Measuring Emotions: Towards rapid and low cost methodologies

Christian Stickel
Research Unit HCI4MED
Medical University Graz
Auenbruggerplatz 2/V
A-8036 Graz

c.stickel@hci4all.at

Andreas Holzinger
Research Unit HCI4MED
Medical University Graz
Auenbruggerplatz 2/V
A-8036 Graz

a.holzinger@hci4all.at

Alexander Felfernig
Graz University of Technology
Institut for Software Technology
Inffeldgasse 16b/II
A-8010 Graz

alexander.felfernig@ist.tugraz.at

ABSTRACT
Emotions are important mental and physiological states influencing both, perception and cognition, consequently influencing decision making. Emotions are a topic of interest in Human-Computer Interaction (HCI) for some time. Popular examples include stress detection or affective computing [27]. The usage of emotional effects for various applications in Computer Science, especially for recommender systems are of increasing interest. Emotional and affective states represent very personal data and might thus enhance user profiling. This paper shows some possible methods for measuring emotions. The described experiment covers ongoing research on aesthetic parameters that influence emotional and affective states. Two versions of an interface have been designed: differing in colors, contrasts and visual complexity – a 'nice' and an 'ugly' interface. Both interfaces had the same underlying functionality. N=16 test persons were randomly divided in two groups, each using one of the interfaces in 8 tasks. There was one impossible task to frustrate the test persons on pupose. The emotional states of the subjects were captured with pictoral and psychophysiological measures for every task. The results suggest an overall positive influence of the aesthetic variables constituting the 'nice' interface, compared to the results of the 'ugly' interface. Positive effects were found concerning the user's effectiveness, efficiency, frustration tolerance, resp. satisfaction and cognitions towards the interface.

Categories and Subject Descriptors
H.5, D.2, J.3, J.4, K.4

General Terms
Design and Human Factors.

Keywords
Human-Computer Interaction, Usability, Aesthetics, Emotion detection, Emotion recognition

1. INTRODUCTION AND MOTIVATION
Reason or Emotion? This is a long question of debate and meanwhile, measuring emotions has gained much interest in Computer Science [17]. Despite great advances in the area of emotion detection, there are many challenges remaining and many unsolved problems in this field. To date, in computer science, emotion recognition is studied manifold with various techniques using e.g. classification methods including Artificial Neural Networks (ANN), Fuzzy Sets, Support Vector Machines (SVM), Hidden Markov Models (HMM) and Rough Set Theory. However, the recognition rate is still not satisfying. The emphasis on visual methods using face expressions for classification becomes understandable when realizing that every laptop nowadays is by default packed with two simple sensors, which can easily be used for emotion detection - webcam and microphone. The theory and model for using facial expressions for emotion detection originates in psychological research done by Ekman and his group dated back to 1972 and has since then been evaluated and refined [9].

However, there are much more possibilities to measure affective and emotional states, e.g. by capturing signals of responses elicited by the central nervous system (CNS) and the autonomous nervous system (ANS). These signals can provide an insight into the arousal part of an emotion, respectively the physical activation of the body. However arousal is just one component of the complex construct called emotion and it will become clearer in the theoretical part of this paper, how every emotion is, on a subjective level, linked to certain cognitions. These cognitions finally influence the action tendencies of the subject, producing preferences or denial, acceptance or trust, disagreement or suspicion, shaping triggers to make a certain decision. The mechanisms are long known and are used in fields including mass media, advertisement or product design. It is important to note that both desires and actions are inseparable linked to each other. The glue is emotion.

Consequently, it is time to make serious use of this knowledge by integrating it into systems. A simple video from a webcam in combination with an online emotion recognition system might be able to enhance user profiles for recommender systems with invaluable additional details. Computer systems which are able to recognize and react dynamical on certain affective states of the user wield the possibility to create very personal user profiles and also elicit certain desired emotional states [28]. This can be used for empowering the end user, as will be shown in this paper, and on the other hand empowering companies to adjust services and products on a very user-centred level.

2. THEORETICAL BACKGROUND

This chapter will outline some theoretical background, which is fundamental for the presented work. The first part will frame the term emotion and describe important components of an emotion, while the second part covers methods for measuring the components. The third part will describe how to map the measurements, respectively the classification of affective and emotional states.

2.1 EMOTIONS – A DEFINITION?

Research in the scientific literature shows, that there exist a multitude of definitions, models and basic terms for emotions, based on different hypothesis of their origin [25]. Therefore we'll look into some of the most influential ones.

2.1.1 THEORIES OF EMOTION

The *James-Lange theory* of emotion states that an occurring event causes physiological arousal first which is then interpreted as arousal. The interpretation of arousal is necessary to experience an emotion. Without noticing the arousal there's no emotion [14][20]. The *Cannon-Bard theory* argues that, given an event, physiological arousal and emotion at the same time. This theory however disregards the role of thoughts or observable behavior [3]. *Schachter and Singer* state that an event causes physiological arousal first. Then a reason must be found for the occurrence of the arousal, in order to experience and also label an emotion [32]. *Lazarus Theory* argues that an event first elicits a thought, which can be conscious or unconscious, before experiencing arousal and emotions. Lazarus model also led to the idea of appraisal [21][22]. A simplified comparision of the different theories can be seen in the box below.

