# The Impact of Gender and Sexual Hormones on Automated Psychobiological Emotion Classification

Stefanie Rukavina, Sascha Gruss, Jun-Wen Tan, David Hrabal, Steffen Walter, Harald C. Traue, and Lucia Jerg-Bretzke

University Clinic for Psychosomatic Medicine and Psychotherapy, Medical Psychology, Ulm
Stefanie.Rukavina@uni-ulm.de

**Abstract.** It is a challenge to make cognitive technical systems more empathetic for user emotions and dispositions. Among channels like facial behavior and nonverbal cues, psychobiological patterns of emotional or dispositional behavior contain rich information, which is continuously available and hardly willingly controlled. However, within this area of research, gender differences or even hormonal cycle effects as potential factors in influencing the classification of psychophysiological patterns of emotions have rarely been analyzed so far.

In our study, emotions were induced with a blocked presentation of pictures from the International Affective Picture System (IAPS) and Ulm pictures. For the automated emotion classification in a first step 5 features from the heart rate signal were calculated and in a second step combined with two features of the facial EMG. The study focused mainly on gender differences in automated emotion classification and to a lesser degree on classification accuracy with Support Vector Machine (SVM) per se. We got diminished classification results for a gender mixed population and also we got diminished results for mixing young females with their hormonal cycle phases. Thus, we could show an improvement of the accuracy rates when subdividing the population according to their gender, which is discussed as a possibility of incrementing automated classification results.

**Keywords:** emotion classification, gender, hormonal cycle, heart rate, facial EMG.

## 1 Introduction

Living in the age of information technology (IT), everyone is nowadays surrounded by technical systems, spreading in all corners of daily living. Users tend to show patterns of behavior when they communicate with technical surfaces including emotional components, generally observed within human-human-interactions [1]. User Interfaces are developed to improve the interaction between humans and the system, however an improvement presumes the identification of the user`s emotion and disposition particular in stressful situations or with people from vulnerable populations. Within

M. Kurosu (Ed.): Human-Computer Interaction, Part V, HCII 2013, LNCS 8008, pp. 474–482, 2013.

the research field of affective computing the aim is to make technical systems more empathetic for user emotions and dispositions presuming the possibility of reliable and valid emotion classification.

Emotions can be described by a three dimensional concept with the dimensions of valence, arousal and dominance (VAD) [2]. Valence reflects the pleasantness of a stimulus, whereas arousal refers to the emotional activation. Because dominance is often found to correlate with valence we used in this study only the two-dimensional theory and did not consider dominance further [3]. Psychophysiological changes occur during emotional behavior and are part of emotions per definition. Therefore, one possibility to recognize emotional behavior is using these psychophysiological parameters. Other ‘emotional channels’ could be speech (semantics, pitch) [4] or mimic expressions [5]. As a disadvantage speech and even mimic expressions can be willingly influenced, whereas psychophysiological behaviors are generally autonomic-regulated and as a second advance they are constantly available. Interestingly, in the field of interactive systems and affective computing the impact of gender has been disregarded, although from a psychophysiological perspective gender and accordingly the menstrual cycle must be a crucial co-factor. To our knowledge, only a few classification studies considered gender as important and used gender specific classification vectors [6-7]. Extending this topic, the hormonal cycle has been shown to result in different emotional responding between young females [8-11]. Therefore, the consideration of differences between males and females can improve the identification of emotions using psychophysiological parameters and should be investigated more precisely.

In the current study we used a Support Vector Machine (SVM) classification based on 5 heart rate parameters (RMSSD, mean NN, SDNN, mean HR, SD HR) and 2 facial EMG parameters (mean amplitude of Corrugator supercilii and Zygomaticus major). The combination of these two physiological parameters has been chosen because they are often found to describe the valence dimension e.g. [12] [13]. The analysis about the influence of gender and especially hormonal cycle effects concerning classification results is the first study done to our knowledge in the field of affective computing.

## 2 Material and Methods

For emotion elicitation we used pictures of the International Affective Picture System and Ulm pictures [14] [15]. All pictures were rated according to the emotional dimensions valence, arousal and dominance, however we focus only the VA-dimensional concept. The valence dimension can be separated roughly in positive (High Valence (HV), negative (Low Valence (LV) and neutral quadrants (see Fig. 1).

