence-arousal ratings for associations between individual musical emotions.[4]

16.4.1 Demographic Information

The online survey was completed by 741 participants, with total of more than 1100 participants participating, but not completing all three parts of the survey—these participants had been removed from the analysis. From 741 participants, 247 are men (33 %) and 494 are women (67 %). The youngest participant is 15 years old, the oldest is 64 years old. More than 75 % of participants fall into age group 30 years old or younger (Mean = 28.45 years). This is most likely due to the use of online survey, social media and public student associations channels for promotion and dissemination of the survey.

Almost 60 % of male and 44 % of female participants have no music education. From the participants with music education, 12 % of women and 6 % of men finished primary music education, which is a standardized 6 year program in Slovenia.

The most popular music genre is Rock, chosen by 31 % of participants. It is followed by Pop, chosen by 17 % of participants, Alternative and Classical, the latter two chosen by 5 % of participants. Other genres received significantly less than 5 % of the votes. As a second favorite genre, 20 % of participants chose Rock, whereas Pop received 14 % of votes. Classical music was the favorite genre in the third-favorite group (13 %), followed by Rock (12 %) and Pop (10 %).

16.4.2 Emotional Mediation of Color and Music

Interactions between auditory (music) and visual (color) modalities significantly depend on user's personality traits, temporary affective state and music context [39, 56, 59, 87]. Experiments conducted by [28, 60] show that cross-modal associations between music and colors use emotional mediation as the underlying mechanism. Their findings "associate specific dimensions of color (saturation, lightness, and yellowness-blueness) with specific high-level musical dimensions (tempo and

[4] Visualization tools for the general overview of the Moodo dataset are available here: http://www.moodo.musiclab.si/#/razplozenjeinglasba.

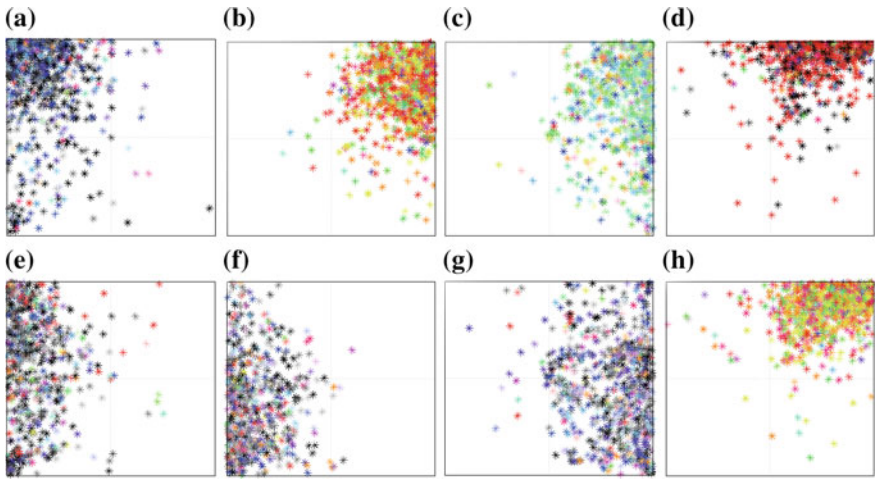

Fig. 16.6 Emotions and their color associations in the valence-arousal space (without music): **a** Anger. **b** Joy. **c** Happiness. **d** Energetic. **e** Fear. **f** Sadness. **g** Relaxation and **h** Liveliness [68]

mode), and show clear evidence of mediation by emotional dimensions (happy-sad and angry-calm)" [60]. Further evidence for direct cross-modal emotional mediation of both modalities has been shown by Pesek et al. [68], based on the analysis of participants' valence-arousal ratings of emotions and their color associations for individual music excerpts. Results show a stark contrast between valence-arousal ratings of individual emotions and their color associations in non-music context (emotion-color associations) and those in music context (emotion-color-music associations), as shown in Figs. 16.6 and 16.7 respectively.[5]

Color associations for individual emotions presented in Fig. 16.6 are in line with previous research [59], with the dark blue-violet and black hues associated with the negative emotions, such as fear and anger (A and E), the light green hues for happiness (C) and the more vibrant red-yellow-green hues for joy and liveliness (B and H), and the distinctly red for energetic (D). One noticeable exception is relaxation (G), occupying the positive position on the valence dimension, but with the hues similar to those of the negative emotions.

In music context (Fig. 16.7), the red hues prevail over the dark blue-violet and the gray-black hues for negative emotions (A and E), the green-yellow hues for positive emotions of joy and happiness (B and C) and on the positive arousal dimension for relaxation (G), while the green hues dominate in relaxation and calmness (G and H). Sadness (F) differentiates itself from the rest of emotions with prevalent blueness. There is an interesting correlation between color associations and the valence-arousal

[5]Note that emotions D: Energetic and H: Liveliness in Fig. 16.6 do not correspond to emotions D: Anticipation and H: Calmness in Fig. 16.7, as the latter are more appropriate in music context (for a discussion on musical and non-musical emotions, see [37]).

