# Bodily sensation maps: exploring a new direction for detecting emotions from user self-reported data

Iván García-Magariño[a,b,*], Luca Chittaro[c], and Inmaculada Plaza[d,b]

[a]*Department of Computer Science and Engineering of Systems, Escuela Universitaria Politécnica de Teruel, University of Zaragoza, c/ Atarazana 2, 44003 Teruel, Spain.*

[b]*Instituto de Investigación Sanitaria Aragón, University of Zaragoza, Zaragoza, Spain*

[c]*Human-Computer Interaction Lab, Department of Mathematics, Computer Science, and Physics, University of Udine, Via delle Scienze, 206, 33100 Udine, Italy*

[d]*Department of Electronics Engineering and Communications, Escuela Universitaria Politécnica de Teruel, University of Zaragoza, c/ Atarazana 2, 44003 Teruel, Spain*

---

Abstract

The ability of detecting emotions is essential in different fields such as user experience (UX), affective computing, and psychology. This paper explores the possibility of detecting emotions through user-generated bodily sensation maps (BSMs). The theoretical basis that inspires this work is the proposal by Nummenmaa et al. (2014) of BSMs for 14 emotions. To make it easy for users to create a BSM of how they feel, and convenient for researchers to acquire and classify users' BSMs, we created a mobile app, called EmoPaint, which includes an interface for BSM creation, and an automatic classifier that matches the created BSM with the BSMs for the 14 emotions. The user study we present aims at evaluating both components of EmoPaint. First, it shows that the app is easy to use, and is able to classify BSMs consistently with the considered theoretical approach. Second, it shows that using EmoPaint increases accuracy of users' emotion classification when compared with an adaptation of the well-known method of using the Affect Grid with the Circumplex Model, focused on the same set of 14 emotions of Nummenmaa et al. Overall, these results indicate that the novel approach of using BSMs in the context of automatic emotion detection is promising, and encourage further developments and studies of BSM-based methods.

*Keywords*: emotion detection, bodily sensation maps, user experience, mobile application

---

- **Corresponding author, Telephone: +34 978645348, Fax: +34978618104*
- *Email addresses: ivangmg@unizar.es (Iván García-Magariño), luca.chittaro@uniud.it (Luca Chittaro), inmap@unizar.es (Inmaculada Plaza)*

## 1 Introduction

The ability of assessing emotions is essential in diverse fields such as affective computing, user experience (UX), or different areas of psychology. Traditional methods often rely on questionnaires, text-based, e.g. the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988) and semantic differentials (Bradley & Lang, 1994), or more visual, e.g. affect grids (Russel et al., 1989) and the self-assessment manikin (Bradley & Lang, 1994). Recent years have witnessed a growing exploration of methods based on physiological signals, e.g. electrodermal activity (Healey & Picard, 1998), electromyography (Gruebler & Suzuki, 2014), cardiac activity (Nardelli et al., 2015), respiratory activity (Gomez et al., 2004), and combinations of different signals (Mandryk & Atkins, 2007; Fleureau et al., 2012; Wu et al., 2010; Chittaro & Sioni, 2014). Physiological methods can have the advantage of supporting an implicit detection of emotions, because they do not require users to identify and report explicitly their emotions. Employing computers in the analysis introduces the opportunity of automatic detection of emotions, a research topic to which the literature is devoting increasing attention, proposing and evaluating classifiers based on physiological signals (Mandryk & Atkins, 2007; Fleureau et al., 2012; Wu et al., 2010; Chittaro & Sioni, 2014; Gruebler & Suzuki, 2014) or the recording of user's behavior based on cameras, e.g. facial expressions (Eleftheriadis et al., 2015; Soleymani et al., 2016), body movement (Castellano et al., 2007) and laughter from body motion (Griffin et al., 2015); microphones, e.g. speech (Deng et al., 2014) and non-verbal sound analysis (Gupta et al., 2016); or written textual communication with other users (Li & Xu, 2014) and conversational agents (Benyon et al., 2013).

This paper begins to explore a new research direction for collecting data from users for emotion detection purposes, by means of user-generated bodily sensation maps (BSMs), which are graphical depictions of the sensations an individual feels in his/her body at a given moment. The basis we start from is the recent theoretical work by Nummenmaa et al. (2014), who established relations between a set of emotions and specific BSMs.

An important advantage of defining a method based on BSMs is that it could support implicit detection of emotions (users do not have to explicitly report their emotions) without the need for physiological sensors (or other sensors such as microphones and cameras), which can be complex, inconvenient, possibly costly, and severely affected by environmental factors (such as noise and light conditions). On the contrary, in the scenario we envision, a user could just "paint" a BSM to represent how his/her body feels, using a common, familiar device (including his/her own smartphone or tablet) and an easy-to-use interface. Then, an automatic classifier could analyze the user-generated BSM and reveal user's emotions. Moreover, BSMs collect a new type of data that could also be used to extend those emotion detection approaches that combine different data sources.

Exploring the feasibility of Nummenmaa et al.'s BSMs for automatic emotion detection requires to study different aspects, which include (i) design and evaluation of the interface with which the user creates his/her BSM, (ii) design and evaluation of a classifier which is able to correctly match user's BSMs with the 14 maps identified by Nummenmaa et al., and (iii) assessment of the extent to which automatic emotion recognition based on such maps produces reliable and practically relevant results. This paper aims at investigating in all three directions, by proposing a mobile tool (called EmoPaint) that includes the BSM creation interface and the classifier, evaluating the effectiveness of both components on users, and evaluating the effectiveness of the supported emotion detection, also in contrast with an adaptation of a well-known user self-report method (Affect Grid with Circumplex Model adapted to focus on the set of 14 emotions of Nummenmaa et al.). The methods we study focus on the feeling component of emotions (i.e. the experiential part), and could complement the techniques commonly used in HCI and UX. Moreover, in addition to its use by researchers for exploring new emotion detection methods, the proposed app could be incorporated also in applications for users, e.g. emotion diaries.

## 2 Related Work

### 2.1 Emotion assessment based on self-report

The assessment of emotions through subjective methods is typically performed through self-report questionnaires that ask users to (i) rate items containing textual emotion labels, (ii) choose a graphical representation of one's emotions among a set of images such as faces or manikins, or (iii) marking the position of one's emotional state on a grid based on two or three emotional dimensions.

As an example of the first type of questionnaire, the Positive and Negative Affect Scale (PANAS) (Watson et al., 1988) includes a list of 20 words: interested, distressed, excited, upset, strong, guilty, scared, hostile, enthusiastic, proud, irritable, alert, ashamed, inspired, nervous, determined, attentive, jittery, active and afraid. Users rate each of the words on a five-point Likert scale. They indicate to what extent the word describes their feelings and emotions. The PANAS scale also provides a positive affect and a negative affect index from the responses. PANAS has been extensively validated in the literature (Seib-Pfeifer et al., 2017). The Geneva Emotion Wheel (GEW) uses labels to determine 20 different kind of emotions (Scherer, 2005). Each kind of emotion is labeled with a pair of words indicating two extremes. It relies on a graphic representation for the intensity of each emotion, with circles of different sizes in a wheel.

As an example of the second type of questionnaire, the Self-Assessment Manikin (SAM) is a graphical scale with cartoon characters (Bradley & Lang, 1994), aimed at assessing the three emotional dimensions of pleasure, arousal and dominance. In particular, it uses five cartoons for each dimension. This scale is effective in cross-cultural measurement of emotional response (Morris, 1995).

As an example of the third type of questionnaire, the Affect Grid (Russel et al., 1989) asks users to indicate their emotional state by selecting a cell from a 9x9 matrix of two emotional dimensions: arousal and valence. The Circumplex model associates such points to a set of categorical emotions (Desmet & Hekkert, 2007).

Broekens and Brinkman (2013) have recently proposed a self-report method that combines the second and the third above mentioned approaches. It uses the same three dimensions of the SAM, but asks users to select an emoticon/mood in a 3D spatial representation through an interactive application called AffectButton.

The existing emotion scales have some limitations. A well-known limitation of methods that use natural language labels such as the PANAS and GEW is the requirement of a single-language instrumentation (Thompson, 2007), because people with different cultures and native languages might interpret the labels differently. Methods based on pictorial information such as the SAM and the AffectButton cannot easily classify user's answers into different categorical emotions. Methods based on choosing a position in a multi-dimensional space make it difficult to accurately categorize the obtained spatial point into an emotion, and proposals such as the Circumplex model are partial solutions to this mapping issue. Moreover, all self-report methods summarized above can be affected by external factors, for example Mesquita (2001) showed that self-report of emotions varies depending whether it is shared in a group or individually.

As a result, there is no agreement about which of the above self-report methods is best, and there is room for proposing and exploring new ones. As an example, in addition to the methods in this section, UX area uses self-report instruments such as emotion cards, emofaces and emotion sampling device (Desmet et al., 2001; Sun & May, 2014; Roseman et al., 1996).

It is worth mentioning that due to their non-verbal nature, self-report methods can be very useful with people who have difficulty expressing emotions verbally such as kids and people with trauma (Becker-Weidman & Hughes, 2008).

### 2.2 Bodily sensation maps and emotions

The bodily sensation maps (BSMs) proposed by Nummenmaa et al. (2014) are a topographical self-report instrument in which an individual paints a human silhouette with different colors based on the sensations (s)he feels in different parts of his/her body. Two different color scales are used to indicate areas of respectively activation (black-yellow scale) and deactivation (black-light blue scale) felt in one's body. Different intensities of activation/deactivation are indicated by the chosen color on the scale. Fig. 1 shows an example of a BSM alongside the scale of colors used. Red represents low levels of activation, orange represents middle levels of activation, and yellow represents the most intense levels of activation. Similarly, deactivation is represented with different shades with dark blue corresponding to low levels of deactivation and light blue to high levels of deactivation. As Nummenmaa et al. explain, such color scales were chosen to associate black to neutral, warm colors to activation, and cold colors to deactivation.