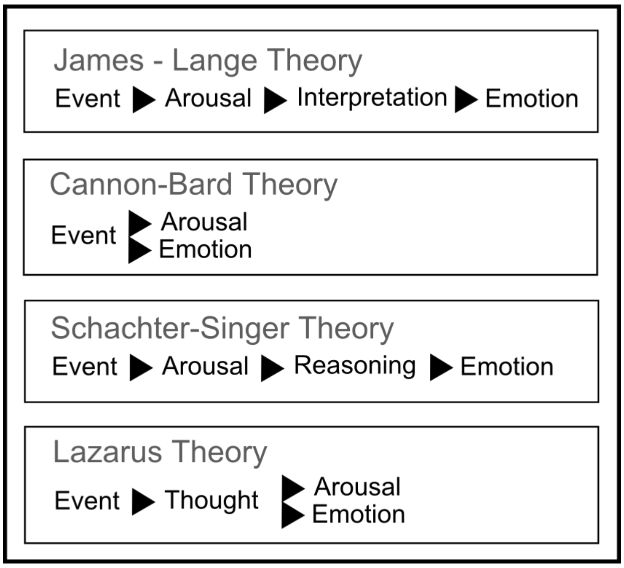

Fig. 1 Popular theories of emotion

2.1.2 BASIC EMOTIONS

The current theory of emotion in neuroscience research argues that the evolution equipped humans with a discrete set of basic emotions [8][26]. Thereby, every emotion is said to be independent. They arise due to the activation of unique neural pathways in the central nervous system (CNS). Further every emotion is said to have a distinct pattern of behavioral, psychological and physiological events. This theory maps every specific emotion to a neural system. However, in order to become more specific, one might ask, which distinct emotions there are. Interestingly there's a big divergence about the number and naming of the fundamental emotions [25]. Fig. 2 shows some sets of basic emotions proposed by several researchers in the field.

Frijda (1986)	desire, happiness, interest, surprise, wonder, sorrow
Tomkins (1984)	anger, interest, contempt, disgust, distress, fear, joy, shame, surprise
Panksepp (1982)	expectancy, fear, rage, panic
Ekman (1982)	anger, disgust, fear, joy, sadness, surprise
Ekman (1982)	anger, disgust, fear, joy, sadness, surprise

Fig. 2 Fundamental emotions (adapted from Ortony, 1990)

2.1.3 AFFECT, MOOD AND EMOTION

The term 'affect' describes an instinctual reaction to events resp. stimuli, which occurs before any cognitive process takes place. Principally an emotion is more complex than an affect and formed by a cognitive process. Affective reactions occur without any extensive perceptual or cognitive encoding and are primary for humans. It's a fast path in the brain that allows making sooner decisions rather than cognitive judgements [41]. Affective states result in affective behaviors, which are observable but neither sufficient nor necessary to characterize emotional states [15][26]. For example, the affective state of anxiety can be felt without any apparent changes in behavior, while an affective behavior like smiling can be elicited without any emotional change. For HCI these affective reactions are valuable information. Measured with the appropriate instruments they provide an insight into fast subconscious reactions of the enduser towards certain stimuli, e.g. a frustrating event like an error message or an aesthetic picture. So far it might also be a good indicator for recommender systems linking affective psychophysiologic changes with certain recommendations.

Moods are also affective states. They are not as directed as an emotions and not as instant occurring as an affect. They can last a day to years and are kind of unfocused and diffuse states. Moods involve 'tone' and 'intensity'. They 'color' an emotion. They further contain a structured set of generalized beliefs concerning the future experience of pleasure and pain [1].

Recent research in HCI uses affective states rather than full blown emotions or moods, because actually it seems that there are no commonly accepted answers, neither on how to model the processes that are causing emotions, nor on classification schemes or dimensions. However for the present work, Lazarus Theory in combination with the Russel's circumplex model of affective states (which will be discussed later in this chapter) proves useful to work with [38].

2.1.4 DEFINITION

Emotions are a psychophysiological process or response which is created every time a conscious or unconscious perception of important changes in the environment or in the physical body appears. It functions in the management of maintaining the balance of information processes and relevant goals in the brain. Every time an event is evaluated as relevant to a goal, an emotion is elicited. The evaluation itself is called appraisal. Positive emotions occur when the goal is advanced; while negative emotions occur when the goal is impeded. Emotions are totally subjective and their core is the expererience of pleasure and pain, which creates the readiness to act in a certain way [11].

2.1.5 COMPONENTS OF EMOTIONS

Emotions consist of several components. A core set of these components was found by the W3C Emotion incubator group as usefull for modeling user emotions in software engineering [40]. See fig. 2 below for an overview on them.

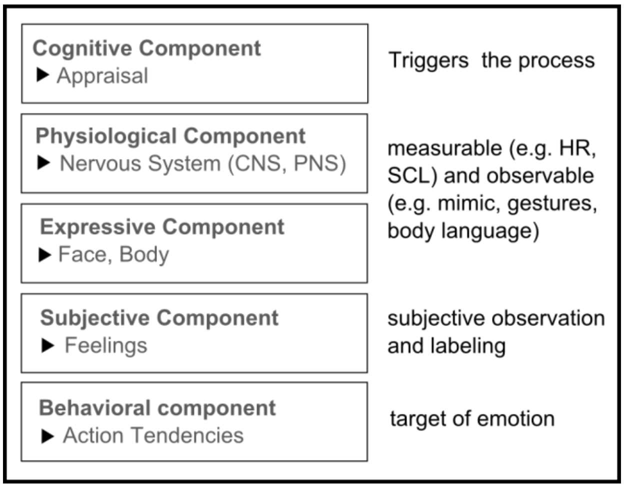

Fig. 3 Components of emotions

2.1.5.1 Cognitive component

Appraisal is the cognitive component of an emotion. It is the process that evaluates all perceptions, resp. occurring situations or events according to the actual values and beliefs of the subject. The process can be triggered consciously or unconsciously [21][22][33]. The cognitive component is measurable by questionnaires and interviews. Interview techniques like the arrow down method can provide an insight into the belief system of subjects. In practice a simple 'like it' button seems to be able to reveal the cognitive component to a usable extend. However depending on the context and the requirements of the user profiling a small list of semantic differentials should provide useful information. The cognitive component also includes the preferences and denials, the so called valence. In the presented study a pictoral method was used to capture this dimension.

