CAPTURING THE MOOD: EVALUATION OF THE MOODSTRIPE AND MOODGRAPH INTERFACES

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

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

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

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

ABSTRACT

This study presents an evaluation of two interfaces for gathering user feedback in online surveys. We evaluated the intuitiveness, usability and time complexity of the proposed interfaces in comparison to the more standard approaches. Over 900 users first participated in an online survey exploring the influence of mood on their emotional responses to music and colors. We included several new interfaces in this survey, so after it was completed, users were asked to complete a second survey where they evaluated various aspects of the interfaces. Our analysis shows reduced time complexity and increased intuitiveness of the new interfaces, compared to standard approaches, resulting in lower mental difficulty and frustration of participants.

Index Terms— color perception, mood estimation, music information retrieval, questionnaire evaluation

1. INTRODUCTION

Human perception is a complex process and as such a subject of study of various scientific disciplines that attempt to model its many aspects. A common denominator of all the studies is that the developed models need to be evaluated. One way of evaluating a model is to annotate a set of percepts that become the ground truth on which the model is evaluated. Although some dataset annotations demand expert domain knowledge (e.g. annotating chord progressions on a set of music excerpts), other studies explore general perceptual tasks where annotation can be performed by any participant (e.g. music

mood estimation, genre recognition, color and mood perception). Nevertheless, an expert is still needed for preparation of the questionnaire and the analysis of the gathered data.

Special care needs to be taken when designing the dataset and gathering annotations [1, 2]. Although very relevant, we find that not enough effort is spent on designing and evaluating the procedure for gathering user annotations in field of music information retrieval. Nonetheless, new datasets are usually warmly accepted by the corresponding research communities, due to the shear complexity of designing and annotating the data.

The focus of our paper is on evaluation of the methodology for collecting a set of annotations relating user's emotional responses to music and colors. Due to the interdisciplinary nature of problem tasks, the collected dataset has potential usage in several scientific fields, including psychology, human-computer interaction and music information retrieval. We performed an online survey to gather annotations from a large number of users. In our survey, we used three new interfaces to capture specific aspects of emotional responses. Two innovative interfaces were used for mood estimation: the *MoodStripe* and the *MoodGraph*. These interfaces were used instead of the more standard inputs, such as *n*-degree scales (implemented by a slider or a group of radio buttons). To capture emotional responses to color, a discrete-scale color wheel was used, instead of the more common continuous version.

In the paper, we evaluate the intuitiveness and ease of use of our new interfaces in comparison to the more standard approaches. After preforming our online survey, we performed a separate survey, where we asked participants to evaluate the

questionnaire used in the original survey. Participants evaluated several aspects of the survey: user experience (UX) [3], complexity of the questionnaire, and aspects of our new interfaces. Our goal was to determine whether the used interfaces represent a step forward in obtaining annotations in an efficient and user-friendly manner. This paper presents an evaluation of the questionnaire using a subset of the NASA load task index [4] evaluation survey and a set of questionnaire-specific questions. Results are presented with an intention to improve the state-of-the-art in gathering of user responses, and thus improve the quality of new annotated datasets.

2. METHODOLOGY OF THE STUDY

Our primary goal was to gather an annotated dataset that captures the interplay of user's mood and their emotional responses to music excerpts and colors, as well as direct relations between music and colors. We gathered data with an online survey to reach a large number of users [5].

In the survey, after filling-in demographic data, participants were given a set of questions related to their current mood. A novel interface, called the *MoodStripe* was used in this step, to capture the user's mood. Users also had to match colors within a color wheel to words describing specific emotions. In the next step, participants were asked to listen to ten 15-second long music excerpts. After each excerpt, the *MoodGraph* interface was displayed, where users indicated their induced and perceived emotions to the music excerpt. Participants were given two sets of words describing emotions (one for induced and one for perceived emotions) which they positioned onto the valence-arousal space in the *MoodGraph*. Additionally, participants were asked to choose a color from the color wheel, which in their opinion best matched the perceived emotions of the music excerpt.

The mood survey was taken by 952 users, providing over 6600 mood/color-perception responses for the 200 music excerpts used in the survey. After the data was collected, we performed a second evaluation survey to rate several aspects of the mood survey. Participants of the mood survey were asked to answer a set of questions presented in Table 1. The questions relate to their experience with the interfaces used in the mood survey. An elaboration of the survey and its results is provided in section 3. The following subsection describes the concept and the implementation of the *MoodStripe* and the *MoodGraph* interfaces.