A relevant result of the studies conducted by Nummenmaa et al. is that the specific BSMs they identified were shown to be culturally universal and language-independent. In particular, such BSMs were consistent across West European (Finnish and Swedish) and East Asian (Taiwanese) samples, in which the participants spoke the corresponding languages. The studies were restricted to the consciously felt emotions, and to 14 categorical emotions. Native speakers selected the appropriate words for the 14 emotions for each language, avoiding figurative language.

Of the 14 BSMs, six concern basic emotions (anger, fear, disgust, happiness, sadness and surprise), seven concern non-basic emotions (love, contempt, depression, anxiety, pride, shame and envy), and one is a neutral state. For instance, Fig. 1 shows the anxiety BSM.

The BSMs of the basic emotions were consistent in five independent experiments that Nummenmaa et al. carried out with different participants. The first experiment allowed the definition of the prototype BSMs from BSMs produced by participants. In the second and third experiments, participants produced the BSMs of their state after emotion-induction procedures, based respectively on short stories and movies. In the fourth experiment, participants produced the BSMs of other people by observing face expressions extracted from a validated dataset. The BSMs of these last three experiments were correlated with the prototype BSMs obtained from the first experiment. Finally, in the fifth experiment, participants were asked to associate the emotions with the prototype BSMs with a multiple-selection test, and all the basic emotions were properly correlated with the BSMs except fear.

Finally, they performed a cluster analysis to check for similarities among BSMs. The existence of similarities among certain BSMs can make classification of a user's BSM more difficult. They found one cluster in positive emotions (happiness, love and pride), four clusters in negative emotions (anger and fear; anxiety and shame; sadness and depression; and disgust, contempt and envy), and another cluster for surprise.

The seminal work by Nummenmaa et al. is inspiring other works that relate topographical body maps with emotions like the social touching maps of Suvilehto et al. (2015), and the body thermal infrared topographical images of Ioannou et al. (2014).

The use of BSMs for detecting user's emotions we explore in this paper is novel, and is quite different from other existing non-verbal, self-report methods (see Section 2.1). While most existing self-report methods select a representation of an emotion from a set (e.g. SAM, emotion cards and AffectButton) or select a point considering several emotional dimensions (e.g. Affect Grid and GEW), the approach we study allows users to represent their body sensations by drawing, without having to explicitly state their emotions or explicitly position themselves in multi-dimensional emotion spaces.

**2.3 Evaluation of emotion recognition systems**

There is no way of evaluating an emotion recognition system by comparing it with a grounded truth, since such thing does not exist in emotions (Hill et al., 2011). As a consequence, emotion recognition systems are typically evaluated by comparing their output with some assumed “real emotions” provided by a self-report method (Ramakrishnan, 2012). Such methods are characterized by the following limitations. First, self-report methods are offline tests. When people are questioned about their emotion, they are distracted from the emotional stimuli (Girodo, 1973). Second, self-report methods can provoke a new emotion in the subject. This effect could be mitigated for example by single-item tests, which avoid the possible appearance of a negative emotion due to the fatigue effect. Third, the repetition of a self-report method can induce secondary emotions. Secondary emotions (also known as meta-emotions) are induced by the awareness of other emotions (Bartsch et al., 2008). For example, a person can realize that (s)he feels anxious about a task, and then (s)he can feel surprised about the fact that the task makes him/her anxious.

Bearing these limitations in mind, an emotion recognition system is evaluated by considering its output as correct when it matches the emotions provided by the selected self-report method. The accuracy of the system could be measured as the ratio of correctly classified emotions over the total number of classifications. Most work about emotion recognition systems assesses the accuracy by calculating Cohen’s kappa coefficient (commonly represented with the *k* letter) (Calvo & Mello, 2010). Kappa coefficient is calculated as $k=(p_0-p_e)/(1- p_e)$, where $p_0$ is the observed rate of agreement and $p_e$ is the theoretical probability of agreement by chance. The advantage of the kappa coefficient is that it avoids the influence of the null error rate, i.e. correct classifications that may happen by chance. This metric is especially relevant for properly interpreting accuracy among emotion recognition systems with different sets of emotions, in which the null error rate varies depending on the size of the sets. This measure is also useful to easily determine whether a system is better than a random classifier, which would obtain theoretically a kappa coefficient of zero.

To compare different emotion recognition systems among them, an intuitive method is to compare their accuracies in terms of the kappa coefficient. Although this comparison is probably the most common (Scherer et al., 2001), some aspects must be taken into account before considering it reliable.

First, it is well-known that different kinds of emotion recognition systems provide different ranges of accuracy. For example, emotion recognition systems based on facial expressions or self-report methods often obtain higher accuracies than physiological methods (Harley, 2016). Thus, the accuracy of a novel emotion recognition system should be compared with the accuracy of a system of a similar kind. In this sense, the BSM-based approach we propose and the traditional method we will compare it with are both self-report methods that do not ask users to explicitly report their emotions, but aim at getting data that can be easier for them to report, and then try to implicitly derive the emotion from that data.

Even within a similar kind of emotion recognition systems, it is highly recommended that the systems are compared on the same data (e.g. the same facial expressions, the same body movements, or the same physiological input). For example, it is well known that facial expression recognition systems commonly obtain higher accuracies when analysing posed photos instead of spontaneous photos (Pantic & Patras, 2006). Thus, a comparison using different kinds of facial photos can be unreliable. However, only a very few studies compare emotion recognition systems using the same data, like the ones based on audio-visual information of spontaneous expressions (Zeng et al., 2009). Ideally, the approach studied in this paper should be compared to another emotion recognition system that uses BSMs, but to the best of our knowledge, no other BSM-based system exists at present.

Moreover, the comparison of emotion recognition systems is usually not representative when they use emotions sets of different sizes (e.g. when the set of emotions of the BSMs and the circumplex have not the same exact size), even though the kappa coefficient corrects the null error rate. For example, it is difficult to compare physiological emotion recognition systems with some self-report systems. Indeed, most physiological systems do not consider more than the six basic emotions proposed by Ekman (1992a), and tend to obtain worse results when trying to classify more than six emotions (Harley, 2016). On the contrary, self-report methods can distinguish among larger sets of emotions, for example up to 19 emotions in MetaTutor (Harley et al., 2013). However, comparisons among self-report systems might still be unreliable if they are not tested on the same set of emotions.

Finally, when comparing the accuracies of emotion recognition systems, the differences should be statistically analyzed to determine their significance. In particular, Cochran's Q test is well accepted by the literature for comparing the accuracies of several classifiers on the same dataset (Looney, 1988).

## 3 The EmoPaint Application

To make it easy for users to create the BSMs that describes their current body feelings, and convenient for researchers to acquire and classify such BSMs, we created a mobile app, called EmoPaint, which includes an interface for BSM creation, and an automatic classifier that tries to match the created BSM with the 14 BSMs identified by Nummenmaa et al. (2014).

The idea of using BSMs originally came from the third and the first author of this paper. EmoPaint was designed by the second author and developed by the first author, starting from three different interaction techniques for painting (see Section 3.1), and following an iterative process. The first prototype was tested by asking three HCI experts (not involved in the project) to use it for painting BSMs. In particular, we showed them Nummenmaa et al.'s 14 BSMs, and they had to reproduce the maps as precisely as possible. Changes were made to the app after interviewing the first expert, and then the app was checked again by the same expert to determine whether suggestions were properly followed. The process continued with the same expert, until he was completely satisfied. The process was repeated in the same way with the second, and finally the third expert. In the following, we describe the main functionalities of the app.

### 3.1 Painting bodily sensation maps

Fig. 2 shows the interface for creating BSMs by painting. The user can tap one of the two icons (flame or snowflake) on the bottom of the screen for painting activations or deactivations, respectively. In each of the two modes, the user can activate (deactivate) any part of the body, going up (down) in the scale of colors shown in Fig. 1.

During the design process, three different ways of painting were considered and tested with prototypes:

- *Touch and spread*: The user can touch any point of the silhouette, and the silhouette reacts as if the finger was a heating (activate) or cooling (deactivate) element. The color starts changing in the touch point following the color scale. If the user keeps touching the point, the color change propagates to an increasingly wider circular area in such a way that the more intense sensation is the one corresponding to the touching point and the color becomes less intense as the distance from that point increases.

- *Drop and push*: The user can touch any point of the silhouette, creating a color-filled circle. Then, (s)he can push the circle (i.e. perform moving gestures from the circle to neighboring areas). The obtained visual effect mimics what happens in the real world when doing the same action on a drop of fresh paint. Moreover, when the user keeps pushing the same specific area, the color of that area changes following the scale of colors. The intensity of the change is higher in the central part of the pushed area.
- *Brush*: The user brushes the screen with the finger, leaving a strip of color. The color in a brushed strip is more intense in the central area, and gradually decreases as the distance from the central area increases. The more the user brushes up and down over the same strip, the more intense the associated colors become.

A general consensus emerged about the third solution, which was considered the most intuitive and easy to use. The final version of the app is thus based on the Brush technique. From a development point of view, a critical part that required special attention was optimizing the response time of the painting actions so that there was no noticeable delay even in low-end smartphones.