2.1.5.2 Physiological component

The physiological component describes all changes of the peripheral nervous system (PNS) e.g. heart rate (HR), skin conductance level (SCL), temperature, blood pressure, respiration or pupillary dilation. This component also includes parameter changes in the central nervous system (CNS) like the neural activity in the brain. The parameter changes of the PNS and CNS are technically measurable with special sensors and devices to some extend and are usually referred to as arousal.

2.1.5.3 Expressive component

The expressive component includes facial expressions, body posture like gestures, movement and also verbal cues like prosody, loudness or voice intonation. The changes in the expressive component are observable, however if using sophisticated methods like computervision or audioanalysis, this component is technically measurable to some extend.

2.1.5.4 Subjective component

The subjective component describes what the subject actually experiences, observes and labels. This is a totally subjective category and hardly generalizable. The subjective component can be measured to some extend with pictoral methods like SAM or emocard. There have been also experiments using a color code for subjective emotion expression.

2.1.5.5 Behavioral component

The behavioral component finally describes action tendencies which have an influence on the motivational state of the subject. Fear for example will most likely create the tendency of avoidance, while joy and happiness will most likely create the tendency of acceptance [11]. The action tendencies can be seen as connector between the result of the appraisal process and the actual observable reaction of the subject. They have a strong influence on the final decision to act. Knowledge about action tendencies can radically improve the quality of a recommender system.

2.2 MEASURING EMOTIONS

As mentioned before there are different methods for measuring each component. The choice of the method will always depend on the nature of the application and the possibilities to acquire the appropriate data. For recommender systems a practical way for emotion aquisition might be to focus on the cognitive and the behavioral component using minimal questionnaires or pictural methods.

2.2.1 PSYCHOPHYSIOLOGIC METHODS

In Biopsychology many physically measurable signals have been found as relevant to emotional or affect responses. These signals can be captured by cameras, microphones, and all kind of other sensors like temperature, infrared, electrodes and so on. Psychophysiologic methods are derived from the physiologic theories. In most cases the observed parameter changes reveal the level of arousal or stress, which is most interesting during an interaction of a user with a certain interface. Tiny affective reactions can reveal distress even before the subject realizes it, as these signals are regulated by the autonomous nervous system (ANS). Psychophysiologic methods include the use of measures such as cardiovascular parameters e.g. heart rate (HR) or heart reate variability (HRV). HR is calculated as number of contractions of the heart in one minute, thus is counted in "beats per minute" (bpm). An adult human heart beats at about 70-75 bpm during a resting period. This rate is varying according to age and fitness among people. When coping with stress the HR increases rapidly. However a better indicator for stress is a constant HRV, which can be calculated from the HR. Another well renowned method is using the skin conductance as indicator of stress e.g. by measuring the skin conductance level (SCL) or the opposite skin conductance response (SCR). Simple stress stimuli are usually followed by a rapid rise of SCL within one to three seconds. SCL depends on the activity of the perspiratory glands, means the more sweat is being produced the more electric current is transported and can be measured.

Another method is observing pupil responses and eye movement e.g. sudden pupil dilations can indicate mental processing resp. mental load. Another possibility is using electroencephalograms (EEG) [27][36]. As mentioned in the introduction videos can be used to detect facial expressions and analyze them for cues of emotions, e.g. by using the widly renowned Facial Action Coding System (FACS) [9][10]. Microphones can be used to record the voice of a subject during an interaction. Changes in the tone of voice are detectable via waveanalysis and can also reveal signs of stress and elevated arousal [34]. The limitation of these methods however is the focus on the physiological and expressional component of an emotion. Without additional measures of the cognitive or subjective component it is very unprecise to derive conclusions on specific emotional or affective states.

2.2.2 QUESTIONNAIRES

A great deal of recent studies using questionnaires or adjective checklists for emotion, mood and affective evaluation comes from consumer behavior literature and human computer interaction. The emotional scales are usually represented as semantic differentials, unipolar scale or free labeling [13][16]. The number of items used, resp. invented in different studies is varying and depends on the context and the focus of the evaluation. While some focus on the mood [31] others focus on affects [4]. Especially mood might be an interesting field for recommnder systems generating user profiles, as its longer lasting and able to reveal emotional dispositions of a subject. A study of Mano [23] focusing on aesthetic sensory aspects of emotional experiences, strives to evaluate the valence as well as the arousal dimension, which is most useful for HCI. This concept is further enhanced with the pictoral methods.

2.2.3 PICTORAL METHODS

A real elegant method for evaluating emotions is the use of pictorial tools. These pictorial tools are primarily based on pictures of faces, body postures or symbolzed dimensions of an emotion. They originate from advertisement and product design; however have been used in HCI and cross-cultural studies as well. The great advantage of this method is the ease of use and relative simplicity of self assessment compared to questionnaires or psychophysiological methods. The most famous of pictoral methods is the so called Self-Assessment Manikin (SAM), which is based on Lang's PAD model of emotion. It maps emotions in three dimensions, which are pleasure (P), arousal (A) and dominance (D) [19]. The SAM method displays every dimension with a simplified graphical representation of a human, evolving along a continuous scale, e.g. the scale for pleasure evolves from a happy smiling face to an unhappy frowning face (see fig.6 on the next page).

