A Regression-based Method for Lightweight Emotional State Detection in Interactive Environments

Pedro A. Nogueira[1], Rui Rodrigues[2], Eugénio Oliveira[2], Lennart E. Nacke[3]

1: LIACC – Artificial Intelligence and Computer Science Lab., University of Porto, Portugal
2: INESC TEC and Faculty of Engineering, University of Porto, Portugal
3: HCI and Game Science Group, UOIT, Canada
{pedro.alves.nogueira, rui.rodrigues, eco}@fe.up.pt, lennart.nacke@uoit.ca

Abstract. With the popularity increase in affective computing techniques the number of emotion detection and recognition systems has risen considerably. However, despite their steady accuracy improvement, they are yet faced with application domain transferability and practical implementation issues. In this paper, we present a novel methodology for modelling individuals' emotional states in multimedia interactive environments, while addressing the aforementioned transferability and practical implementation issues. Our method relies on a two-layer classification process to classify Arousal and Valence based on four distinct physiological sensor inputs. The first classification layer uses several regression models to normalize each of the sensor inputs across participants and experimental conditions, while also correlating each input to either Arousal or Valence. The second classification layer then employs decision trees to merge the various regression outputs into one optimal Arousal/Valence classification. The presented method not only exhibits convincing accuracy ratings – 89% for Arousal and 84% for Valence - but also presents an adaptable and practical approach at emotional state detection in interactive environment experiences.

Keywords: Affect recognition, regression analysis, affective computing, games, physiology, galvanic skin response, heart rate, electromyography.

1 Introduction

Video games have pioneered the most recent breakthroughs in various computer science fields, such as computer graphics, artificial intelligence and human-computer interaction. This popularity, along with their considerable emotional influence potential [1], games and digital media in general have also become popular study cases for affective computing experiments. However, despite the consecutive advances in gaming technology, there is still a distinct lack of affective experience evaluation tools. These tools are not only needed to perform usability tests on traditional applications, but are also a crucial and necessary first step for more complex emotionally reactive applications [2,3]. Given this current immediate need and complex design of future applications using these systems, means that they must not only be sufficiently accu-

rate (depending on each applicational case), but also present a general approach that can easily be adapted to various situations, while requiring minimal calibration steps.

In our research we are interested in how to develop a physiologically based emotion detection system that offers a balanced trade-off between accuracy, adaptability and calibration. Throughout this paper we justify the need for such a system, while also describing the various approaches taken in the literature, how they relate to one another and, finally, the process through which we designed the proposed method.

2 Emotional Recognition Techniques

Various taxonomies exist for emotional detection systems and each one has its own dedicated (and in some cases extensive) literature. These taxonomies rely on diverse input signals, such as interaction features, body motion, facial expressions or physiological measures. However, the latter ones have proven to be the most reliable and adaptable and, as such, are the ones we have based our system on.

2.1 Physiological Emotion Recognition Techniques

Physiological emotion recognition techniques attempt to explore possible correlations between game (or otherwise) events and physiological alterations. This type of approach usually employs multiple input modalities for real-time applications [4]. These techniques can be further divided into model-based and model-free types. While model-based approaches link physiological changes to popular models derived from emotion theories (e.g. Russell's popular arousal and valence dimensions [5] or Plutchik's Emotion Wheel [6]), model-free techniques build their mappings based solely on (almost exclusively subjective) ground truth annotations. However, systems may not exclusively rely on these two types. For instance, a hybrid approach may be chosen instead of adopting either a model-based approach that builds on a theoretical framework's pre-determined mappings between affect and physiological metrics, or a mode-free approach that seeks these mappings in the annotated data but assumes nothing about the underlying structure of this function. In fact, many known systems use the latter approach; they assume a theoretical model of emotion as their structure and build the mappings by asking users to rate their experiences on it [4].