Although it is not the central topic of this study, it is worth mentioning that we also added an optional emotion diary functionality to the app (see Fig. 3), which allows users to see an animated history of their BSMs recorded over time, with smooth transitions between each BSM and the following one. The functionality is controlled through a familiar video player interface supporting play, pause, skip forward and backward, and slider dragging. The emotion diary also includes a stats section that displays the number of times the user felt each emotion in a particular month.

### 3.2 Classifying bodily sensation maps

In the proposed app, each BSM is internally represented with 162,358 points within the silhouette. These points can take real values that represent different grades of the scale of colors proposed by Nummenmaa et al. (2014) for BSMs. From each BSM, the app derives a signature representation that allows to efficiently manage comparisons (Datta et al., 2008). Signatures are lossless representations of color histograms (Serratosa & Sanfeliu, 2006). A color histogram represents the distribution of colors of an image or a part of it, with the number of pixels in each color range from a fixed list. Signatures and color histograms have been widely used for comparing images in the image recognition area (Hafner et al., 1995; Swain & Ballard, 1991).

Each BSM generated by the user is classified by comparing it with the signatures of the 14 BSMs of Nummenmaa et al. More concretely, we use a classifier that has been trained with the 14 BSMs, exploiting the well-known nearest neighbor pattern classification (Cover & Hart, 1967). A key choice in designing a classifier concerns the similarity function used in the comparison. Since an essential feature of BSMs is the location of sensations, we have divided the body into the following regions: head, shoulders, arms, hands, chest, belly, hips, and legs. These regions of interest are treated separately to increase classification performance as recommended by Vu et al. (2003).

To compare each user-painted body region with the 14 reference BSMs, we considered two possibilities: average color distance (Hafner et al., 1995) and color histogram intersection (Swain & Ballard, 1991). To choose between the two, we tested them in the EmoPaint app. To get a dataset for the test, three different users were asked to paint each of the 14 BSMs of Nummenmaa et al. (2014) with EmoPaint, generating a set of 42 BSMs The color histogram intersection obtained better results (i.e. a higher amount of BSMs were properly classified) on the set, and was used in the final version of the app. A color histogram intersection is calculated as the sum of the common numbers of points in each color range between two histograms.

The intersection similarities of the body regions are combined by summing them. In this way, the similarity between two BSMs represents the number of pixels that coincide in the same color range and body region.

To support easily possible changes and extensions to the set of recognized BSMs, we have integrated a Case-based Reasoning (CBR) component in the app. When launching the app for the first time, the case memory is made by the 14 BSMs of Nummenmaa et al. When a user paints a new BSM, the system retrieves the most similar case from the memory, and displays the emotion associated with that BSM (see example in Fig. 4). Enabling the CBR component allows the user to change the database in a supervised learning fashion. If the option is enabled, a "Wrong" button offers the possibility to reject the classification proposed by the system, retaining the currently painted BSM as a new case and to specify to which emotion it should be associated. In this way, the classifier can be personalized to better handle similar cases involving the individual user. When a new case is retained, it replaces the previous one associated to the same emotion.

## 4 User Study

We carried out a user study to assess the feasibility of recognizing emotions from self-reported BSMs. The study was designed by the second and first author of the paper, and carried out by the first and third author. Four research questions were considered: (i) does the EmoPaint interface allow users to easily create their BSMs?, (ii) to what extent the app is able to recognize the 14 BSMs of Nummenmaa et al. (2014)?, (iii) to what extent the proposed approach is able to recognize the actual emotion of participants?, and (iv) does the new approach improve on traditional self-report approaches adapted to recognize the same set of emotions? For the last question, the traditional approach we considered was the well-known Affect Grid (Russel et al., 1989) in combination with the Circumplex Model (Desmet & Hekkert, 2007; Zagalo et al., 2005) adapted to focus only on the 14 emotions of Nummenmaa et al.

We also collected qualitative feedback from the participants to highlight possible design opportunities for improving the app.

### 4.1 Participants

We recruited 54 participants (31 male, 23 female) through personal contact. The participants had different occupations and received no compensation. The nationalities of participants were Italian (27), Spanish (25), French (1), and Mexican (1). Age ranged from 18 to 66 (M=31.87, SD=11.39). All participants were familiar with using smartphones.

### 4.2 Measures

#### *4.2.1 Painting and Recognition of the 14 BSMs*

To measure the capability of the app to support painting and recognition of the 14 BSMs of Nummenmaa et al. (2014), we asked participants to paint each of the 14 BSMs with the app. Then:

- The capability of supporting participants in painting BSMs was measured by calculating the similarities between the user-painted version of each BSMs and the original version of Nummenmaa et al. We used the similarity function between signatures described in Section 3.2. We calculated the average similarity and its standard deviation for each of the 14 BSMs. Since the similarity represents a number of points, we calculated the percentage of similarity by dividing it by the total number of BSM points. The percentage of similarity between two BSMs is the sum of the color histogram intersections for the eight body regions, divided by the total number of points of a BSM (i.e. 162,358). The resulting ratio is represented as a percentage by multiplying it by 100.

- The percentage of BSMs correctly recognized by the classifier over the total number of classifications was determined. Since the classifier is non-binary, Cohen's kappa coefficient (Cohen, 1960) was calculated from this accuracy. The accuracy and the kappa coefficient were then calculated independently for each of the 14 BSMs. The performance of the classifier was also measured with a confusion matrix (Stehman, 1997).

*4.2.2 App usability*

To measure system usability and ease of learning of the painting interface for creating BSMs, we employed well-known instruments. In particular, we used the System Usability Scale (SUS) (Brooke, 1996) for measuring the usability, and the dimension of ease of learning of the Usefulness, Satisfaction and Ease of Use (USE) instrument (Lund, 2001).

The SUS scale provides a result in the 0-100 range. The results of USE are in the 1-7 range, but we converted them to the 0-100 range, to make it easy to compare them with the SUS results.

*4.2.3 Recognition of actual emotions*

To assess the capability of the proposed app to recognize actual participant's emotions, a second task was assigned to participants: they were asked to relive a recent situation of their choice that provoked them an emotion, and represent their current body sensations by generating a BSM with the app. Reliving memories to influence the current emotion of subjects is a technique widely supported by the psychology literature (Joormann et al., 2007; Labouvie-Vief et al., 2003). Participants were explicitly told that they had to paint the sensations they felt at the moment of painting, which were not necessarily the same as the ones they felt in the past situation they relived. After they completed their BSM, the app classified the emotion and they had to say if that was or not the emotion they were feeling while painting. The performance of the app in detecting actual emotions was measured with accuracy and Cohen's kappa coefficient.

*4.2.4 Comparison with a traditional self-report method on the same set of emotions*

The presented approach was compared with a well-known self-report instrument. We selected the combination of the Affect Grid and the Circumplex Model because they are frequently used for measuring emotions in UX (Colomo-Palacios et al., 2011). However, we adapted the Circumplex Model to focus on the 14 categorical emotions of the proposed approach. We discarded pictorial affective measurement methods such as SAM and Affect-Button since there was not a way to obtain from them the set of 14 categorical emotions.

Participants were asked to represent how they felt using the Affect Grid. In this scale, participants marked a cross in the 9x9 space shown in Fig. 5 to represent their valence-arousal state. Then, we used the Circumplex Model

(Desmet & Hekkert, 2007; Zagalo et al., 2005) to map the values of the Affect Grid into the 14 emotions of the BSM approach with the mapping shown in Fig. 6. The Circumplex model associates each emotion with a point in the 9x9 valence-arousal space of the AffectGrid (represented with bold numbers in Fig. 6). In this 9x9 space, the remaining points were associated with the emotion of the nearest point among those directly associated by the Circumplex model (with the non-bold numbers in Fig. 6). We used a combination of the angular distance and Euclidian distance to determine the proximity between two points in this space. This is the most common way of adapting the Circumplex Model as one can observe also in the recent literature (Wittig et al., 2016). This association into 14 categorical emotions was necessary so that both emotion detection systems used the same set of emotions, making results comparable (Harley, 2016), as discussed in section 2.3. From the BSM approach of Nummenmaa el al., we inherited the limitation that the 14 emotions do not fully cover the entire valence-arousal space. In particular, Fig. 6 highlights the points that are distant from any of the 14 emotions with a grey background color. The Circumplex model would associate these points with other emotions such as bored, lethargic, calm, relaxed, tranquil, sympathy, and comforted. To consider the fact that the adapted Circumplex excludes some of the emotions of the original Circumplex, after comparing the BSM and the adapted Circumplex approach on all user-provided cases, we also repeated the analysis, excluding the cases in which the adapted Circumplex Model used associations that were different from the original Circumplex (i.e. the ones highlighted in Fig. 6).

*4.2.5 Qualitative feedback*

During the first task (painting each of the 14 BSMs with the app), participants were invited to think-aloud about the possible difficulties encountered with the interface as well as the features they appreciated. In addition, participants were also interviewed at the end of the task to get possible suggestions for improving the app.

**4.3 Procedure**

The experimenter asked each participant to seat in a chair facing a table. Firstly, the experimenter showed each participant how to use the painting interface of EmoPaint on a smartphone. The experimenter used both the activation and deactivation modes, and went through the scale of colors in different body parts. Then the experimenter told the participant to try the painting interface to check if (s)he understood how to use it.

After this, the experimenter showed participants a figure with all the 14 BSMs of Nummenmaa et al. (2014) together. These BSMs were presented to the participant just as colored silhouettes without any reference to any emotion. The experimenter asked participants to paint each silhouette with the app as accurately as possible, in a given order, considering both the color levels and their position. The experimenter asked each participant to be aware of the small differences between some silhouettes. To counterbalance learning effects, participants were asked to paint the 14 BSMs in different orders, selected in such a way that each BSM was painted a similar number of times as first, second, third, and so on. Participants were asked to think-aloud while doing the task, reporting difficulties encountered with the interface as well as the features they appreciated.

