On-body Sensing and Signal Analysis for User Experience Recognition in Human-Machine Interaction

Roya Haratian
Faculty of Science and Technology
Bournemouth University
Poole, UK, BH12 5BB
e-mail: rharatian@bournemouth.ac.uk

Tijana Timotijevic
School of Electronic Eng. and Computer Science
Queen Mary University of London
London, UK, E1 4NS
e-mail: t.timotijevic@qmul.ac.uk

***Abstract*—In this paper, a new algorithm is proposed for recognition of user experience through emotion detection using physiological signals, for application in human-machine interaction. The algorithm recognizes user's emotion quality and intensity in a two dimensional emotion space continuously. The continuous recognition of the user's emotion during human-machine interaction will enable the machine to adapt its activity based on the user's emotion in a real-time manner, thus improving user experience. The emotion model underlying the proposed algorithm is one of the most recent emotion models, which models emotion's intensity and quality in a continuous two-dimensional space of valance and arousal axes. Using only two physiological signals, which are correlated to the valance and arousal axes of the emotion space, is among the contributions of this paper. Prediction of emotion through physiological signals has the advantage of elimination of social masking and making the prediction more reliable. The key advantage of the proposed algorithm over other algorithms presented to date is the use of the least number of modalities (only two physiological signals) to predict the quality and intensity of emotion continuously in time, and using the most recent widely accepted emotion model.**

Keywords-on-body sensing; signal analysis; user experience; dynamic neural field

I. INTRODUCTION

Experience of emotion is one of the key aspects of user experience affecting to all aspects of the human-machine interaction (HMI), including utility, ease of use, and efficiency [1], the machine's ability to recognize user's emotion during user-machine interaction would improve the overall HMI usability. The machines that are aware of the user's emotion could adapt their activity features such as speed based on user's emotional state. This paper focuses on emotion recognition through physiological signals, as this bypasses social masking and the prediction is more reliable.

Emotions are represented through three models, discrete or categorical, dimensional and appraisal models. In discrete or categorical model emotions are represented by discrete labels such as happiness, anger, disgust, sadness, anxiety and surprise which are the basic emotions. In the dimensional model, emotions are labelled in a two or three dimensional space. It is shown in [2] that three independent and bipolar dimensions, valance (pleasure-displeasure), degree of arousal and dominance-submissiveness are necessary and sufficient to adequately define emotional states. The validity and reliability of two-dimensional emotion space (valence-arousal) is examined in [3]. In the appraisal model, emotions are defined as processes. Appraisal of significance of an event is shown by emotion quality, intensity and duration while considering dynamics of emotion.

To measure the pleasure, arousal and dominance associated with a person's affective states the Self-Assesment Mankin (SAM) is used in literature where subjective reports were measured to a series of pictures that varied in the valance, arousal and dominance dimension [4]. Geneva Emotion Wheel (GEW), Fig. 1, which is software-based instrument for measuring emotion in real-time application, is introduced in [5]. The GEW presented on a computer screen, and all members of an emotion family were identified by a specific label, which became visible when moving the mouse across a circle. The users are asked to rate the intensity of an experienced or imagined emotion on the basis of the distance from the hub of the wheel and the size of the circles. The Geneva Emotion Wheel can be considered the first instrument to design the dimensional layout of the emotion qualities on pure appraisal dimensions (arrangement of emotion terms in two dimensional space) and the intensity of the associated subjective feeling (distance from origin). The last version of GEW is introduced in [6] which is the model used in this paper.

Physiological signals are collected through on-body sensors and used to identify user experience through the predicted wearer's emotion. The sensors gather physiological parameters like heart rate, body temperature, motion, galvanic skin response, etc. The correlation of bio-signals with user's emotion wearing the bio-sensors is shown in [7]. The type, position and number of sensors used depend on the application of the wearable systems. In this context, the wearable system design needs to consider the wearers' comfort and fitting requirements while considering measurement performance. For example, the weight and the size of the system need to be kept small and the system should not interfere with the user's movements or actions [8]. Considering the correlation of bio-signals with emotions, the user's experiences is recognized using physiological data through signal monitoring, analysis and development of advanced algorithms [9].