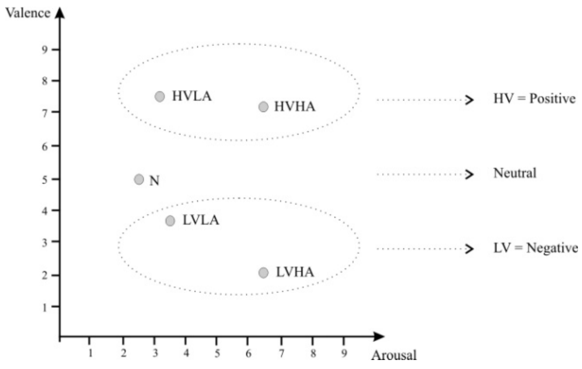


**Fig. 1.** Valence - Arousal concept with the ratings of the used stimuli

According to their arousal they can be distinguished between high and low arousal. To intensify the emotion induction we used a prolonged presentation according to [16-17]. One stimuli sequence (picture block) consisted of 10 pictures with similar valence and arousal ratings, with each picture being displayed for 2 sec without a pause between the following pictures. In total, 100 pictures were displayed in terms of 10 picture blocks, including 5 different affective states (see Fig. 2).

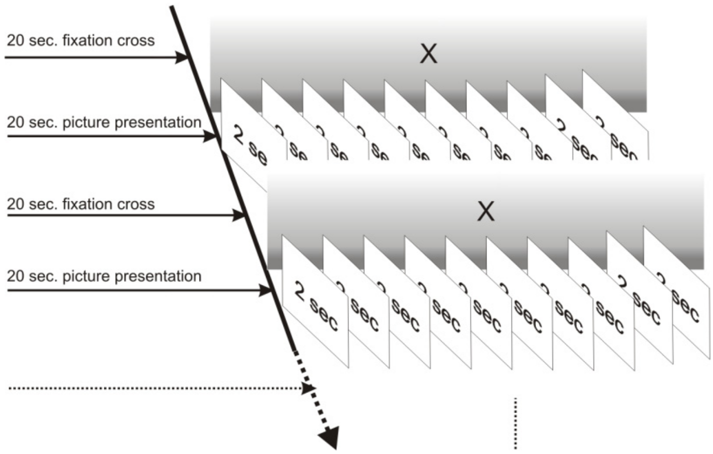


**Fig. 2.** Stimuli presented in a blocked version, with each block consisting of 10 pictures, shown without a pause within between

## 3 Subjects

All right-handed subjects we recruited over bulletins scattered around the University of Ulm and the University of Magdeburg for participation in our experiment.

They were healthy and had normal vision or corrected normal vision. The experiment was designed and implemented according to the ethical guidelines of the University of Ulm (Approval by the Ethical Committee #245/08-UBB/se).

Our sample in total n= 122 subjects was subdivided in different subgroups according to different objectives (to measure gender and menstrual cycle differences as well as age differences). We recruited 73 young participants (n=50 females; n=25 follicular (Ø age=23.72 years, SD=3.52); n=25 luteal women (Ø age = 25.88 years, SD= 4.94)) and 49 older subjects (n=28 females (Ø age = 61.5 years, SD=6.14) and n=21 males (Ø age = 68.29 years, SD=6.40)). For measuring the influence of the menstrual cycle, all young females took part at a specific day according to their menstrual cycle phase. For further information about determining the menstrual cycle phase and recruiting young females please see [18].

## 4 Procedure

All physiological signals were recorded with a NeXus-32 (*NeXus-32, Mind Media, The Netherlands)* and the trigger data was recorded with the software Biobserve Spectator (*BIOBSERVE GmbH, Germany*).

The heart rate information was measured with a BVP (blood volume pulse)-sensor, which was attached to the left middle-finger of every right handed subject and measured the blood volume running through the blood vessels via infrared light within each heart period (plethysmography).

The facial EMG signals were captured by using bipolar miniature silver/silver chloride (Ag/AgCl) skin electrodes with a 4 mm diameter. Both electrodes were placed on participants' left *corrugator supercilii* and *zygomaticus major* muscle regions, according to the guidelines for EMG placement recommended by Fridlund and Cacioppo [19].