Various successful attempts have been made in the field of emotion recognition using the various types of objective modelling techniques aforementioned – although the most popular ones are clearly the hybrid ones. For instance, Chanel [7] was able to classify arousal using naïve Bayes classifiers and Fisher Discriminant Analysis (FDA), based on Electroencephalogram (EEG), skin conductance (SC), blood volume pressure (BVP), heart rate (HR), skin temperature and respiration rate measures. Complementary to this work, Leon [8] proposes the classification of valence in three intensity levels, using different measures (SC, its time gradient and derivative, HR and BVP), along with auto-associative neural networks.

Also using neural networks, Haag et al. [9] propose employing EMG, SC, skin temperature, BVP, ECG and respiration rates to classify emotional states, reporting en-

couraging results (89% accuracy for arousal and 63% for valence, with a 10% error margin). A more detailed study has also found proof that moderately complex features extracted from HR, SC and BVP can be used to predict emotions, such as "fun" [10]. Within the proposed line of low calibration approaches, work by Vinhas et al. [11] proposes a system capable of measuring both arousal and valence in real-time, using the subject's SC response and HR derivative. A key factor of this work is that it introduced a continuous classification of arousal and valence, thus increasing the state of the art's maximum granularity. Finally and similar to Vinhas, Mandryk presents an approach based on fuzzy logic that classifies ECG, EMG and SC measurements in terms of both arousal and valence [12].

2.2 Employed Physiological Metrics

Based on the discussed literature, we conducted a survey on the most successful physiological channels and features. In decreasing order of importance, four main factors were taken into account: (1) Precision: How accurately can the signal be interpreted in terms of either arousal or valence? (2) Sensor Reliability: How likely is the sensor to fail and how much calibration does it require? (3) Signal Processing: How much signal processing (e.g., filtering, noise reduction) does the channel require to allow extracting meaningful features and is this processing cost affordable in a real-time scenario? (4) Intrusiveness: Would the required apparatus interfere with the gameplay experience, potentially biasing it in any significant way?

This survey led us to adopt skin conductance, facial electromyography and electrocardiography-based metrics. All of these channels have become popular in the affective computing literature by providing the most accurate and interpretable measurements. A brief description of each metric follows.

Electrodermal Activity

Electrodermal activity (EDA), usually measured in the form of skin conductance (SC), is a common measure of skin conductivity. EDA arises as a direct consequence of the activity of eccrine (sweat) glands. Some of theses glands situated at specific locations (e.g. palms of the hands and feet soles) respond to psychological changes and thus EDA/SC measured at these sites reflects emotional changes as well as cognitive activity [13]. In terms of psychophysiological correlations, SC has been linearly correlated with arousal [7,8,9] and extensively used as stress indicator [11], in emotion recognition [8,12] and to explore correlations between gameplay dimensions [14]. We measured SC using two Ag/AgCL surface sensors snapped to two Velcro straps placed around the middle and index fingers of the non-dominant hand [13].

Cardiovascular Measures

The cardiovascular system is composed by the set of organs that regulate the body's blood flow. Various metrics for its activity currently exist, among which some of the most popular ones are: blood pressure (BP), blood volume pulse (BVP) and heart rate (HR). Deriving from the HR various secondary measures can be extracted, such as inter-beat interval (IBI) and heart rate variability (HRV). HR has been correlated with arousal [11,12], anxiety [14] and immersion [1]. Along with the HR's derivate, the HRV has also been suggested to distinguish between positive and negative

emotional states (valence) [11]. In our experiments HR and HRV were inferred from the participants' BVP readings using a finger sensor.

Electromyography

Electromyography (EMG) is a method for measuring the electrical potentials generated by contraction of skeletal muscles [13]. Facial EMG has been successfully used to distinguish valence in gameplay experiences [12]. In the former experiences, Hazlett describes the zygomaticus major (cheek) muscle as significantly more active during positive events and the corrugator supercilii (brow) muscle as more active in negatively-valenced events. Conati's more recent attempts at the identification of user emotions using this same approach further validate the previously discussed results [15]. We measured facial EMG through surface electrodes on the corrugator supercilii and zygomaticus major muscles.

