Mapping User Engagement to States of Affect via an Unobtrusive Biofeedback Device

A Dynamic Visualization of Real-time Assessment

Anthony Psaltis, Charalampos Rizopoulos and Constantinos Mourlas
Faculty of Communication and Media Studies, National and Kapodistrian University of Athens, 5 Stadiou str, Athens, Greece

Keywords: Biofeedback, Physiological Measurement, Engagement, Affective Interactions, User Evaluation.

Abstract: The elicitation of affect can be regarded as an influencing factor upon a person’s cognition, emotional state, mood, attention and motivation. It is also recognizable as an inhibited physiological process expressed by the human brain as induced or suppressed hormonal and neural stimulation that subsequently instigates a physical and mental level of attentiveness attractiveness or aversiveness. Physical reactions to emotion causing events and stimuli in affective computing are classified by direct mapping of facial and postural expressions to corresponding patterns, by using visual and postural observation methods. Despite the fact that physiological assessment is generally more reliable and less error-prone, a higher amount of research has been devoted to visual and postural methods due to the greater complexity and specific knowledge requirements of the former. Concentrating more on the physiological aspect of assessing affect, we have developed a biofeedback device, sensing reactions instigated by emotion-causing events and results have been assessed in real-time using suitable visualization methods. In previous attempts to acquire this type of measurements, human subjects were physically and psychologically impaired by the electrodes and wiring attachments used for the acquisition of signals and therefore validity of data was to some extent in question. In order to achieve an uncompromising assessment environment we designed a system that acquires heart rate and stress measurements via an ordinarily looking computer mouse. Certain combinations of heart rate precipitation and tonic level / phasic response of stress levels were investigated as reactions to emotion-inducing events. Corresponding patterns of physiological measurements to a real time affect allocation model have reached interesting correlations of events with respective states of engagement to an impressive degree of coincidence.

1 INTRODUCTION

Various studies of emotion and affect ranging from the areas of psychology (Eysenck, 1982; Frijda, 1986; Strongman, 2003; Scherer, 2005; Eysenck, 2006), clinical physiology (Lykken & Venables, 1971; Hassett, 1978; Venables & Christie, 1980; Damasio, 1994/2006; Tsatsou, 2006), affective computing (Picard, 1997), and HCI (Norman, 2004; Brave & Nass, 2008; Fairclough, 2009; see also the vast literature on User Experience [UX] design and game studies – e.g. Kuniavsky, 2010; Calleja, 2011; Law & Sun, 2012 etc.) have demonstrated that physiological stimulations of the human brain elicited by emotional events are direct, instant, measurable and quantifiable by using fMRI, PET, electrodermal response, respiration, heart rate or combinations of the above. Previous work addressing this issue (e.g. Ark et al., 1999) required the user to be impaired by wires and electrodes. This requirement did not only restrict the user’s posture and mobility, but also had important psychological consequences, thus affecting the validity of the measurements. In the implementation described in this paper, users were completely untethered, since they only had to use an ordinary computer mouse in the usual manner; therefore, they were unaffected by such limitations.

We assessed the concurrent excitation of two of these quantities, namely electrodermal response (skin conductance [SC]) and heart rate (HR), via a specifically designed interface providing strong emotional stimulation in an attempt to answer the following questions:
What is the interpretation of the psychosomatic state of the subject when both heart rate and stress level are increasing simultaneously?

Is there a correlation between events that occur during the interaction with the system and the aforementioned quantities?

Would the combinations provided by the two states of each of the measured quantities show a predictable pattern in response to certain emotional events?

Could this system produce an adequately reliable mechanism to map a psychosomatic condition onto a descriptive model that would ideally be in accordance with one of the established models of affect, namely Russell’s (1980, 2003) circumplex model?

Finally, can this system be used efficiently as a component of an intelligent interface?