Another tool called Product Emotion Measurement Instrument (PrEmo) was recently developed by Desmet [6] in order to enrich the process of product design. It is able to measure 14 distinct emotions, seven pleasant and seven unpleasant. The novel approach is using expressive cartoon animations; thereby each of the emotions is represented by an animation of dynamic facial expressions, body posture and vocal expressions. Each item is accompanied by a three point scale, representing the ratings 'I feel the emotion', 'I feel the emotion to some extend' and 'I do not feel the emotion expressed by the animation'. The subjects are first shown a picture of a prodct and then asked to self-asses their feelings.

2.3 MAPPING EMOTIONS

The data gained from psychophysiological methods, questionnaires or pictoral methods needs to be mapped in some kind of emotion space, in order determine which affective states or emotions occurred. Dimensional models of emotions have a long history in psychology [35]. Especially the two dimensional model of the valence arousal space is attractive to be used, due to reduced dimensionality. Nowadays it is known as circumplex model of affect [29][30]. The model proposes that all affective states emerge from two neurophysiological systems linked to valence and arousal. Every emotion is thereby understood as a linear combination of the both dimensions. Fig. 4 depicts the model.

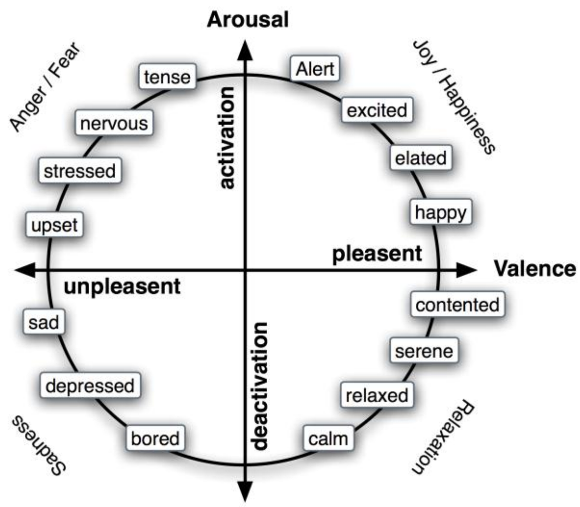

Fig. 4 Circumplex model of affect (ad. from Russel, 1980)

The circumplex model of affect was later adapted and modified by Desmet [7] for product design purposes as can be seen in fig.5. It is also possible to map target qualities of User Experience (UX) into this model [38].

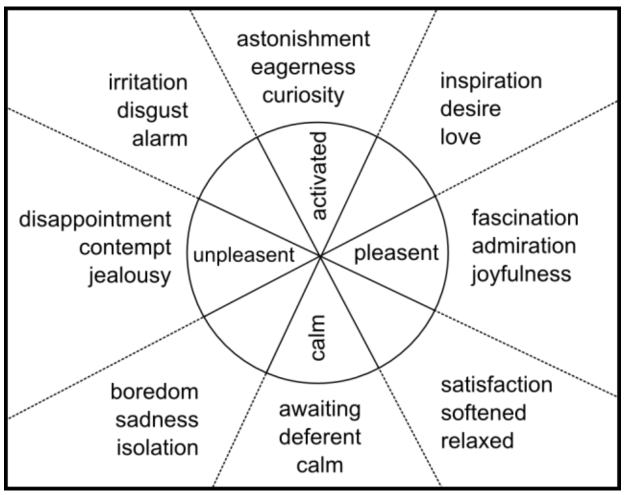

Fig. 5 Circumplex model of affect with product relevant emotions (from Desmet, 2007 & Russel, 1980)

For SAM the underlying space is threedimensional (PAD). Mehrabian and Russel correlated SAM values with an emotion adjective list and established standard values for different emotions on in the PAD space [24].

3. METHODODOLOGY

This chapter shows a practical example of emotion measures from our current work. Aesthetic stimuli are said to take part in shaping affective and emotional states, therefor it's interesting to find out how this applies to interface design. What is the difference between a 'nice' and an 'ugly' interface with the same underlying functionality?

3.1 RESEARCH QUESTIONS

The scope of the described experiment was to examine the influence of aesthetics on the user's affective and emotional states, interactions and cognitions towards a system. The overall hypothesis was that a beautiful 'nice' design has a positive impact on the participants.

H1: Users are more patient and have a higher frustration tolerance towards a 'nice' interface.

H2: Even if errors occur the 'nice' system is higher rated in the final SUS than the same system with an 'ugly' interface, showing that satisfaction and frustration are in connection with the look of the given application.

H3: The user works more efficient with a 'nice' design than with an 'ugly' design.

3.2 EXPERIMENTAL DESIGN

The experiment was primarily designed as thinking aloud test, for collecting as much subjective impressions, resp. cognitions as possible. In such a test the subjects express their thoughts by speaking them out loudly. Usually people are not used to speak out loudly about every single thought and so they tend to stop speaking. A moderator supports the subjects in order to keep up the speaking flow. The moderator also introduces the experiment and aborts tasks in case the participants exceed the given time. There were eight different tasks, which are not necessary to be described in detail here. These tasks were alternating in their level of difficulty, ranging from getting to know the interface and the elements to more complex interactions with different applications. Important to know is that the third task was designed to frustrate the user on purpose, by demanding something that was technically impossible to solve. After each task the subject got a printed questionnaire for evaluating the current emotional state. The questionnaire contained a modified SAM score und an EmoCard score, both pictoral methods, which will be described later. Additional to these, measures of the heartrate (HR) and the skin conductance level (SCL) were recorded, to monitor the arousal, resp. physiological component. At the end of the experiment the subjects rated their overall experience with the system usability scale (SUS).

