Spontaneous Extrapolation and Detection of Stress Level Based on Human Computer Interaction and Physiological Response

*Fatima Isiaka  
Kassim Mwitondi  
Department of Computing  
Sheffield Hallam University  
Sheffield, United Kingdom

Adamu Ibrahim  
School of Computing  
University of Leeds  
Leeds, United Kingdom

Abstract — User Experience (UX) is of high importance for design specifications on web applications. Peering into the user’s mind is of high value. We always like to understand user cognitive processes, why they do certain things online, and whether they see some elements on the webpage that we cannot. Methods such as interview and surveys normally gather this information by simply asking the users subjectively. ther methods such as the observational process mainly rely on viewing or observing the users’ behaviour. In this paper we try to approach this in an objective perspective by directly observing what goes on in the users mind, through non-obstructive technique without interrupting the data collection thereby obtaining qualitative and quantitative data of high imension for modeling. To further test for bias in variance, prediction and confident intervals, boostrapping was employed for measures of accuracy to sample estimates.

Keywords — UX, stress detection, extrapolate, task performance, boostrapping, cognitive process, human physiological response.

I. INTRODUCTION

Recent years have seen that interrelating Human Computer Interaction (HCI) and Human Physiological Response (HPR) is of high value to understanding UX [6]. Most UX investigations is based on or limited to overt observable behavior [22]. There has been very few work done in the use of physiological response with eye tracking in HCI. The latency in some physiological response measure is slow, e.g. skin conductance response (SCR). Response normally appear between 1 - 3 seconds after stimulus onset [1]. This means that the eyes have left the area of interest or fixation on a visual stimulus that caused the spikes in the SCR effect, long before the effect was registered in data. This latency is sometimes difficult to take into account [13][28][29]. Modeling the quantitative data obtained can tell a different story. A common finding in cognitive neuroscience states that user subjective perception of their attitude or behavior doesn’t always correspond to their neural activity [6][5], this is a short form of saying that users doesn’t always know what is going on in their minds. An example is of a participant eye fixation directed on a particular content on a webpage that correspond to certain spikes in their emotional arousal state observed in real-time, while subjectively they admitted to not really affected by the said element that occupied their visual attention [31][6]. In this paper we try to approach this through an objective perspective and make use of the quantitative data collected from an eye tracker and physiological response measure sensors. The main goal is to investigate the possibility of applying modeling techniques on high dimensional Human Computer Interaction and Human Physiological Response (HCI-HPR) data can predict affect states and recognize patterns. These patterns could be used to ultimately detect users stress level and associate it to an element. A modelling approach is the best technique [16][9][18], when considering this.

II. RELATED WORK

Though there has been quite a lot of work done on HCI where eye tracking is concerned, there are actually quite a few on eye tracking combined with physiological measures in HCI. We continually try to find methods in which to properly understand UX that allows us to gain more insight into user’s unconscious desire as they interact with elements since acknowledging that visual attention and engagement are not necessarily linked to perceptions of experience [15][10][17]. Physiological response measurements allow investigations on this area by collecting objective measures of performance [30][25].

Rather than asking participants if they find a particular tasks difficult or if their attention was distracted on the sudden appearance of stimuli on screen, their physiological response can be measured to give an objective rather than subjective report in an invariant process. The physiological measures considered here are the skin conductance, pupil dilation, skin temperature and eye movement, which are briefly discussed in the following subsections.
A. Skin Conductance Measure (SCR)

The SCR is a measure of the electrical changes of the skin as a result of sweat[1], usually caused when the user is involved in an activity[27]. It is a measure of the emotional arousal state of a person, and can easily be measured with electrodes non-invasively placed on the fingertips, wrist or palm (Figure 1a). The SCR grows higher during states such as stress, interest, attention or nervousness and lower during states such as relaxation or boredom (Eq 1), depending on the state of the user and the task the user is involved in[26]. Figure 2 shows the SCR with typically computed features such as the latency (time from stimulus to an onset), amplitude (height of an arousal state), peak (maximum arousal state), SCL (tonic phase) e.t.c.

$$f(\alpha) = \begin{cases} \text{High} & \alpha \in \varepsilon \\ \text{Low} & \end{cases}$$

where $$\varepsilon$$ can be any emotional arousal state, either elated or relaxed.