2.1. Capturing the mood

Compared to previous surveys relating music and mood [6], we introduced two novel interfaces for capturing user responses. The interfaces provide a substitution for the standard scales. The *MoodStripe* (Figure 2) is implemented as a canvas with one horizontal dimension. The dimension represents the presence of an emotion, ranging from *completely absent*

on the left, to *significantly expressed* on the right. The user is provided with a set of words describing different emotions and by dragging them to different positions on the canvas, expresses his or hers own mood. Since there is no strict marking of the dimension, the provided scale is continuous. Positions of placed emotions can also be quantized to discrete values if desired for analysis. The amount of information retrieved by the *MoodStripe* interface is at least equal, when compared to a set of radio buttons commonly used to capture similar information - an example of a 7-degree scale in shown in Figure 1.

Our second new interface, the *MoodGraph* (Figure 3) integrates three similar yet distinctive tasks: obtaining the perceived and induced emotions of the user, and positioning both types of emotions onto the valence-arousal space. An alternative to the *MoodGraph* would be to implement the tasks with a set of two radio groups or check boxes for perceived/induced emotions and a canvas to position each emotion onto the valence-arousal space. However the *MoodGraph* enables us to merge all tasks into one compact form. It provides two groups of emotions: the first group, marked with an icon of a person, defines the set of induced emotions; the second group, marked with an icon of a note, defines the set of perceived emotions. At least one emotion from each category should be placed onto the valence-arousal space. By dragging emotions onto the canvas, an icon representing the category appears next to the dragged word, as seen in Figure 3.

In our evaluation survey, we also evaluated the usefulness of a discrete color wheel containing 49 color tones. The 49 tones were chosen as a trade-off between the number of choices and the complexity of the full continuous color wheel. By clicking on the color wheel, the user can pick the best matching color to a music excerpt or emotional label. We evaluated how appropriate this limited set of colors is to capture the intended user feedback.

	se ne pojavlja		je srednje izraženo		je zelo izrazito
Igriv	○ 1 ○ 2	○ 3 ○ 4	○ 5	○ 6 ○ 7
Aktiven	○ 1 ○ 2	○ 3 ○ 4	○ 5	○ 6 ○ 7
Boječ	○ 1 ○ 2	○ 3 ○ 4	○ 5	○ 6 ○ 7
Buden	○ 1 ○ 2	○ 3 ○ 4	○ 5	○ 6 ○ 7
Dremav	○ 1 ○ 2	○ 3 ○ 4	○ 5	○ 6 ○ 7
Energičen	○ 1 ○ 2	○ 3 ○ 4	○ 5	○ 6 ○ 7
Entuziastičen	○ 1 ○ 2	○ 3 ○ 4	○ 5	○ 6 ○ 7

Fig. 1. A set of radio buttons commonly used in standard questionnaires. Each row represents a specific emotion, radio buttons define the scale from 1 (not present) to 7 (significantly expressed).

No. Question	Range	Comments
1 Mental difficulty of solving the questionnire	7-degree scale	*1 - simple, 7 - extremely complex*
2 Physical difficulty of solving the questionnire	7-degree scale	*1 - simple, 7 - extremely complex*
3 Time complexity of solving the questionnaire	7-degree scale	*1 - extremely brief, 7 - extremely time consuming*
4 Intuitiveness of the *MoodGraph* interface	7-degree scale	*1 - not intuitive at all, 7 - extremely intuitive*
5 Time complexity of the *MoodGraph* interface	7-degree scale	*1 - extremely brief, 7 - extremely complex*
6 Number of available emotions in categories	7-degree scale	*1 - categories are lacking emotions, 7 - categories include too many emotions*
7 Intuitiveness of the *MoodStripe* interface	7-degree scale	*1 - not intuitive at all, 7 - extremely intuitive*
8 Intuitiveness of the *MoodStripe* interface compared to a radio group	7-degree scale	*1 - radio group is far more intuitive, 7 - MoodStripe is far more intuitive*
9 Time complexity of the *MoodStripe* interface compared to a radio group	7-degree scale	*1 - radio group is faster by far, 7 - MoodStripe is faster by far*
10 Number of color tones on the color wheel interface	7-degree scale	*1 - Color wheel lacks colors, 7 - color wheel includes too many colors*
11 Time spent while solving the questionnaire	7-degree scale	*time spent in minutes -* *{ < 3 , 3-5, 5-7, 7-9, 9-11, 11-15, > 15 }*
12 Frustration induced while solving the questionnaire	7-degree scale	*1 - extremely frustrating, 7 - not frustrating at all*

Table 1. The structure of the evaluation survey. The user was presented an image of the MoodGraph (Fig. 3) before the 4th question, images of the *MoodStripe* (Fig. 2) and a radio group (Fig. 1) before the 7th question, and an image representing a color wheel before the 10th question.
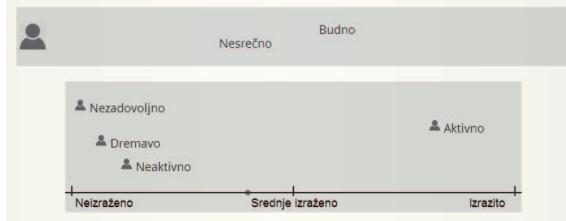
Fig. 2. The *MoodStripe*: users position emotions on a continous scale from not present on the left to very expressed on the right.