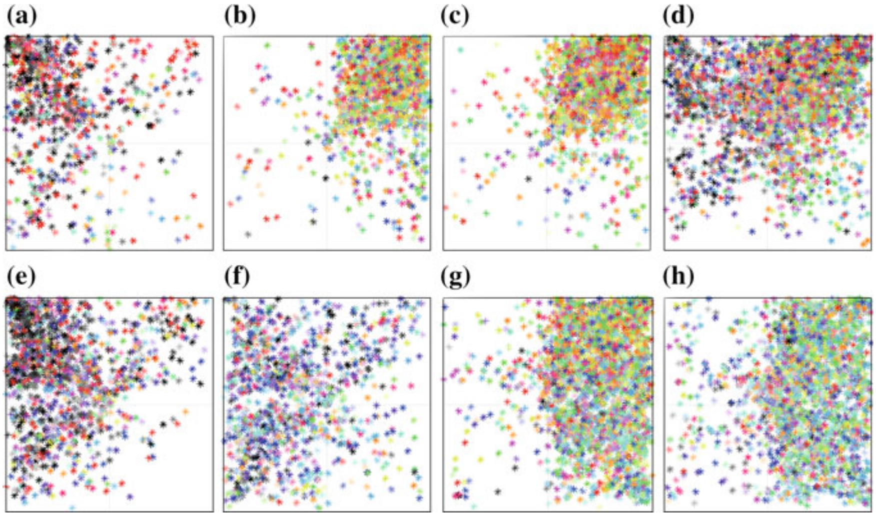

Fig. 16.7 Emotions and their color associations in the valence-arousal space (with music): **a** Anger. **b** Joy. **c** Happiness. **d** Anticipation. **e** Fear. **f** Sadness. **g** Relaxation and **h** Calmness

space for anticipation (D), where hues associated with the negative emotions dominate the negative pole of valence space, whereas green-blue hues, associated with more passive but pleasant emotions (such as G: Relaxation and H: Calmness), dominate in the quadrant of positive valence and negative arousal.

In general, emotions in music context occupy a more central position in the valence-arousal space, whereas the distribution of individual ratings is significantly larger than that of the emotion ratings in non-music context. This shows that musical emotions are being perceived differently from their non-music counterparts, and can at times occupy semantically opposite positions [108]; especially along the valence dimension, where for example sad music can sometimes be perceived as pleasant (compare E: Fear and F: Sadness in Figs. 16.6 and 16.7 for the distribution of valence-arousal ratings).

16.4.3 Genre Specificity of Musically Perceived Emotions

Emotional mediation of color and music associations is to some extent further constrained by the music genre. Beyond specific sets of musical characteristics and styles that differentiate one genre from another, genres also convey particular sets of musical emotions, or more precisely, emotions are perceived differently among individual genres because of the underlying musical characteristics and style represented by a particular genre [16].

Table 16.1 Typical emotions selected by participants for individual genres

Genre/emotion	Anger	Anticipation	Calmness	Fear	Happiness	Joy	Relaxation	Sadness
Country					✓	✓	✓	
Dance/Disco					✓	✓		
Easy listening			✓				✓	
Electronic		✓						
Hip hop/rap	✓			✓				
Metal	✓			✓				
New age		✓	✓	✓	✓		✓	✓
Pop		✓	✓			✓		✓
Rock	✓							✓

Comparing participants' ratings of music excerpts for different genres we can argue that genre specificity of musically perceived emotions is stronger for the musically more obscure genres (e.g., Metal), i.e. those genres that share least of common features with more mainstream genres, such as Pop or Dance/Disco, for example. Typical emotions (most used by participants) for selected genres are presented in Table 16.1, while Fig. 16.8 shows the valence-arousal ratings of top three genres (represented by music excerpts) and the prevalent color association for each of musical emotions (introduced in Fig. 16.7).

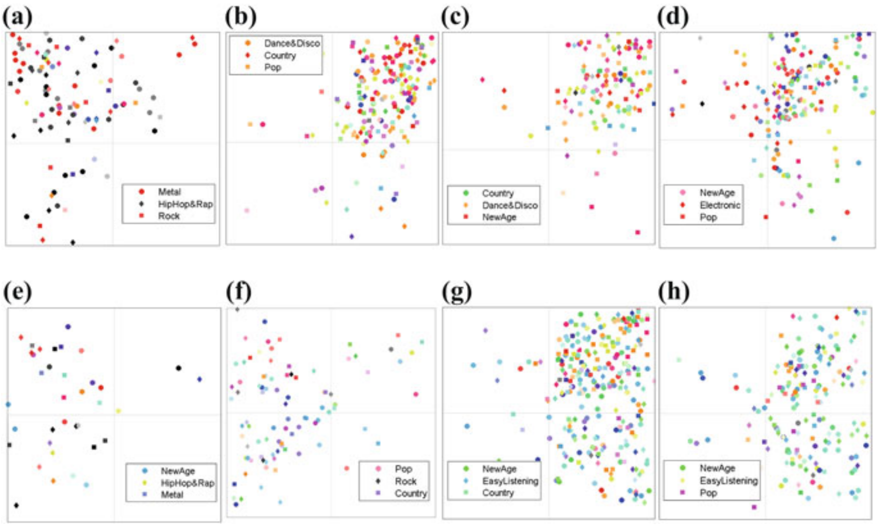

Fig. 16.8 Top three genres for individual emotions: **a** Anger. **b** Joy. **c** Happiness. **d** Anticipation. **e** Fear. **f** Sadness. **g** Relaxation. and **h** Calmness

As Table 16.1 and Fig. 16.8 show, different genres convey different sets of emotions, and even the same emotion might be perceived differently across musically and stylistically heterogeneous genres.