B. Skin Temperature (ST)

ST measure increased exertion, excitement and stress. Temperature at the surface of the skin changes accordingly to blood circulation through body tissue [21]. Data from ST can also back SCR data in terms of interpreting stress [4][3]. The temperature of the body changes according to activities and other external factors just like the SCR.

C. Eye Movement

Eye movement is the voluntary or involuntary movement of the eyes, which helps in acquiring, fixating and tracking visual stimuli. The three most important attributes in eye movement behaviour are the fixation (location of a user’s gaze), Fixation Duration (length of time that a user fixates) and Saccade (the angular distance between fixations) (Figure 1b). The fixations are mapped to the x and y coordinates that pinpoint where the user looked on a given display [6]. User experience can be reflected and determined through their attitude or behaviour [8][20]. Negative attitude towards a complex application shows a poor experience while positive attitude towards a complex application shows a good experience [23]. The three possible emotional states discussed here that can be experienced by a user during an interactive session could be expressed as stress, relaxation and neutral mood which can be demonstrated in the following concept.

$$f(\alpha) = \begin{cases} (1-\varepsilon) \times 0 + (\varepsilon \times \alpha) & \text{if relaxed (positive)} \\ (1-\varepsilon) \times 0 + (\varepsilon \times \alpha) & \text{if stressed (negative)} \\ \varepsilon & \text{if neutral} \\ \end{cases}$$

Where

$$\varepsilon = \begin{cases} 0 & \text{if positive effect} \\ 1 & \text{if negative effect} \end{cases}$$
III. METHODS

The method for data collection was based on usability evaluation testing, with 44 participants aged between 18-44. The users are both average users and constant user of webpages. They were seated in front of an eye tracker with their fingertips placed on a wireless SCR measure, while we collect both their physiological data and eye movement data recorded by an eye tracker. They interacted with a set of six webpages. The tasks assigned were both contrived and non-contrived tasks. Two of the webpages had both the Auto-suggest List (ASL) deactivated. The purpose is to see if the absence or presence of ASL increased or decreased their stress level. While the dynamic contents on the other pages where used as a form of emotional incitement to induce or evoke responses from the users and gauge if the sudden appearance of a picture or advisement distract them from their given task. The wireless SCR sensor (Affectivia Q-Sensor) (Figure 1a) was used to collect the SCR data. It also has the capability to measure and record the ST of the participants. The data collection was approach in two ways:

- Real Time
- Delayed

In real time (Figure 1a), we were able to investigate and observe the spikes in SCR, PD and ST while they interact with elements on the webpages. The delayed enabled us to access each session (Figure 3 ) of the experiment revisit, extract time sessions and register delay in latency of physiological readings that correspond to latency of eye movement and also for further analysis. This helps to identify a particular instance in time that might have been missed. Most of the behavior data such as “Looking”, “Interest” and “Clicking”, are logged in the eye tracker during real time and also during delay. Using this method, we can be able to obtain real-time feedback without interrupting collection of data. The quantitative data was exported to a spreadsheet, where a total of 264 instances were generated; with each participant’s data generating six instances. A short term interval processing was adopted for the sync HCI-HPR.
IV. SIMULATIONS AND ANALYSIS

The method for data analysis was approached by extracting features with a custom algorithm that helps to run the simulations that detect increase in stress level based on the average amplitudes response detected for each webpage and the response duration. The steps for the algorithm is stated below (Figure 4). The feature extraction were considered based on significant variables often used in literature, when considering a multimodal approach such as combining eye tracking data and physiological measures [28][6], the novel approach involves appending saccade size of fixations. The diagram below illustrates these features. Each participant’s baseline differs and thus increase in amplitude depends on variations in median threshold used in extracting event correlates to produce a matrix Z(m, n). Participants generates instances each based on the number of webpages viewed. Savistsky filter [24] Eq 3 was applied for removal of noise and other artifacts on the physiological measures. The baseline (skin conductance level (SCL)) was estimated based on point interpolation moving average technique (Eq 4). Peaks on the SCR were detected on a given threshold that corresponds to a participant’s median SCR.

$$f_k = \frac{\sum_{l=m}^{m} y_{k-1}}{\sum_{l=n}^{n} a_l}$$  \hspace{1cm} (3)

$$Y_{k+1} = A_1 + A_2 X + A_3 X^2 + \ldots + A_n X^n$$

$$A_k = \text{Coefficients (Scalers)}$$

$$Y_{k1}$$ refers to the smooth SCR.

