Unveiling the Perceptual Differences in Mood and Color Perception in Non-Mainstream Music Styles

Pesek, Matevž and Marolt, Matija

Abstract: ***The article explores the differences in listeners' perception of non-mainstream music styles, uncovering the influence of the music context on their responses. To tackle this problem, the Moodo dataset—which links emotional and color perception to music, and includes user context—is used. Our focus is on two parts of the dataset that were not previously analyzed: a collection of Slovenian folk songs and a collection of electro-acoustic music pieces. The analysis of user responses to these non-mainstream music styles revealed strong induced emotions, not otherwise present in more mainstream genres. We show how this knowledge can be applied to the development of every-day music applications.***

Index Terms: ***music information retrieval, music emotion recognition, affective MIR, mood estimation, Moodo dataset***

1. Introduction

To analyze, retrieve and organize information in music, the field of music information retrieval (MIR) emerged in the last two decades [6]. It has grown since its early beginnings to encompass a number of topics and bring together researches from several well-established fields, from computer science and signal processing on one side, to psychology and music theory on the other. Several standardized MIR tasks have been established, which makes the comparison of different approaches possible. In the topics more closely related to signal processing, polyphonic transcription [15, 11, 23] and chord estimation [22, 26] have dominated the general interest of the MIR community. More recently, a variety of user-related MIR tasks emerged, taking the user's specific taste and opinion as the benchmark (e.g. [5, 29]), starting with music genre classification [37, 3]. With a growing number of available annotated datasets, mood estimation [14] has also emerged as an independent task [18, 16]. These user-centered topics gave rise to *affective music information retrieval*, which includes systematic analysis and modelling of affect in music [13, 8], user preference elicitation based on facial expression [34], study of mood and emotions [38], and others.

With the growing interest of both industry and academia [4], the emerged affective MIR subfield can be viewed as a consequence of the shift of interest from the *system as the goal* paradigm towards the *system as the tool*, where the goal is to accommodate each user individually. Affective MIR is thus focused on user-awareness—its main goal is to incorporate the user's demographic-oriented, personality-oriented and preference-oriented information into the system. The most significant barrier to achieve a truly *affective* MIR system is the lack of regularized and diverse annotated data. To overcome this barrier, Pesek et al. [24] proposed a new multi-modal dataset—named the Moodo dataset—that includes listeners' emotional ratings of music pieces, as well as their demographic data and music preferences.

By analyzing data from the Moodo dataset, the authors have previously shown how colors and music are mediated by emotions and how emotion-color-music associations are fur-

Manuscript received Sep, 2017 and accepted for the journal in Nov. 2017.

The authors are with the University of Ljubljana, Faculty of Computer and Information Science, Slovenia (e-mail: matevz.pesek@fri.uni-lj.si).

ther constrained by the user context of mood, gender, age, and music preferences [27]. The authors also showed that the inclusion of user context can significantly improve prediction of emotional responses in music.

In this paper, we continue our analyses, and study the influence of music styles on emotional perception of music. For this purpose, we analyzed emotion-color relations within two non-mainstream music collections included in the Moodo dataset: a collection of Slovenian folk songs, consisting of 20 songs from field recordings of Slovenian traditions; and a collection of electro-acoustic music, consisting of 20 contemporary works. Our analyses expose the differences in users' emotional and visual perception between both collections. The results are compared to previous analyses of this dataset, that were focused on mainstream and film music. We elaborate how our findings can aid in discovery of non-mainstream music within user-context-dependent music recommendation applications.

The paper is structured as follows. Section 2 presents the current state-of-the-art in MIR and an overview of the Moodo dataset. Section 3 reports on the gathered responses to the non-mainstream music subsets, and presents the results of our analyses, focusing on emotional and color responses to individual collections within the Moodo dataset. The results are discussed in Section 4. We conclude the paper with a discussion on the differences in emotional perception of music in different music styles and its influence on every-day applications involving music.

2. Related Work

The expansion of digitally available audio collections has boosted research in music emotion recognition, both for emotion recognition from audio, as well as from lyrics and music tags. The approaches are commonly benchmarked using datasets with reference annotations. In many cases, the annotations are produced explicitly for the dataset; in others, an online service is used as the source of annotations through user tags. The latter can be gathered with less difficulty and in larger amounts, however the amount of noise is significantly higher as there are no strictly defined rules for annotation by tagging. Tags also commonly include not only emotions involved in a song—either perceived or induced by both music and lyrics—but also music genre, instruments, the number and types of vocals etc. On the other hand, explicitly provided annotations contain less noise and more information, yet involve a lot of effort to gather, and are limited in size.