16.4.4 Perceived and Induced Emotions in Music

Emotional processing of music generally involves two types of emotions [24], those that are conveyed by music (perceived) and those evoked in the listener (induced). Preliminary analysis of the dataset showed that participants do differentiate between perceived and induced emotions. We found significant variance in participants' ratings of individual emotions from both categories, especially on the valence dimension, where variance was largest on the music perceived as unpleasant [68]. This finding shows that in certain music contexts, especially those with the perceived negative connotation, music can produce a variety of, sometimes polar, perceived-induced emotion responses in the listener (see Fig. 16.9). This is in line with the previous research on perceived and induced emotions in music [80], and while the differentiation between the two categories is not always clear (for the analysis of possible interactions, see [26]), the results show both aspects should be accounted for when integrating emotions into user-aware MIR systems.

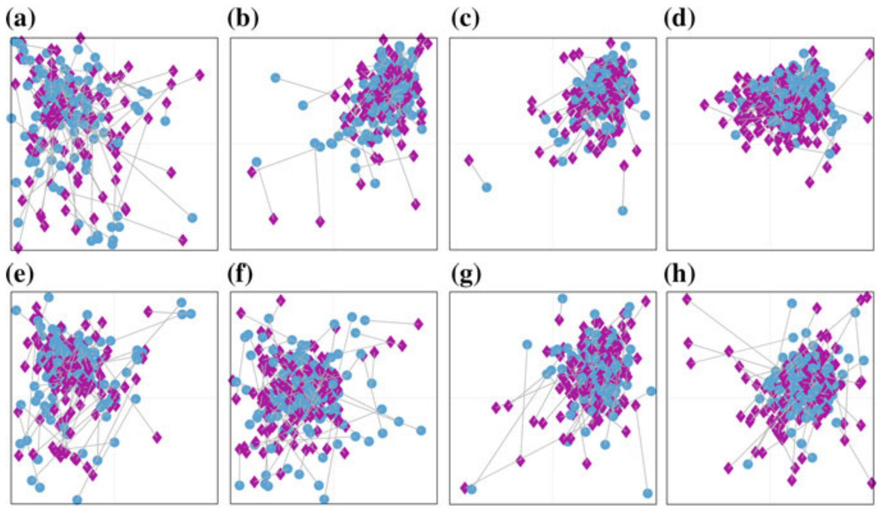

Fig. 16.9 Variance between induced (diamond) and perceived (circle) emotions in valence-arousal space [68]. Here, the average participants' ratings (centroids) for individual music excerpts are shown as induced-perceived emotion pairs for the following emotions: **a** Anger. **b** Happiness. **c** Joy. **d** Anticipation. **e** Fear. **f** Sadness. **g** Relaxation and **h** Calmness

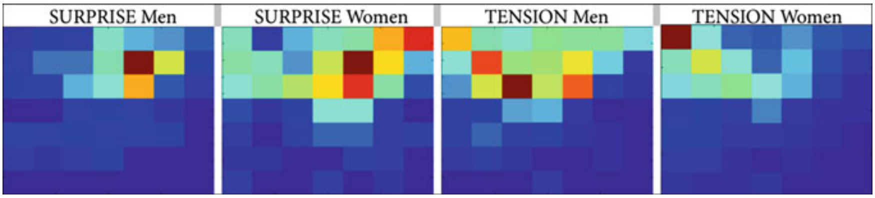

Fig. 16.10 Gender differences in the perception of musical emotions *Surprise* and *Tension*

16.4.5 Additional User Context: Mood, Gender and Musical Education

Beyond general correlations found in participants' emotional ratings of color and music there are additional user contexts that can give us further insight into our perception of music. Here, the effects of mood, gender and musical education, as well as the effects of different emotion combinations, are shortly presented through the analysis of quantized valence-arousal spaces, similar as used by [102]. These findings further emphasize the need for a more comprehensive MIR research and analysis of general user contexts.

Figure 16.10 shows differences in the perception of musical emotions between male and female participants for *surprise* and *tension*. These emotions represent the dynamic properties of music [21] and are perceived as such on the arousal dimension, with both groups rating them as active. However, there are significant differences in the perception of *surprise* and *tension* on the valence dimension, with females' ratings expressing significant variability on *surprise*, whereas males' ratings are more distributed on *tension*. This shows that at least in certain music contexts, the perception of *surprise* and *tension* is gender dependent. However, no significant differences were found in the male and female ratings for other musical emotions in the dataset.