The platform for this experiment was the Personal Learning Environment (PLE) of Graz University of Technology (http://ple.tugraz.at). The PLE is an advanced E-Learning Environment for students. Technically it's an AJAX webdesktop that connects different online services of the university through widgets, e.g. the LMS or newsgroups. The PLE includes also external webservices like a dictionary and games. All widgets are developed by students of the IT faculty. The PLE is customizable and styleable by CSS, which made it easy to create two interface designs with the same underlying functionality. The development of the 'nice' and the 'ugly' style was done with the "JQuery Theme Roller". Both styles were designed to have the same readability and accessibility. Due to precedent research in aesthetic variables, the differences between the designs were based on the parameters color, contrast and visual complexity [37][39].

The test was performed in under laboratory conditions using a 13,3" wide screen notebook with a native resolution of 1366x768 pixels and an attached mouse. Several cameras recorded the face and body postures of the subjects. Techsmith Camtasia Studio (http://ww.techsmith.com/camtasia.asp) was used for screen recordings for later analysis. 16 participants (eight male, eight female) took part in the experiment, with the age ranging from 18 to 30. N=16 is a typical study size for this kind of usability research - it was not the aim of this research to make a large scale statistical study, instead to gain insight into the related issues. All participants belonged to the target group of students and had no or just few experience with the PLE. They were randomly split in two groups. Eight participants did the test on the 'ugly' version of the interface while the the other eight tested the 'nice' design of the PLE.

3.3 METHODS

The experiment was conducted as "Thinking Aloud Test" in combination with SUS, SAM, EmoCards and additional psychophysiological measures. The following paragraphs describe these methods.

3.3.1 SYSTEM USABILITY SCALE

The System Usabiltiy Scale (SUS) is a simple questionnaire with ten questions, originally developed in 1986 by John Brooke [2]. The questions cover the measurement of the usability of a system. The main aspects of the questions are effectiveness, efficency and satisfaction. The answer gets a value for the evaluation. The sum of every score is multiplied by 2.5. The range for the values is from 0 to 100. The advantage of this score is a normalized compareable value for the usability of a system.

3.3.2 SAM

The Self Assessment Manikin (SAM) is originally a three dimensional rating system based on pleasure (valence), arousal and dominance [19]. In this experiment we just used the dimensions for valence and arousal. We also used a seven point scale instead of the original nine point scale to reduce the decision time for the subjects. Fig. 6 shows the used graphical representaion.

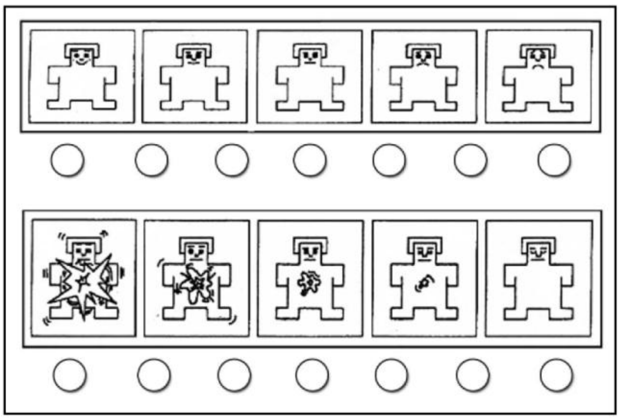

Fig. 6 Modified SAM score used in the experiment

The upper scale evaluates the valence, while the lower evaluates the arousal. Both of the scales have seven different values. For the evaluation the different values got numbers from -3 to 3, in order to map them into the valence-arousal space.

3.3.3 *EMOCARDS*

The Emocards are like the SAM Score a pictoral method to evaluate the emotions of users. They originate from Desmet as predecessor of the PrEmo [5]. There are eight different cartoon faces to represent an emotion. For every value there are two faces, a male and a female face. The subject chooses one of the faces which identifies most likely the current emotional state. The subject does not know the excact meaning of the faces, as there's no labelling. The Emocards had following eight categories (fig.5): - excited neutral - excited pleasant - average pleasant - calm pleasant - calm neutral - calm unpleasant - average unpleasant - excited unpleasant The categories are like the SAM score based on pleasantness (valence) and arousal. An advantage of this method is the simplicity and quickness to learn how to use these Emocards [5].

3.3.4 *PSYCHOPYSIOLOGICAL MEASURES*

As additional measures of the physiological component of emotion, the skin conductance level (SCL) and the heartrate (HR) of the subjects were recorded. For the SCL and HR recordings we used a Lightstone device from the wild divine project (http://www.wilddivine.com). It acquires the data with non-invasive sensors on the fingertips. This however interferes with tasks that depend on extensive keyboard input, so the tasks were designed with minimal use of the keyboard. The HRV was later on calculated from the raw HR data.

4. RESULTS

The analysis of the SAM scores shows that the average values for valence are higher for the 'nice' group as can be seen in fig.8. Although the valence curves correlate to some extend, the 'ugly' group rated after the 'frustration' more negative. Most interestingly the overall values of the arousal score are higher for the 'ugly' group, see fig.9.