For this purpose, a mapping of physiological signals variation into emotion's model is required. Physiological or bio-signals are used in research field of affect-sensing to identify emotions. Galvanic Skin-conductivity Response, GSR, provides a measurement of the of skin conductance which increases linearly with a person's level of overall arousal or stress. Blood Volume Pulse, BVP, is indicator of blood flow. Since each heart beat (or pulse) presses blood through the vessels, BVP can also be used to calculate heart rate and inter-beat intervals. Heart rate increases with negative valence emotions such as anxiety or fear. Electromyography, EMG, measures the muscle activity or frequency of muscle tension, correlate with negative valence emotions. Respiration Rate, RSP, measures how deep and fast a person is breathing. Slow and deep breathing indicates a relaxed resting state while irregular rhythm, quick variations, and cessation of respiration corresponds to more aroused emotions like anger or fear. Electrocardiogram, ECG, signal measures activity of the heart. It can be used to measure heart rate and inter-beat intervals to determine the heart rate variability (HRV). A low HRV can indicate a state of relaxation, whereas an increased HRV can indicate a potential state of mental stress or frustration [7].

The proposed algorithm in this paper for the purpose of affective states recognition uses only two bio-signals: skin conductivity and pulse rate. As the intensity of arousal is linearly correlated with skin conductivity and valence is correlated negatively with blood volume pulse rate, it is possible to extract arousal-relevant features from GSR and valance-relevant features from BVP.

In this paper, we first review different emotion models, and the correlation of physiological signals with emotions is introduced. Section II, reviews emotion recognition algorithms developed to date. Section III introduces the new signal processing algorithm that implements GEW framework and uses only two physiologic signals. Results and discussion for the algorithm is presented in section IV and the paper is concluded in section V.

II. Literature Review

The algorithms used in the literature for recognition of emotion through physiological signals are mainly based on classification approaches. In [10], an emotion recognition system is developed to classify emotional states from physiological signals gathered from one subject over many weeks of data collection. The emotions were induced through musical performances. Features such as mean, standard deviation of raw signals and first and second difference of the raw signal is extracted and are classified into 8 discrete emotion states by using Feature Projection method. The accuracy of 81 percent is achieved on eight classes of emotions. In another study, three physiological signals such as EEG, skin conductance and pulse were collected from subjects while using audio-video contents as a stimulus. For discrete emotion recognition, Support Vector Machine is used to design discrete emotion classifier. The recognition rate of 41.7% for five emotions and 66.7% for three emotions were attained [11].

An emotion recognition approach based on physiological changes in music listening is introduced in [12]. Multiple subjects over many weeks were participant of musical induction while four-channel biosensors were used to measure electromyogram, electrocardiogram, skin conductivity, and respiration changes. A wide range of physiological features from various analysis domains, including time/frequency, entropy, geometric analysis, sub-band spectra, multiscale entropy, etc., were used in order to find the best emotion-relevant features and to correlate them with emotional states. Classification of four discrete musical emotions (positive/high arousal, negative/high arousal, negative/low arousal, and positive/low arousal) is performed by using an extended linear discriminant analysis. An improved recognition accuracy of 95 percent and 70 percent for subject-dependent and subject-independent classification, respectively, is achieved.

An emotion recognition system based on physiological signals is introduced in [13]. It uses five physiological signals such as blood volume pulse, electromyography, skin conductance, skin temperature and respiration. Features such as the mean and the standard deviation of the raw signals and first differences of the raw signals and the second differences of the raw signals are calculated. Six basic emotions such as Amusement, Contentment, Disgust, Fear, No emotion (Neutrality) and Sadness are induced through the international affective picture system (IAPS) while for each emotion ten images are presented during 50 seconds. Two pattern classification methods, Fisher discriminant and SVM method are used and compared for emotional state classification which reaches classification performance as 92% over six emotional states.