Influence of mood is evident in the color perception of negative musical emotions, as shown in Fig. 16.11. Here, the color perception of *sadness* is perceived differently depending on the mood of the two groups of participants: participants of one group are in a satisfied mood, whereas participants of the other are discontent. The overall valence-arousal ratings are similar, but the difference in color ratings is obvious, with black only present in color ratings of participants feeling discontent. The effects of mood on the color perception of music have been found for the negative (unpleasant) emotions such as *sadness*, but not for the positive emotions, such as *happiness*.

Figure 16.12 shows the influence of music education on the perception of musical emotions *dreamy* and *surprise*. Participants with music education (years 1–20) exhibit a significantly higher variance on both valence and arousal dimension, compared to the participants with no music education. However, the influence of music education on the perception of music should be further investigated by mapping participants' valence-arousal ratings of emotions and colors to the underlying musical parameters.

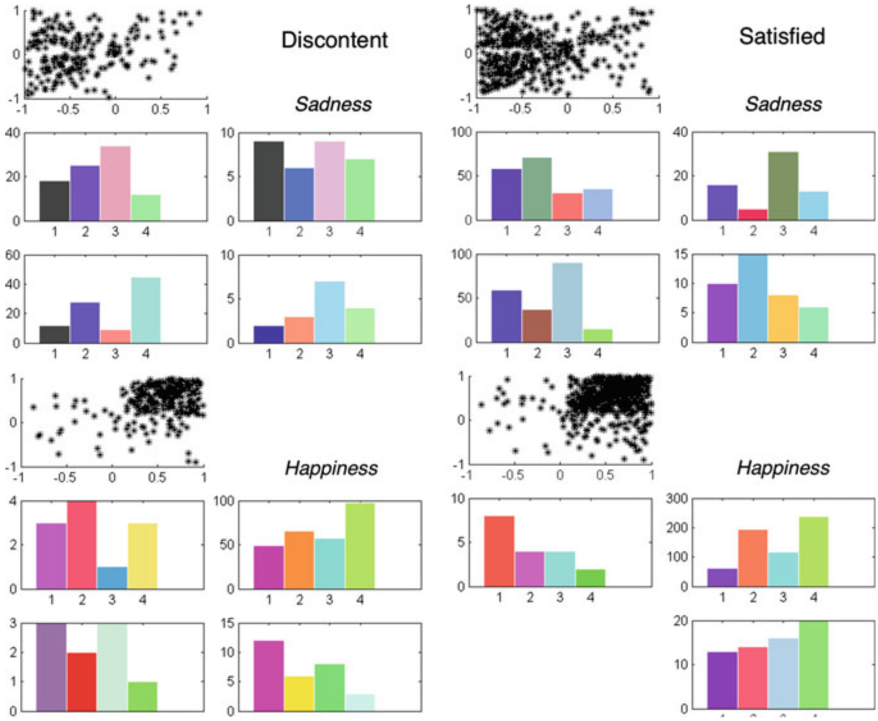

Fig. 16.11 Effects of participant's mood (Discontent vs. Satisfied) on the color perception of *sadness* and *happiness*

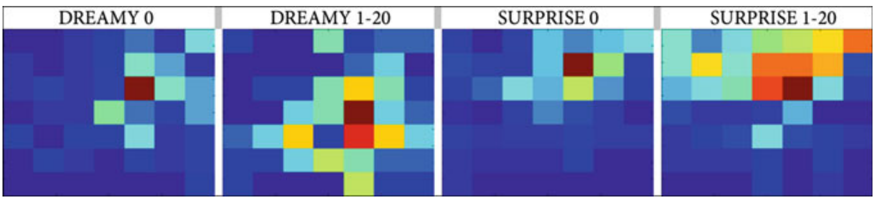

Fig. 16.12 Influence of music education on the perception of *dreamy* and *surprise* in music. Figure shows two groups of participants: participants with no music education ('DREAMY 0' and 'SURPRISE 0') and participants with music education ('DREAMY 1–20' and 'SURPRISE 1–20')

More widespread are the effects of emotion combinations presented in Fig. 16.13. The effects of negative and positive emotion combinations are shown through the variations in valence-arousal ratings for *anticipation*, *liveliness* and *tension*. Common to all three is the positive position on the arousal dimension, with *tension* leaning towards the negative and *lightness* towards the positive valence. Figure shows how the associated negative emotions (*Anger*, *Fear*) affect valence-arousal ratings of all three emotions towards the negative valence, whereas the associated positive

