A Simple Algorithm for Emotion Recognition, Using Physiological Signals of a Smart Watch

David Pollreisz and Nima TaheriNejad
Institute of Computer Technology, TU Wien
Gusshausstrasse 27-29, 1040 Vienna, Austria
nima.taherinejad@tuwien.ac.at

***Abstract*— Recently, it has become easier and more common to measure physiological signals through wearable devices such as smart watches. Extracting emotional states of individuals with problems expressing it, such as autistic individuals, can help their parents, friends, and therapists to obtain a better understanding of what they feel throughout their day. Although emotion recognition methods based on physiological signals have been studied for many years, there is a smaller body of literature about systems working with data obtained from wearable devices. In this paper, we present an emotion recognition system with a small footprint suitable for limited resources of wearable devices. Other than identifying the emotions (with a success rate of 65%), The proposed system also tags each recognition with a confidence value (on average 57%).**

I. INTRODUCTION

Many groups have tried to acquire various data including physiological signals [1] [2], to identify emotions out of some special characteristics of the input. Emotion recognition could improve the quality of life in different situations, for example, in therapeutic contexts. It is difficult for some individuals, e.g. children who suffer from Autism Spectrum Disorder (ASD), to express their feelings or ask for help when they need it. Many of them suffer from speech impairment [3] or show an atypical facial expression [4]. On the other hand, people in the social environment of these individuals often can not correctly recognize their expressions which intensifies the misunderstandings. Due to the fact that the parents of children that suffer from ASD have a higher stress level [5], alternative methods can be very helpful for them in understanding their child. This can be facilitated nowadays thanks to the possibility of building multi-modal emotion recognition systems and devices -such as smart watches- which measure physiological signals, and are significantly smaller and more affordable [6].

With our work, we hope to contribute to the -rather sparse- literature of analyses based on smart watch data and show some new possible ways to extract the information about emotions. Our approach has two main distinctive characteristics. First, it uses a simple and compact architecture suitable for implementation in systems with limited resources such as smart watches as well as mobile and Embedded Systems (ESs). Having a small foot print is more important when considering multi-modal emotion recognition on such devices. Second, we provide a confidence level for each recognized emotion. A feature that can be particularly useful in multi-modal emotion recognition systems. Especially when different modes provide conflicting results. The rest of this paper is organised as the following: In the next section, a brief overview of the literature is given. In Section III, the proposed method and some alternatives are described. In Section IV, we discuss our results and finally, Section V concludes the paper.

II. LITERATURE OVERVIEW

Lee et al., [7] used a Neural Network (NN) for identification of emotions. They analysed the Heart rate (HR) and the Electrodermal activity (EDA) values of different test subjects in the time and frequency domain. In their research, they mention that if the Root-Mean Square of Successive Differences (RMSDD) and the Standard Deviation Normal-to-Normal-Intervals (SDNN) rises, but the mean value of the HR falls, there is an increased activity in the sympathetic nervous system which is a sign of fear. This and similar features were implemented in the NN. The accuracy of the results were reported to be 80.2%.

Jang et al., [8], [1] have also tried to find the best features to identify emotions out of physiological signals -such as max. skin temperature (SKT), mean SKT and mean EDA - using various machine learning methods. They have tested five machine learning algorithms: Linear Discriminant Function (LDF), Classification And Regression Tree (CART), Self-Organising Map (SOM), Naive Bayes, and Support Vector Machine (SVM). They came to the conclusion that for their study [8] SVM was the best fit with 99.04% success.

Dai et al., [9] used an advanced version of the SVM, called the Reputation-driven Support Vector Machine (RSVM). They extracted a few features of Electrocardiography (ECG) signals and analysed them. First, they used an SVM algorithm on this data and afterwards the RSVM. With SVM they achieved an accuracy of 66.95% and with RSVM they were able to increase the accuracy to 75.19%.

On the other hand, there are approaches which are based on statistical analyses. For example, Jang et al., in their more recent work, in addition to the aforementioned five machine learning algorithms, worked on a statistical method called Discriminant Function Analysis (DFA) [1]. A method which in their study appeared to be the best way to identify boredom, pain, and surprise [1].