$$V_k = T_k \left( \frac{\sum_{l=m}^{m} y_{k} - 2a_{n+1}}{2a_{n+1}} \right)$$ \hspace{1cm} (4)

$$\bar{k}$$ is the resulted data-points by resampling the smoothen SCR data-points, taking a moving average of 2n-1 window size on each points in $$\bar{k}$$, in each time interval $$T_k$$. The smoothed data still maintained its shape accordingly. Each physiological measure undergoes this process depending on how noisy the data are. The eye movement data obtained includes the PD and fixations captured by the eye tracker. The derived variable here is the saccade size $$D$$ that gives the Euclidean distance between two points ($$x_{n1} - y_{n1}$$) and ($$x_{m2} - y_{m2}$$).

$$D = \sqrt{(x_{11} - y_{11})^2 + \ldots + (x_{n2} - y_{m2})^2}$$ \hspace{1cm} (5)

where ($$x_{11}, y_{11}$$), $$(x_{21}, y_{21})$$, ..., $$(x_{n2}, y_{m2})$$ are fixation points on the vertical plain of a webpage and ($$x_{m1}, y_{m1}$$) are the fixations on the horizontal plain. The sync data are exported and written to a spreadsheet using MATLAB (Figure 6). If the SCR falls below median range, the participants is at the tonic place, hence nothing is actually happening and he/she is relaxed. Once there is increase in amplitude that exceeds the median level, the participants are either considered stressed or excited and given the nature of the task the former is the classified affect state. Hence we integrate between “Stress”, “Neutral”, and “Relaxed” state of the users. The SCR is the physiological measure that serves as the major constant response for this case, given its ability for detecting spontaneous and evoked reaction in literature. The algorithm (Figure 4) detects each increase in amplitude and predicts the maximum SCR (peak) within a certain interval that correspond to a particular event. The final output data resulted into a higher dimensional dataset Z(m, n) developed for the purpose of modelling. We considered applying classification algorithms to see if patterns can be recognized from the integrated multimodal data. The patterns identified are trained based on the labelled class (stress level). This will provide a reasonable view of the input datasets and perform the most likely matching inputs based on the variations on the dataset. The peak detection algorithm provides means to organize the data from the sensors and detect events based on the peaks identified.

![Algorithm](image)

**Fig. 4: Algorithm to detect associated user interaction to web events and stress level**
The custom algorithm was able to detect affects corresponding to events during interaction with a performance of 80%. The next stage is to detect outliers that might be present using the least median of squares for regression or least mean of squares (LMS) adopted in [2] for outlier detection. This is based on the mean and standard deviation of each variable. It estimates parameters by solving the nonlinear minimization problem.

To carry this out, two variables the URL (webpage) and Status (affect state) were removed from the scene, this provide a means to search for structures and try and find out whether natural structures exist in the dataset as opposed to physiological measures obtained from the participants. Using this method, we were able to detect that visiting webpages with dynamic contents especially where the Auto-Suggest List (ASL) is disabled and also where picture and video contents were found, increases the stress level of some participants because they were distracted (Figure 3 and 8). The absence of ASL disrupts the participants expectations of the task, thereby increasing their stress level. The chart in Figure 7 shows the number of instances there are increase in stress level of participants for the six webpages.

Instances like observation 45 and 215 had fixation duration of 4087 and 1534 secs, while looking at Yahoo and iGoogle page, even with application of filters before feature extraction. This is quite a large variation that totally exceed the average fixation duration (250ms) and normal fixation between 100ms- 500ms [12]. The LMS was able to detect this unusual occurrence on multiple simulation runs and this was traced to both pages. These outliers were suppressed. The chart also shows visiting the ASL enabled pages reduces the stress level of users than visiting the ASL disabled pages in Figure 8a and 8b.
The pupil dilation during task searching process shows difficulty on the ASL area, the greater the difficulty of task, the greater the pupil dilation. Variations in affect state shows greater number of increases in stress level at start of stimulus presentation (Latency), that implies difficulty in task performance for most participants (Figure 9).