The structure of this paper is as follows: a brief description regarding the biofeedback principles, our purpose built electronic device and our system setup is given. Subsequently, the experimental design methodologies adopted and the development of the front-end interface are outlined. Presumptions, limitations, and predictions are clarified in detail in the psychosomatic analysis section, followed by a conclusive interpretation of our experimental data and suggestions as to how the system may be improved and used in broader areas of interactive applications.

2 BIOFEEDBACK ACQUISITION

Derived from basic principles of biofeedback measurements (Venables & Christie, 1980), psychological stress is expressed in the human body by a physiological response of brain induced vasodilatation of sweat glands of the skin. This alteration causes measurable changes in electrical skin conductance that produces quantifiable indicators of stress levels. Our purpose-built electronic device used for the acquisition of HR and stress measurements (SC) has been presented in previous publications of the authors (Psaltis, Mourlas, 2011). HR is acquired based on the principle of Infrared Spectroscopy, effectively using as sensing elements two deflecting infrared sensors acquiring HR pulses, for optimised error cancellation and a more reliable reading. For stress measurement the skin conductance method has been adopted as opposed to skin resistance, thus providing a more reliable indication of the measured quantity by eliminating inaccuracy problems due to perspiration on the part of the subject.

Stress level is identified by two silver-silver chloride contact rings measuring skin conductance, placed on the sides of a computer mouse and at the points where the thumb and middle fingers are resting during the typical use of the mouse by a user. The two HR sensors are situated in the centre of the SC rings. Hardware redundancy is embedded in the system allowing for two channel acquisition of signals of HR and SC simultaneously, eliminating artefacts and discontinuities caused by movement of the fingers unless contact with the mouse is completely lost. In case of lost contact the system preserves the trait values until the next valid measurement is detected. Minute losses of contact with the sensors or erratic movement of the fingers onto the mouse could produce an error in HR reading in the scale of milliseconds between pulse readings, while the SC measurement is unaffected. Considering that our system does not take into account detailed pulse shape and cardiac arrhythmias, but is instead designed for detecting heart rate and heart rate variability, this error is within acceptable margins of tolerance of the system, as it produces an estimated attenuation of ±0.18 of a pulse per reading. The mouse, including a signal preconditioning circuit, is connected to a computer via an electronic interfacing circuit that filters and conditions the above primary signals and converts them into a form that can be read by the computer via the two channel audio input. A purpose-built software suite comprising the appropriate components required for a system configuration console, signal processing and display is the final component of the system. The use of this console is essential for the initial settings required, such as time interval between measurements, audio card selection and settings, as well as data storage configuration.

HR detection and correction, as well as SC auto-calibration algorithms, have been implemented, providing a relative baseline for each subject independent of the actual stress levels of individuals. Since our interest is not exactly how much stress is the user experiencing but instead the state of stress in relation to the previous level of stress, we devised a method for auto calibration, envisaging our reference point (“baseline”) at the mid distance of the difference between highest and lowest measured stress value weighed by the trait values. The baseline is continuously updated based on the measurements and the deviation from the mean value of SC (“tonic level”). This was important primarily because it
eliminated problems like the need for initial calibration and similarly alleviated continuation inconsistencies during measurements once the user released the mouse momentarily. Moreover, as an additional threshold, the baseline provided a method of distinguishing additional details in the attributes of our measurements, effectively indicating the zero point of transition during instantaneous reactions of the users (“phasic response”).

3 ASSESSMENT STUDY

3.1 General Description and Hypotheses

The primary objective of our research was to assess the conditions in which a certain physiological response is indicative of a strong emotional stimulus. In our system, given that we had two important quantities (HR and SC), a reasonable assumption was to examine conditions where both physiological quantities measured showed a common tendency compared to their previous measurement and also follow the same directional pattern, i.e. both either increasing or decreasing at the same time. From the two measured quantities (HR and SC), the two main system parameters, namely the SC and the HR gradient, are derived. The SC gradient refers to the vector form of SC measurements in relation to the baseline, whereas the HR Gradient represents the relationship between two consecutive HR measurements, weighed by a factor of heart rate variability (HRV). As such, the gradient value indicates the tangent of the curve that depicts the user’s psychosomatic state. A gradient can be positive (i.e. raising pattern with regard to previous value), negative (i.e. falling pattern when the new value is lesser than the previous) or zero representing a flat pattern essentially indicating no change of the previous and current quantities.