A comparison table of reviewed previous works is available in [14]. According to the literature, several bio-signals such as Electromyogram, Electrocardiogram, Electro-dermal, Activity Skin, Temperature, Blood Volume, Pulse and Respiration are used to recognize discrete emotions such as Sad, Anger, Stress, Surprise and so on. The stimuli used are audio-visual such as IAPS, music, movies. The extracted features from bio-signals are typically mean, standard deviation of raw signals and their first derivative, high frequency and low frequency powers and so on. The classification approaches are Support Vector Machine, Linear Discriminant Analysis, K-Nearest Neighbour, Bayesian Networks, and Neural Network and so on varying in accuracy depending on whether the approach is user-dependant or independent. An approach going a step further from discrete emotion recognition toward recognition of emotion in a continuous space is proposed in [15]. The algorithm for recognition of affective valance and arousal uses a multiclass arousal/valance classifier. By dividing arousal and valance axes into four regions, sixteen classes of emotions are formed. Similar approach using different classification algorithms is proposed in [16] and [17].

A continuous emotion recognition algorithm is introduced in [18] using fusion of facial expression, gesture and audio cues. In the paper, level of arousal and valance is predicted over time continuously. Using long short-term memory (LSTM) Neural Network as building unit for layers of a recurrent neural network (RNN), the longer temporal dependency features are learnt in the algorithm. It enables continues prediction of valance and arousal over time. In [19] a dynamic model is proposed based on dynamic field theory (DFT) enabling prediction of emotion intensity. DFT can be seen as a generalization of recurrent neural networks to continuous dimensions, adding a functional interpretation to each layer. It shows improvements in emotion recognition accuracy to the previous common approaches to emotion intensity estimation.

This review of literature on the emotion recognition using physiological signals shows that the developed algorithms use more than two bio-signals. In the algorithm proposed in this paper only two bio-signals are used. The emotion models used in the literature are mainly discrete and the classification of emotions is based on discrete classes or on finding out the levels of valence or arousal. The continuous emotion recognition approaches are either need many modalities of physiologic signals, or only predicts the intensity of emotion. The algorithm proposed here uses the most recent emotion model, GEW, Fig. 1, which allows to map emotions in a two dimensional continues space. The proposed algorithm is able to predict intensity and quality of emotion continuously as well as duration of emotion.

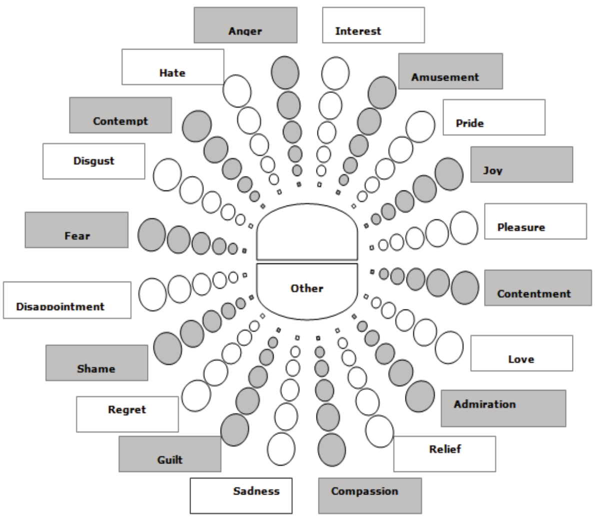

Figure 1. Geneva Emotion Wheel, 3rd version [6]

III. Method

For the purpose of developing an algorithm that can continuously predict affective state, we need to consider a dynamic model for signal processing of gathered physiological signals through the designed experiments. In this section, the proposed model followed by the signal processing approach is explained.

A. The Model

The proposed model in this paper is based on dynamic system theory providing the language in which embodied and situated stance can be developed into a scientific theory. Dynamic field theory (DFT) is a mathematical and conceptual framework which was developed to model embodied human cognition [20]. The basic computational element of DFT is Dynamic Neural Field (DNF). The response of a neural population is modelled through mathematical representation of DNFs. The dynamics of DNF are mathematically formulize as follows.