Valderas et al., [10] also did a statistical analyses of their signals which used a Mann-Whitney U-test to evaluate statistical differences between relaxation, fear and joy. With

978-1-5090-2809-2/17/$31.00 ©2017 IEEE

TABLE I

SAMPLES MARKED WITH AN EMOTION BY SUBJECTS [11]

Solicited Emotions	Samples Marked as Such
Happiness	16
Sadness	15
Anger	8
Pain	8

the calculation of some statistical values like the root mean square or the transformation into the frequency domain, they looked for differences in each emotion they measured. However, to the best of our knowledge, so far they have not proposed any system to automatically identify these emotions based on those statistical analyses.

A different approach was tried in [2]. They analysed the signals in four different ways: they looked at the signal in the time domain, frequency domain, with a Poincare Plot, and a differential plot. For classification, they used an SVM and a Genetic Algorithm (GA). The GA extracted the feature characteristics out of the four different analysis and combined them to extract an emotion. With the combination of these two components, an accuracy of 90% was achieved.

III. PROPOSED SYSTEM

For our analysis, we use the bio-signals acquired during our data collection campaign [11]. In this campaign, we used the Empatica E4 smart watch to collect EDA, SKT and HR values of ten subjects that were between 20 to 25 years old. The subject were asked to watch an emotion stimulant video and then fill a Self-Assessment Manikin (SAM) to tag the emotions they felt during the experiments. Overall, we obtained 49 samples, summarized in Table I [11]. Characteristics of various physiological signals and their correlation to the emotions under study were studied in depth in [11]. Table II shows a summary of the results of the measurements and what characteristics of the signals can be related to each emotion. For the quantitative definitions of the terms used in this table and how they were obtained the reader is kindly referred to [11]. What comes here after, are the new contributions concerning the automatic identification of emotions based on and in addition to our previous work [11], focused on data collection and analysis.

1) Overview: The high level flow chart of the proposed system, is shown in Fig. 1. In the preprocessing step of the algorithm, first the baseline values of the subject are calculated. Then, in the feature extraction, based on the baseline, HR, SKT, and EDA signals are analysed regarding their changes (increase/decrease). Additionally, the EDA signal is analysed in terms of size and number of peaks.

TABLE II

OBSERVED TRENDS IN THE COLLECTED DATA [11].

Emotion	HR	EDA	EDA Peaks	SKT
Happiness	slight increase	increase	small & few	slight decrease
Sadness	decrease	increase	small & many	slight decrease
Anger	slight decrease	increase	big & some	slight decrease
Pain	no change	increase	big & one	no change

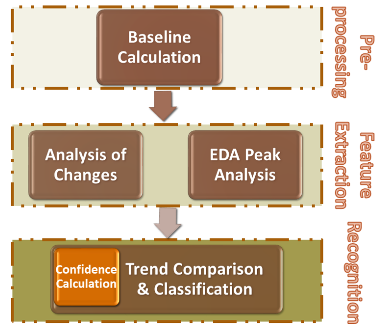

Fig. 1. Overview of the proposed algorithm.

In the recognition step, these features are compared to the trends observed in Table II. Based on the number of trends associated with each emotion, and observed in the extracted features, the sample is classified. Within the same process, a confidence value is calculated for each classification. In the rest of this section, we provide more details about the feature extraction and classification.

2) Baseline Calculation: To this end the mean value from 10 seconds before the stimulus (i.e. a video) started to the beginning of the stimulus was calculated. This is shown in Fig. 2, where the stars mark the beginning and end of the application of the stimulus and baseline is also marked.

3) Analysis of Increase and Decrease: For this purpose, the baseline was then compared with each value of the signal (sampled at 1Hz) after the beginning of the application of the stimulus, that is the first star in Fig. 2. If any change is observed in a signal, a counter associated with that type of change in that signal is increased. For the HR for example, if the mean value is 3bmp lower than the value of the baseline signal, the number stored in the counter for the "small increase" was increased by one. If the mean value is 9bmp higher than the value of the baseline signal, the counter for the "big increase" was increased by one. The same model was applied to the values smaller than the baseline, this time with decrease counters. At the end, all the counters were compared to calculate the overall change. If the increase counter was higher than the decrease counter, it was considered that the signal had increased during the application of the stimulus. The same procedure was applied to the EDA and the SKT signals.