3 Methods

In order to gather enough ground truth data to determine whether we could build an emotion recognition system for gameplay and multimedia content, we conducted a series of controlled experiments with a total of twenty-two healthy participants. In line with the literature, the applied experimental protocol was initially tested and refined in an iterative prototypical cycle using several pilot studies comprising a total of 10 participants. The results reported in this paper apply to the data collected and processed for the remaining twelve participants in the final iteration of the experimental procedure [7,8,9,11,12]. Participants ranged from undergraduate students to more senior researchers and were aged 22 to 31 (M=24.83, SD=2.29). As with other studies [7,11,12], given the population distribution, we limit our findings to this specific demographic. Seven of the participants reported playing video games at least monthly, while the remaining ones reported sporadic activity, whenever big titles came out.

3.1 Experimental Conditions

Sessions were divided into three conditions designed for obtaining the necessary data samples to train our system. The first two conditions were aimed at eliciting extreme arousal and valence ratings: the first one was a long session of relaxing music; the second was playing the horror video game Slenderman, by Parsec Productions; and the third one aimed at eliciting neutral to mild reactions using 36 emotionally-charged images from the International Affective Picture System (IAPS) library [16].

In each of the experimental conditions, the participant's SC, facial EMG and BVP readings were recorded. SC was measured at the subject's index and middle fingers using two Ag/AgCL surface sensors snapped to two Velcro straps. BVP was measured at the thumb using a clip-on sensor. Facial EMG was measured at the zygomaticus major (cheek) and the corrugator supercilii (brow) muscles and, as previously mentioned, correlated with positive and negative valence, respectively [13].

Sessions were timed and conducted in a room with acoustic insulation, controlled lighting and temperature conditions. Participants were left alone in the room at the

beginning of each section. The only human interaction was during the relaxation and briefing periods in between conditions.

3.2 Experimental Procedure and Apparatus

After signing an informed consent form, participants underwent the experimental protocol. Each condition was preceded by a relaxation period of approximately 5 minutes, through which baseline (averaged) values for each channel were extracted. The participants then underwent each of the experimental conditions, whilst reporting their affect ratings. Experimental conditions were sub-divided into its constituent training samples, each with the same length as the event, plus a buffer length of 5 seconds at both its extremities. Regarding training samples, since we were interested in the participant's lowest emotional activation values for the relaxing music condition, in this case the sample was equal to the whole condition. On the remaining two conditions, each event – images for the IAPS condition and gameplay events for the Slenderman condition – was isolated and independently rated by the participants.

Regarding the annotation procedure, participants rated the training samples immediately after their presentation. The exception to this was the Slenderman condition, since interrupting the gameplay activity to rate each event was not only intrusive; it also implied our physical presence, which could contaminate the experience. As such, for this condition, the gameplay session was recorded by a commercial frame grabber (Fraps) at 30Hz and analysed in conjunction with the participant in a post-gameplay interview. Participants reported their absolute maximum arousal and valence ratings in a 10-point Likert scale, ranging from -5 to 5. Since it is harder for individuals to rate their mean affect over a 10 or 20-second time window than to isolate a single, more intense emotional peak, we chose to ask participants to rate each event according to their absolute maximum, as it introduced the least noise in the obtained ratings.

All of the sessions were performed on a MacBook Pro computer running Windows 7. The monitor was a 17" LCD display running at a resolution of 1080p. Physiological data was collected using the Nexus-10 hardware by Mind Media. Each condition had an average duration of 10 to 12 minutes, with the exception of the terror videogame, which usually took 15 to 20 minutes. Overall, from setup to debriefing, the experiment had an approximate duration of 2 hours.