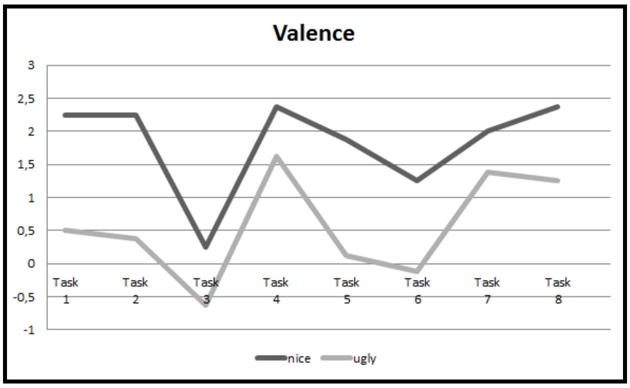

Fig. 7 Average SAM valence is higher for the 'nice' group

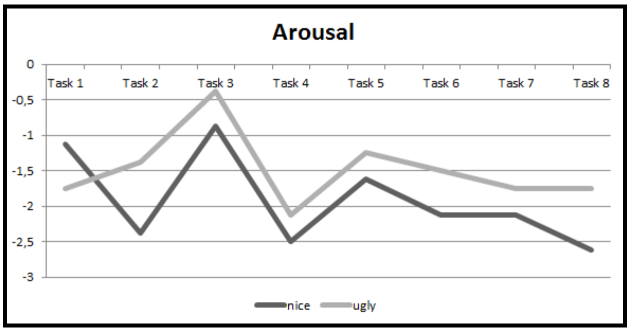

Fig. 8 Average SAM arousal is higher for the 'ugly' group

The results of the Emocards were counted as votes for a certain category of emotional state (fig.5). All together there have been 64 votes per group. The summarized category votes over all tasks show significant differences between the 'nice' and the 'ugly' group.While for the nice design the majority chose pleasent and calm-pleasant emotional states, the results for the ugly design show more variance. Fig. 9 depicts these results, thereby every field represents 5 votes (for the data visualization the results were rounded). Mapping the SAM scores into this model shows almost the same results.

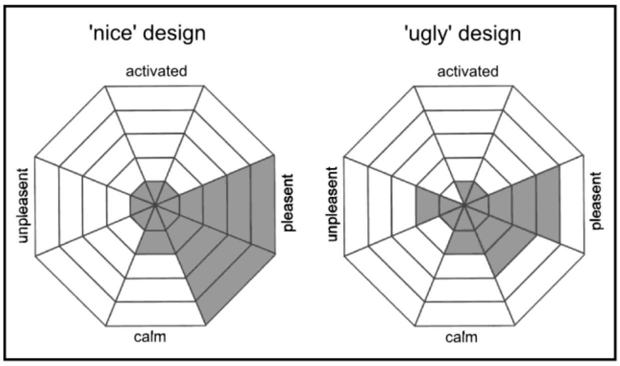

Fig. 9 Emocards show more positive emotions for nice design

The System Usability Scale (SUS) questionnaire has 10 items, with a hidden value between 0 and 4. These values are summed up and multiplied with 2.5 for every subject. The final score is the average over all subjects, with a maximum of 100. Fig. 10 shows the comparison between the ugly and the nice design for every single question of the SUS questionnaire (Q1 to Q10). For further details on the questions have a look at Brooke's work, which is available for free [2].

Interesting here is that the 'nice' design is always evaluated better or equal than the 'ugly' design, although both groups experienced the same underlying system, functionality and the same frustrating task. The overall score of the SUS depicts a better usability for the nice design (83/100) compared to the ugly design (66/100).

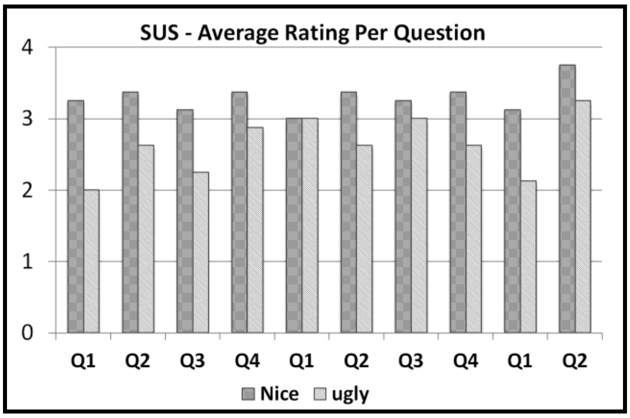

Fig. 10 The 'nice' interface got an overall higher SUS rating

The overall averaged task times show that the 'nice' group had significantly shorter task times, as can be seen in fig.11. The task times sum up to an average of 5 minutes more for the 'ugly' group in the whole trial. Compared to the predefined maximum trial time the 'ugly' group needed 14% more time than the 'nice' group. Looking at the frustration task times, we found that the 'nice' group had more patience than the 'ugly' group, before stating that the task is impossible to solve. The 'nice' interface got in average almost 40 seconds more attention to solve the task – and attention is a key for a usable system.

Effectiveness describes the ability to solve a task to some extend. For the dimension of effectiveness we found, that the 'nice' group had overall 3 aborted tasks, while the 'ugly' group had 7 aborted tasks. Efficiency is the product of effectiveness and time. A high efficiency describes the best effectiveness in the shortest time. Regarding the results from the task times and the effectiveness, the 'nice' group had a higher efficiency.

