Measuring Observers' EDA Responses to Emotional Videos

Conference Paper · December 2019

DOI: 10.1145/3369457.3369516

Jessica Rahman
Australian National University

Md Zakir Hossain
Australian National University

Measuring Observers' EDA Responses to Emotional Videos

Jessica Sharmin Rahman*
Research School of Computer Science
The Australian National University
Canberra, Australia
jessica.rahman@anu.edu.au

Md. Zakir Hossain
Research School of Computer Science
The Australian National University
Canberra, Australia
zakir.hossain@anu.edu.au

Tom Gedeon
Research School of Computer Science
The Australian National University
Canberra, Australia
tom.gedeon@anu.edu.au

ABSTRACT

Future human computing research could be enriched by enabling the computer to recognize emotional states from observers' physiological activities. In this paper, observers' electrodermal activities (EDA) are analyzed to recognize 7 emotional categories while watching total of 80 emotional videos. Twenty participants participated as observers and 16 features were extracted from each video's respective EDA signal after a few processing steps. Mean analysis shows that a few emotions are significantly different from each other, but not all of them. Our generated arousal model on this dataset with these participants using their EDA responses also differs a little from the abstract models proposed in the literature. Finally, leave-one-observer-out approach and neural network classifier were employed to measure the performance, and the classifier reaches up to 94.8% correctness at the seven-class problem. The high accuracy inspires the potential of this system to use in future for recognizing emotions from observers' physiology in human computer interaction settings. Our generation of an arousal model for a specific setting has potential for investigating potential bias in dataset selection via measuring participant responses to that dataset.

CCS CONCEPTS

• **Human-centered Computing → Interaction Techniques**;
• **User Interface Design**;

KEYWORDS

Electrodermal Activity, Emotion Recognition, Arousal Model, Neural Network

ACM Reference Format:
Jessica Sharmin Rahman, Md. Zakir Hossain, and Tom Gedeon. 2019. Measuring Observers' EDA Responses to Emotional Videos. In *31ST AUSTRALIAN CONFERENCE ON HUMAN-COMPUTER-INTERACTION (OZCHI'19), December 2–5, 2019, Fremantle, WA, Australia.* https://doi.org/10.1145/3369457.3369516

1 INTRODUCTION

Emotions are a vital aspect of human life and they highly influence our day-to-day activities and decisions. In addition, they play a crucial role in daily social interaction. Therefore, emotion recognition is becoming an increasingly popular topic among researchers. There are various verbal and non-verbal behaviors that helps to recognize different emotions. Some of the common nonverbal human behaviors are: facial expressions, gestures, postures, body movements etc. However, these behaviors often do not always reflect true human emotions as they can be manipulated or controlled. In this regard, physiological signals can be considered as a strong measurement to detect emotions as they cannot be easily modified. Physiological signals such as electrodermal activity (EDA) and respiratory activity have shown sensitivity to change of emotions [11]. Thus, emotion recognition using physiological signals has become a topic of interest for the last few years. This research has a range of applications such as stress detection, anxiety measurement, human computing, evaluating human-computer interaction, and so on.

A number of papers in the literature have shown the utility of physiological signals in identifying different emotions. Studies have been conducted using images, videos, music and text to elicit different emotions. Jang et. al. [10] used different types of sound to elicit three different types of emotion (boredom, pain, surprise) and analyzed a set of physiological signals from participants' to achieve 84.7% classification accuracy. Picard et al [14] used personalized imagery to elicit emotion on one participant using a number of physiological signals and achieved an accuracy of 81% for eight emotion

ACM ISBN 978-1-4503-7696-9/19/12…$15.00
https://doi.org/10.1145/3369457.3369516

categories. Kim et.al. [12] uses a combination of audio and visual stimuli to predict three categories of emotion using electrocardiogram, skin temperature variation and EDA. They were able to achieve 78.4% accuracy for classifying three emotion categories. Signals such as skin temperature, heart rate have shown to achieve high accuracy of 98% in text based stress recognition model [18]. Emotion recognition using other nonverbal behavior such as facial expression and gestures have demonstrated a comparatively lower accuracy than physiological signals [5, 13].

In this paper, we analyze the effects of participants' EDA responses while watching a set of emotional videos as observers are good at categorizing emotions [21] and at recognizing facial expression [9]. After processing the collected EDA data, 16 linear and non-linear features are extracted and finally, evaluated by mean analysis and classified using neural network (NN).