2.1 Retrieving The Mood

Researches in music emotion recognition can be divided into two groups, employing either discrete or dimensional emotion models for gathering user input. For the latter, a number of variations of the Likert intensity scales [17] and Russell's circumplex model of affect [28] have been introduced. One of the biggest drawbacks of dimensional models is a limited number of dimensions used for modeling emotions, with the low-dimensional affective space typically reduced to valence and arousal (e.g. [9]).

The discrete models aim to overcome this limitation by introducing multiple-category ratings of emotions (e.g. [14]). Several models have been well-accepted within the MIR community: the Geneva Emotional Music Scale (GEMS) [40] was developed as an emotion model for musically induced emotions. The GEMS model is based on 45 terms, although shorter versions have been introduced, resulting in the 25 and 9 term models, GEMS-25 [35] and GEMS-9 [2]. The Music Information Retrieval Evaluation eXchange (MIREX) community accepted a five-cluster model derived from the AllMusicGuide mood repository [10], with label sets consisting of 5 to 7 labels per cluster, resulting in a total of 29 labels.

Neither type of the model has been proven superior in all aspects. Therefore, authors of the Moodo dataset developed a new hybrid annotation model, which integrates features from both modeling approaches: the discrete (induced and perceived) emotion labels and the continuous 2-dimensional valence-arousal space. By selecting emotion labels for either perceived or induced categories, and placing them onto the valence arousal space, dimensional and discrete data are obtained simultaneously. Such an approach also yields insights into participants' perception of emotional labels and their position in the valence-arousal space, which cannot be captured by traditional approaches.

Several datasets with mood annotations have been gathered in the past. The MoodSwings Turk Dataset provides on average 17 valence-arousal ratings of perceived emotions for 240 clips of popular music [30], where a collaborative game [32] was partially used to obtain the data. The Cal500 gathered emotional labels for 500 popular songs [36], averaging three annotations per song. The extended dataset CAL10k also provided several acoustic features and tags along with 153 genre tags for more than 10.000 songs [33]. The MTV Music Dataset [31] contains a set of valence-arousal ratings for 192 popular songs annotated by 5 annotators with different musicological backgrounds. [1] gathered over 8000 responses by 1778 participants on a set of 400 excerpts of classical, rock, pop and electronic music. The Emotify game had been used in a data gathering procedure, which additionally included the basic demographics. Lykartsis et al. tested the German version of GEMS-25 model for electro-acoustic and popular music, expanding the original model with three additional categories (GEMS-28-G) [19]. They gathered annotations from 245 participants on a set of 20 music pieces of classical and popular instrumental music and electro-acoustic music. [12] used the GEMS-9 model on 4000 participants and 12000 music excerpts from 53 songs in order to measure participants' emotional responses. Several other datasets were presented, predominantly containing popular music, while some also included non-mainstream music (e.g.: [39, 7]).

The multi-modal Moodo dataset was proposed by Pesek et al. [27]. It overcomes several drawbacks of the existing datasets and additionally enhances the auditory modality with visual information and extensive information on the users.

2.2 Gathering the Moodo Dataset

The dataset was gathered through a survey, which was conducted online in three parts.

In the first part, the participants provided the following basic demographic data: age, gender, native language and area of living (city/rural area). Additionally, they were asked about their musical experience and taste, specifically the number of years of music education, experience with playing an instrument, amount of time listening to music per day and their preferred music genres.

In part two of the survey, the participants were asked to provide ratings of their current mood and emotions, and their perception of individual emotions and associated colors. The participants were first asked to report their current mood and emotions in the valence-arousal space [28], and to associate their current mood with a color. Next, the participants were asked to describe their current mood using a continuous emotion annotation interface called the MoodStripe [25]. In this way, they rated their mood by placing labels for 17 emotions onto a continuous axis spanning from complete absence to total presence of the positioned emotion. Finally, the participants were asked to select a color representing each of the 17 emotions, and to place the emotion labels onto a 2-dimensional valence-arousal space of the MoodGraph interface [23]. The MoodGraph is a hybrid emotion annotation model that combines the properties of discrete and dimensional models by integrating multiple-category annotations with the dimensionality of the valence-arousal space. The model allows draggable actions, enabling the user to position the labels within the continuous space using a mouse.

In the third part of the survey, the participants' emotional and color responses to music were gathered. A set of 200 music excerpts, each 15 seconds long, 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 [7], 20 excerpts from a collection of Slovenian folk songs, and 20 excerpts from a contemporary electro-acoustic music collection.