3.3 Data Analysis and Feature Extraction

Regarding data analysis and feature extraction, HR, HRV and SC readings were collected at 32Hz, while facial EMG was collected at 2048 Hz. HR and HRV (R-R intervals) readings were computed from the raw BVP readings using the BioTrace+ software suite. All of the physiological signals were then exported to a tab-delimited text file sampled at 32 Hz for future analysis.

The exported raw data was then filtered for anomalies, which were deleted. The exception to this rule was the HR readings, which were filtered using an irregularity detection method that estimated the past variation of the HR signal and only allowed a 25% variation, as described in [9]. HRV was then recomputed from the corrected HR

values. Raw SC were corrected by subtracting baseline values and EMG amplitude values were then extracted from the raw EMG readings using the Root-Mean Square procedure. Subsequent signal analysis revealed no additional filtering was necessary. Sensor readings were then smoothed using a moving average filter [12,13]. HR and HRV were smoothed over a 2-second moving window, SC over a 5-second window and EMG over a 0.125-second window [12,13].

As previously mentioned, each condition was segmented into several training samples that were independently rated by participants using a 10-point Likert scale. As participants were asked to rate their absolute emotional peaks, so were these values extracted from each channel's training sample. Overall, an average of 230 data points were collected per participant: the minimum values for the relaxing music condition (5 data points, one per channel), 36 samples for the IAPS condition (180 data points), an average of 6 samples (30 data points) in the terror videogame condition (number of gameplay events varied) and 3 neutral baseline samples (15 data points).

4 Detecting AV States

This section details how we used the annotated ground truth from the three previously described experimental conditions to detect Arousal and Valence (AV) states from the participants' physiological data. The developed method categorizes participants' AV ratings through a two-layer classification process (Fig. 1). The first classification layer applies several regression models to each of the four physiological inputs, which allow us to simultaneously normalize these inputs and correlate them to the AV dimensions. The second classification layer then combines the arousal and valence ratings obtained from the previous step into one final rating by minimising their intrinsic error margins.

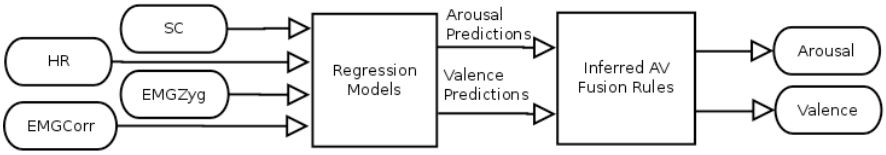

Fig. 1. High-level overview of the proposed system. The system accepts four pre-processed inputs and generates two AV outputs.

4.1 Physiological Input Regression Models

One of the most common issues with emotional recognition systems is the difficulty in obtaining an absolute scaling for the reported measures [8,9,11,12]. This is due to either the used material's inability to elicit meaningful arousal and/or valence alterations or by failing to annotate how each recorded event actually impacted the participant (i.e. assuming that a full emotional range elicitation occurred); usually, a combination of both these factors. In our experiments we tackled this challenge by exposing the participants with a wide range of emotional content; ranging from relaxing music

sessions, to a large representative sample from the IAPS library, and the psychological terror videogame Slenderman. The aforementioned issue is traditionally addressed by normalising recorded values across participants; a process that introduces a fairly high amount of noise in the data since not only assumes that all participants experienced the same overall emotional ranges, but more importantly assumes all of them experienced these ranges to their fullest extent. To counter this problem we approached it from a different perspective. Instead of a simple normalisation, we explored the correlation functions between each of the physiological channels and the participants' ratings using regression models. Regression analysis has been successfully used in the past for identifying emotional responses to both music excerpts [17] and audio-visual stimuli in PET scans [18], thus strengthening our choice.

Table 1. Ten-fold cross validation fitness values obtained for each regression model that obtained significance ($p<0.05$). Model complexity was controlled using stepwise-regression.