Previous to the start of each experiment, each biosignal was visualized with the BioTrace software (appertained to the NeXus-32) and corrected to avoid bad signals or other influences.

## 5 Signal Processing and Feature Extraction

In total, 19 subjects had to be excluded, either because of technical problems (n=2) or due to a large amount of artifacts within the heart rate recordings (n=13). If the HR during one stimuli block contained artifacts, the whole subject was removed. Additionally, the final EMG subject group was conformed to the final HR population, thus 4 subjects had to be excluded, which lead to a total population of n=53 young subjects (n=37 females) and n=36 older subjects (n=21 females). However, for this gender study only the younger participants have been used for the analysis.

**a) Heart Rate**

All parameters were calculated after the offline identification of the Inter-Beat-Interval of the following NN-intervals. A Matlab script extracting these intervals almost automatically was developed, however every signal segment was displayed on the screen for visual correction and for having the chance to correct the identification points of the NN-intervals or to delete the signal segment for excluding artifacts. We extracted the following parameters of the time-domain of the heart rate signal:

Mean NN [ms], SDNN [ms], mean HR [bpm], SD HR [bpm], RMSSD [ms].

$$\text{Mean NN [ms]: } mean\ NN = \frac{1}{n}\sum_{k=1}^{N} IBI_k \quad (1)$$

$$\text{SDNN [ms]:} SDNN = \sqrt{\frac{1}{n}\sum_{k=1}^{N}(NN_k - \overline{NN})^2} \quad (2)$$

$$\text{RMSSD [ms]: } RMSSD = \sqrt{\frac{1}{n}\sum_{j=1}^{N}(NN_{j+1} - \overline{NN}_j)^2} \quad (3)$$

Mean HR [bpm] and SD HR [bpm] were calculated from the mean NN. These parameters served as the initial basis for different classification topics. We did not use frequency domain features, due to problems with the validity of short HRV recordings [20].

**b) Facial EMG**

The raw facial EMG of the *corrugator supercilii* and *zygomaticus major* were offline filtered by a 20–250 Hz band-pass Butterworth filter (order = 4) to exclude motion related components and an adaptive filter was applied to deal with 50 Hz power line interference [21]. The signals were then rectified and smoothed by the root mean square (RMS) technique with a 125 ms sliding window. Facial EMG changes were derived from subtracting baseline activity (i.e., the mean of the RMS of two seconds before each picture block onset) from the respective picture block viewing periods (i.e., the mean of the RMS). Subsequently, we standardized (i.e., Z score) EMG changes within each participant and within each site (according to [22]) to remove variability that might exist in different stimulus conditions. This was done to directly compare the signals from the distinct groups.

All the processing and analyses were conducted using the MATLAB software package (*version R2009a, Mathworks Inc., USA*). Every classification was performed with the data mining software RapidMiner (http://www.rapid-i.com).

# 6 Results

At first, the five extracted raw heart rate features (mean HR, SD HR, mean NN, SD NN and RMSSD) were used to classify between the valence dimensions positive vs. negative. For each classification we used a 10fold validation (see Table 1).

**Table 1.** Classification results with the use of 5 heart rate features

| Group | classes | HR | Acc. [%] |
|---|---|---|---|
| follicular | pos/neg | 5 | 60.88 |
| luteal | pos/neg | 5 | 59.38 |
| young females | pos/neg | 5 | 54.19 |
| young males | pos/neg | 5 | 58.81 |
| gender mixed | pos/neg | 5 | 53.90 |

As you can see in Table 1 the classification rates differ in respect to gender. For example, the valence dimension could be identified within males for 58%, whereas the accuracy for females was about 54%. Interestingly, the classifications separately for the specific menstrual cycle phases were higher than the classification for all young females. The same effect can be seen in mixing gender, where only 53.9% accuracy rate could be achieved.

However, these classification results were still not satisfying, considering the chance level of about 50% for two classes. Therefore, we calculated for another classification each heart rate parameter again but baseline corrected, which reflected the cardiac reactivity compared to the baseline situation, see Table 2 (left third). Before we tried the combination of heart rate and facial EMG features, we classified the data with solely EMG. The classification accuracy was even worse, except for the male group. Afterwards, the combination of both feature sets were tested for each subgroup but the results differed for each group. Whereas the accuracy for the follicular group was improved, the accuracy for the luteal and male group got worse.