After participants had painted each BSM, the experimenter took the phone and used the app to classify the emotion associated with the painted BSM, without showing the screen to the participant. The experimenter noted the output of the app to determine whether the app recognized or not the painted BSM as the corresponding BSM by Nummenmaa et al.

After completing this task, the experimenter interviewed the participant to get possible suggestions for improving the app. Then, the participant filled the SUS and USE questionnaires.

After this, the experimenter explained the association between the scale of colors and bodily sensations. Then, participants carried out the second task (reliving a past experience and painting a BSM of his/her current sensations), already described in Sections 4.2.3 and 4.2.4. Participants subsequently marked the Affect Grid. To say whether their current emotion in the second task was recognized correctly, participants used the EmoPaint app (the "Right" and "Wrong" buttons in Fig. 4). When participants chose "Wrong", they were asked by the app to select from the list of the 14 emotions which was the most similar to their felt one (see Fig. 7).

When participants chose "Wrong", we informally tested the possibility of learning the new BSM. The experimenter selected on the app the CBR option to memorize the case in the case memory, and then asked the participant to paint the same sensations again. Finally, the app classified the emotion of the new painted BSM, and the participant was asked to indicate whether the output of the app was correct.

## 5 Results

### 5.1 Painting and Recognition of the 14 BSMs

Table 1 shows the average similarity percentages (introduced in section 4.2.1) and the standard deviations, between the user-painted BSMs and the corresponding BSMs of Nummenmaa et al., and the total result. These average similarity percentages were in the interval 69-82%, and these results reveal the ability of the app to support proper representation of BSMs.

| Emotion | Similarity (%) | |
|---|---|---|
| | Average | SD |
| Anger | 75.0 | 10.7 |
| Anxiety | 77.2 | 10.8 |
| Contempt | 76.0 | 8.3 |
| Depression | 77.3 | 7.0 |
| Disgust | 81.7 | 5.4 |
| Envy | 81.6 | 6.6 |
| Fear | 80.7 | 6.9 |
| Happiness | 73.4 | 20.1 |
| Love | 74.3 | 17.1 |
| Neutral | 70.0 | 13.8 |
| Pride | 80.1 | 10.2 |
| Sadness | 69.7 | 13.5 |
| Shame | 73.3 | 8.5 |
| Surprise | 81.2 | 8.3 |
| *Total* | 76.5 | 12.0 |

*Table 1 Similarities between painted BSMs and the 14 BSMs*

Table 2 shows the accuracy (and Cohen's kappa coefficient) of the app in classifying the BSMs, for each of the 14 BSMs of Nummenmaa et al., while Table 3 provides the confusion matrix revealing some pairs of emotions with some recognition power limitations. The app had a high (i.e., 87% or over) accuracy in recognizing the BSMs of the 6 basic emotions (anger, fear, disgust, happiness, sadness, and surprise). The BSM of the neutral state and of most of the complex emotions were also classified properly with high accuracy, with a few exceptions (the worst case is given by the BSM for the envy non-basic emotion).

| | Anger | Fear | Disgust | Happiness | Sadness | Surprise | Neutral | Anxiety | Love | Depression | Contempt | Pride | Shame | Envy | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Accuracy (%) | 92.6 | 90.7 | 94.4 | 96.3 | 87.0 | 94.4 | 88.9 | 96.3 | 96.3 | 100 | 51.9 | 87.0 | 77.8 | 33.3 | 84.8 |
| Cohen's Kappa | .920 | .900 | .940 | .960 | .860 | .940 | .880 | .960 | .960 | 1.00 | .481 | .860 | .761 | .282 | .836 |

*Table 2 Accuracy and Cohen's kappa coefficient for detection of the 14 BSMs*

| | | Predicted by the app | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Anger | Fear | Disgust | Happiness | Sadness | Surprise | Neutral | Anxiety | Love | Depression | Contempt | Pride | Shame | Envy |
| Actual according to the BSMs of Nummenmaa | Anger | .93 | .06 | .00 | .02 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| | Fear | .00 | .91 | .09 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| | Disgust | .00 | .02 | .94 | .00 | .00 | .02 | .00 | .02 | .00 | .00 | .00 | .00 | .00 | .00 |
| | Happiness | .00 | .02 | .00 | .96 | .00 | .00 | .00 | .00 | .02 | .00 | .00 | .00 | .00 | .00 |
| | Sadness | .00 | .00 | .00 | .00 | .87 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .13 | .00 |
| | Surprise | .00 | .00 | .02 | .00 | .00 | .94 | .00 | .02 | .00 | .00 | .00 | .00 | .00 | .02 |
| | Neutral | .00 | .00 | .00 | .00 | .00 | .02 | .89 | .04 | .00 | .00 | .00 | .00 | .06 | .00 |
| | Anxiety | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .96 | .00 | .00 | .02 | .02 | .00 | .00 |
| | Love | .00 | .04 | .00 | .00 | .00 | .00 | .00 | .00 | .96 | .00 | .00 | .00 | .00 | .00 |
| | Depression | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | 1.00 | .00 | .00 | .00 | .00 |
| | Contempt | .00 | .00 | .00 | .00 | .00 | .48 | .00 | .00 | .00 | .00 | .52 | .00 | .00 | .00 |
| | Pride | .02 | .04 | .04 | .00 | .00 | .00 | .00 | .00 | .04 | .00 | .00 | .87 | .00 | .00 |
| | Shame | .00 | .00 | .00 | .00 | .07 | .02 | .00 | .13 | .00 | .00 | .00 | .00 | .78 | .00 |
| | Envy | .00 | .00 | .00 | .00 | .00 | .57 | .00 | .02 | .00 | .00 | .06 | .00 | .00 | .33 |

*Table 3 Confusion matrix for detection of the 14 BSMs*

## 5.2 App usability

The average values of usability and ease of learning are presented in table 4. Both features ranked as 80 or more in average in the 0-100 range.

| | Average | SD |
|---|---|---|
| Usability | 79.6 | 13.5 |
| Ease of Learning | 88.4 | 11.6 |

Table 4 Usability and ease of learning in the 0-100 range

## 5.3 Recognition of actual emotions and comparison with a traditional self-report method

In the second task, the actual emotions reported by participants were: anxiety (15 participants), happiness (11), neutral (8), fear (6), anger (5), love (3), disgust (2), pride (1), sadness (1), shame (1), and surprise (1).

Table 5 shows the global accuracy (and Cohen's kappa coefficient) of the proposed approach in classifying actual emotions of the participants, alongside with the results obtained with the Affect Grid combined with the adapted Circumplex Model. The proposed approach obtains a higher performance in classifying participant's actual emotions as one can observe in both the accuracy and the Cohen's kappa results. In particular, the accuracy of the proposed approach is almost double the other. In addition, the obtained kappa coefficient is more than double with the proposed approach.

| | EmoPaint app | Affect Grid with Circumplex Model | Ratio |
|---|---|---|---|
| Accuracy (%) | 38.9 | 20.4 | 1.91 |
| Kappa | .342 | .142 | 2.40 |

*Table 5 Detection of participant's actual emotion with the adapted Circumplex Model for the 14 emotions*

It is worth mentioning that 83.3% of the associations made by the adapted Circumplex model are the same as the ones that would have been made by the original Circumplex model. To compare the performance of the BSM and the Circumplex more thoroughly, we performed also a second analysis, restricted to the cases that are directly covered by the original Circumplex, with no need for adaptation. The results of the second analysis are even better the previous ones: the EmoPaint app further improves over the Circumplex in accuracy as well as Cohen's kappa (Table 6).

| | EmoPaint app | Affect Grid with Circumplex Model | Ratio |
|---|---|---|---|
| Accuracy (%) | 37.8 | 17.8 | 2.13 |
| Kappa | .330 | .115 | 2.88 |

*Table 6 Detection of participant's actual emotion excluding the cases in which the adapted Circumplex Model provided a different output from the adapted one.*

Table 7 shows the table of contingency between the proposed approach and the Affect Grid with the adapted Circumplex Model. The number of cases in which the proposed approach provides a better output than the other one is higher than the opposite cases (i.e. 17 vs. 7). Again, we repeated the analysis excluding the cases in which the adapted Circumplex Model used a different association from the original Circumplex, and Table 8 presents the results.

| | | | Affect Grid with Circumplex Model | | Total |
|---|---|---|---|---|---|
| | | | wrong | correct | |
| EmoPaint | wrong | Count<br>% of total | 26<br>48.1% | 7<br>13.0% | 33<br>61.1% |
| | correct | Count<br>% of total | 17<br>31.5% | 4<br>7.4% | 21<br>38.9% |
| Total | | Count<br>% of total | 43<br>79.6% | 11<br>20.4% | 54<br>100% |

*Table 7 Table of contingency between EmoPaint and Affect Grid with the adapted Circumplex Model*

| | | | Affect Grid with Circumplex Model | | Total |
|---|---|---|---|---|---|
| | | | wrong | correct | |
| EmoPaint | wrong | Count<br>% of total | 22<br>48.9% | 6<br>13.3% | 28<br>62.2% |
| | correct | Count<br>% of total | 15<br>33.3% | 2<br>4.4% | 19<br>37.8% |
| Total | | Count<br>% of total | 37<br>82.2% | 8<br>17.8% | 45<br>100% |

*Table 8 Table of contingency between EmoPaint and Affect Grid with Circumplex Model restricted to the cases that are directly covered by the original Circumplex.*

To determine whether these differences are significant, Cochran's Q test was applied, as routinely done in the literature for comparing the accuracies of several classifiers (Looney, 1988). The result was statistically significant ($\chi^2 = 4.167$, p=0.041). If we consider only the cases in which the adapted Circumplex model used original associations, the result was still statistically significant ($\chi^2 = 3.857$, p=0.050).