Physiological Channel	**AV Space Dimension**	**Adjusted-R^2 Model Values (μ, σ)**	
		Linear	**Polynomial**
SC	Arousal	0.90 ± 3.8^{-2}	0.95 ± 3.0^{-2}
HR	Arousal	0.68 ± 7.1^{-2}	0.74 ± 8.9^{-2}
EMG_{Zyg}	Valence (positive)	0.84 ± 1.4^{-2}	0.92 ± 1.6^{-1}
EMG_{Corr}	Valence (negative)	0.83 ± 7.9^{-2}	0.95 ± 7.5^{-2}
HR	Valence	0.88 ± 1.0^{-1}	0.96 ± 6.4^{-2}

By using the annotated data as our ground truth we were able to reflect each participant's characteristic physiological activation functions in their own model and, at the same time, relate them to their corresponding AV dimensions. We proceeded to explore the correlations of each annotated physiological channel to the AV dimensions and, apart from the HRV channel; have confirmed each of the correlations referred in the literature. However, despite HR negatively correlating with Valence, this was not observed for all three experimental conditions, as HR did not significantly fluctuate in the IAPS condition. As such, the results reported in Table I for the HR-Valence correlation refer only to the training samples extracted in the first two experimental conditions, not all three as per the remaining described correlations. In this exploratory phase we used both linear and non-linear (third degree polynomial) models. Model complexity was kept in check using bidirectional stepwise regression. This procedure was based on their adjusted-R^2 values in order to minimise the effects of a large number of predictors on the polynomial models. The final models were then re-evaluated for correctness using a 10-fold cross-validation scheme, as shown above on Table 1.

Although there are multiple accounts of a linear correlation between SC and Arousal [7-9,10,14], there is no common evidence that any of the remaining collected metrics correlate linearly with any of the AV dimensions. In fact, this seems highly unlikely. For example, it is not possible that the HR's signal distribution varies symmetrically with rising and decreasing arousal. An individual with an average of 80 beats per minute (BPM) may easily reach 140 BPM (60 BPM above his baseline)

when severely excited, but cannot naturally reach 20 BPM (the same 60 BPM bellow his baseline), even if completely relaxed.

Table 2. Statistical comparison between linear and third degree polynomial regression models.

Correlation Model	*t*-statistic	*p*-value
SC-Arousal	-2.397	0.035
HR-Arousal	-2.393	0.036
EMG_{Zyg}-Valence (positive)	-2.396	0.038
EMG_{Corr}-Valence (negative)	-2.825	0.018
HR-Valence	-3.297	0.007

Upon a statistical analysis between the linear and polynomial models we found non-linear correlations are indeed supported by our data. One-tailed paired t-tests using the models' adjusted-R^2 values as within-subject conditions revealed statistically significant ($p<0.05$) differences between the linear and polynomial models for: SC-Arousal (t = -2.397, p = 0.035), HR-Arousal (t = -2.393, p = 0.036), EMG_{Zyg}-Valence (t = -2.396, p = 0.038), EMG_{Corr}-Valence (t = -2.825, p = 0.018) and HR-Valence (t = -3.297, p = 0.007). Upon closer inspection, we found that although the polynomial SC-Arousal models presented a significant improvement over the linear ones, they were marginally different from the latter and only presented a better fit (5% better), while the same ratio on the remaining models presented improved fitness values from 9 to 14%. Resulting from this analysis we decided to maintain the linear model for SC-Arousal and opt for the polynomial models in the remaining ones.

4.2 AV Rating Fusion Models

Having obtained the various AV ratings from the regression models in the previous step it becomes then necessary to fuse them in order to obtain a final AV classification. To this end we are faced with two possible approaches: pure statistical/machine learning (ML) techniques or rule-based methods grounded in a theoretical emotional framework [5,6]. Given that our method is a hybrid of these two approaches and that we have already established the sensor mappings to our theoretical model's dimensions, we choose to leverage the benefits of a more data-driven technique (i.e. ability to find more complex relationships in the data and provide better generalization capabilities across multiple subjects without performing any strong assumptions on the data structure or properties) [4].