The participants were asked to provide the following information about 10 randomly chosen music excerpts from the dataset: a color best reflecting the excerpt; any number (minimally one) of emotion labels that best represent the perceived emotions of the music excerpt (by dragging them onto the valence-arousal space in the MoodGraph interface); and any number (minimally one) of emotion labels that best represent the induced emotions of the music excerpt (by dragging them onto the valence-arousal space in the MoodGraph interface).

The dataset and its visualization is publicly available[1]. It includes the 200 song excerpts, contextual data for 741 participants and 7000 responses to music excerpts (each consisting of one or more perceived and induced emotion labels, positioned in the valence-arousal space, accompanied by the selected color). Additionally, annotations of all songs by a musicological expert are provided, including rhythm, tempo, tunefulness, dynamics, beats per minute, mode, harmonic complexity, consonance and metrum.

3. Analysis of Responses to Non-Mainstream Music

A thorough analysis of the Moodo dataset was conducted by Pesek et al. [27], where an analysis of responses to popular and film music excerpts is provided. In this paper, we focus on the analysis of responses to the two non-mainstream music styles.

The set of Slovenian folk songs includes 20 music excerpts. The songs are part of the Slovenian oral folk song tradition, and were as such transferred and changed through generations, their renditions preserved on ethnomusicological field recordings. The songs are mostly melodic with a clear rhythm and contain one or more singers and accompanying instruments. The set of electro-acoustic works contains 20 music excerpts, generated through the use of computers, typically avoiding Western music rules and patterns. An analysis of musicological annotations of both subsets shows significant differences in their harmonic complexity and tunefulness in comparison to the mainstream music genres (Figure 1).

[1]http://moodo.musiclab.si

3.1 Differences in Perception of Emotions and Colors

Based on our previous studies of the Moodo dataset, there is a clear distinction between the valence-arousal positions and chosen colors of emotions in different music contexts. Figures 2 and 3 show valence-arousal ratings and colors for a set of emotion labels describing induced emotions in folk and electro-acoustic contexts. As shown in these Figures, the context significantly influences the chosen colors and valence-arousal positions.

Analysis of responses shows that users rate induced emotions in both non-mainstream genres quite differently. In folk music (Figure 2), the light green and blue hues dominate the dark blue and black colors in all emotion graphs, except for *fear* and *sadness*, where the dark blue and red hues prevail. *Fear* and *anger* also share similar distributions in the valence-arousal space. *Sadness* is dominated by light and dark blue colors. Responses to electro-acoustic music (Figure 3) are mostly dark (gray, dark red and blue hues prevail). *Fear* and *anger* share both dark colors and valence-arousal distributions. *Fear* is also more frequently labelled in the Electro-acoustic collection. In mainstream genres (Figure 4), emotions are on average positioned closer to the electro-acoustic works than to folk songs.

Table 1 contains average valence-arousal ratings for the selected music styles in addition to ratings in the non-music context, and average values across the whole Moodo dataset. The averages confirm our observations of Figures 2 and 3. *Anger* and *fear* have significantly smaller negative valence for folk songs when compared to electro-acoustic music, and are significantly less exciting according to the arousal dimension. *Happiness* is perceived as more positive (valence dimension) in folk songs, but not significantly different in the arousal position. On the contrary, *relaxation* is perceived a significantly more exciting (arousal position) in folk songs.