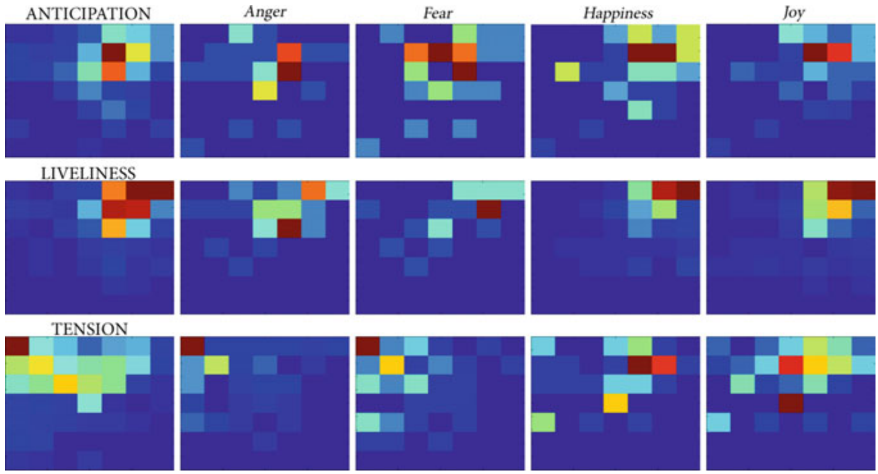

Fig. 16.13 Effects of emotion combinations on the perception of musical emotions for *anticipation*, *liveliness* and *tension*. Figure shows differences in valence-arousal ratings influenced by the associated negative (*anger*, *fear*) and positive (*happiness*, *joy*) emotions

emotions (*Happiness*, *Joy*) lean towards the positive valence. All three, *anticipation*, *liveliness* and *tension*, are often considered as alternative third dimension to valence and arousal, for modeling the dynamic aspects of music (such as tempo and rhythm) [17, 18].

16.5 Conclusions

The multidisciplinary nature of MIR research and the need for integration of the three fundamental aspects—music, the user and interaction—pose a number of exciting challenges for the future. The aim of this chapter was to address some of the current issues and argue towards user-aware MIR. And while the research on emotional and color perception of music is only a small piece in the overall 'user-aware' puzzle, we believe it is nevertheless important.

Presented analyses show that correlations between emotions, color and music are largely determined by context. There are differences between the emotion–color associations and valence-arousal ratings in non-music and music context, with the effects of genre preference evident for the latter. Participants were able to differentiate between perceived and induced musical emotions, and furthermore, between genre specific emotions. Results also show gender plays no major role and that female and male ratings of musical emotions and associated colors correlate, apart from differences in the perception of musical *tension* and *surprise*. The influence of mood is evident in the color perception of unpleasant emotions, such as *sadness*, but not for the positive emotions, such as *happiness*. More evident are the effects of emotion combinations describing music. Here, the associations between individual musical emotions affect their valence-arousal ratings, depending on the negative or the positive potency of individual emotion. The influence of music education has been shown for *dreamy* and *surprise*, and should be further investigated, especially in regard to the emotions that relate to the underlying musical parameters, such as *surprise*, *tension*, *anticipation* and *liveliness*. We assume that users with music education or skills might perceive music differently due to a better understanding of musical concepts [57], but other aspects, such as mood and personality, should be considered as well [99].

The underlying mechanisms governing emotion induction are not unique to music and most of the findings presented here can be extended to other domains of affective computing. The benefits of multimodal integration are particularly relevant to the user-aware systems research, both in terms of improving the overall user experience, with audio-visual modalities acting as the dominant features of interface design, as well as in improving the existing recommendation algorithms.

References

1. Albert, W., Tullis, T.: Measuring the user experience: collecting, analyzing, and presenting usability metrics (Google eBook). Newnes (2013)
2. Aljanaki, A., Bountouridis, D., Burgoyne, J.A., van Balen, J., Wiering, F., Honing, H., Veltkamp, R.C.: Designing games with a purpose for data collection in music research. Emotify and hooked: two case studies. Lecture Notes Computer Science (2014)
3. Aljanaki, A., Wiering, F., Veltkamp, R.C.: Computational modeling of induced emotion using GEMS. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 373–378. Taipei (2014)
4. Barthet, M., Fazekas, G., Sandler, M.: Multidisciplinary perspectives on music emotion recognition: implications for content and context-based models. In: CMMR, pp. 492–507. London (2012)
5. Barthet, M., Marston, D., Baume, C., Fazekas, G., Sandler, M.: 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)
6. Bergstrom, T., Karahalios, K., Hart, J.C.: Isochords: visualizing structure in music. In: Proceedings of Graphics Interface, pp. 297–304 (2007)
7. Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., Dacquet, A.: Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts. Cogn. Emot. **19**(8), 1113–1139 (2005)
8. Bulkin, D.A., Groh, J.M.: Seeing sounds: visual and auditory interactions in the brain. Curr. opin. neurobiol. **16**(4), 415–419 (2006)
9. Calvert, G.A.: Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb. Cortex **11**(12), 1110–1123 (2001)
10. Canazza, S., De Poli, G., Rodà, A., Vidolin, A., Zanon, P.: Kinematics-energy space for expressive interaction in music performance. In: Proceedings of MOSART. Workshop on Current Research Directions in Computer Music, pp. 35–40 (2001)
11. Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., Lepore, F.: Audio-visual integration of emotion expression. Brain Res. **1242**, 126–135 (2008)
12. De Gelder, B., Bertelson, P.: Multisensory integration, perception and ecological validity. Trends Cogn. Sci. **7**(10), 460–467 (2003)