The high dimension data can tell us more about the relation between the participants interaction and their emotional response to task they were assigned. For this, Pearson correlation was adopted to see if there is any significant relation among variables. This will determine the modelling approach to consider for the classification of the emotional arousal states. The next section discusses the results.

V. RESULTS

The variables were selected based on feature importance for HCI-HPR. From the correlation test performed, the output displays a heatmap (Figure 10) indicating little correlation among parameters, except for the mean increase in amplitude (MPeak) and the affect class (Status) with ($r^2 = 0.80, p < 0.05$). This clearly excepts the concept of emotional response arousal dependency on increase and decrease in amplitudes. The mean skin temperature (MST) also shows significant correlation with skin potential ($r^2 = 0.64, p < 0.05$). But there is very little correlation between most eye movement data and physiological response data, even with appending of saccad size of eye movement response. This consequently reduces the effect of correlation among features.
This form of behavior on datasets is most appropriate for the Naive Bayes classifier [14]. Even though some of the features exhibit significant correlation. However, it is quite unclear whether the correlated occurrence is as a result of suppressed outliers or the saccade measures. Other classifiers (Support Vector Machine [19], J48 [7] and Bagging [11]) were adopted to improve training algorithm, aggregates the model performances (Figure 17) and test behavior of data. The 70/30 split partitioning was used to separate the dataset into a training and test set. This partitioning can ensure values of the variables to be automatically divided across the training and test set.

### A. Model Performance

Based on aggregates of the model performance in classifying the affect state, the number of instances correctly classified as stress is high enough to justify that stress is the main emotion variable to be considered since it proved to be consistently high throughout the data analysis process. The number of instances that were correctly classified as "Stress" are 37, while 12 instance were correctly classified as "Relaxed", and 7 instance were either between stress level or tonic phase (Neural).

From literature, Bagging always seem to perform the most on several cases, and didn’t fail in this instance given the uncorrelated diversity and latency contingency of HCI-HPR.

The ROC plot shows the diagnostic outcome of the model performance, the proximity of the model performance to 100%, illustrates that the diagnostic accuracy though not 100% accurate but close enough to ascertain that patterns can be recognized and detected in relation between HCI-HPR. It is feasible then to say or accept that there is possibility in automatically detecting stress in users and associate it to elements on a webpage, and possible to make better design decisions and produce apps that seek to reduce or decrease stress level by increasing number of ASL display and decreasing number of dynamic elements such as picture and video contents present on a page depending on the type of pages on the website.
VI. SIMULATING PHYSIOLOGICAL DATA

The purpose of simulating data that satisfies constraints based on the physiological data collected, is to perform further validation on our model performance and results. Most form of simulated data acquisition method is based on specific mean and standard deviation used to generate random numbers that satisfies the condition of a specific dataset. For our purpose, we make the mean and variance exactly to the physiological pre-defined values. The variables are fitted, slackened and scaled to the right proportion as to emulate the original data. Given that \( X_1, X_2, \ldots, X_n \) are the original variables, then the new variable is obtained as:

\[
\text{New } X_i = c_1^2 \left( \frac{X_i - \bar{X}}{S} \right) + e_2
\]

Where \( \bar{X} = \frac{1}{n} \sum X_i \) is referred to as the simple mean and \( S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 \) is the sample variance such that the given sample mean of the set \( \text{New } X_i \) is simply in direct proportion to the value \( e_2 \) with equal variance of \( e_1 \). This also implies that an example of the form:

\[
D_i = n + (m-n) \left( \frac{X_i - m_{\text{min}}(X_{\text{min}}, X_{\text{max}})}{m_{\text{max}}(X_{\text{min}}, X_{\text{max}}) - m_{\text{min}}(X_{\text{min}}, X_{\text{max}})} \right)
\]

produces a dataset of \( D_1, \ldots, D_n \) restricted to interval \( (n, m) \). These form of scaling generally alters the distribution pattern of the data, even for spatially distributed datasets. For the purpose of simulating normal multivariate data, the R package was used. It allows for the simulation of datasets that can emulate the behavior of the original data within the context of normally distributed data. The data are simulated specifically from the conditional distributed data of a normally distributed variable, given that the sample and mean and the co-variance are equal to a pre-specified value. One important thing to note is that the marginal distributions are not exactly normalized. The following sections demonstrates the results obtained by carrying the above stated processes.