Taking into account the mentioned desired properties for our classification model, we selected regression trees for this task. It is also worth mentioning that while the regression models are participant-dependent the models used to fuse their classifications are not, since the regression models were meant to account for the participants' own physiological traits and normalise the captured readings. Thus, the amount of training data available for these methods was substantially higher and their results present a much stronger proof of generality towards AV classification.

The rationale behind this step was that the fusion models would be able to leverage each channel's error margins (Fig. 2) to decide how each one should be weighted with at each given time in the final classification. In other words, according to the regression model's error function at the current value, the sensor fusion model would be able to decide which channel had the highest probability of being at fault and thus minimise its contribution towards the final AV classification value.

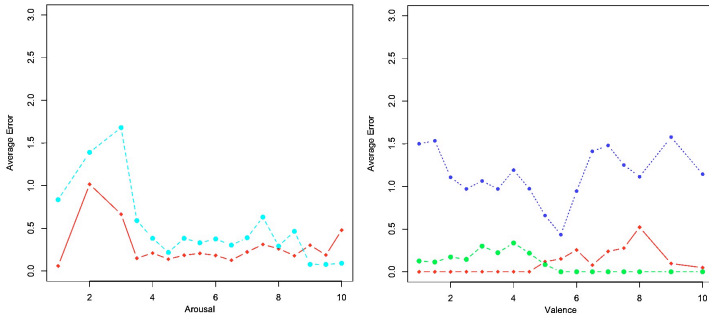

Fig. 2. Average error values for each of the five regression models employed in the first classification layer. *Left panel:* SC (red) shows a much smaller error than HR (blue) in classifying arousal, except for high arousal values, perhaps due to the HR's more immediate response or saturation issues on some participants. *Right panel:* HR (blue) exhibits much higher relative error than both the EMG channels (Zygomaticus in red and Corrugator in green). However, since EMG values are null a large percentage of the time for some participants HR presents itself as a potentially vital fallback. Notice that each EMG channel was only used to classify either positive or negative valence, hence the symmetrical nil error rates between them.

Regression trees were trained from the classified data samples collected in our experiments (one per AV dimension), using a standard CART implementation. The splitting criterion was determined using the sum of squares reduction at each node. Furthermore, the trees were pruned to avoid over-fitting – and, to a lesser extent, interpretability – using the expected increase in the fitness estimate (see Fig. 3), and validated using 3-fold and 10-fold cross-validation schemes (Table 3).

Having built the regression trees we were interested in two main factors: a) whether the created models fitted the data correctly, and b) if these models did indeed combine the classification power of the involved channels across multiple participants.

Table 3. Accuracy ratings for the arousal and valence fusion trees.

Error Margin (AV points)	Arousal Accuracy (%)		Valence Accuracy (%)	
	3 fold CV	10 fold CV	3 fold CV	10 fold CV
0.5	84.5	89.1	84.6	76.2
1.0	98.2	99.0	86.3	85.7

The results presented in Table 3 show that we were able to identify Arousal with as much as 89% accuracy and Valence as high as 84%, when using an error margin be-

neath 1 points in the AV scale. Accuracy results improve considerably if the error margin is increased to two points, but given the limited resolution this implies, we consider it too wide to be of significance. A quick analysis of the average classification error values for each classifier also reveals a well-documented aspect in the field of emotional recognition: that the error pertaining Arousal classification is significantly lower than the one involved in classifying Valence. This is perhaps due to the more primal aspects of Arousal being easier to detect physiologically, as opposed to the higher cognitive complexity (and subjectivity) of Valence. Although this increased classification difficulty is not as perceivable in the higher error margin results, it is far more so in the lower ones. Despite this, we consider the obtained results as highly positive and adequate for our purposes. Regarding the second factor, our results show that the decision tree models are able to maintain high classification ratings even when generalising across all our participants.