$$\tau\dot{u}(x,t) = -u(x,t) + h + \int f(u(x,t'))\omega(x-x')dx' + S(x,t)$$

$$\omega(x-x') = c_{exc}\exp\left[-\frac{(x-x')^2}{2\sigma_{exc}^2}\right] - c_{inh}\exp\left[-\frac{(x-x')^2}{2\sigma_{inh}^2}\right]$$

$$f(u(x,t)) = \frac{1}{1+exp[-\beta u(x,t)]} \qquad (1)$$

In (1) formula, $u(x,t)$ is the activation of DNF over dimension x, to which the underlying neural population is responsive. h is a negative resting level and $S(x,t)$ is an external input. The lateral interactions in DFT are modelled through a short-range excitation and a long-range inhabitation where σ_{exc}, σ_{inh}, c_{exc} and c_{inh} are width and amplitude of the excitatory and inhibitory parts of interaction kernel. $f(u(x,t))$ shapes the output of DNF through a sigmoid function where β is the slope of sigmoid [21].

In the proposed DNF, there are two one dimensional input neurons, U_1 and U_2, where the two bio-signals are the inputs to them. The model inputs are GSR which is correlated with arousal and Pulse rate which is correlated with valance in the emotion model. To stabilise the signals there are two memory neurons, M_1 and M_2, where their outputs are the inputs of the emotion neuron, E. The two-dimensional emotion neuron is modelling the emotion space with valance and arousal dimensions. The proposed neural population is shown in Fig. 2.

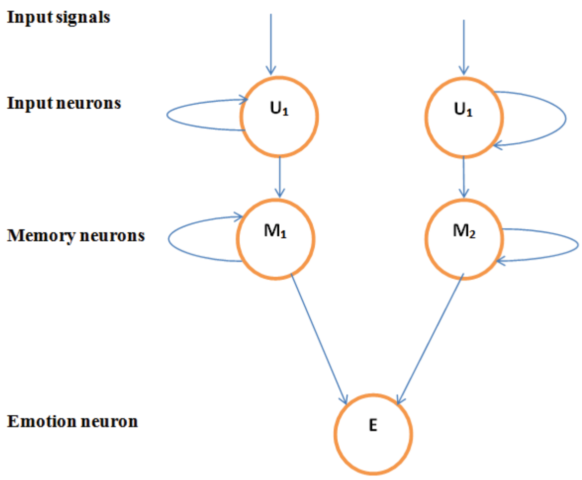

Figure 2. The proposed neural population for the model

In the proposed neural population, the input bio-signals are fed into the input one-dimensional fields and passing through the middle input neurons acting as memory neurons. The output signals of this layer are considered as two inputs of the two dimensional emotion neural field with valance and arousal axes where the length of each axe is assumed 100.

Transferring the emotion representation from valance-arousal Cartesian space into polar quality and intensity of emotion representation is formulised through equation (2). In the polar system emotion is modelled as a vector, $I(t)$, where the length of the vector is emotion intensity and the vector's angle, $\theta(t)$, is the quality, $Q(t)$, of emotion as shown in Fig. 3. Considering a two dimensional neural filed with size of 100 for each axes, the vector transformation of emotion will be as (2) where the centre of the polar coordinate system would be in the middle of the Cartesian space. In the 2D GEW space, the centre of Cartesian coordinate system with valance, $v(t)$, and arousal, $a(t)$, axes is in the left bottom side of the space.

$$v(t) = 50I(t)\cos\theta(t) + 50 \qquad (2)$$

$$a(t) = 50I(t)\sin\theta(t) + 50$$

$$where\ \ 0 < I(t) < 1, \quad 0 < \theta(t) < 360$$

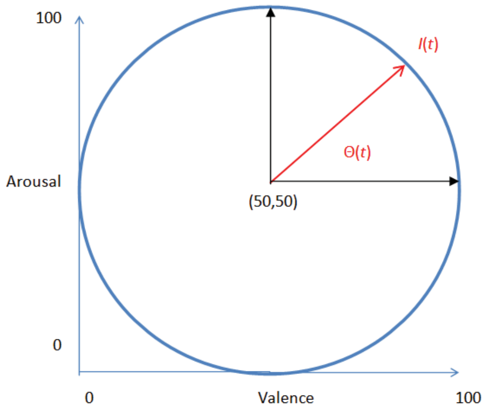

Figure 3. Transformation of emotion arousal-valance representation in Cartesian space into quality and intensity vector in polar space

B. *Signal Processing*

As it was discussed in the literature review, the intensity of arousal is linearly correlated with skin conductivity and valence is correlated negatively with pulse rate. Therefore, the GSR and pulse rate are inputs of the system. The extracted feature from GSR is the rate of change and from Pulse rate is its inverse as the pulse rate increases with negatively valance emotions.