2 METHODS

Observers

Twenty participants (14 female and 6 male) took part voluntarily in this experiment. The mean age was 23 years old with a standard deviation of 5.8. All the participants were asked to sign a written consent form before their voluntary participation in the study. This study was approved by the Human Research Ethics Committee of The Australian National University (ANU).

Experiment Design

The Acted Facial Expressions In The Wild (AFEW) dataset [3] has been used for the purpose of this experiment. Each participant watched a total of 80 videos which were divided into 7 categories. They are: Anger, Disgust, Fear, Happy, Neutral, Sad and Surprise. All the videos were around 2-3 seconds in length. Participants were asked some general demographic questions at the beginning of the experiment. After watching each video, they were asked to rate the genuineness of the video in a 5-point rating scale ('Completely fake', 'Surface acted', 'Don't know', 'Deep acted', and 'Completely real'). They were also asked to rate their confidence level on their answer using a 5-point scale (1 being not confident at all and 5 being very confident). They were also asked if they had seen the video before or not.

Data Collection and Processing

EDA data was collected using Empatica E4 wristband with a sampling rate of 4 Hz [4]. Data collected in these kinds of experiments are highly susceptible to noise due to movements by the participants. Therefore, we normalized the data using min-max normalization technique and smoothed using median filter [17].

Feature Extraction

A total of 16 features (linear and nonlinear) were extracted [1, 2, 19] from the processed data to reduce the computational cost of analysis [20] as listed below.

- Mean - Mean is calculated from the filtered EDA signals over the length of watching one video category.
- Root Mean Square - This indicates the mean power of EDA signals over a fixed length.
- Variance of EDA Signals - Shows the average value of the power of EDA signals.
- Integrated EDA - This feature represents the absolute summation of EDA value over the fixed window length.
- Simple Square Integral - It indicates the summation of the absolute power of EDA signals.
- Average Amplitude Change - This feature is calculated by first finding the difference between two consecutive samples, then averaging them over the fixed window length.
- Hjorth Parameters - These parameters are commonly used in signal processing of physiological signals. We calculate the parameter mobility which measures the standard deviation of the slope of signals with respect to standard deviation of the amplitude of the signal.
- Hurst Exponent - This non-linear feature indicates the complexity and correlation properties of physiological signals.
- Entropies - Entropy represents the randomness in physiological signals. We calculate 5 types of entropy for the features. They are:
  - Sample Entropy - Measures complexity without taking self-similarity into account (uses distance between two signals)
  - Approximate Entropy - Measures complexity using self-similarity of signals
  - Shannon's Entropy - Measures randomness uncertainty in signals.
  - Permutation Entropy - Calculates complexity of time-series data based on analyzing various permutation patterns.
  - Fuzzy Entropy - A measurement to calculate ambiguous uncertainty in signals
- Log Detector - This is an important non-linear feature, which is calculated by the average logarithm over the fixed window length.
- Difference Absolute Standard Deviation Value - It is calculated by the standard deviation between two consecutive signals.
- Detrended Fluctuation Analysis - Determines the self-similarity level of physiological signals.

3 RESULTS AND DISCUSSION

Mean Analysis

From the set of extracted features we chose mean values of all participants in every emotion category and performed some statistical tests on that data. Figure 1 shows the mean values for the 7 emotion categories:

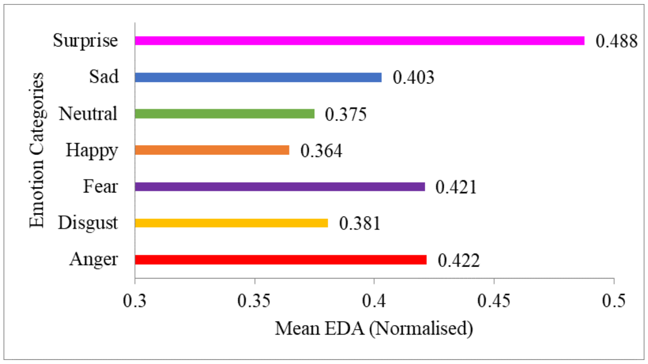

Figure 1: Mean values of EDA for 7 Emotion Categories (Range 0.3-0.5 chosen for better visualization)