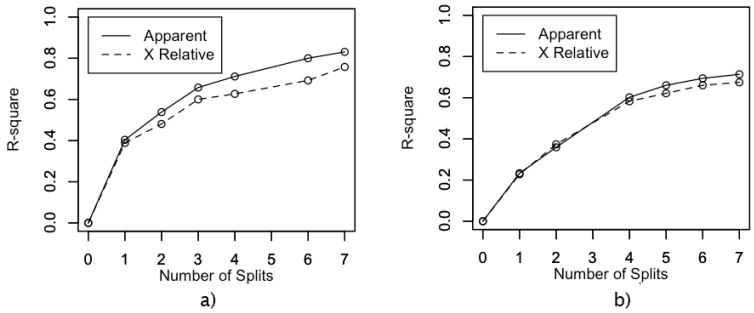

Fig. 3. Apparent and X relative goodness of fit (measured through the R^2 value) for the AV rating fusion trees (a) Arousal, b) Valence).

5 Discussion

Having presented the results for each component of our system, we now discuss our findings, along with some of the system's conceptual features. Our first finding was made as early as our pilot data collection experiments and refers to the emotional elicitation potentials of the material used in each of the data collection experiments' conditions (see section 3.2). We observed that, although the IAPS subjective ratings showed a noticeable variation for all participants, their elicitation potential was considerably limited. This effect was further amplified when participants were first exposed to the other two conditions, which featured more intense stimuli. When exposed to the experimental conditions in this particular order, participants reported a significantly smaller variance in their emotional stimulation ratings: approximately 25% to 40% less. Since the IAPS library was uniformly sampled, we thus consider it may not possess sufficiently strong stimuli to elicit extreme physiological changes and should be complemented by stronger emotional stimuli in future studies.

Our second finding relates to the observed correlation indexes between recorded physiological metrics and AV dimensions. Although these metrics showed high correlation indexes across participants and despite doing so for the remaining two condi-

tions, HR did not significantly correlate with Valence in the IAPS images condition. We consider this is possibly due to two factors: *1)* their low emotional elicitation potential, and *2)* due to the nature of the elicited events. It seems the Slenderman condition was unable to elicit events with both positive arousal and valence, thus, presenting this particular interplay between HR and Valence. The same seems to have occurred in the relaxing music condition as only negative arousal and positive valence were elicited. Drachen et al. have previously found this same correlation in first-person shooter videogames [14], and thus we believe it is an interesting parallel research avenue. This finding also has a significant impact on this particular implementation of the system, as it implies that while it is possible to estimate Valence based on HR for the present scenario, the correlation must be verified in future adaptations – possibly outside of the first-person shooter game genre, given Drachen's results.

Our third and final consideration relates to the created regression trees' classification capabilities. Regression trees are a specific sub-type of decision trees suited for handling continuous data inputs. However, while they are able to deal with continuous data inputs, such as in our case, their output is not continuous. In other words, these models are able to distinguish between a number of classes that is – at most – equal to the number of leaf nodes in the tree. This limitation can be somewhat eliminated by increasing the number of classes, but implies a trade-off between the classification accuracy and the size of the required training dataset. In sum, when considering a practical implementation, the degree of classification granularity should be considered and the appropriate experimental conditions set beforehand. An alternative to this approach is considering statistical models that provide a continuous classification (e.g. neural networks) or applying a normalized weighting scheme on the regression models' error functions to estimate their optimal contributions towards the final output. In future work, we will report on these results via direct comparison.

6 Conclusions

We have presented a data-driven method to interpret selected psychophysiological measures in terms of the Arousal/Valence theoretical model of emotions. The exhibited results show that we are able to perform this process with adequate accuracy for both these dimensions, while maintaining a low sensor calibration threshold. The employed classification methods are also easily interpreted, thus contributing to the system's re-applicability. Furthermore, the low computational cost involved in the classification process contributes towards its adequacy in real-time scenarios.

Using regression models, we have also addressed the pressing issue of correct inter and intra-participant signal normalization in psychophysiological data. Although these methods require a small amount of calibration, they enable a much clearer comparison between emotionally distinct experiences. This approach has also allowed us to build our final classification layer in a participant-independent fashion, thus contributing to the system's overall predictive power.