The algorithm used for the classification of tendencies on each response acquired by the biofeedback device is illustrated in figure 1.

We classified the users’ behavioural patterns measured via our experimental platform into four areas corresponding to four dominant states of engagement, which in turn we hypothesized as corresponding to the quadrants of Russell’s circumplex model of affect. It should be noted that no single gradient is correlated to either dimension of the circumplex model (i.e. arousal, valence); rather, data are mapped to the four quadrants in relation to the states defined below.

- **State of Focused Involvement / Engagement** (positive arousal and valence), where the user is satisfied for fulfilling the task successfully.
- **State of Contentment** (negative arousal, positive valence), where the user is unable to fulfil a task but maintains a high level of activation.
- **State of Perceived Difficulty** (positive arousal,
negative valence), where the subject’s focus on a task does not change significantly, while satisfaction diminishes.

- State of Non-involvement / Apathy (negative arousal and valence), where an uninterested and inattentive person performs a task in negative valence and arousal levels.

We adopted these patterns of interpretation so that the stimuli produced by our experimental scenarios are a more realistic representation of the users’ psychosomatic state compared to attempting to infer emotion in detail. We reached this decision because our system does not have the necessary accuracy to detect specific emotions with precision.

During the experiments, processed data were displayed on-screen in real time, representing user responses mapped as loci onto a screen area subdivided into four quadrants. Each of those quadrants dynamically represented the states described above as interpreted by the indices produced by measurements of HR and SC. One locus point was created every two seconds. The relationship between the measurements and the pattern of the projected mapping is shown in figure 3.

![Figure 3: Mapping of measurements.](image)

4 DESCRIPTION OF EXPERIMENTAL PROCEDURE

Facing the interesting challenge of designing a front-end that would take full advantage of the capabilities of our system to perform emotional assessment in a thorough and convincing manner and at the same time explore to an optimal degree the latest concepts of emotion-inducing techniques in user interaction, we came up with a visual environment, in which all stimulating events and reactions of the user, as well as processed data from our system, were displayed in real-time. A recording of data, as well as factors and parameters used for generic data transformation, were also produced for tracing and post-processing. Thus the assessment and verification of the effectiveness of our system was facilitated by the ability to reproduce as well as observe user responses at a later stage.

For the experiment, an ordinary desktop PC (Intel Dual Core processor, 2 GB memory) was used. The experimental sessions took place in a computer lab at the University of Athens, Greece, under stable environmental conditions (approx. 20 degrees Celsius) for all subjects.

Each session of the experiment included two parts with a total duration of twenty minutes. The first part consisted of a puzzle game (“Liquid Measure”, available at [http://www.friv.com/](http://www.friv.com/)). The game was intended to elicit positive affect as a result of successful task completion, and negative affect in cases in which the subject was unable to progress. It provided an unknown environment with tasks of progressively increasing difficulty. The first part had a duration of 10 minutes. No prior instructions were given; thus, subjects needed to concentrate and improvise in order to complete each of the twelve levels of the game. Not all levels were completed by the participants, as the time limit was not sufficient in most cases.

The second part consisted of a video of motor traffic accidents found on Youtube ([available at https://www.youtube.com/watch?v=26gTIQ1FDW4](https://www.youtube.com/watch?v=26gTIQ1FDW4)). It was intended to elicit instinctive physiological responses to negatively valenced stimuli (essentially indicative of distress and fear). Some of the accidents were predictable, whereas others occurred suddenly. Two separate categories of emotion-inducing stimuli were included in this video. The first category consisted of accidents occurring at a distance, while the second category consisted of accidents (or near misses) involving the car on which the camera was located. These two categories of stimuli were assumed to be evaluated differently. More specifically, accidents that occurred further away were assumed to be evaluated as less threatening compared to accidents occurring virtually at a first-person view.