**Table 2.** Comparison of different classifications between positive and negative with different feature combinations (left column: HR_change features; middle: EMG features; right: combination of both features)

| Group | HR | Acc. [%] | EMG | Acc.[%] | HR+EMG | Acc. [%] |
|---|---|---|---|---|---|---|
| follicular | 5 | 60.38 | 2 | 54.40 | 7 | 64.51 |
| luteal | 5 | 59.05 | 2 | 63.14 | 7 | 59.76 |
| Young females | 5 | 54.10 | 2 | 59.46 | 7 | 52.74 |
| Young males | 5 | 60.83 | 2 | 69.49 | 7 | 63.21 |
| Gender mixed | 5 | 53.07 | 2 | 60.83 | 7 | 55.38 |

In the last step, we classified the valence dimension again, but we used the raw heart rate features and combined them with the EMG data (see Table 3).

These two feature sets resulted in the best classification results, with accuracy rates above 60% for each subgroup. The classification achieved still within the follicular group the highest rates and mixing gender (gender mixed) or both menstrual cycle phase (females) the accuracy drops. Classifying the menstrual cycle phase within all negative and in another classification within all positive resulted in the highest accuracy rates of about 72-72%.

**Table 3.** Final classification with 5 HR and 2 EMG features

| Group | classes | HR and EMG | Acc. [%] |
|---|---|---|---|
| follicular | pos/neg | 7 | 68.52 |
| luteal | pos/neg | 7 | 67.38 |
| young females | pos/neg | 7 | 61.78 |
| young males | pos/neg | 7 | 66.35 |
| gender mixed | pos/neg | 7 | 61.56 |
| all negative | follicular/luteal | 7 | 74.29 |
| all positive | follicular/luteal | 7 | 72.86 |

## 7 Discussion

Interactions between humans and technological systems happen nowadays more frequently, because humans are surrounded by technical systems like computers or automats. Although a lot have been done to improve these interactions (e.g. specific user interfaces) the identification of the users´ emotions and dispositions is still not clearly solved. One possibility to improve such classifications or the emotion identification is to consider gender and within young females the menstrual cycle phase as important co-factors. We analyzed the impact of these two factors on psychobiological emotion classification with a SVM and the use of heart rate and facial EMG features.

At first, we could show that for the use of heart rate features it is not necessary to take baseline-corrected features, in contrary this lead to a diminished identification of the valence information between positive and negative. We could also show that the use of solely heart rate features or solely EMG features to discriminate the valence is not enough, although both parameters have been found to be valence sensitive. Interestingly we found a higher accuracy rate for the male group with the EMG features, which indicates a specific EMG pattern for males but it somehow is not consistent for the follicular group. This demonstrates the need for a gender specific use of feature extraction.

General differences concerning gender were that the classification of men and women seems to differ. The reason for the diminished accuracy rate within young females seems to be their menstrual cycle phase. Because classifying positive vs. negative conditions within the follicular and luteal group lead to higher classification results of about 67% or 68%. Combining the female group the accuracy rate drops to about 62%. The same happens when gender is mixed in the gender-mixed- classification.

Another interesting result was that we achieved the highest accuracy rates for distinguishing the menstrual cycle phase within all negative/positive conditions. It means that the differentiation between the two cycle phases is easier than the differentiation between positive and negative for the SVM, which confirms our hypotheses of being influencing on psychobiological classification.

To improve the presented classification rates, another psychobiological parameter like the skin conductance should be included in the future. This parameter is innervated only by the sympathetic nervous system and therefore should serve well for predicting the arousal dimension. During the further procedure, a feature selection is planned to analyze gender specific features. Maybe the so far used features can be reduced specifically for gender and menstrual cycle phases. In respect to the aim of real-time classification we tried to use as less features as possible achieving still satisfying identification results. However, one has to keep in mind that under such circumstances the classification results might be not as high as after processes taking too much time (e.g. feature selection processes, different calculations of parameters).