The model parameters are τ as the time constant of each neuron, β as the slope parameter of output function for each neuron, as well as the coupling coefficients and kernel excitation and inhabitation parameters. After initialisation of model parameters, a cost function will be calculated using the training data. The model simulator estimates the quality and intensity with regard to the input training data of multiple trials (Fig. 4). To identify model parameter, the cost functions need to be minimised.

The cost function is calculated for training trials and the average of all mean square errors, MSEs, of multiple trials is used as the function which should be minimized to derive the model parameters.

$$Cost_{quality} = \frac{1}{T*L}\sum_l \sum_t \left|\hat{Q}_l(t) - Q_l(t)\right|^2 \qquad (3)$$

$$Cost_{intensity} = \frac{1}{T*L}\sum_l \sum_t \left|\hat{I}_l(t) - I_l(t)\right|^2 \qquad (4)$$

In (3) and (4), the model quality and intensity estimate is represented by $\hat{Q}_l(t)$ and $\hat{I}(t)$ while the targeted ones are represented by $Q_l(t)$ and $I(t)$. The total recording time, t, for each trial is represented by T and total number of trails, l, by L. Following the explained procedure enables estimation of quality and intensity continuously. The model has 32 parameters which have to be estimated through optimization of cost function. For solving the optimization problem Genetic Algorithm was used.

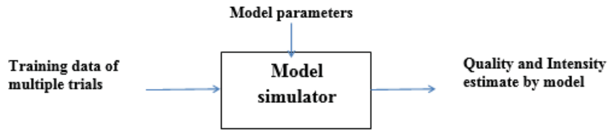

Figure 4. Model parameter estimation process

The designed experiment consists of several trials of showing series of pictures which are rated by subjects while physiological signals are collected. The pictures are from IAPS, International Affective Picture System, database designed to provide a standardized set of pictures for studying emotion and attention. The dataset has been widely used in psychological research and was developed by the National Institute of Mental Health Centre for Emotion and Attention at the University of Florida.

During the time that the picture is on the screen, the rating is made on a 2D labelling system. Each series of pictures belongs to one emotion quality labelled on the outer borders of the 2D labelling system. Subjects can rotate a knob to the emotion quality label and move the indicator on radial line of the circle according to the emotion intensity they feel during viewing the pictures, the closer to the centre, the less intense emotion felt during viewing the picture based on GEW emotion model. The time length of showing each picture is 15s and it takes about 10s for an emotion to be induced using pictures. By the end of each trail a black screen is shown to help the subject reach neutral state for the felt emotion.

IV. RESULTS

The preliminary results of algorithm validation on a single subject trial are presented in this section to demonstrate the feasibility and utility of the algorithm. The extracted feature from GSR and Pulse rate as the input bio-signals are the rate of change and the inverse, respectively. The raw bio-signals are filtered to remove the low frequency noise contents and then normalised. Fig. 5 shows the neural field population initial response to the input signals. It shows the activation of the field while the Gaussians are interacting and competing with each other to reach to a steady state with one dominant Gaussian on emotion field. After the initial phase, during the estimation process it appears that one of the Gaussians tends to dominate others as it shown in Fig. 6. By the end, one of the Gaussians dominates others and its location on the field with valance and arousal axes indicating the estimated emotion (Fig. 7). In this example, the vector length is %43 and the angle is 45 degree which indicates the estimated emotion is Pride with %43 intensity based on the emotion model in Fig. 1. We need to explore the introduced approach further on wider population range of emotion with more number of subjects.