In addition to the physiological measurements obtained, participants were able to express their estimate of psychosomatic condition at any time by selecting one of the four states displayed on the screen throughout the duration of the test session.

It should be noted that additional information
regarding user preferences on pictorial or textual assessment or psychological profiles was neither requested nor taken into account in this study. Apart from demographics (age and gender), the only type of personal information requested from the subjects concerned their driving experience.

The sample (N=17) consisted of university students ranging from 1st-year to masters students (10 women and 7 men). The age ranged from 18 to 47 years of age with a mean value of 26.2 years. The participants had no prior knowledge of our experiment setup, were free to decline participating, and also retained the option to withdraw or refuse to complete a part at any point during the experiments. Two subjects were not included in the final assessment: one female user was inconsistent as far as the contact with the computer mouse is concerned and one male participant was very talkative during the experiment. Although the consistency of the results obtained from the latter user was better than 83%, he exhibited respiration arrhythmias that we judged it could have affected stress measurements considerably; therefore, this participant was excluded from analysis.

5 DATA ASSESSMENT AND EVALUATION

Data analysis was performed for each individual subject in three steps. First, the visual content (both game and video) was weighed according to the perceived emotional impact of each particular event, providing a table of predicted state of engagement, event number, and timestamp. The participants’ response to the emotional stimuli was then assessed and the predicted state of engagement for each specific event was compared to the state of engagement measured by the system. The accuracy rate for each participant was determined by the degree of similarity between the predicted and observed states of engagement during the course of the experiment.

In fact, the correlation of the corresponding values of emotional intensity was assessed in comparison with values acquired as responses from our system. For each subject, data were analysed in order to formulate a distribution table that provided a more extensive indication of the convergence or divergence of emotional patterns derived from our system, thus deducing its accuracy expressed as a percentage.

During the development of the interface, we integrated data post-processing algorithms and produced results in their final form ready for analysis. For example, instead of ending up after each experiment with a vast amount of raw data that then had to be transformed into a meaningful form, we integrated display of processed data proactively. As such, the outcome of post-processed data was presented to the users in real time. At the same time, detailed data recordings allowed for further exploration and tracing as required. Data obtained from all 15 subjects are presented in Table 1.

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Age</th>
<th>No of Events (Average)</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>F18a</td>
<td>18</td>
<td>51</td>
<td>91.38%</td>
<td>91.38%</td>
</tr>
<tr>
<td>F18b</td>
<td>18</td>
<td>62</td>
<td>98.13%</td>
<td>98.13%</td>
</tr>
<tr>
<td>F18c</td>
<td>18</td>
<td>69</td>
<td>89.08%</td>
<td>89.08%</td>
</tr>
<tr>
<td>F23a</td>
<td>23</td>
<td>55</td>
<td>94.23%</td>
<td>94.23%</td>
</tr>
<tr>
<td>F23b</td>
<td>23</td>
<td>60</td>
<td>87.76%</td>
<td>87.76%</td>
</tr>
<tr>
<td>F24</td>
<td>24</td>
<td>60</td>
<td>88.91%</td>
<td>88.91%</td>
</tr>
<tr>
<td>F25</td>
<td>25</td>
<td>57</td>
<td>89.40%</td>
<td>89.40%</td>
</tr>
<tr>
<td>F27b</td>
<td>27</td>
<td>67</td>
<td>94.68%</td>
<td>94.68%</td>
</tr>
<tr>
<td>M32a</td>
<td>32</td>
<td>53</td>
<td>95.41%</td>
<td>95.41%</td>
</tr>
<tr>
<td>M32b</td>
<td>32</td>
<td>59</td>
<td>90.26%</td>
<td>90.26%</td>
</tr>
<tr>
<td>M35</td>
<td>35</td>
<td>55</td>
<td>86.74%</td>
<td>86.74%</td>
</tr>
<tr>
<td>M37a</td>
<td>37</td>
<td>60</td>
<td>91.33%</td>
<td>91.33%</td>
</tr>
<tr>
<td>M37b</td>
<td>37</td>
<td>51</td>
<td>90.61%</td>
<td>90.61%</td>
</tr>
<tr>
<td>M47</td>
<td>47</td>
<td>49</td>
<td>94.71%</td>
<td>94.71%</td>
</tr>
<tr>
<td>Median</td>
<td>27</td>
<td>59</td>
<td>91.33%</td>
<td>91.38%</td>
</tr>
</tbody>
</table>