Anger						
Disgust	0.234					
Fear	0.492	0.141				
Happy	0.093*	0.203	0.067*			
Neutral	0.126	0.407	0.105	0.287		
Sad	0.336	0.157	0.214	0.007**	0.157	
Surprise	0.073*	0.022**	0.029**	0.005**	0.015**	0.021**
	Anger	Disgust	Fear	Happy	Neutral	Sad

Table 1: T-test Values for All Pairs of Emotions

It has been shown in Figure 1 that observers feel high and low cognitive load while watching *Surprise* and *Happy* videos compared to other emotional categories. To find the differences between emotion pairs, we performed a two-tailed permutation test. The test is applied here to identify the time points where EDA is different between two emotions. Over the analysis, we found 4 emotion pairs (Disgust-Surprise, Happy-Sad, Happy-Surprise, Neutral-Surprise) significantly differ from one another ($p < 0.05$). The analysis technique is very similar as found in [8]. We also analyzed every pair of emotions using T-test and the results showed statistical significance ($p < 0.05$) for 6 pair of emotions. They are: Happy-Sad, Happy-Surprise, Disgust-Surprise, Fear-Surprise, Neutral-Surprise and Sad-Surprise. Table 1 shows the significance value for all pair of emotions. The numbers with $*$ and $**$ are the pairs that show meaningful differences. The numbers with $**$ show significant labels at 95% confidence interval, while numbers with $*$ shows significance only at the 90% confidence interval (illustrated to show these values are close to our normal significant interval of 95%).

We can visualize this relationship using an emotion model, which is a two dimensional model based on valence and arousal level of emotions frequently used in the area of affective computing [16]. Valence refers to intrinsic goodness or badness while arousal corresponds to alertness/response readiness [15]. The arousal model for the 7 emotions is shown in Figure 2. We call it an arousal model, because we did not consider valence as a measurement reference as yet, but hopefully will consider it in the future. The valence levels are kept same as the original model.

From Figure 2, we can see that surprise has a very high arousal value compared to all other emotions. Therefore, it is able to show significant difference with all other emotion except another high arousal emotion anger, and there it is close to being significant (p=0.73). The feature is also able to differentiate between a high valance high arousal emotion (Happy) and a low valence low arousal emotion (Sad). We can also see from Figure 2 that if we consider neutral as a reference line then happy and sad are not located in the position proposed for arousal models widely in the literature. Happy is a high arousal high valence emotion and sad is a low arousal low valence emotion according to the literature. But based on our data we can see happy is showing low arousal and sad is showing high arousal. This is possible from observers' perspective perhaps because when they are watching sad videos they feel sad, but maybe when watching happy videos (more common than seeing sad videos) they take them normally. Comparing it with more emotions in all of the categories will help us to understand this phenomenon in greater detail.

Evaluation Measures

We report the classification accuracy of the system, that shows the percentage of the system predicting the video category correctly. Along with the accuracy, we also calculate F-measure, which is a commonly used evaluation measure represented by the harmonic mean of precision and recall. Precision refers to the fraction of the predicted labels matched while recalls refers to the fraction of reference labels matched. We also report the precision, recall (also referred as sensitivity), specificity (true negative rate) and geometric mean values (measure of central tendency).

Classification

The classification process was done using MATLAB R2018a software with an Intel(R) Core(TM) i7-5200U processor with 3.60 GHz, 16.00 GB of RAM and Microsoft Windows 10 Enterprise 64-bit operating system. The labels were given according to the 7 video categories mentioned in the experiment