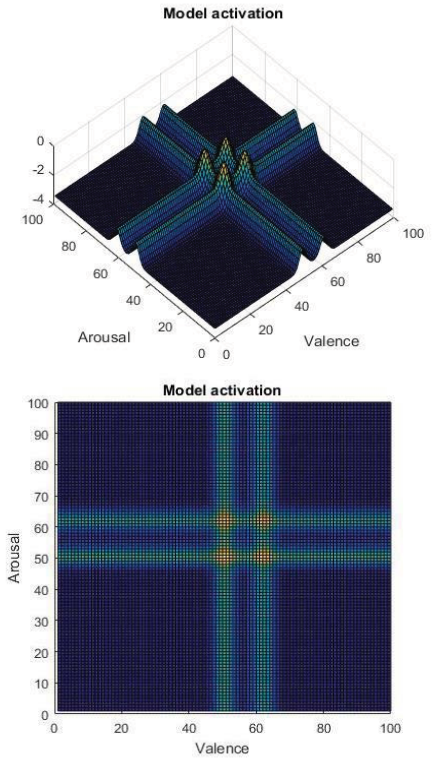

Figure 5. Snapshot of the neural field initial response to the bio-signals, the three dimensional view followed by a two dimensional one

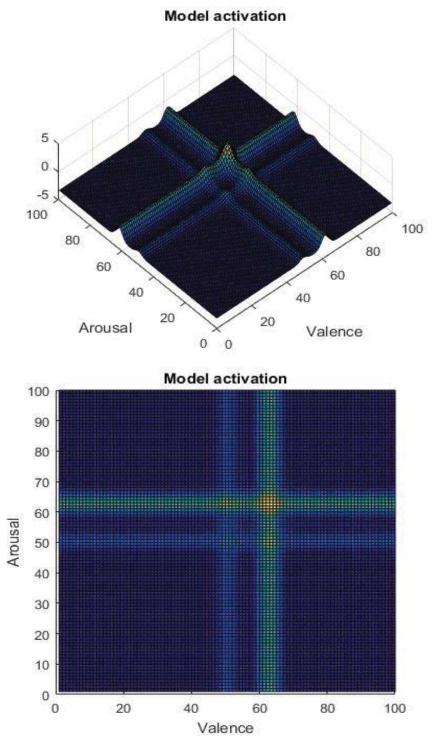

Figure 6. Snapshot of the neural field response to the bio-signals during the estimation process, the three dimensional view followed by a two dimensional one.

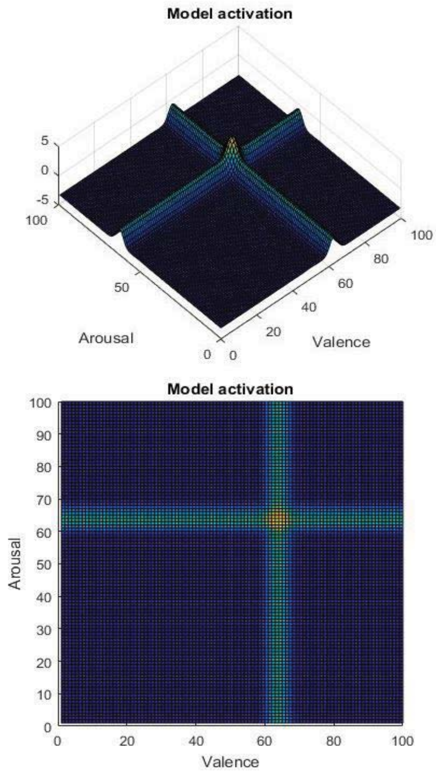

Figure 7. Snapshot of the estimated emotion in the two dimensional neural field, the three dimensional view followed by a two dimensional one

V. Conclusion

The paper proposes a new algorithm to recognize human affective state continuously given physiological signals and reports preliminary results that demonstrate the proof of concept. It is motivated by the context of a continuous human-machine interaction in which the machine is expected to continuously adapt to human emotional state. The two key novelties here are the use of the minimum number of modalities (only two physiologic signals) for emotion estimation, which make the algorithm feasible in a wide range of practical user scenarios; and estimation of emotion quality and intensity in a time and space continuum of a two-dimensional space of valence and arousal. Because of high number of parameters that are required to be adjusted in the training phase of the algorithm, there is need for high power processors to increase the speed of process. Despite of the existing drawback, development of the new proposed algorithm will enhance the interaction between human and machine in a more coordinated and synchronous manner. It enables machines to interact with the users in a more conscious environment with wide range of applications from human-machine interaction to healthcare and entertainment industry.