As emotional detection is not only a critical component of a wider range of affective computing applications, but also a highly complex task, we expect this method

will contribute to the standardization of their development guidelines. This will thus translate into a quickening of their implementation cycle, ultimately allowing for a higher percentage of the time to be allocated to the creation of more complex affective computing systems or affect-related experimental studies.

7 References

1. Ermi, L., Mäyrä, F. 2005. Fundamental Components of the gameplay experience: Analysing immersion. Proc. of DiGRA 2005.
2. Leite, I., Pereira, A., Mascarenhas, S., Castellano, G., Martinho, C., Prada, R., Paiva, A. 2010. *Closing the Loop: From Affect Recognition to Empathic Interaction.* Proc. 3rd Int. Workshop on Affect Interaction in Natural Environments.
3. Cavazza, M., Pizzi, D., Charles, F., Vogt, T., and André, E. 2009. *Emotional input for character-based interactive storytelling*. AAMAS 2009, 1, 313-320.
4. Yannakakis, G. N., Togelius, J. 2011. Experience-driven Procedural Content Generation. IEEE Transactions on Affective Computing, 2(3), 147-161.
5. J. A. Russel, 1980. "A Circumplex Model of Affect". *Personality and Social Psychology*, 39(6), 1161-1178.
6. Plutchik, R. 1980. A General Psychoevolutionary Theory of Emotion. Emotion: Theory, research and experience, 1(1), 3-33.
7. Chanel, G., Kronegg, J., Grandjean, D., Pun, T. 2006. *Emotion Assessment!: Arousal Evaluation Using EEG's and Peripheral Physiological Signals*. Int. Workshop on Multimedia Content Representation, 530–537.
8. Leon, E., Clarke, G., Callaghan, V., Sepulveda, F. 2007. *A user-independent real-time emotion recognition system for software agents in domestic environments*. Engineering Applications of Artificial Intelligence, 20(3), pp. 337-345.
9. Haag, A., Goronzy, S., Schaich, P., Williams, J. 2004. *Emotion recognition using bio-sensors: First steps towards an automatic system*. Affective Dialogue Systems.
10. Yannakakis, G. N., Hallam, J. 2008. *Entertainment Modeling through Physiology in Physical Play*. Int. Journal of Human- Computer Studies, 66(10), 741-755.
11. Vinhas, V., Silva, D., Oliveira, E. Reis, L. 2009. *Biometric Emotion Assessment and Feedback in an Immersive Digital Environment*. Int. Journal of Social Robotics, 307-317.
12. Mandryk, R., and Atkins, M. 2007. *A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies*. International Journal of Human-Computer Studies, 65(4), 329-347.
13. Stern, R., Ray, W., Quigley, K. 2000. *Psychophysiological Recording*. Oxford Un. Press.
14. Drachen, A., Nacke, L., Yannakakis, G., Pedersen, L. 2010. *Correlation between Heart Rate, Electrodermal Activity and Player Experience in First-Person Shooter Games*. Proc. 5th ACM Siggraph Symposium on Video Games, 49-54.
15. C. Conati, "Modeling User Affect from Causes and Effects," in *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization*, 2009.
16. P. J. Lang, M. M. Bradley & B. N. Cuthbert. 2008. "International affective picture system (IAPS)". University of Florida, Gainesville, FL.
17. Y. Yang. 2008. A Regression Approach to Music Emotion Recognition. IEEE Trans. On Audio, Speech and Language Processing, 16(2), 448-457.
18. S, Aalto, E, Wallius, P, Näätänen, J, Hiltunen, L, Metsähonkala, H, Sipilä, H., Karlsson. 2005. Regression analysis utilizing subjective evaluation of emotional experience in PET studies on emotions. Brain Res Brain Res Protoc, 15(3), 142-154.