In an additional test conducted with a subset of the above participants whereby the users were requested to look away and divert their attention away from the computer, we observed that values mapped onto the four quadrants representing the four states of engagement changed to the states of Perceived Difficulty and Non-involvement with an accuracy of 98.1%. As soon as the users redirected their line of sight back to the computer, their state of engagement returned to the positively valenced state of Active Involvement.
6 FINDINGS OF EXPERIMENTS

6.1 Data Interpretation

Data from the experiments indicated a high degree of coincidence (91.35%) between optimally assigned values and those produced by the system with very few No-Reading errors. Users described their psychosomatic state by selecting manually one of the states displayed as follows: ‘Focused Involvement / Engagement’ 71%, ‘Contentment’, 9%, ‘Perceived Difficulty’ 13%, ‘Non-involvement’ 7% of the time of the experiment. The deviation between the values obtained through self-report and those obtained through measurements was approximately 17%, although it is worth noting that users were not inclined to use the manual selection feature very often. Positive user involvement during the experiments is represented in our model by the states of Focused Involvement and Contentment. The opposite states were actually mapping either the state where users were changing level of difficulty between scenarios or while they spent time waiting for the next event with diminished or diminishing involvement. Overall results classified all users in the above four cases at 44.2% (Focused Involvement), 31.9% (Contentment), 12.6% (Perceived Difficulty), and 11.3% (Non Involvement / Apathy) of the time of the experiment. Detailed data regarding allocation per quadrant and gender is shown in table 2.

The difference in responses between male and female participants was small (<1%), although event-to-response evaluation has shown coincidence between 87.76% for female and 98.13% for male users. Coincidences in the game task were similar for both male and female participants. However, differences were observed in the video session. In our view, this small difference has its origins in the profiles of male participants being affected more by aversive driving experiences (age of 32-47 with driving experience) than female participants (age 18-27 with little or no driving experience whatsoever). Differences in both male and female participants between gaming and video sessions indicated less accurate event / response matching during the gaming session data than that during visual observation, which was more accurate by 4.1%. This is explained by the fact that, during the gaming session, users had to dislocate their fingers from the mouse irregularly, although during the video session the contact of their fingers with the sensing elements of the mouse was uninterrupted and the quality of biofeedback acquisition was therefore nearly ideal.

A caption of a typical real time model representing the mapping during an experiment is shown in figure 4.

Table 2: Mapping distributions per quadrant and gender (AI = Active Involvement, C = Contentment, PD = Perceived Difficulty, NI = Non-involvement).

<table>
<thead>
<tr>
<th>Emotional Mapping Distribution (%)</th>
<th>AI</th>
<th>C</th>
<th>PD</th>
<th>NI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>44.6</td>
<td>32.2</td>
<td>11.6</td>
<td>11.1</td>
</tr>
<tr>
<td>Female</td>
<td>43.9</td>
<td>31.6</td>
<td>11.0</td>
<td>14.2</td>
</tr>
<tr>
<td>Mean</td>
<td>44.2</td>
<td>31.9</td>
<td>11.3</td>
<td>12.6</td>
</tr>
</tbody>
</table>