References

[1] R. Setchi, O. Asikhaia, “Exploring User Experience with Image Schemas, Sentiments, and Semantics” *IEEE Transactions on Affective Computing*, early access, 2018

[2] J Russell, "Evidence of a three-factor theory of emotions" *Journal of Research in Personality*, pp. 273-294, 1997.

[3] E. Schubert, "Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space" *Australian Journal of Psychology*, pp. 154-165, 1999.

[4] M.Bradley, and P. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential" *Journal of behavior therapy and experimental psychiatry*, pp. 49-59, 1994.

[5] K. R Scherer, "What are emotions? And how can they be measured?" *Social science information*, pp. 695-729, 2005.

[6] K. R. Scherer, V. Shuman, J. R. J. Fontaine & C. Soriano, "The GRID meets the Wheel: Assessing emotional feeling via self-report. In: Components of emotional meaning: A sourcebook" Oxford: Oxfor University Press, 2013.

[7] H. Gunes, "Automatic, dimensional and continuous emotion recognition" *International Journal of Synthetic Emotions*, pp. 68-99, 2010.

[8] R. Haratian, T. Timotijevic, and C. Phillips, "Reducing power and increasing accuracy of on-body sensing in motion capture application" *IET Signal Processing Journal*, no.10. vol. 2, pp. 133-139, 2016.

[9] O. C. Santos, R. Uria-Rivas, M. C. Rodriguez-Sanchez, and J. G. Boticario, "An Open Sensing and Acting Platform for Context-Aware Affective Support in Ambient Intelligent Educational Settings" *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3865, 2016.

[10] R Picard, E. Vyzas, and E. Healey, "Toward Machine Emotional Intelligence:Analysis of Affective Physiological State" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1175-1191, 2001.

[11] K. Takahashi, "Remarks on emotion recognition from multi-modal bio-potential signals", pp. 1138-1143, 2004.

[12] J. Kim, and E. Andre, "Emotion Recognition Based on Physiological Changes in Music Listening" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2067-2083, 2008.

[13] C. Maaoui, and A. Pruski, "Emotion recognition through physiological signals for human-machine communication". *INTECH* Open Access Publisher, 2010.

[14] S. Jerritta, M. Murugappan, R Nagarajan, and K. Wan, "Physiological Signals Based Human Emotion Recognition: A Review*". Penang*, pp. 410-415, 2011.

[15] M. Nardelli, G. Valenza, A. Greco, A. Lanta, E. Scilingo, "Recognizing Emotions Induded by Affective Sounds through Haert Rate variability " *IEEE Transactions on Affective Computing*, vol. 6, no. 4, 2015.

[16] G. Valeza, A. Lanata, E. Scilingo, "the role of nonlinear dynamics in affective Valence and arousal recognition", IEEE Transactions on Affective Computing, Vol 3, no. 2, 2012.

[17] W. Wen, G. Liu, N. Cheng, P. Shangguan and W. Huang, "Emotion recognition based on multi-variant correlation of physiological signals" *IEEE Transactions on Affective Computing*, vol. 5, no. 1, 2014.

[18] M. Nicolaou, H. Gunes, M. Pantic, "Continiuse Prediction of spontaneouse affect from multiple cues and modalities in valance-arousal space" IEEE Transactions on Affective Computing, vol. 2, no. 2, 2011.

[19] R. Jenke, A. Peer, "A cognitive architecture for modeling emotion dynamics: Intensity estimation from physiological signals" Cognitive Systems Research, Vol. 49, pp. 128-141, 2018

[20] Y. Sandamirskaya, "Dynamic neural fields as a step toward cognitive neuromorphic architectures" *Frontiers in Neuroscience*, 2014.

[21] G. Schoner, "Dynamical systems approaches to cognition" *The Cambridge Handbook of Computational Psychology*, 2007.