Table 9 compares the accuracies of EmoPaint and Affect Grid for detecting each actual emotion, alongside the frequency of that emotion in the sample. Table 10 shows the same analysis but restricted to the cases that are covered by the original Circumplex without need for adaptation.

| Actual emotion | | Accuracy (%) | |
|---|---|---|---|
| Emotion | Frequency | EmoPaint | Affect Grid with Circumplex Model |
| anxiety | 15 | 46.7 | 20.0 |
| happiness | 11 | 9.1 | 0.0 |
| neutral | 8 | 37.5 | 75.0 |
| fear | 6 | 83.3 | 0.0 |
| anger | 5 | 0.0 | 20.0 |
| love | 3 | 33.3 | 0.0 |
| disgust | 2 | 50.0 | 0.0 |
| pride | 1 | 0.0 | 100 |
| sadness | 1 | 100 | 100 |
| shame | 1 | 100 | 100 |
| surprise | 1 | 100 | 100 |

*Table* 9 *Accuracies for detecting each actual emotion*

| Actual emotion | | Accuracy (%) | |
|---|---|---|---|
| Emotion | Frequency | EmoPaint | Affect Grid with Circumplex Model |
| anxiety | 14 | 50.0 | 21.4 |
| happiness | 8 | 12.5 | 0.0 |
| neutral | 4 | 0.0 | 75.0 |
| fear | 6 | 83.3 | 0.0 |
| anger | 5 | 0.0 | 20.0 |
| love | 3 | 33.3 | 0.0 |
| disgust | 2 | 50.0 | 0.0 |
| pride | 1 | 0.0 | 100 |
| shame | 1 | 100 | 100 |
| surprise | 1 | 100 | 100 |

*Table* 10 *Accuracies for detecting each actual emotion restricted to the cases that are directly covered by the original Circumplex.*

### 5.5 Qualitative feedback

The most frequent comments about the painting interface are presented in Table 11, together with their individual frequencies. It is worth mentioning that participants were able to make any comment without restrictions. We gathered very similar comments into categories for the sake of brevity and clarity in presentation.

| Comment | Frequency |
|---|---|
| You might add a width selector for the brush. | 16 |
| I would like to have an eraser tool to return colored areas to black (i.e. the neutral sensation). | 14 |
| I would expect tapping on the screen to produce small painted circles. | 10 |
| I suggest a different scale of colors for representing sensations (the suggested scales were different). | 9 |
| I would like to undo my last action(s). | 8 |
| It would be useful to zoom in some parts of the body (most referred to was the head). | 6 |
| The app could incorporate a tutorial or similar assistance. | 4 |
| I feel several emotions at the same time. Why the app provides only one? | 4 |
| I would like to receive advice for changing negative emotions when they are detected. | 2 |
| I would like to be able to represent the sensations of the backside of my body. | 2 |
| The app could be useful for representing and tracking pain for medical purposes. | 2 |

*Table 11 Most frequent comments of participants*

### *5.5 Informal testing of the CBR option*

As mentioned in the Procedure section, we informally tested the CBR option of the app with participants who did not agree with the classification of their actual emotion. The second try at painting the BSM was always classified correctly (accuracy 100%, Cohen's kappa coefficient 1.00). Table 12 shows the accuracy of the EmoPaint app in classifying each actual emotion in the second painting, alongside the frequencies.

| Actual emotion | Frequency | Accuracy (%) |
|---|---|---|
| happiness | 10 | 100 |
| anxiety | 8 | 100 |
| anger | 5 | 100 |
| neutral | 5 | 100 |
| disgust | 1 | 100 |
| fear | 1 | 100 |
| pride | 1 | 100 |
| *Total* | 33 | 100 |

*Table 12 Accuracy of EmoPaint after updating the CBR database and painting the same sensations again (for those cases in which participants did not agree with the first classification)*

## 6 Discussion

The current work has shown that the BSMs could be the basis for a promising research line in the area of self-report methods and emotion recognition.

To make BSMs practically useful in emotion recognition, we have presented the EmoPaint app, which allows users to just "paint" their bodily sensations with a familiar device (including their own smartphone or tablet), without needing physiological sensors (or other sensors such as microphones and cameras).

EmoPaint has shown to be highly accurate in both representing and recognizing BSMs that are similar to the 14

BSMs of Nummenmaa et al. First, the app allowed participants to paint the 14 BSMs with high similarity rates with the original 14 BSMs. Hence, the study has shown the capability of the app to adequately support painting of the BSMs.

Second, in analyzing the 14 BSMs painted by each user, the app recognized all 6 basic emotions, the neutral state, and most non-basic emotions (i.e. anxiety, love, depression, pride and shame) with high recognition rates. Two non-basic emotions (i.e. envy and contempt) were instead identified in a less satisfactory way. This cannot be due to the lack of representation capacity of the app, since the app allowed participants to paint BSMs with high average similarities to the two BSMs for envy and contempt, and these similarities were far from being the lowest ones. The lack of accuracy for these two non-basic emotions is due to the fact that their BSMs are very similar to the BSMs for other emotions. As the confusion matrix reveals, both envy and contempt are usually confused with surprise. This is simply confirmed by looking at the original BSMs, for example Fig. 8 shows the BSMs of envy and surprise side-by-side. This similarity requires one the ability and attention to paint them with high fidelity, because even small imperfections can seriously blur the already small difference between the two. The fact that two of the 14 BSMs are difficult to distinguish from others is thus a limitation of the set proposed by Nummenmaa et al. Although contempt and envy may commonly be manifested together in some people (Hahn, 1994), these emotions are not usually psychologically related with surprise. Thus, the the small differences between the BSM for envy (or contempt) and the BSM for surprise in Nummenmaa et al. makes it difficult to recognize this psychological distinction.

Usability and ease of learning results confirm that the interface was able to properly support participants in creating BSMs. Indeed, all participants were able to use the app immediately to carry out the assigned tasks without the need for experimenter's help. Nevertheless, we plan to add an optional tutorial feature, in case a future version of the app is publicly distributed.

A limitation of the current study is that we recruited only people who are familiar with smartphone use. It might be interesting to extend the study to people that do not use smartphones, to assess whether the app could be easily used for collecting emotion-related data from anyone.

The accuracy in recognizing actual emotion that the participant might feel was relatively low with the proposed approach as well as with the Affect Grid with the adapted Circumplex Model. This is probably due to the large number of classes in the classification (14 different emotions). Indeed, most automatic emotion recognition methods use a smaller number of emotion classes (Zeng et al., 2009). However, the proposed approach obtained higher accuracy than the adapted traditional method and the difference was statistically significant. Since we used an adaptation of the Circumplex Model, we also analyzed the results excluding the cases in which this adaptation used associations different from those of the original Circumplex. In this second analysis, the performance of the BSM approach actually improved, and the improvement was statistically significant. However, the p-value of the statistical test became borderline for a .05 significance level, probably due to the reduction in the number of cases.

Another possible limitation of the study concerns how we assessed if the recognition of participant's actual emotion was correct. We relied on an evaluation provided by the participants, who indicated their correct emotion in the list of 14 emotions. Most participants found the list adequate, but a few of them mentioned that they would have described their emotion differently (e.g., relaxed). In these few cases, they selected the emotion from the list which they believed was more similar to their actual one. An issue common to other emotion classification research is that some participants might have denied their actual emotion as they had convinced themselves of not having it, an unconscious mechanism of self-repairing (Salovey et al. 1995). For example, one participant explicitly mentioned "even though I broke up with my boyfriend yesterday, now I am feeling neutral". Despite these drawbacks, the selection from a list of emotions has been widely used as an effective validation mechanism in emotion recognition (Besel & Yuille, 2010; Russell et al., 2003). Although some physiological variables could have also been measured for reinforcing the validation, we ruled this measurement out for keeping the use of the app as natural as possible (for example, skin conductance sensors on the fingers would have made using the app less natural, and probably less easy).

For some participants, the app might have been useful to uncover actual emotions that they were not aware of. Indeed, mobile apps can be useful for increasing participants' self-awareness of emotions, as in the work of Morris et al. (2010), who showed this fact for their mobile app.

The informal results about the effectiveness of the CBR option suggest to study in more depth a customizable version of the app that could start with the 14 BSMs by Nummenmaa et al., and then provide users with a mechanism to increase accuracy for their specific body sensations as well as a way to collect datasets of corrected BSMs from sample users, aiming at improving and refining the original BSMs by Nummenmaa et al.

The presented accuracy for actual emotion recognition might not be representative for those emotions that occurred with very low frequency, even just once. The emotional states were caused by memories freely selected by the participants, and consequently the numbers of felt emotions were not balanced. Future study could follow a different emotion induction protocol, for example by using stories, photos or videos. This might allow us to collect an adequate number of cases for each emotion. Moreover, EmoPaint could be further evaluated by comparing its accuracy with other non-verbal instruments for detecting emotions, such as the AffectButton or SAM.