3. EVALUATION OF THE QUESTIONNAIRE

The online evaluation survey was filled-in by 125 users, who all took part in our mood survey. The number is high enough so that we can draw strong conclusions on various aspects of the survey and the proposed new user feedback gathering interfaces.

The participants involved in the questionnaire evaluation survey were presented a set of twelve questions. Some of the questions were drawn from the existing evaluation standard NASA load task index, while others were intended to evaluate different aspects of our interfaces. The questions are presented in Table 1. We performed the evaluation in Slovenian language in order to avoid potential misunderstanding of English expressions. Our mood survey was also only distributed to Slovenian population. The questions in Table 1 are translations of original questions.

Results are displayed in Figure 4. Although responses to the first question show balanced mental difficulty of the questionnaire (1), the physical difficulty seems to be more uniformly distributed across participants (2). Thus, it can be
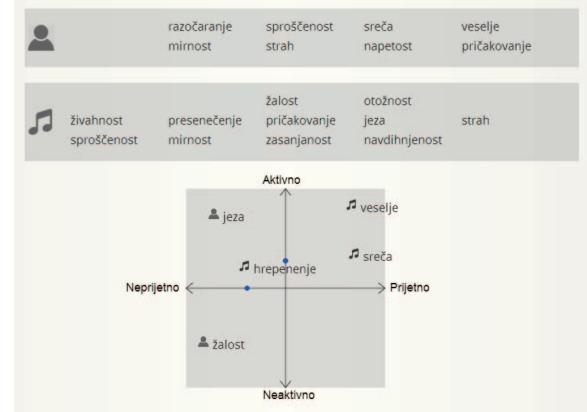

Fig. 3. The *MoodGraph*: users can position induced (person icon) and perceived (note icon) emotions onto the valence-arousal space. The shown graph has marked axes in Slovenian language with the words *Neprijetno*(unpleasant) and *Prijetno* (pleasant) for the abscissa, and *Neaktivno* (inactive) and *Aktivno* (active) for the ordinate values. Blue dots indicate the position of the selected emotion on the valence/arousal axes.

speculated that the listening part of the questionnaire represents a physical challenge to a significant number of participants. The presented *MoodGraph* interface was quite intuitive (4); however, it was also time demanding (5). Considering the task load of the interface (combining three distinctive tasks as described in section 2.1), this was expected. The number of emotions in *MoodGraph* categories was slightly unbalanced (6), thus the categories should be extended in our future work. The *MoodStripe* interface represents a significant improve-

Fig. 4. The twelve cumulative histograms represent the distribution of answers to twelve evaluation questions.

ment over 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 enlarging the set of color tones in our future work.

A subset of participants experienced some frustration when solving the questionnaire (12). The majority spent 11-15 minutes or even more than 15 minutes solving the questionnaire (11). However, the cumulative responses show an opinion of a balanced questionnaire (3) regarding the estimated time spent solving.

4. CONCLUSIONS AND FUTURE WORK

Results of evaluation demonstrate the usefulness of the proposed *MoodGraph* and *MoodStripe* interfaces. The key advantages, compared to the standard input types, were proven to be in reduced time complexity and increased intuitiveness

of the interfaces, thus resulting in lower mental difficulty and frustration of participants. We also obtained some useful clues for improving the interfaces and survey, which we intend to consider in our next English version of the mood survey, which will be made public shortly.

5. REFERENCES

[1] Julie Yu and Harris Cooper, "A Quantitative Review of Research Design Effects on Response Rates to Questionnaires," *Journal of Marketing Research*, vol. 20, no. 1, pp. 36–44, 1983.

[2] E McColl, A Jacoby, L Thomas, J Soutter, C Bamford, N Steen, R Thomas, E Harvey, A Garratt, and J Bond, "Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and pa-

tients.," *Health technology assessment (Winchester, England)*, vol. 5, no. 31, pp. 1–256, Jan. 2001.

[3] William Albert and Thomas Tullis, *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Google eBook)*, Newnes, 2013.

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

[5] Authors, "Submission to the 2nd Empire Workshop ID 2," in *Supplied as additional material empire2014.pdf*, 2014.

[6] Cyril Laurier, Olivier Lartillot, Tuomas Eerola, and Petri Toiviainen, "Exploring Relationships between Audio Features and Emotion in Music," in *Proceedings of the Conference of European Society for the Cognitive Sciences of Music (ESCOM)*, 2009.