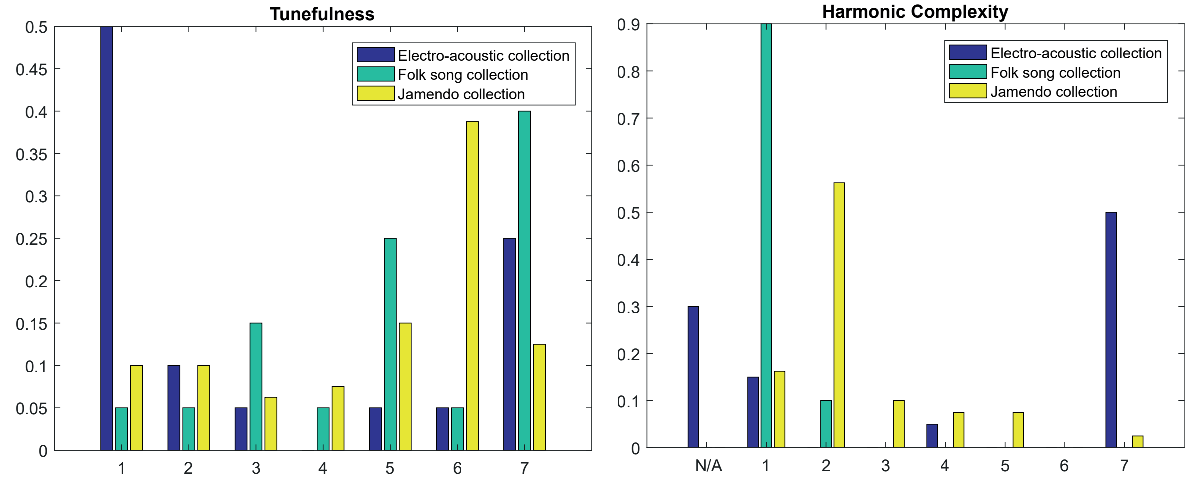

Figure 1: A comparison of electro-acoustic, folk and mainstream (Jamendo) collections with respect to tunefulness and harmonic complexity on a scale from 1,to 7. Tunefulness reflects the melodic aspects of an excerpt, which is high for folk songs and low for electro-acoustic works. Harmonic complexity is naturally low for folk songs and can vary from extremely complex to simple or even absent (N/A) in electro-acoustic works. The mainstream collection (Jamendo) consists of a variety of music genres and shows mostly above average tunefulness and low harmonic complexity.

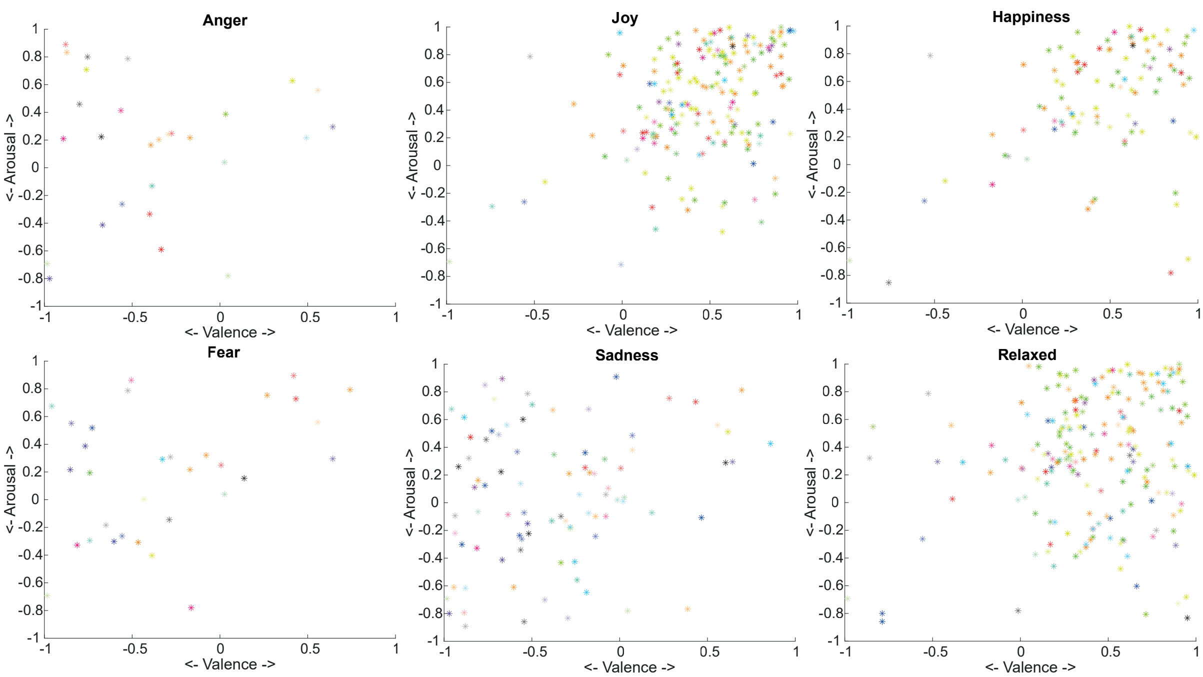

Figure 2: Emotion labels and their positions in the valence-arousal space for emotions induced by folk music. The colors associated with each emotion are also shown. There is a clear difference in color vividness and brightness when compared to the electro-acoustic collection in Figure 3. Moreover, the average valence-arousal ratings change significantly especially with negative emotions (e.g. *fear*, *anger*).