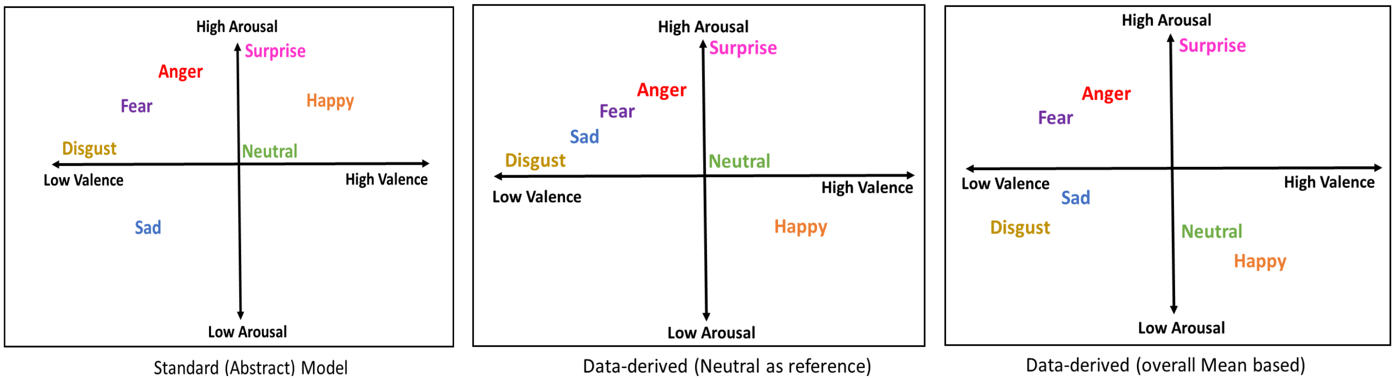

Figure 2: Arousal models of emotion: left = Standard abstract model; middle/right = data derived (Neutral/Mean as reference)

design. A leave-one-observer-out process was performed to distinguish among the 7 emotion categories. A simple pattern recognition network was employed which consisted of one input layer, one hidden layer and one output layer. The hidden layer was constructed using 30 hidden nodes. The model achieves a total of 94.8% accuracy based on the average of 20 runs. Figure 3 shows the accuracies of all 7 emotional categories.

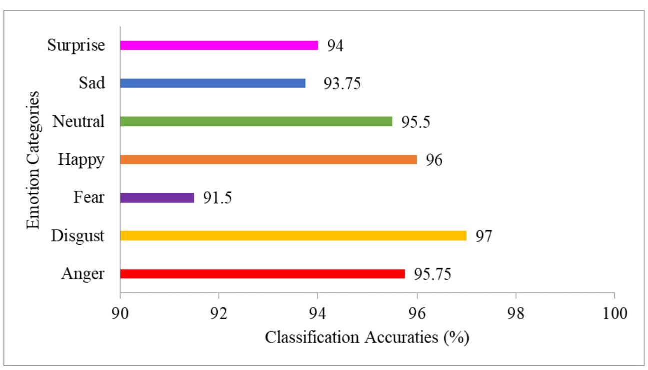

Figure 3: Classification performances while observers were recognizing seven emotions (Range 90-100 chosen for better visualization)

We also demonstrate all evaluation measures, which are calculated on the average result of all categories (please see Table 3). Geometric mean and harmonic mean values provide more useful information than arithmetic mean when comparing groups having different properties [7].

From Table 3 it is evident that the neural network model achieves high score in both F-measure and Geometric mean. So this model is effective for EDA signal based emotion recognition problem.

F-measure	Precision	Recall	Specificity	G-mean
0.842	0.753	0.958	0.947	0.952

Table 2: Evaluation measures for classifying among seven emotional categories

4 CONCLUSION

In this paper, we looked into the effects on participants' EDA activity while they watched a set of videos of 7 different emotional categories. Signals were collected from participants' in an experimental setting. Collected signals were normalized, filtered, and then a set of 16 features were extracted. Classification using a simple neural network showed a high accuracy of 94.8% in identifying the 7 different emotional video categories.

The initial analysis also shows some noticeable difference of our data driven arousal model from our observers' perspective, when compared to the (abstract) standard models in the literature. The data-derived model with neutral as the baseline is quite similar to the standard abstract model, with the only changes being Happy and Sad changing sides as Low/High arousal. Further analysis will be conducted and evaluated to identify the reasons. Questions to be answered are whether the dataset was biased, whether our 20 participants were somehow different from the expected population reaction, or whether the abstract model is just incorrect. It is also important to point out that EDA activity can vary according to the difference in stimuli types, participants' age, gender etc [6]. Also the number of samples might be considered small, although experiments have shown that it is reasonable [9]. Arguably, it makes more sense to use the overall average reaction to be the baseline between high and low arousal, which spreads the emotional reactions over a wider range. This differs more from the standard model. We will also analyze other physiological signals to cast more light on these issues and to investigate their use in predictions.

REFERENCES

[1] U Rajendra Acharya, Yuki Hagiwara, Sunny Nitin Deshpande, S Suren, Joel En Wei Koh, Shu Lih Oh, N Arunkumar, Edward J Ciaccio, and Choo Min Lim. 2018. Characterization of focal EEG signals: a review. *Future Generation Computer Systems* (2018).