**Acknowledgements.** This research was supported by the Transregional Collaborative Research Center SFB/TRR 62 "Companion-Technology for a Cognitive Technical System" funded by the German Research Foundation (DFG), a PhD scholarship by the University of Ulm for Stefanie Rukavina and a doctoral scholarship by the China Scholarship Council (CSC) for Jun-Wen Tan.

## References


1. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Center for the Study of Language and Information. Cambridge University Press, New York (1996)
2. Lang, P.J., et al.: Looking at pictures: affective, facial, visceral, and behavioral reactions. Psychophysiology 30(3), 261–273 (1993)
3. Burriss, L., Powell, D.A., White, J.: Psychophysiological and subjective indices of emotion as a function of age and gender. Cognition & Emotion 21(1), 182–210 (2007)
4. Cowie, J.M., et al.: A review of Clinical Terms Version 3 (Read Codes) for speech and language record keeping. Int. J. Lang. Commun. Disord. 36(1), 117–126 (2001)
5. Limbrecht, K., et al.: Advantages of mimic based pre-analysis for neurophysiologic emotion recognition in human-computer interaction. In: F.-M.A. (ed.) Emotional Expression: The Brain and The Face. UFP Press (2012)
6. Frantzidis, C.A., et al.: Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. IEEE Transactions on Information Technology in Biomedicine 14(3), 589–597 (2010)
7. Bailenson, J.N., et al.: Real-time classification of evoked emotions using facial feature tracking and physiological responses. International Journal of Human-Computer Studies 66, 303–317 (2007)
8. Ossewaarde, L., et al.: Neural mechanisms underlying changes in stress-sensitivity across the menstrual cycle. Psychoneuroendocrinology 35(1), 47–55 (2010)
9. Goldstein, J.M., et al.: Sex differences in stress response circuitry activation dependent on female hormonal cycle. Journal of Neuroscience 30(2), 431–438 (2010)
10. Andreano, J.M., Cahill, L.: Menstrual cycle modulation of medial temporal activity evoked by negative emotion. Neuroimage 53(4), 1286–1293 (2010)
11. Farage, M.A., Osborn, T.W., MacLean, A.B.: Cognitive, sensory, and emotional changes associated with the menstrual cycle: a review. Archives of Gynecology and Obstetrics 278(4), 299–307 (2008)
12. Tan, J., et al.: Repeatability of facial electromyography (EMG) activity over corrugator supercilii and zygomaticus major on differentiating various emotions. J. Ambient Intell. Human Comput. 3(3), 3–10 (2012)
13. Bradley, M.M., et al.: Emotion and motivation II: sex differences in picture processing. Emotion 1(3), 300–319 (2001)
14. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (IAPS): Technical manual and affective ratings. University of Florida, Center for Research in Psychophysiology, Gainesville (1999)
15. Walter, S., et al.: The influence of neuroticism and psychological symptoms on the assessment of images in three-dimensional emotion space. Psychosoc. Med. 8, Doc04 (2011)
16. Smith, J.C., Bradley, M.M., Lang, P.J.: State anxiety and affective physiology: effects of sustained exposure to affective pictures. Biological Psychology 69(3), 247–260 (2005)
17. Valenza, G., Lanata, A., Scilingo, E.P.: The Role of Nonlinear Dynamics in Affective Valence and Arousal Recognition. IEEE Transactions on Affective Computing 99 (2011) (PrePrints)
18. Rukavina, S., et al.: The influence of the menstrual cycle on gender differences in personality items. In: Annual International Conference of Cognitive and Behavioral Psychology (CBP). Global Science & Technology Forum (GSTF), Singapore (2012)

19. Fridlund, A.J., Cacioppo, J.T.: Guidelines for human electromyographic research. Psychophysiology 23(5), 567–589 (1986)
20. Nussinovitch, U., et al.: Reliability of Ultra-Short ECG Indices for Heart Rate Variability. Ann. Noninvasive Electrocardiol. 16(2), 117–122 (2011)
21. Andrade, A.O., et al.: EMG Decomposition and Artefact Removal, in Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges, I.-.-.-.-. Ganesh R. Naik, Editor. 2012, InTech.
22. Neta, M., Norris, C.J., Whalen, P.J.: Corrugator muscle responses are associated with individual differences in positivity-negativity bias. Emotion 9(5), 640–648 (2009)