4. Discussion

The differences in perception of emotions in the non-music and music contexts have been previously thoroughly explored on this dataset and concurs with other research (e.g. [20, 21]).

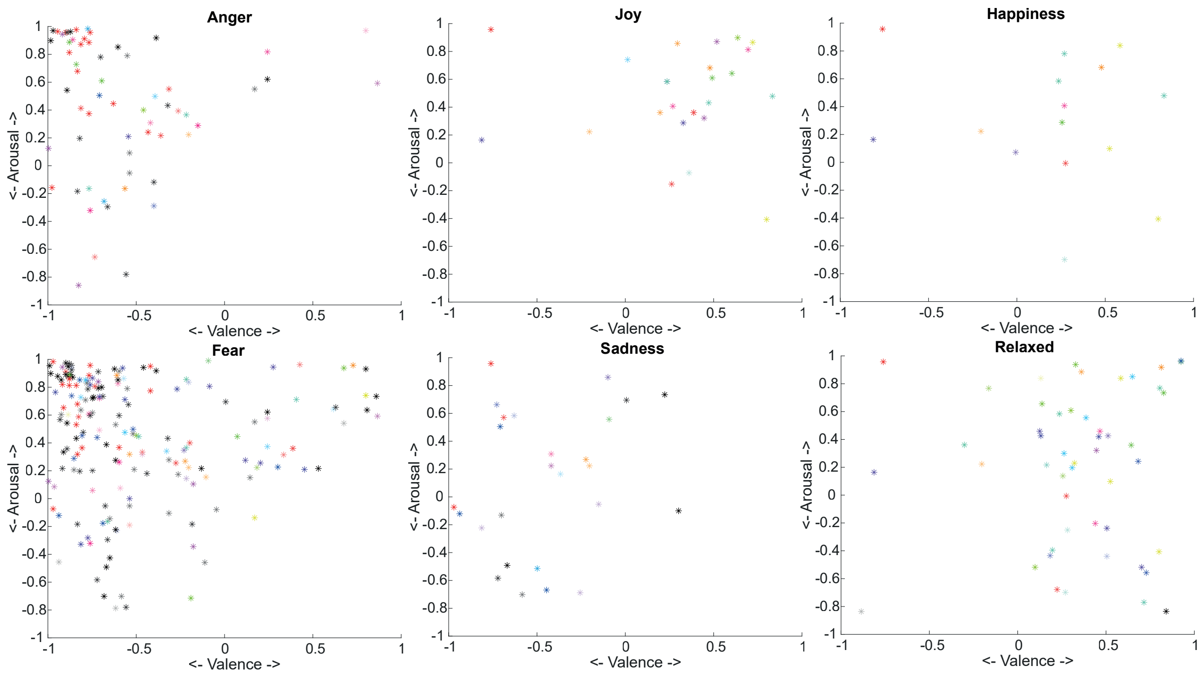

Figure 3: Emotion labels and their positions in the valence-arousal space for emotions induced by electro-acoustic music. The colors associated with each emotion are also shown. The selected colors are darker when compared to the folk song excerpts. 'Negative' emotions are prevalent (e.g *anger*, *fear*, but not *sadness*), 'positive' emotions are less common (e.g. *happiness*).