13. Dibben, N.: Emotion and music: a view from the cultural psychology of music. In: Proceedings of the 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009 (2009)
14. Doehrmann, O., Naumer, M.J.: Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. Brain Res. **12**(42), 136–150 (2008)
15. Donaldson, J., Lamere, P.: Using visualizations for music discovery. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR). Tutorial (2009)
16. Eerola, T.: Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. J. New Music Res. **40**(4), 349–366 (2011)
17. Eerola, T.: Modeling listeners' emotional response to music. Top. Cogn. Sci. **4**, 607–624 (2012)
18. Eerola, T.: Modelling emotional effects of music: key areas of improvement. In: Proceedings of the Sound and Music Computing Conference 2013, SMC 2013. Stockholm, Sweden (2013)
19. Eerola, T., Vuoskoski, J.K.: A comparison of the discrete and dimensional models of emotion in music. Psychol. Music **39**(1), 18–49 (2010)
20. Eerola, T., Vuoskoski, J.K.: A review of music and emotion studies: approaches, emotion models, and stimuli. Music Percept. **30**(3), 307–340 (2013)
21. Eerola, T., Lartillot, O., Toiviainen, P.: Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 621–626 (2009)
22. Ekman, P.: An argument for basic emotions. Cogn. Emot. **6**, 169–200 (1992)
23. Ernst, M.O., Bülthoff, H.H.: Merging the senses into a robust percept. Trends Cogn. Sci. **8**(4), 162–169 (2004)
24. Evans, P., Schubert, E.: Relationships between expressed and felt emotions in music. Musicae Sci. **12**, 75–99 (2008)
25. Evans, K.K., Treisman, A.: Natural cross-modal mappings between visual and auditory features. J. Vis. **10**(1), 6 (2010)
26. Gabrielsson, A.: Emotion perceived and emotion felt: same or different? Musicae Sci. **5**(1 suppl):123–147 (2002)
27. Gingras, B., Marin, M.M., Fitch, W.T.: Beyond intensity: spectral features effectively predict music-induced subjective arousal. Q. J. Exp. Psychol. 1–19 (2013) [ahead-of-print]
28. Griscom, W.S., Palmer, S.E.: The color of musical sounds: color associates of harmony and timbre in non-synesthetes. J. Vis. **12**(9), 74–74 (2012)
29. Grohganz, H., Clausen, M., Jiang, N., Mueller, M.: Converting 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)
30. Hart, S.G.: Nasa-task load index (NASA-TLX); 20 years later. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **50**(9), 904–908 (2006)
31. Herrera-Boyer, P., Gouyon, F.: MIRrors: music information research reflects on its future: special issue foreword. J. Intell. Inf. Syst. **41**, 339–343 (2013)
32. Hu, X., Downie, J.S.: Exploring mood metadata: relationships with genre, artist and usage metadata. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR). Vienna (2007)
33. Isaacson, E.: What you see is what you get: on visualizing music. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 389–395. London (2005)
34. Jaimovich, J., Coghlan, N., Knapp, R.B.: Emotion in motion: a study of music and affective response. In: From Sounds to Music and Emotions, pp. 19–43. Springer (2013)
35. Jiang, N., Mueller, M.: Automated methods for analyzing music recordings in sonata form. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR). Curitiba (2013)
36. Julia, C.F., Jorda, S.: SongExplorer: a tabletop application for exploring large collections of songs. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 675–680. Kobe (2009)