The four numbers in the periphery of the graph (in figure 4 the numbers 45, 251, 117 and 121 respectively) represent the actual number of occurrences of the biofeedback data acquired from the participant and classified by our system. The number in the centre of the graph is the index of measurement helping to identify the exact time and event during post data assessment. At the top of the graph Tonic Level, Phasic Response as well as Arousal, Stress Level Baseline and Gradients were displayed on each cycle of measurements. The latest value is indicated with a characteristic border that fades out when the next value is mapped. The relationship of each measurement to the baseline in each quadrant can be thought as the imaginary diagonal axis drawn from the centre of the graph towards the corners of the rectangle. This additional information may in future development be scaled so as to provide a more detailed analysis of the psychosomatic state of the user into eight subdivisions rather than the existing four.
7 DISCUSSION

State of engagement was the most important psychophysiological condition we attempted to identify and quantify in this research as it is highly regarded in Affective Computing research (e.g. Picard, 1997). Results have verified our views that user engagement transitions between affective activation states can be reliably detected by simultaneous changes in HR and SC gradients. Indication is immediate as expressed by previous research work assessing similar aspects by means of fMRI and SC (Tsatsou, 2006). We may conclude that our system produces a valid and accurate snapshot of the user’s state of engagement because, as reported earlier, whenever the users looked away or otherwise disengaged their attention from the screen, the system immediately detected this shift of focus correctly by labelling their state of attention as either “Perceived Difficulty” or “Non-involvement” - both of which are negatively valenced when represented as coordinates for the dimensions of Russell’s circumplex model of affect.

Throughout the entire experimental process, a transition from the state of Active Involvement to that of Contentment immediately after stressful stimuli was most frequently observed. This transition is interpreted as the effect of the slow decay of the stress level, which produces a negative SC gradient, effectively indicating a reduction of arousal.

Taken together, the aforementioned points indicate that a mapping of the users’ state of engagement onto Russell’s circumplex model of affect is accurate, at least with respect to placing the identified state in the correct quadrant (i.e. positive/negative valence, positive/negative arousal).

From the above, as well as the overall results of the experiments, we have indications from our system that the correlation of common gradients of present and past values of HR and SC shows high probability of success in determining various states of engagement. The time interval for each measurement was crucial for the accuracy of our system and optimised accordingly in order to provide enough time for the subjects’ physiological response to settle into a detectable timeframe. Additionally, this timeframe allowed for the normalization of artefacts that could have been misleadingly accounted as spontaneous reactions of the participant.

The time frame was chosen following optimisation deduced by the assessment of data produced by added measurements into a larger buffer of data. Time over 2 seconds has shown a smoother transition between states but also slowed down the detection of user response and therefore rejected.

8 CONCLUSIONS AND FUTURE WORK

Beginning with the somewhat simplistic but sensible assumption that coinciding gradients of HR and SC may express some relation to an emotional state, we assessed the possibilities of detecting basic cognitive processes such as engagement, which may be indicative of underlying emotional states. Results from our experiments have shown that a correlation exists; however, we are not clear as to the exact emotion expressed. It is possible that the excitement state detected and interpreted by our system as a state of engagement is in fact another possible combination of emotional state producing similar measurements - simply coinciding with the interpretation of our algorithm. Although our system at this stage is rudimentary for detailed mapping of various emotional states such as frustration, sadness, depression etc., we believe that our system can be used effectively and efficiently as a component of an intelligent interface for detecting the user’s degree of involvement in various applications (e.g. educational assessment, distance learning etc.); however, future research may lead to further improvements in the identification of more detailed psychosomatic states.

An additional avenue we are planning to explore is the validation of the output of our system with the aid of standardized validation methodologies and instruments, such as IADS and IAPS (Bradley & Lang, 2007), as well as appropriate emotion recognition software (e.g. facial emotion recognition). Furthermore, we are planning to implement a visual pupil size detection component, which is expected to increase the validity of our system with respect to focusing intensity.

Our research may be applicable to fields such as e-learning, educational assessment, virtual environments, and further areas requiring remote assessment of user psychosomatic condition.

REFERENCES