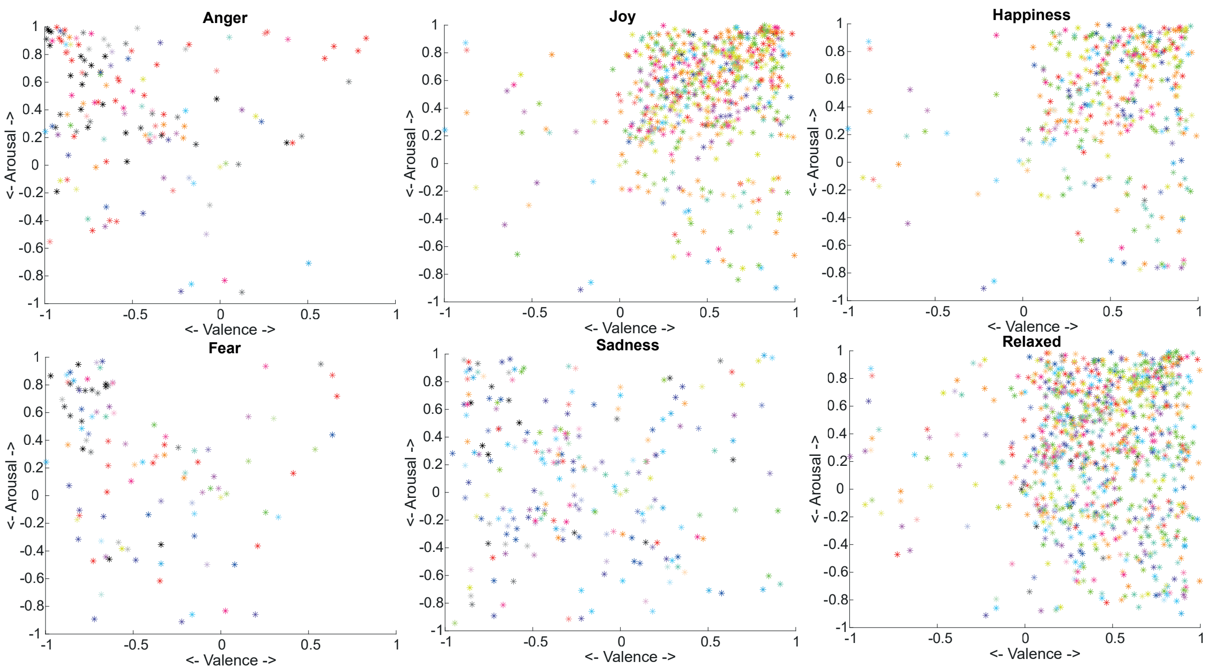

Figure 4: Emotion labels and their positions in the valence-arousal space for emotions induced by mainstream music. The colors associated with each emotion are also shown. The valence-arousal distributions are more similar to electro-acoustic works in comparison to folk music. There are significantly more points on the graphs due to a higher numbers of excerpts in these collections.

Table 1: A comparison of valence-arousal ratings (each component separately) for perception of emotions in non-music context, music context across the Moodo dataset, and separately for folk songs and electro-acoustic works.

	Non-music context				Music context overall			
	$\mu(V)$	$\mu(A)$	$\sigma(V)$	$\sigma(A)$	$\mu(V)$	$\mu(A)$	$\sigma(V)$	$\sigma(A)$
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

	Folk songs				Electro-acoustic music			
	$\mu(V)$	$\mu(A)$	$\sigma(V)$	$\sigma(A)$	$\mu(V)$	$\mu(A)$	$\sigma(V)$	$\sigma(A)$
Anger	-0.37	0.16	0.47	0.50	-0.57	0.43	0.38	0.48
Joy	0.47	0.53	0.30	0.33	0.40	0.45	0.31	0.38
Happiness	0.49	0.49	0.34	0.40	0.27	0.41	0.44	0.31
Fear	-0.18	0.18	0.51	0.49	-0.48	0.46	0.45	0.40
Sadness	-0.41	0.00	0.43	0.51	-0.45	-0.05	0.33	0.52
Relaxation	0.43	0.33	0.36	0.45	0.35	0.16	0.37	0.55

In this study we focused on relations between induced emotions and colors, mediated by the music context by looking at two different music styles. The analysis unveiled significant differences in the users' perception of emotions on several layers. First, the distribution of emotions varies between both collections—folk songs induce more positive emotions, such as *joy* and *happiness*, when compared to electro-acoustic music. Second, listeners' perception of positions of emotion labels in the valence-arousal space is different over both styles. Second, the neutral emotions are considered more positive for folk songs, and third, the choice of colors differs, as discussed in Section 3.

When deciding on the musical contents of the Moodo dataset, care was taken to uniformly distribute excerpts of different music styles. The dataset thus contains a range of different mainstream genres, from metal to classical music. Similarly, the folk songs and electro-acoustic works contain excerpts from different performers and composers, varying in topics, tunes, instruments etc. This analysis therefore reveals interesting patterns in users' perception which differ between the music styles included in the dataset. The harmonic, melodic or lyric content naturally influences the users' perception of a specific song. However, the visual and affective perception is significantly biased by the music style in the non-mainstream collections.

5. Conclusions

In this paper we analyzed users' emotional responses to two non-mainstream music collections included in the Moodo dataset, the Slovenian folk song collection and the electro-acoustic music collection. We explored the perception of emotions and colors in these two collections, which unveiled a strong influence of the non-mainstream style on the frequency of induced emotions, their valence-arousal positions and the corresponding colors. An analysis of the data, similar to the one presented in this paper, could also be replicated on mainstream genres. However, the Jamendo collection in the Moodo dataset contains 80 songs in 20 genres, resolving in very small subsets of songs, from which strong conclusions are difficult to make.