37. Juslin, P.N., Sloboda, J.A.: Music and Emotion: Theory and Research. Oxford University Press (2001)
38. Juslin, P.N., Laukka, P.: Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. J. New Music Res. **33**(3), 217–238 (2004)
39. Juslin, P.N., Västfjäll, D.: Emotional responses to music: the need to consider underlying mechanisms. Behav. Brain Sci. **31**(5), 559–575 (2008)
40. Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J.A., Turnbull, D.: Music emotion recognition: a state of the art review. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 255–266. Utrecht (2010)
41. Koelsch, S.: Towards a neural basis of music-evoked emotions. Trends Cogn. Sci. **14**(3), 131–137 (2010)
42. Kohonen, T.: The self-organizing map. In: Proceedings of the IEEE **78**(9) (1990)
43. Kreutz, G., Ott, U., Teichmann, D., Osawa, P., Vaitl, D.: Using music to induce emotions: influences of musical preference and absorption. Psychol. Music **36**, 101–126 (2007)
44. Kurabayashi, S., Imai, T.: Chord-cube: music visualization and navigation system with an emotion-aware metric space for temporal chord progression. Int. J. Adv. Internet Technol. **7**(1), 52–62 (2014)
45. Lamere, P., Eck, D.: Using 3D visualizations to explore and discover music. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 173–174 (2007)
46. Laurier, C., Sordo, M., Serrà, J., Herrera, P.: Music mood representations from social tags. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 381–386 (2009)
47. Lee, J.H., Cunningham, S.J.: The impact (or non-impact) of user studies in music information retrieval. Ismir 391–396 (2012)
48. Lee, J.H., Cunningham, S.J.: Toward an understanding of the history and impact of user studies in music information retrieval. J. Intell. Inf. Syst. (2013)
49. Levitin, D.J., Tirovolas, A.K.: Current advances in the cognitive neuroscience of music. Ann. N.Y. Acad. Sci. **1156**, 211–231 (2009)
50. Lykartsis, A., Pysiewicz, A., Coler, H., Lepa, S.: 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 and Emotion (ICME3), pp. 1–15. Jyväskylä (2013)
51. Mardirossian, A., Chew, E.: Visualizing music: tonal progressions and distributions. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 189–194. Vienna (2007)
52. Marin, M.M., Gingras, B., Bhattacharya, J.: Crossmodal transfer of arousal, but not pleasantness, from the musical to the visual domain. Emotion **12**(3), 618 (2012)
53. Marks, L.E., Ben-Artzi, E., Lakatos, S.: Cross-modal interactions in auditory and visual discrimination. Int. J. Psychophysiol. **1**, 125–145 (2003)
54. McGurk, H., MacDonald, J.: Hearing lips and seeing voices. *Nature* **264**, 746–748 (1976)
55. Mcvicar, M., Freeman, T., De Bie, T.: Mining the correlation between lyrical and audio features and the emergence of mood. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 783–788. Miami (2011)
56. Meyer, L.B.: Emotion and Meaning in Music. University of Chicago Press. Chicago (1956)
57. Müllensiefen, D., Gingras, B., Musil, J., Stewart, L.: The musicality of non-musicians: an index for assessing musical sophistication in the general population. PLoS ONE **9**(2) (2014)
58. Omez, P., Danuser, B.: Relationships between musical structure and psychophysiological measures of emotion. Emotion **7**(2), 377 (2007)
59. Ou, L.-C., Luo, M.R., Woodcock, A., Wright, A.: A study of colour emotion and colour preference. Part I: colour emotions for single colours. Color Res. Appl. **29**(3) (2004)
60. Palmer, S.E., Schloss, K.B., Zoe, X., Prado-León, L.R.: Music-color associations are mediated by emotion. Proc. Natl. Acad. Sci. **110**(22), 8836–8841 (2013)

61. Pampalk, E., Dixon, S., Widmer, G.: Exploring music collections by browsing different views (2004)
62. Pampalk, E.: Islands of music analysis, organization, and visualization of music archives. OGAI J. (Oesterreichische Ges. Artif. Intell.) **22**(4), 20–23 (2003)
63. Panda, R., Malheiro, R., Rocha, B., Oliveira, A., Paiva, R.P.: Multi-modal music emotion recognition: a new dataset. In: Proceedings of the Methodology and Comparative Analysis CMMR (2013)
64. Parise, C.V., Spence, C.: 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. PLoS One **4**(5), e5664 (2009)
65. Pearce, M., Rohrmeier, M.: Music cognition and the cognitive sciences. Top. Cogn. Sci. **4**(4), 468–484 (2012)
66. Peretz, I., Coltheart, M.: Modularity of music processing. Nat. Neurosci. **6**(7), 688–691 (2003)
67. Pesek, M., Godec, P., Poredoš, M., Strle, G., Guna, J., Stojmenova, E., Pogačnik, M., Marolt, M.: 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 and Expo (ICME), pp. 1–4 (2014)
68. Pesek, M., Godec, P., Poredos, M., Strle, G., Guna, J., Stojmenova, E., Pogacnik, M., Marolt, M.: Introducing a dataset of emotional and color responses to music. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 355–360. Taipei (2014)
69. Pressing, J.: Cognitive complexity and the structure of musical patterns. Noetica **3**, 1–8 (1998)
70. Remmington, N.A., Fabrigar, L.R., Visser, P.S.: Reexamining the circumplex model of affect. J. Pers. Soc. Psychol. **79**(2), 286–300 (2000)
71. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6):1161–1178 (1980)
72. Saari, P., Eerola, T.: Semantic computing of moods based on tags in social media of music. IEEE Trans. Knowl. Data Eng. **26**(10), 2548–2560 (2014)
73. Schedl, M., Flexer, A.: Putting the user in the center of music information retrieval. In: Proceedings of the 13th International Society for Music Information Retrieval Conference, (Ismir), pp. 416–421 (2012)
74. Schedl, M., Knees, P.: Personalization in multimodal music retrieval. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7836, LNCS, pp. 58–71 (2013)
75. Schedl, M., Flexer, A., Urbano, J.: The neglected user in music information retrieval research. J. Intell. Inf. Syst. **41**(3), 523–539 (2013)
76. Scherer, K.R.: Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? J. New Music Res. **33**(3), 239–251 (2004)
77. Scherer, K.R., Zentner, M.R.: Emotional effects of music: production rules. In: Juslin, P.N., Sloboda, J.A. (eds.) Music and emotion. Oxford University Press, New York (2001)
78. Schimmack, U., Reisenzein, R.: Experiencing activation: energetic arousal and tense arousal are not mixtures of valence and activation. Emotion (Washington, D.C.) **2**(4), 412–7 (2002)
79. Schmidt, E.M., Kim, Y.E.: Modeling musical emotion dynamics with conditional random fields. In: ISMIR, pp. 777–782 (2011)
80. Schubert, E.: Emotion felt by listener and expressed by music: a literature review and theoretical investigation. Frontiers Psychol. **4**(837) (2013)
81. Schuller, B., Hage, C., Schuller, D., Rigoll, G.: 'Mister DJ, cheer me up!': musical and textual features for automatic mood classification. J. of New Music Res. **39**(1), 13–34 (2010)
82. Serra, X., Magas, M., Benetos, E., Chudy, M., Dixon, S. Flexer, A., Gómez, E., Gouyon, F., Herrera, P., Jordà, S., Paytuvi, O., Peeters, G., Vinet, H., Widmer, G.: Roadmap for music information research. Jan Schlüter (2013)
83. Soleymani, M., Caro, M.N., Schmidt, E.M., Sha, C.-Y., Yang, Y.-H.: 1000 songs for emotional analysis of music. In: Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia—CrowdMM '13, pp. 1–6. ACM Press, New York, USA (2013)