4) Analyses of the Peaks: To extract more knowledge from the signals, according to [11], information about the EDA peaks needed to be acquired. Although the EDA baseline values vary in the relatively large range of 0.1 μSiemens to 12 μS, all the peaks almost have the same range

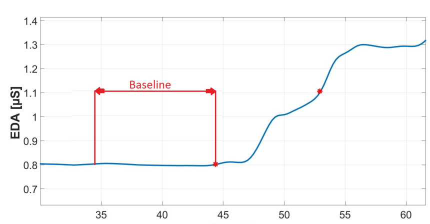

Fig. 2. Calculation of the mean value of the baseline

TABLE III

CLASSIFICATION FOR HAPPINESS. EACH COLUMN SHOWS A SAMPLE AND PROBABILITY OF BEING ASSOCIATED WITH EACH EMOTION.

Emo./sub.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
Anger	0	50	25	0	**36**	**36**	**36**	0	0	0	0	**52**	**50**	0	**36**	0
Happiness	0	**50**	**50**	**57**	*32*	*32*	*32*	**67**	**40**	**40**	**57**	*24*	**50**	**67**	*32*	**67**
Pain	0	0	0	0	0	0	0	0	30	30	0	0	0	0	0	0
Sadness	**100**	0	25	43	32	32	32	33	30	30	43	24	0	33	32	33

of height. That is, regardless of the range in which the EDA value starts, the peaks have always similar heights. Hence, the peaks were categorized into two groups with different heights. If a peak height is between 0.05 to $0.4\mu S$ it is classified as a small peak and if a peak height is more than $0.85\mu S$, it is considered as a big peak. An example is shown in Fig. 3, where (a) big and (b) small peaks are automatically tagged by the proposed system. Thus, we analysed all EDA signals and whenever a small or large peak was detected, its respective counter was increased. These statistical analyses were then used to categorize the signals into three groups: (i) Signals with over 10 peaks, (ii) signals with more than 2 and less than 10 peaks, and (iii) signals with exactly 1 peak.

5) Classification: The classification is done via a decision tree based on statistical information obtained in [11]. First, it starts by analysing the peaks in the EDA signal. For example, if there are less than two big peaks in the measured signal, a counter is increased for the emotion Pain. However, if there are more small peaks in the signal, the counter for the happiness, anger, and sadness is increased. The rationale behind it being that different emotions sometimes show similar behaviours. However, we note that statistically, sometimes one emotion shows such a trend more often than others. Therefore, we calculated the probability for each emotion showing certain characteristics and added different weights to the counters. For example, the emotions anger and sadness demonstrated similar characteristics for small peaks but the statistical analysis shows that in the presence of many small peaks, it is more likely that the test subject is sad. As a consequence, in such a case, the counter for sadness is increased by a coefficient of 1.1, whereas that of anger is increased by 1. After the system processes all the EDA peaks, it starts to analyse the changes in the HR, SKT and EDA in a similar manner. That is, each time a change displayed in the Table II is found, a weighted counter associated with the respective emotion is increased. At the end, the probability of occurrence of each emotion is calculated in percentage based on the value of the counters for each emotion.

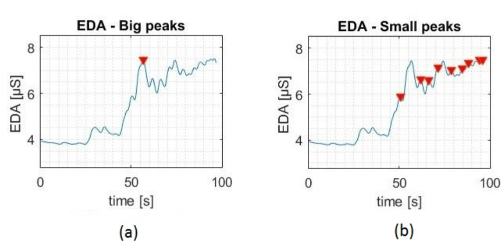

Fig. 3. (a) Shows the identification of a big peak and (b) shows the identification of a few small peaks.