The approach we studied used the set of 14 emotions from the original research by Nummenmaa et al. However, this set might not be an ideal one from a broad emotion research perspective. For example, depression is usually regarded as an emotion disorder with clinical connotations rather than an emotion (Christensen, 2017). In addition, one could also observe that the number of positive and negative emotions in the set are not balanced. Furthermore, this set does not include emotions that combine high pleasantness and low activity, leaving some gaps in the valence-arousal space. In some cases, one could need to assess the presence of emotions that are not present in the set, such as guilt, lust, interest and boredom. The proposed app could be easily extended to include more emotions, possibly introduced by users themselves, by adding a BSM for each new emotion. A further study with this extension would be needed to determine if there are more emotions (either positive, negative or neutral) that can be detected in terms of BSMs. Such kind of studies could have theoretical implications because they could be used to extend and refine the set of BSMs by Nummenmaa et al.

It is worth mentioning that some participants explicitly said that they felt a combination of emotions. This problem is shared with other emotion recognition methods. Emotions could be represented as weighted combinations of some basic ones, but there is controversy about this. In psychology, some articles argue that there is no coherent and simple way of representing complex emotions as combinations of basic ones (Ortony & Turner, 1990), while others highlight the utility of representing emotions as combinations of some basic ones (Ekman, 1992b).

## 7 Conclusions and future work

This paper has begun to explore a new research direction for collecting emotion-related data from users by means of user-generated BSMs, i.e. topographical self-report data about user's body sensations. To the best of our knowledge, this is the first proposal and evaluation of an easy-to-use approach that allows users to "paint" their

BSMs, and incorporates an automatic classifier of the user-generated BSMs.

The study in this paper showed that the proposed interface is easy to learn and use, and that the classifier is able to classify user-generated BSMs consistently with the considered theoretical approach by Nummenmaa et al. (2014). In addition, the presented approach improved the recognition of actual emotions of participants in comparison to an adaptation of the well-known method of using the Affect Grid with Circumplex Model, increasing accuracy with a statistically significant difference. This adaptation was performed to focus on the same categorical 14 emotions of the BSM approach. However, we also performed the analysis a second time, excluding the cases in which the adapted Circumplex Model used associations different from the original Circumplex, and the results actually improved: the BSM showed an increase in accuracy, and the difference between the two approaches remained statistically significant.

In addition to the studies mentioned in the discussion, the next steps of our project involve the exploration of possible practical uses of the EmoPaint concept. Automatic recognition of emotions with a familiar device can increase emotional self-awareness of users in a wide range of contexts. For example, we will explore it in the context of mindfulness exercises that emphasize self-awareness of emotions and bodily sensations.

Another direction of future work is to evaluate EmoPaint in the large, for example by making the app available on app stores (Chittaro and Vianello, 2016). This could allow us to study whether the supervised learning functionality actually improves the accuracy of emotion detection in naturalistic use. It is a well-known fact that CBR classifiers can improve their performance by retaining new cases (Mantaras et al., 2005). On-line distribution could also allow us to reach users in several countries, supporting interesting comparisons, because culture might affect the association between bodily sensations and some emotions (Breugelmans et al., 2005).

## Acknowledgments

This research has been carried out during a stay of the first author at the HCI Lab of the University of Udine, Italy. The stay has been supported by the program "Estancias de movilidad en el extranjero José Castillejo para jóvenes doctores" funded by the Spanish Ministry of Education, Culture and Sport with reference CAS16/00075. We also acknowledge support from “Universidad de Zaragoza”, “Fundación Bancaria Ibercaja” and “Fundación CAI” in the “Programa Ibercaja-CAI de Estancias de Investigación" with reference IT6/15. This work has also been supported by the Spanish Ministry of Economy and Competitiveness (ref. TEC2013-50049-EXP).

We thank Nicola Zangrando (HCI Lab. University of Udine) for his help with graphics resources for the EmoPaint app.

## References

Bartsch, A., Vorderer, P., Mangold, R., & Viehoff, R. (2008). Appraisal of emotions in media use: Toward a process model of meta-emotion and emotion regulation. Media Psychology, 11(1), 7-27.

Becker-Weidman, A., & Hughes, D. (2008). Dyadic developmental psychotherapy: an evidence-based treatment for children with complex trauma and disorders of attachment. *Child & Family Social Work*, 13(3), 329-337.

Benyon, D., Gambäck, B., Hansen, P., Mival, O., & Webb, N. (2013). How Was Your Day? Evaluating a Conversational Companion. *IEEE Transactions on Affective Computing*, *4*(3), 299-311.

Besel, L. D., & Yuille, J. C. (2010). Individual differences in empathy: The role of facial expression recognition. Personality and individual differences, 49(2), 107-112.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

Breugelmans, S. M., Ambadar, Z., Vaca, J. B., Poortinga, Y. H., Setiadi, B., Widiyanto, P., & Philippot, P. (2005). Body sensations associated with emotions in Raramuri Indians, rural Javanese, and three student samples. *Emotion*, 5(2), 166.

Broekens, J., & Brinkman, W. P. (2013). AffectButton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies*, 71(6), 641-667.

Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, 1(1), 18-37.

Castellano, G., Villalba, S. D., & Camurri, A. (2007, September). Recognising human emotions from body movement and gesture dynamics. In International Conference on Affective Computing and Intelligent Interaction (pp. 71-82). Springer Berlin Heidelberg.

Chittaro, L., & Sioni, R. (2014). Affective computing vs. affective placebo: Study of a biofeedback-controlled game for relaxation training. *International Journal of Human-Computer Studies*, *72*(8), 663-673.

Chittaro L., Vianello A. (2016). Evaluation of a mobile mindfulness app distributed through on-line stores: a 4-week study, *International Journal of Human-Computer Studies* 86, 63-80.

Christensen, G. T., Maartensson, S., & Osler, M. (2017). The association between depression and mortality–a comparison of survey-and register-based measures of depression. *Journal of Affective Disorders*, 210, 111-114.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20 (1), 37–46.

Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., & García-Crespo, Á. (2011). Using the affect grid to measure emotions in software requirements engineering. *Journal of Universal Computer Science*, 17(9), 1281-1298.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, *40*(2), 5.

De Mantaras, R. L., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., ... & Keane, M. (2005). Retrieval, reuse, revision and retention in case-based reasoning. *The Knowledge Engineering Review*, *20*(03), 215-240.

Deng, J., Zhang, Z., Eyben, F., & Schuller, B. (2014). Autoencoder-based unsupervised domain adaptation for speech emotion recognition. *IEEE Signal Processing Letters*, *21*(9), 1068-1072.

Desmet, P., & Hekkert, P. (2007). Framework of product experience. *International journal of design*, *1*(1), 57-66.

Desmet, P.M.A., Overbeeke, C.J. & Tax, S. J. E. T. (2001). Designing Products with Added Emotional Value: Development and Application of an Approach for Research through Design. *The Design Journal*, 4(1), 32-47.

Ekman, P. (1992a). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169-200.

Ekman, P. (1992b). Are there basic emotions?, *Psychological Review*, 99(3), 550-553

Eleftheriadis, S., Rudovic, O., & Pantic, M. (2015). Discriminative shared Gaussian processes for multiview and view-invariant facial expression recognition. *IEEE transactions on image processing*, *24*(1), 189-204.

Fleureau, J., Guillotel, P., & Huynh-Thu, Q. (2012). Physiological-based affect event detector for entertainment video applications. *IEEE Transactions on Affective Computing*, *3*(3), 379-385.

Girodo, M. (1973). Film-induced arousal, information search, and the attribution process. *Journal of Personality and Social Psychology*, *25*(3), 357.

Gomez, P., Stahel, W. A., & Danuser, B. (2004). Respiratory responses during affective picture viewing. *Biological Psychology*, *67*(3), 359-373.

Griffin, H. J., Aung, M. S. H., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., & Bianchi-Berthouze, N. (2015). Perception and automatic recognition of laughter from whole-body motion: continuous and categorical perspectives. *IEEE Transactions on Affective Computing*, *6*(2), 165-178.

Gruebler, A., & Suzuki, K. (2014). Design of a wearable device for reading positive expressions from facial EMG signals. *IEEE Transactions on Affective Computing*, *5*(3), 227-237.

Gupta, R., Audhkhasi, K., Lee, S., & Narayanan, S. (2016). Detecting paralinguistic events in audio stream using context in features and probabilistic decisions. *Computer Speech & Language*, *36*, 72-92.

Hafner, J., Sawhney, H. S., Equitz, W., Flickner, M., & Niblack, W. (1995). Efficient color histogram indexing for quadratic form distance functions. *IEEE transactions on pattern analysis and machine intelligence*, *17*(7), 729-736.

Harley, J. M., Bouchet, F., & Azevedo, R. (2013, July). Aligning and comparing data on emotions experienced during learning with MetaTutor. In Artificial Intelligence in Education, vol. 7926 of Lecture Notes in Computer Science (pp. 61-70). Springer Berlin Heidelberg.

Harley, J. M. (2016). Measuring emotions: A survey of cutting-edge methodologies used in computer-based learning environment research. In S. Tettegah & M. Gartmeier (Eds.). *Emotions, Technology, Design, and Learning* (pp. 89-114). London, UK: Academic Press, Elsevier.

Healey,  J.A., Picard, R.W., (1998). StartleCam: a cybernetic wearable camera, *Proceedings of the Second International Symposium on Wearable Computers (ISWC 1998)*. IEEE Computer Society, Washington, DC, USA, pp.1–8.

Hill, E., Dumouchel, P., & Moehs, C. (2011). An evidence-based toolset to capture, measure and assess emotional health. *Journal of CyberTherapy and Rehabilitation*, 4(2), 188-191.

Ioannou, S., Gallese, V., & Merla, A. (2014). Thermal infrared imaging in psychophysiology: potentialities and limits. *Psychophysiology*, 51(10), 951-963.