84. Song, Y., Dixon, S., Pearce, M.: A survey of music recommendation systems and future perspectives. In: Proceedings of the 9th International Symposium Computer Music Modelling and Retrieval (CMMR), pp. 395–410. London (2012)
85. Speck, J.A., Schmidt, E.M., Morton, B.G., Kim, Y.E.: A comparative study of collaborative vs. traditional musical mood annotation. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 549–554. Miami (2011)
86. Spence, C.: Audiovisual multisensory integration. Acoust. Sci. Technol. **28**(2), 61–70 (2007)
87. Spence, C.: Crossmodal correspondences: a tutorial review. Atten. Percept. Psychophys. **4**(1), 971–995 (2011)
88. Spence, C., Senkowski, D., Röder, B.: Crossmodal processing. Exp. Brain Res. **198**(2), 107–111 (2009)
89. Stalinski, S.M., Schellenberg, E.G.: Music cognition: a developmental perspective. Top. Cogn. Sci. **4**(4), 485–497 (2012)
90. Stevens, C.J.: Music perception and cognition: a review of recent cross-cultural research. Top. Cogn. Sci. **4**, 653–667 (2012)
91. Tingle, D., Kim, Y.E., Turnbull, D.: Exploring automatic music annotation with "acoustically-objective" tags. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 55–62. New York (2010)
92. Torrens, M., Hertzog, P., Arcos, J.L.: Visualizing and exploring personal music libraries. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR). Barcelona (2004)
93. Torres-Eliard, K., Labbé, C., Grandjean, D.: Towards a dynamic approach to the study of emotions expressed by music. Lect. Notes Inst. Comput. Sci. Soc. Inform. Telecommun. Eng. **78**, 252–259 (2011)
94. Turnbull, D., Barrington, L., Torres, D., Lanckriet, G.: Semantic annotation and retrieval of music and sound effects. IEEE Trans. Audio Speech Lang. Process. **16**(2), 467–476 (2008)
95. Typke, R., Wiering, F., Veltkamp, R.C.: A survey of music information retrieval systems. In: Proceedings of the International Symposium on Music Information Retrieval, ISMIR, pp. 153–160 (2005)
96. Van Gulik, R., Vignoli, F.: Visual Playlist generation on the artist map. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR). London (2005)
97. Van Gulik, R., Vignoli, F., Van de Wetering, H.: Mapping music in the palm of your hand, explore and discover your collection. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR). Barcelona (2004)
98. Vroomen, J., de Gelder, B.: Sound enhances visual perception: cross-modal effects of auditory organization on vision. J. Exp. Psychol. Hum. Percept. Perform. **26**(5), 1583–1588 (2000)
99. Vuoskoski, J.K., Eerola, T.: Measuring music-induced emotion: a comparison of emotion models, personality biases, and intensity of experiences. Musicae Sci. **15**(2), 159–173 (2011)
100. Vuoskoski, J.K., Eerola, T.: The role of mood and personality in the perception of emotions represented by music. Cortex **47**(9), 1099–1106 (2011)
101. Wang, J.C., Yang, Y.H., Chang, K., Wang, H.M., Jeng, S.-K.: Exploring the relationship between categorical and dimensional emotion semantics of music. In: Proceedings of the Second International ACM Workshop on Music Information Retrieval with User-centered And Multimodal Strategies—MIRUM '12, p. 63. ACM Press, New York, USA (2012)
102. Wang, J.-C., Wang, H.-M., Lanckriet, G.: A histogram density modeling approach to music emotion recognition. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE (2015)
103. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. **54**(6), 1063–1070 (1988)
104. Weigl, D., Guastavino, C.: User studies in the music information retrieval literature. Ismir 335–340 (2011)
105. Witten, I.B., Knudsen, E.I.: Why seeing is believing: merging auditory and visual worlds. Neuron **48**(3), 489–496 (2005)

106. Yang, Y.H., Chen, H.H.: Machine recognition of music emotion (2012)
107. Yoshii, K., Goto, M.: Music thumbnailer: visualizing musical pieces in thumbnail images based on acoustic features. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 211–216. Philadelphia (2008)
108. Zentner, M., Grandjean, D., Scherer, K.R.: Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion **8**(4), 494 (2008)