The research in the affective MIR can significantly influence future development of music recommender systems. Further analysis of emotional processing and interactions between visual and auditory modalities will help tackle a number of challenges, such as better user interfaces and consequently improved user experience in user-aware systems. Music recommendation is not the only topic which deals with the auditory and visual perception in interfaces. There is as a variety of action-related scenarios, which play an important role in most information systems, for example, the auditory supporting notifications in user actions and event occurrences in most information systems in everyday use. In our view, an immediate step towards the improvement of the interfaces lays in the annotated multi-modal datasets. Until recently, the lack of such datasets posed a significant obstacle. Consequently, the information systems development largely ignored the importance of the interfaces, especially in the auditory domain. We have shown, the Moodo dataset can offer significant insights into the multi-modal interrelations and can aid tackling these challenges in the future.

References

[1] 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*, page 2014, 2014.

[2] Anna Aljanaki, Frans Wiering, and Remco C. Veltkamp. Computational Modeling of Induced Emotion Using GEMS. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 373–378, Taipei, 2014.

[3] Amelie Anglade, Rafael Ramirez, and Simon Dixon. Genre Classification Using Harmony Rules Induced from Automatic Chord Transcriptions. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 669–764, Kobe, 2009.

[4] V. Blagojević, D. Bojić, M. Bojović, M. Cvetanović, J. Đorđević, Đ. Đurđević, B. Furlan, S. Gajin, Z. Jovanović, D. Milićev, V. Milutinović, B. Nikolić, J. Protić, M. Punt, Z. Radivojević, Ž. Stanisavljević, S. Stojanović, I. Tartalja, M. Tomašević, and P. Vuletić. Chapter one - a systematic approach to generation of new ideas for phd research in computing. In Ali R. Hurson and Veljko Milutinović, editors, *Creativity in Computing and DataFlow SuperComputing*, volume 104 of *Advances in Computers*, pages 1 – 31. Elsevier, 2017.

[5] Dmitry Bogdanov, Martín Haro, Ferdinand Fuhrmann, Anna Xambó, Emilia Gómez, and Perfecto Herrera. Semantic audio content-based music recommendation and visualization based on user preference examples. *Information Processing and Management*, 49:13–33, 2013.

[6] J. Stephen Downie, Andreas F. Ehmann, Mert Bay, and M. Cameron Jones. The Music Information Retrieval Evaluation eXchange: Some Observations and Insights. In Wieczorkowska A.A. and Ras Z.W., editors, *Advances in Music Information Retrieval*, pages 93–115. Springer-Verlag, Berlin, 2010.

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

[8] Tuomas Eerola. Modelling emotional effects of music: Key areas of improvemnt. *Proceedings of the Sound and Music Computing Conference (SMC)*, pages 269–276, 2013.

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

[10] Xiao Hu and J Stephen Downie. Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, 2007.

[11] Eric J. Humphrey, Taemin Cho, and Juan P. Bello. Learning a Robust Tonnetz-Space Transform for Automatic Chord recognition. In *Acoustics, Speech and Signal Processing (ICASSP)*, pages 453–456, New York, 2012.

[12] Javier Jaimovich, Niall Coghlan, and R.Benjamin Knapp. Emotion in Motion: A Study of Music and Affective Response. In *From Sounds to Music and Emotions*, pages 19–43. Springer, 2013.

[13] Patrik N Juslin. What does music express? Basic emotions and beyond. *Frontiers in Psychology*, 4(September):1–14, 2013.

[14] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull. Music emotion recognition: A state of the art review. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 255–266, Utrecht, 2010.

[15] A.P. Klapuri. A perceptually motivated multiple-F0 estimation method. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, pages 291–294. IEEE, 2005.

[16] Cyril Laurier, Owen Meyers, Joan Serrà, Martin Blech, Perfecto Herrera, and Xavier Serra. Indexing music by mood: design and integration of an automatic content-based annotator. *Multimedia Tools and Applications*, 48(1):161–184, oct 2009.

[17] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[18] Yu-Ching Lin, Yi-Hsuan Yang, and Homer H Chen. Exploiting online music tags for music emotion classification, 2011.

[19] Athanasios Lykartsis, Andreas Pysiewicz, Henrik

Coler, and Steffen 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)*, pages 1–15, Jyväskylä, jan 2013.

[20] Li-Chen Ou, M. Ronnier Luo, Andrée Woodcock, and Angela Wright. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, 29(3), 2004.

[21] Stephen E. Palmer, Karen B. Schloss, Zoe Xu, and Lilia R. Prado-León. Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences*, 110(22):8836–8841, 2013.