Joormann, J., Siemer, M., & Gotlib, I. H. (2007). Mood regulation in depression: Differential effects of distraction and recall of happy memories on sad mood. *Journal of abnormal psychology*, 116(3), 484-490.

Labouvie-Vief, G., Lumley, M. A., Jain, E., & Heinze, H. (2003). Age and gender differences in cardiac reactivity and subjective emotion responses to emotional autobiographical memories. *Emotion*, 3(2), 115-126.

Li, W., & Xu, H. (2014). Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications*, *41*(4), 1742-1749.

Looney, S. W. (1988). A statistical technique for comparing the accuracies of several classifiers. *Pattern Recognition Letters*, 8(1), 5-9.

Lund, A. M. (2001). Measuring usability with the USE questionnaire. *Usability Interface*, 8(2), 3-6.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International journal of human-computer studies*, *65*(4), 329-347.

Mesquita, B. (2001). Emotions in collectivist and individualist contexts. *Journal of personality and social psychology*, 80(1), 68.

Morris, J. D. (1995). Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response. *Journal of advertising research*, 35(6), 63-68.

Morris, M. E., Kathawala, Q., Leen, T. K., Gorenstein, E. E., Guilak, F., DeLeeuw, W., & Labhard, M. (2010). Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. *Journal of medical Internet research*, *12*(2), e10.

Nardelli, M., Valenza, G., Greco, A., Lanata, A., & Scilingo, E. P. (2015). Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, *6*(4), 385-394.

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences*, *111*(2), 646-651.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions?. *Psychological review*, *97*(3), 315.

Pantic, M., & Patras, I. (2006). Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(2), 433-449.

Ramakrishnan, S. (2012). Recognition of emotion from speech: a review. In S. Ramakrishnan (Ed.) *Speech Enhancement, Modeling and recognition–algorithms and Applications*. Rijeka (Croatia): InTech

Roseman, I.J., Antoniou, A.A., Jose, P.E. (1996) Appraisal determinants of emotions: constructing a more accurate and comprehensive theory, *Cognition and Emotion*, 10(3), 241-277.

Russell, J. A., Bachorowski, J. A., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual review of psychology*, 54(1), 329-349.

Russel, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493-502.

Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the Trait Meta-Mood Scale. *Emotion, disclosure, and health*, *125*, 154.

Seib-Pfeifer, L. E., Pugnaghi, G., Beauducel, A., & Leue, A. (2017). On the replication of factor structures of the Positive and Negative Affect Schedule (PANAS). *Personality and Individual Differences*, 107, 201-207.

Scherer, K. R. (2005). What are emotions? And how can they be measured?. *Social science information*, 44(4), 695-729.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural psychology*, *32*(1), 76-92.

Serratosa, F., & Sanfeliu, A. (2006). Signatures versus histograms: Definitions, distances and algorithms. *Pattern recognition*, 39(5), 921-934.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62 (1): 77–89

Soleymani, M., Asghari-Esfeden, S., Fu, Y., & Pantic, M. (2016). Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, *7*(1), 17-28.

Swain, M. J., & Ballard, D. H. (1991). Color indexing. *International journal of computer vision*, *7*(1), 11-32.

Suvilehto, J. T., Glerean, E., Dunbar, R. I., Hari, R., & Nummenmaa, L. (2015). Topography of social touching depends on emotional bonds between humans. *Proceedings of the National Academy of Sciences*, 112(45), 13811-13816.

Sun, X. and May, A., (2014). Design of the user experience for personalized mobile services. *International Journal of Human Computer Interaction*, 5 (2), 21 - 39.

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of cross-cultural psychology*, 38(2), 227-242.

Vu, K., Hua, K. A., & Tavanapong, W. (2003). Image retrieval based on regions of interest. *IEEE Transactions on knowledge and data engineering*, *15*(4), 1045-1049.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect:

the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063.

Wittig, S., Kloos, U., & Ratsch, M. (2016). Emotion model implementation for parameterized facial animation in human-robot-interaction. *Journal of Computers*, *11*(6), 439-446.

Wu, D., Courtney, C. G., Lance, B. J., Narayanan, S. S., Dawson, M. E., Oie, K. S., & Parsons, T. D. (2010). Optimal arousal identification and classification for affective computing using physiological signals: virtual reality Stroop task. *IEEE Transactions on Affective Computing*, *1*(2), 109-118.

Zagalo, N., Torres, A., & Branco, V. (2005, November). Emotional spectrum developed by virtual storytelling. In *International Conference on Virtual Storytelling* (pp. 105-114). Springer Berlin Heidelberg.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, *31*(1), 39-58.

**Iván García-Magariño** is a PhD Assistant Professor (AyD) of Languages and Information Systems (LIS) in the Department of Computer Science and Engineering of Systems in the University of Zaragoza, Spain, where he is a member of the EduQTech research group (http://webdiis.unizar.es/~ivangmg/). He has authored or co-authored over 80 academic international publications. His major research interests are in multi-agent systems, agent-based simulation, software engineering processes and development of mobile applications. He has received research grants from several organizations such as the Spanish Ministry of Education, Culture and Sport, the Spanish Ministry of Economy and Competitiveness, and the Foundation Antonio Gargallo, and companies like Boeing and the Santander Bank group.

**Luca Chittaro** is full professor of Human Computer Interaction (HCI) in the Department of Mathematics, Computer Science, and Physics of the University of Udine, Italy, where he heads the HCI Lab (http://hcilab.uniud.it). He has authored or co-authored about 190 international publications, and he is an ACM Distinguished Speaker. His major research interests are in virtual reality, mobile HCI, serious games, persuasive technology and their applications in health and safety. He has received research grants from a wide range of organizations, including the US Federal Aviation Administration (FAA), the European Union (EU), the Italian Ministry of University and Research (MIUR), and companies such as the Benetton Group and the Intesa Sanpaolo Bank group.

**Inmaculada Plaza** obtained Degree in Physics and PhD in Electronic Engineering by the University of Zaragoza. She is currently a senior lecturer (TU) in the Department of Electronic Engineering and Communications at this University. She heads the EduQTech R&D&I group (http://eduqtech.unizar.es/en/). She has authored or co-authored over 150 research publications. She is a Senior Member of IEEE, in which she was in charge (Chairwoman) of the Spanish Chapter of the IEEE Education Society (2010-2012). She has been awarded by public and private organizations (http://eduqtech.unizar.es/en/premios/). Her research interests include technology to improve quality of life (mHealth, wearable sensors) and quality in education.

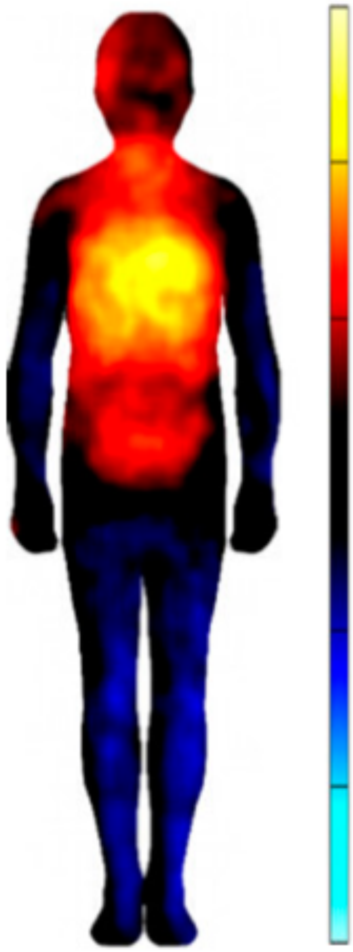

*Fig. 1 Example of a BSM. Activation is represented with the black-yellow color scale, deactivation with the black-light blue color scale*