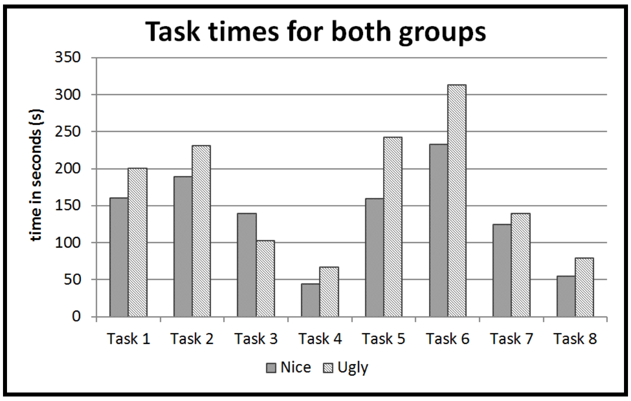

Fig. 11 Average task time for the 'ugly' group is higher

From the psychophysiological measures we found that especially the HRV of the subjects correlates to some extend with their SAM arousal scoring. Finally the average SCL values of both groups differ significantly, as can be seen in fig.12, indicating a higher stress level in the 'ugly' group.

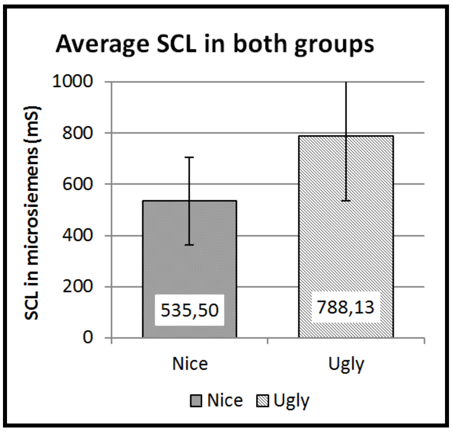

Fig. 12 Average SCL was higher for the 'ugly' group

5. DISCUSSION

The overall results clearly show that the 'nice' design had a positive effect on the participants. Regarding the three hypothesis, we found that the group using the 'nice' design had more patience in the frustration task [H1]. They gave it an average of 40 more seconds to find a solution. The results of the SUS for the 'nice' design show very good scores for usability, while the score of the ugly design is significantly lower, although both interfaces had the same functionality [H2]. Regarding effectiveness and efficiency we found that the group using the 'nice' interface was much faster and less error-prone [H3]. The Emocard method revealed general positive emotions towards the 'nice' interface, while the scores for the 'ugly' interface show a more diverse distribution of emotions. The structure of the designed difficulty of the tasks can clearly be seen in the valence and the arousal of the SAM results. The first two tasks had an easy introductory reason. The third task was designed to frustrate the subjects, while the fourth task was designed to be very easy in order to lift the subjects' spirit again. Task 5 and 6 were challenging, while Task 7 and 8 were easy again. Looking on the graphs of arousal and valence, as well as on the psychophysiological measures, the sequence of these tasks is clearly visible, revealing episodes of stress and dissatisfaction. The subjects were moderately stressed and dissatisfied with the first two tasks, more so with the frustration task and less with the easy fourth task. The challenging tasks increased stress and dissatisfaction. The seventh and eighth tasks were more satisfying and less stressful than the tasks before. A look at the overall average SCL values shows that the 'ugly' group was overall more stressed than the 'nice' group.

6. CONCLUSION AND FUTURE WORK

A most interesting fact shown in this experiment is how aesthetic interface variables can influence the endusers preferences, frustration tolerance, efficiency and emotional states in a positive way. This is fundamental for any decsion support systems, i.e. recommender systems. We presented different methods to measure emotions and showed their application in a practical example and conclude that emotional data of endusers measured in usability tests can help to improve user interfaces. Thinking beyond Usability Engineering, affective and emotional data reveal and include very personal and subjective patterns that can very well be integrated into recommender systems, which either deal with emotionally charged content like music or movies, or improve recommendations by getting a clue of unconscious affective reactions concerning denial or acceptance.

7. ACKNOWLEDGEMENT

We cordially thank the following students of the lecture 706.046 'AK Human-Computer Interaction: Applying User Centered Design' at Graz University of Technology for their enthusiasm and hard work in preparing of the material and carrying out the experiment under supervision: Thomas Altmann, Angelika Moore, Philipp Rausch and Kurt Schlager.

8. REFERENCES

[1] E. A. Blechman. *Moods, Affect, and Emotions*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1990.

[2] J. Brooke. SUS - A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland. *Usability Evaluation in Industry*, 189-194. Taylor and Francis, London, 1996.

[3] W. B. Cannon. The James-Lange theory of emotion: A critical examination and an alternative theory. *American Journal of Psychology*, 39:10-124, 1927.

[4] A. Chaudhuri, R. Buck. Affect, Reason, and Persuasion. *Human communications research*, 21(3):422-440, 1995.

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

[6] P.M.A. Desmet. Measuring emotion; development and application of an instrument to measure emotional responses to products. In M.A. Blythe, A.F. Monk, K. Overbeeke, & P.C. Wright, *Funology: from usability to*

enjoyment, 111-123. Kluwer Academic Publishers, Dordrecht, 2003.

[7] P. M. A. Desmet, P. Hekkert. Framework of product experience. *International Journal of Design*, 1(1): 57-66, 2007.

[8] P. Ekman. An argument for basic emotions. *Cognition and Emotion*, 6:169 –200, 1992.

[9] P. Ekman, W.V. Friesen, P. Ellsworth. *Emotion in the human face: Guideline for research and an intergration of findings*. Pergamon Press, New York, 1972.