[2] Rubana Chowdhury, Mamun Reaz, Mohd Ali, Ashrif Bakar, Kalaivani Chellappan, and Tae Chang. 2013. Surface electromyography signal processing and classification techniques. *Sensors* 13, 9 (2013), 12431–12466.

[3] Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. 2011. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. *Proceedings of the IEEE International Conference on Computer Vision*, 2106–2112. https://doi.org/10.1109/ICCVW.2011.6130508

[4] E4 [n.d.]. *E4 Wristband from Empatica.* https://www.empatica.com/research/e4/

[5] Yin Fan, Xiangju Lu, Dian Li, and Yuanliu Liu. 2016. Video-based Emotion Recognition Using CNN-RNN and C3D Hybrid Networks. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16).* ACM, New York, NY, USA, 445–450. https://doi.org/10.1145/2993148.2997632

[6] Elia Gatti, Elena Calzolari, Emanuela Maggioni, and Marianna Obrist. 2018. Emotional ratings and skin conductance response to visual, auditory and haptic stimuli. In *Scientific data.*

[7] Geometric & Harmonic Means in Data Analysis [n.d.]. *On Average, You're Using the Wrong Average: Geometric & Harmonic Means in Data Analysis.* https://towardsdatascience.com/on-average-youre-using-the-wrong-average-geometric-harmonic-means-in-data-analysis-2a703e21ea0

[8] MZ Hossain, T Gedeon, R Sankaranarayana, D Apthorp, and A Dawel. [n.d.]. Pupillary responses of Asian observers in discriminating real from fake smiles: A preliminary study.

[9] Md Zakir Hossain, Tom Gedeon, and Ramesh Sankaranarayana. 2018. Using temporal features of observers' physiological measures to distinguish between genuine and fake smiles. *IEEE Transactions on Affective Computing* (2018).

[10] Eun-Hye Jang, Byoung-Jun Park, Mi-Sook Park, Sang-Hyeob Kim, and Jin-Hun Sohn. 2015. Analysis of physiological signals for recognition of boredom, pain, and surprise emotions. *Journal of physiological anthropology* 34, 1 (2015), 25.

[11] Jonghwa Kim and Elisabeth André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence* 30, 12 (2008), 2067–2083.

[12] K. H. Kim, S. W. Bang, and S. R. Kim. 2004. Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing* 42, 3 (01 May 2004), 419–427. https://doi.org/10.1007/BF02344719

[13] Chuanhe Liu, Tianhao Tang, Kui Lv, and Minghao Wang. 2018. Multi-Feature Based Emotion Recognition for Video Clips. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18).* ACM, New York, NY, USA, 630–634. https://doi.org/10.1145/3242969.3264989

[14] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 10 (2001), 1175–1191.

[15] J. S. Rahman, T. Gedeon, S. Caldwell, R. Jones, M. Z. Hossain, and X. Zhu. 2019. Melodious Micro-frissons: Detecting Music Genres From Skin Response. In *2019 International Joint Conference on Neural Networks (IJCNN).* 1–8. https://doi.org/10.1109/IJCNN.2019.8852318

[16] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[17] Jerritta Selvaraj, Murugappan M, R Nagarajan, and Wan Khairunizam. 2011. Physiological signals based human emotion Recognition: a review. *Proceedings - 2011 IEEE 7th International Colloquium on Signal Processing and Its Applications, CSPA 2011* (03 2011). https://doi.org/10.1109/CSPA.2011.5759912

[18] Nandita Sharma and Tom Gedeon. 2013. Computational models of stress in reading using physiological and physical sensor data. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining.* Springer, 111–122.

[19] Triwiyanto Triwiyanto, Oyas Wahyunggoro, Hanung Adi Nugroho, and Herianto Herianto. 2017. An investigation into time domain features of surface electromyography to estimate the elbow joint angle. *Advances in Electrical and Electronic Engineering* 15, 3 (2017), 448–458.

[20] Johannes Wagner, Jonghwa Kim, and Elisabeth André. 2005. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *2005 IEEE international conference on multimedia and expo.* IEEE, 940–943.

[21] Robert Sessions Woodworth and Harold Schlosberg. 1954. *Experimental psychology.* Oxford and IBH Publishing.