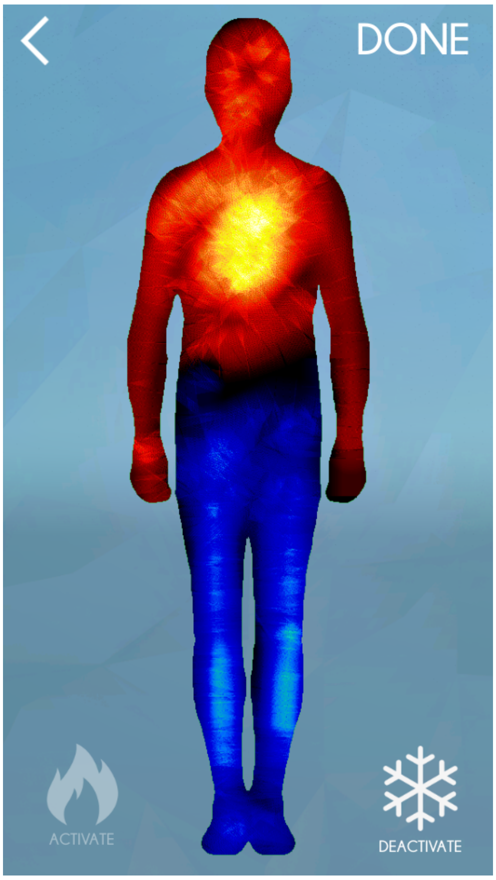


*Fig. 2 Painting interface of the EmoPaint app*

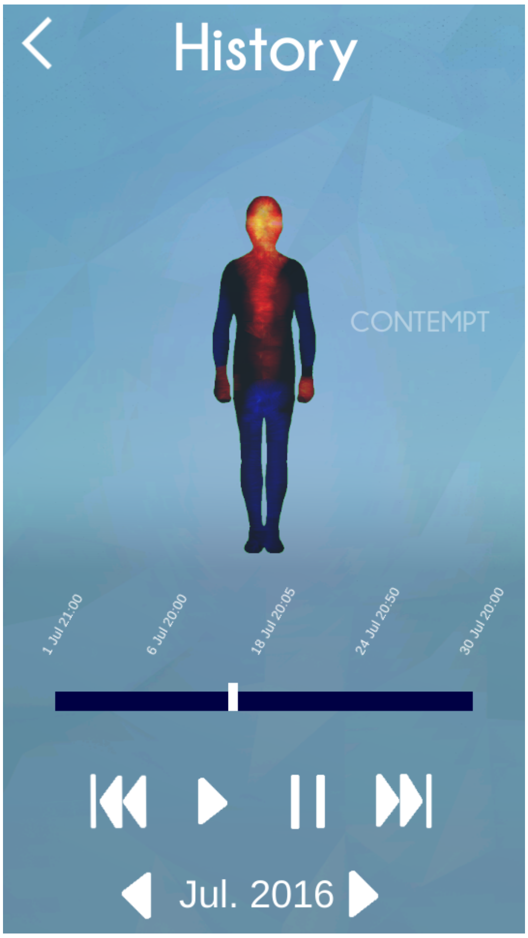


*Fig. 3 Emotion diary based on BSMs*

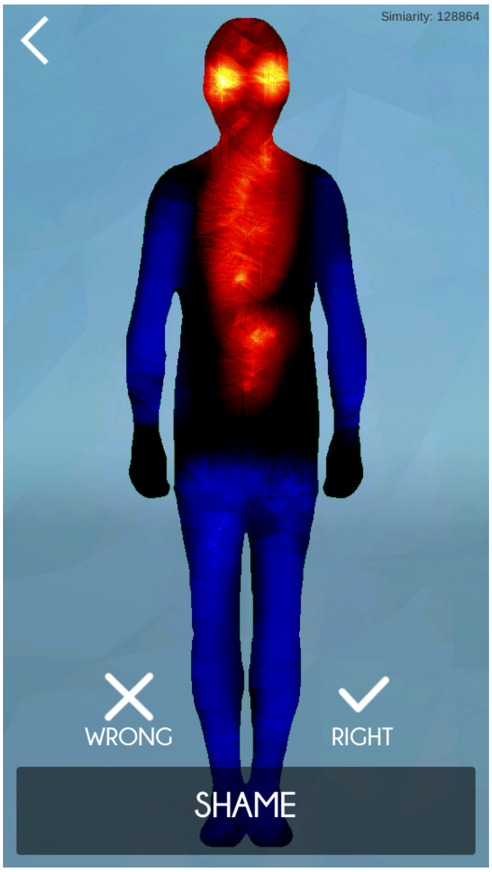


*Fig. 4 Classification of a BSM*

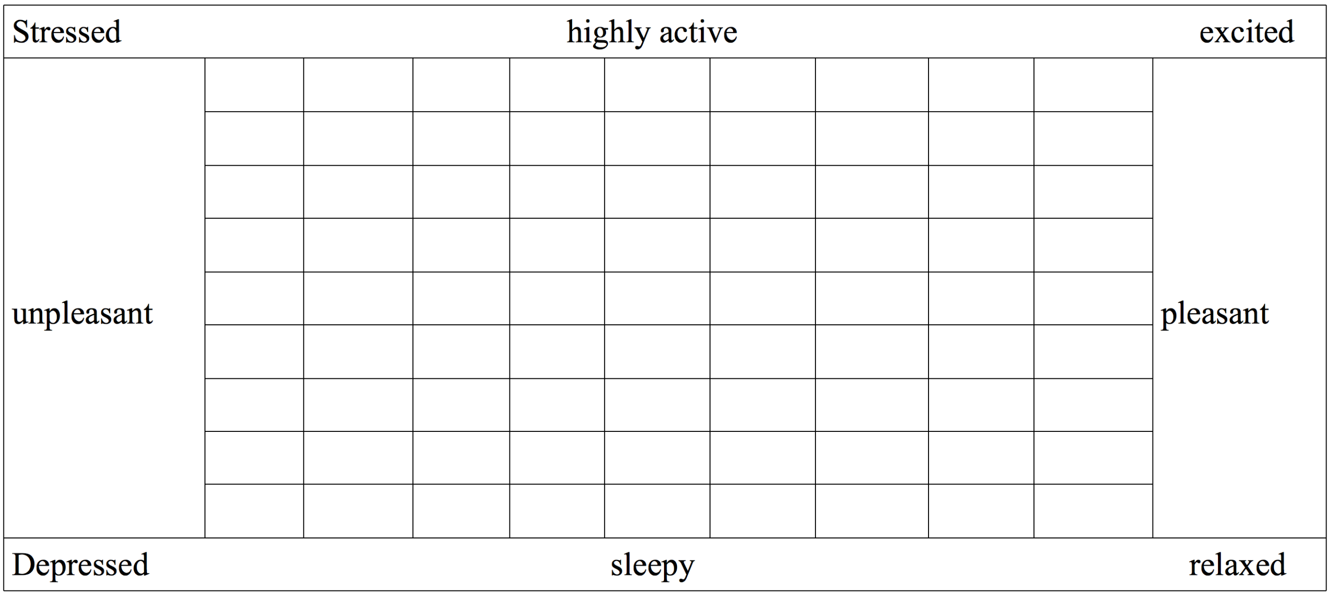


*Fig. 5 Affect Grid*

| Stressed | | | | | highly active | | | | | excited |
|---|---|---|---|---|---|---|---|---|---|---|
| | 7 | 7 | 7 | 14 | **14** | 14 | 11 | 9 | **9** | |
| | 1 | **7** | 7 | 14 | 14 | 14 | 11 | **11** | 9 | |
| | **1** | 1 | 2 | 2 | 14 | 14 | 11 | 11 | **8** | |
| | **6** | **3** | 3 | **2** | 10 | 10 | 8 | 8 | 8 | |
| unpleasant | 13 | **13** | 13 | 10 | **10** | 10 | 8 | 8 | 8 | pleasant |
| | **5** | 5 | 5 | 10 | 10 | 10 | 10 | 10 | 10 | |
| | **12** | 12 | 4 | 4 | 10 | 10 | 10 | 10 | 10 | |
| | 12 | **4** | 4 | 4 | 10 | 10 | 10 | 10 | 10 | |
| | 4 | 4 | 4 | 4 | 10 | 10 | 10 | 10 | 10 | |
| Depressed | | | | | sleepy | | | | | relaxed |

Anger (1), anxiety (2), contempt(3), depression(4), envy(5), disgust(6), fear(7), happiness(8), love(9), neutral (10), pride (11), sadness (12), shame(13), and surprise(14).

*Fig. 6 Association of the Affect Grid points with the 14 categorical emotions by means of the adapted Circumplex model*

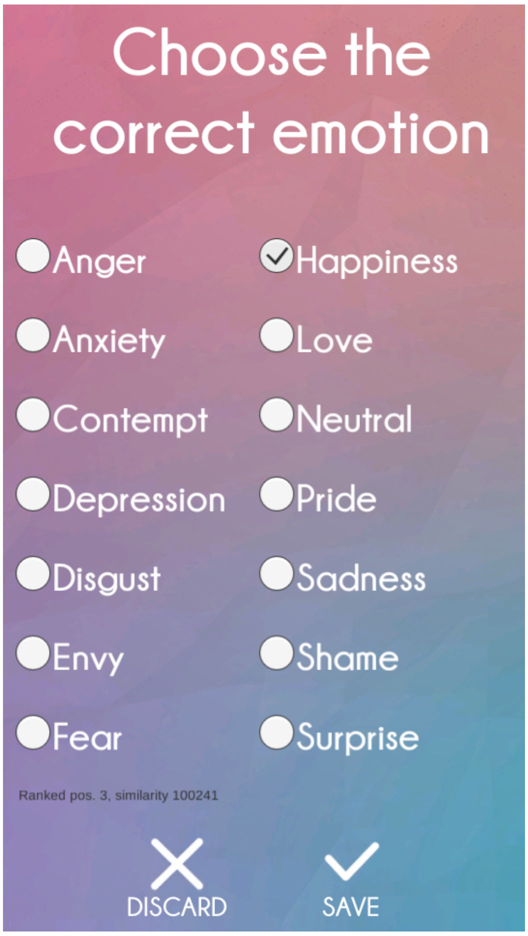


*Fig. 7 Selection of the correct emotion*

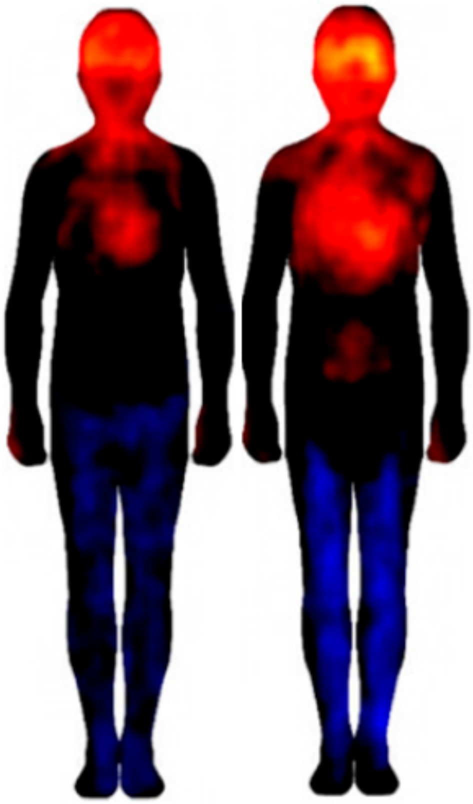

*Fig. 8 BSMs of envy (left) and surprise (right)*