VII. DISTRIBUTION OF SIMULATED DATA

From the hypothetical data generated, we have \( z \) has the Baseline, Amplitude, MST, MSP, MappedFX and Status as the best selected variables from our original model. The Status is the only categorical variable from the entire dataset selected. From the data set, amplitude and MST of the response has the lowest p-values (\( p_1 = 7.33 \times 10^{-6} \) and \( p_2 = 0.000175 \)) and are best predictors for the stress level. Looking at the density and histogram graph (Figure 12), it is noted that censoring in the amplitude values have far more reaction at 20 than the rest of the distribution. The simulated data also emulates the associative characteristics of the original data based on Pearson correlation, where there is obvious absence of correlation among attribute has shown in Figure 14, thought the Naive Bayes classifier is suitable for such distribution since all variable seem to be independent of each other. We resolve to the bootstrapping process to produce normal proportionally distributed data to determine its significance of data generalization. As a form of model comparison using linear model, the best fit also turn out to be the original model that comprises all attributes with confidence interval at 1.522. The receiver operating characteristics (Figure 15) shows with area under the curve as 0.57, which is far less when compared to the original model with area under the curve 0.91. This implies the generalization of the model is significant enough to be recognized as valid.
Figure 13 describes the relative likelihood for the simulated variables to take on a given value, based on the stress level of the entire distribution, the integral over the space for stressed used of equal to one. This is relative to the number of variables selected for best model, such that the domain in the n-dimensional space of the values of the variables results to the probability that a realization of the dataset falls with the given domain of stress level.

Fig. 13: Probability density function for simulated data

For a given distribution of the form above, the density plot shows that there is no significant overlap, which implies that the users were close to been relaxed than stress, so it is more realistic to investigate what contributes to the relaxed state of users during HCI events. The simulated data can also tell us or describe how the relaxed state of users is actually a better status to defining the stress levels than neutral or stress mode.

Fig. 14: Correlation matrix of Simulated attributes

VIII. STANDARD ERRORS AND CONFIDENCE INTERVALS FOR BOOSTRAPPING

To construct a confidence interval for a populated parameter, an interval centered at a point estimate with a margin of error equal to twice the standard error was utilized. The size of the standard error was estimated by applying bootstrap and resampling many samples with substitutions from the original or simulated dataset. Each size has the original sample computed at a point estimate for each variable, in order to find the standard deviation of the bootstrap stats. Figure 16 shows that the distribution is nearly symmetrical and also bell-shaped. The standard error is given as 0.61 and the constructed confidence interval with margins of error rounded to one decimal place as -1.8 and 1.8. This then interprets that the 95% confident that the mean latency time for response peaks among users of the webpages who definitely did not experience stress is in the interval of approximately 2 minute latency. The accuracy of this diagnostics depends basically on the original data used for modelling and as part of representation from the population of importance. As the experiment conducted is based on non-obstructive method, it seems realistic.
IX. CONCLUSION

To further validate, we can examine the references to gain more understanding about the conditions associated with sampling procedure like for example, investigate on whether users drawn to aesthetics of the webpages reduces stress levels, or changing method of data collection like adopting to obstructive and subjective measures. Today it is possible to actually understand users’ cognitive and unconscious thought processes, due to constant struggle to find the best methods to allow us to understand this concept. Implicit reactions and emotions can be measured, while users interact with elements on a webpage, this is because the benefit of physiological data is primarily that the measure of task performance is due to objective rather subjective measures. Data can be collected and examined in either delayed or real time without interrupting data collection. The reason for application of modeling is to see if pattern can be identified from the quantitative data obtained from physiological measures and eye tracking of...
participants. Based on the performance of models, it is fit to conclude that patterns can be learned in HCI-HPR. Though the custom algorithm used here might be complex and traditionally exemplified, it will be more feasible to produce a prototypical tool of its form for easy data analyses and user friendly. One of the future contributions will be to develop an integrative window based model for simulation of human physiological response to both static and dynamic contents of a webpage. This would facilitate analysis and evaluation protocol of HCI-HPR and this could contribute and help achieve better design decisions in shorter period of time than the normal traditional methods.

ACKNOWLEDGMENT

The authors would like to the Web Ergonomics Lab in University of Manchester, where the experiment was conducted.

REFERENCES