[10] A. Fridlund, P. Ekman, H. Oster. Facial expresions of emotion. In A. Siegman & S. Feldstein (Eds.), *Nonverbal behavior and communication*, pages 143-224. Erlbaum, Hillsdale, NJ, 1987.

[11] N. H. Frijda. *The emotions*. Cambridge University Press, Cambridge (UK), 1986.

[12] X. Gong, et al. Expression Detection Based on a Novel Emotion Recognition Method. *International Journal of Computational Intelligence Systems*, 4:44-53, Feb 2011.

[13] M. Hassenzahl, A. Platz, M. Burmester, K. Lehmer. Hedonic and Ergonomic quality aspects determine a software appeal. *Proceedings of conference on human factors in computing system CHI'2000*, DenHag, pages 201-208. 2000.

[14] W. James. What is emotion?, *Mind*, (ix): 189, 1884.

[15] J. Kagan. Behavioral Inhibition as a temperamental category. In R. J. Davidson, K. R. Scherer, & H. Hill Goldsmith, *Handbook of affective sciences*, pages 320-331. Oxford University Press, New York 2003.

[16] J. Kim, J.Y. Moon. Designing towards emotional usability in customer interfaces trustworthiness of cyber-banking system interfaces. *Interacting with computers*, 10:1-29, 1998.

[17] X. L. Kong, Y. Yang. Measuring Emotions in Interactive Contexts. In Y. H. Pan, et al., *E-Business, Creative Design, Manufacturing – Caid & Cd'2009*. 526-530. IEEE, New York, 2009.

[18] P. J. Lang. Behavioral treatment and biobehavioral assessment: Computer applications. In J. Sidowski, J. Johnson, & T. Williams, *Technology in mental health care delivery systems*. 119-137. Ablex, Norwood, 1980.

[19] P.J. Lang. *The cognitive Psychophysiology of emotion: anxiety and the anxiety disorders.* Lawrence Erlbaum, Hillsdale, NJ, 1985.

[20] C. G. Lange. *Ueber Gemuethsbewegungen*, 3:8, 1887.

[21] R.S. Lazarus. *Psychological stress and the coping process*. McGraw-Hill, New York, 1966.

[22] R.S. Lazarus. *Emotion and Adaptation*. Oxford University Press, New York, 1991.

[23] H. Mano. Assessing emotional reactions to TV ads: a replication and extension with a brief adjective checklist. *Advances in consumer research*, 23:63-69, 1996.

[24] A. Mehrabian, J. Russel. Evidence for a Three-Factor Theory of emotions. *Journal of Research in Personality*, 11:273-294, 1977.

[25] A. Ortony, T.J. Turner. What's basic about emotions? *Psychological Review*, 97: 315-331, 1990.

[26] J. Panksepp. *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press, New York, 1998.

[27] R. Picard. *Affective computing*. The MIT press, Cambridge, 1998.

[28] R. Picard, J. Klein. Computer that recognises and responds to user emotion: theoretical and practical implications. *Interacting with computers*, 14:141-169, 2002.

[29] J. A. Russell. Core affect and the psychological construction of emotion. *Psychological Review*, 110:145-172, 2003.

[30] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 1980.

[31] C. Rusting, R.Larsen. Moods as sources of stimulation: relationships between personality and desired mood states. *Personal individual differences*, 18(3):321-329, 1995.

[32] S. Schachter, J. E. Singer. Cognitive, social and physiological determinants of emotional states. *Psychological Review*, 69:379-399, 1962.

[33] K.R. Scherer. Appraisal Considered as a Process of Multilevel Sequential Checking. In K. R. Scherer, A. Schorr & T. Johnstone, *Appraisal Processes in Emotion: Theory, Methods, Research*, pages 92-120, Oxford University Press, 2001.

[34] K.R. Scherer. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40:227-256, 2003.

[35] H. Schlosberg. The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology*, 44:229-237, 1952.

[36] C. Stickel, J. Fink, A. Holzinger. Enhancing Universal Access – EEG based Learnability Assessment. In C. Stephanidis, *Universal Access to Applications and Services*. Lecture Notes in Computer Science (LNCS 4556), pages 813-822, Springer, 2007.

[37] C. Stickel, K. Maier, M. Ebner, A. Holzinger. The Modelling of Harmonious Colour Combinations for improved Usability and User Experience (UX). In V. Luzar-Stiffler, V.H. Dobric, Z. Bekic, *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces*, pages 323–328. IEEE, New York, 2009.

[38] C. Stickel, M. Ebner, S. Steinbach-Nordmann, G. Searle, A. Holzinger. Emotion Detection: Application of the Valence Arousal Space for Rapid Biological Usability Testing to Enhance Universal Access. In C. Stephanidis, *Universal Access in Human-Computer Interaction*, Lecture Notes in Computer Science (LNCS 5614), pages 615–624. Springer, Berlin, Heidelberg, New York, 2009.

[39] C. Stickel, M. Ebner, A. Holzinger. The XAOS Metric: Understanding Visual Complexity as measure of usability. In G. Leitner, M. Hitz, A. Holzinger, *Human-Computer Interaction in Work & Learning, Life & Leisure*, Lecture Notes in Computer Science (LNCS 6389), pages 278-290. Springer, Berlin, Heidelberg, New York, 2010.

[40] W3C Incubator Group Report 10 July 2007. http://www.w3.org/2005/Incubator/emotion/XGR-emotion visited last on 2011/07/29

[41] R. B. Zajonc. Feelings and Thinking: Preferences Need No Inferences. *American Psychologist*, 35(2):151-175, 198