[22] Helene Papadopoulos and Geoffroy Peeters. Large-case Study of Chord Estimation Algorithms Based on Chroma Representation and HMM. *Content-Based Multimedia Indexing*, 53-60, 2007.

[23] M. Pesek, A. Leonardis, and M. Marolt. Robust Real-Time Music Transcription with a Compositional Hierarchical Model. *PLoS ONE*, 12(1), 2017.

[24] Matevz Pesek, Primoz Godec, Mojca Poredos, Gregor Strle, Joze Guna, Emilija Stojmenova, Matevz Pogacnik, and Matija Marolt. Introducing A Dataset Of Emotional And Color Responses To Music. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 355–360, Taipei, 2014.

[25] Matevž Pesek, Alja Isaković, Gregor Strle, and Matija Marolt. Improving the usability of online usability surveys with an interactive Stripe scale. In *Proceedings of the 19th International Multiconference Information Society*, pages 21–24, Ljubljana, 2016.

[26] Matevž Pesek and Matija Marolt. Chord estimation using compositional hierarchical model. In *6th International Workshop on Machine Learning and Music, held in conjunction with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML/PKDD 2013*, 2013.

[27] Matevž Pesek, Gregor Strle, Alenka Kavčič, and Matija Marolt. The Moodo dataset: Integrating user context with emotional and color perception of music for affective music information retrieval. *Journal of New Music Research*, 46(3):1–15, 2017.

[28] James A Russell. A circumplex model of affect. *1Journal of personality and social psychology*, 39(6):1161–1178, 1980.

[29] Markus Schedl, Arthur Flexer, and Julián Urbano. The neglected user in music information retrieval research. *Journal of Intelligent Information Systems*, 41(3):523–539, jul 2013.

[30] Erik M Schmidt and Youngmoo E Kim. Modeling Musical Emotion Dynamics with Conditional Random Fields. In *ISMIR*, pages 777–782, 2011.

[31] Björn Schuller, Clemens Hage, Dagmar Schuller, and Gerhard Rigoll. 'Mister DJ, Cheer Me Up!': Musical and textual features for automatic mood classification. *Journal of New Music Research*, 39(1):13–34, 2010.

[32] Jacquelin A Speck, Erik M Schmidt, Brandon G Morton, and Youngmoo E Kim. A Comparative Study of Collaborative vs. Traditional Musical Mood Annotation. In *2Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 549–554, Miami, 2011.

[33] Derek Tingle, Youngmoo E. Kim, and Douglas Turnbull. Exploring automatic music annotation with "acoustically-objective" tags. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 55–62, New York, mar 2010.

[34] Marko Tkalčič, Nima Maleki, Matevž Pesek, Mehdi Elahi, Francesco Ricci, and Matija Marolt. A Research Tool for User Preferences Elicitation with Facial Expressions. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 353–354, Como, Italy, 2017. ACM.

[35] Kim Torres-Eliard, Carolina Labbé, and Didier 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*, 78:252–259, 2011.

[36] Douglas Turnbull, Luke Barrington, David Torres, and Gert Lanckriet. Semantic annotation and retrieval of music and sound effects. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(2):467–476, 2008.

[37] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, jul 2002.

[38] Jonna K. Vuoskoski and Tuomas Eerola. The role of mood and personality in the perception of emotions represented by music. *Cortex*, 47(9):1099–1106, 2011.

[39] Yi-Hsuan Yang and Homer H Chen. Machine Recognition of Music Emotion, 2012.

[40] Marcel Zentner, Didier Grandjean, and Klaus R. Scherer. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, 8(4):494, 2008.

Matevž Pesek is a doctoral student at the Faculty of Computer and Information Science, University of Ljubljana. He received his B.Sc. in computer science in 2012, where he is currently working on his Ph.D. and has been working as a teaching assistant since. He is a member of the Laboratory of Computer Graphics and Multimedia since 2009. His research interests are biologically-inspired models, deep architectures including compositional hierarchical modelling and music multi-modal perception, including human-computer interaction, and visualization for audio analysis and music generation.

Matija Marolt received the B.S. and Ph.D. degrees, both in computer science, from University of Ljubljana, Faculty of Computer and Information Science, Slovenia in 1995 and 2002 respectively. He is associate professor with Faculty of Computer and Information Science, University of Ljubljana, where he has been working since 1995. He is head of Laboratory of Computer Graphics and Multimedia, where his research interests include music information retrieval, specifically semantic description of audio, retrieval and organization of ethnomusicological archives and audio-visual interaction.