6) Confidence: Confidence is an important aspect of observations of a system which is often simply overlooked [12]. This can be particularly more problematic in multi-modal emotion recognition systems where different modes of recognition provide conflicting results [13]. Lack of any measure to resolve such conflicts is a problem that to the best of our knowledge is not addressed by far. We contend that the confidence, can be a solution to this problem.

One of the advantages of the proposed system is that providing confidence values comes at no additional computation cost. The probabilities calculated for each emotion are used and provided as the confidence value of the recognition system. Hence, we do not show only that an emotion is tagged, but also how trustworthy this assessment is. We emphasise that confidence is not the same as the success rate of the algorithms, which is often provided by various systems as a measure of the reliability and trustworthiness. We will observe the difference between the two in Section IV.

IV. RESULTS AND DISCUSSIONS

1) Performance: Table III shows an example of the outputs of the system for all the measurements that were marked as happy by the test subjects. Each column represents a different set of measurement. The percentages written in bold are the first choice of the system (as the tagged emotion) and the italic numbers present the second choice of the algorithm for each data set. We can see that the system has correctly classified 10 out of 16 samples of happiness (63%). Considering the cases where the correct emotion was the second choice of the algorithm, this percentage is increased to 94%. Table IV summarizes the results of the proposed system for other emotions. In our Matlab simulations[1], on a PC with 2 processor cores running at 2GHz and 16 GB RAM, the algorithm takes approximately 0.6s to process a one minute long sample. Thus, it meets the desired real-time requirements.

2) Confidence: Looking at Table III, we can see that often, when the system has a false recognition, the confidence value of the tagged emotion (which is the probability of its association with that emotion) is below 50% (in many cases 36%). On the other hand, correct assessments have often a confidence of 60% and above. Therefore, confidence can be used as an indicator of trust. For example, in the case of a conflict, if this mode is tagging an emotion with low confidence, it may be better to consider other modes of emotion recognition, especially, if other modes provide a higher confidence value. Furthermore, we note that our

[1]We note that for the ES usage the algorithm needs to be implemented in other platforms, which in turn will increase its speed and efficiency too.

TABLE IV

SUMMARY OF RESULTS FOR ALL EMOTIONS.

	First Choice	Second Choice	Confidence
Sadness	47%	87%	68 %
Pain	100%	100%	67%
Happiness	63%	94%	55%
Anger	50%	100%	40%
Average	65%	95%	57%

system was able to correctly tag "Pain" unanimously, that is with a 100% success rate. However, it has done so with an average confidence of only 67%. This conforms with the continuity of emotion spectrum and shows that despite the 100% success rate of the algorithm in recognizing pain in our experiments, it can not be always and 100% trusted in recognizing this emotion. This demonstrates how the success rate of an algorithm is different from its confidence and how the former is not necessarily a good indicator of the latter.

3) Discussion: As displayed in Table IV, the system has successfully detected the right emotion (as its first choice) 65% of the times with an average of 57% confidence in tagging the right emotion. As seen in Table II, there are some similarities in the characteristics between anger, sadness and happiness. Therefore, in some cases it is more difficult to detect the right emotion. As shown in Table III, in many cases the values of the probabilities are nearly the same for happiness, anger and sadness. Due to the fact that these emotions show similar characteristics, a successful classification for these three emotions has occurred between 47 to 63% of the cases, whereas it has happened unanimously (100%) for Pain. This shows that -as observed in other works too, e.g., [1]- certain characteristics, and consequently algorithms based on them, are more suitable for certain emotions and less for other emotions. It is also interesting to note that although sadness has the smallest success rate, at the same time it has (even though by small margin) the highest confidence when tagging the right emotion. Anger, however, despite having a slightly better success rate (50%) compared to sadness (47%), has an average of only 40% confidence which makes anger the least trustworthy classification of the proposed system.

4) Comparison: To compare the proposed system with other works in the literature, we have summarized their reported performance in Table V[2]. In this table we can see that even though SVM is more complex than the one presented in this paper (especially if some features such as Reputation-driven part are added), the performance improvement is not all that significant (e.g., only 2.29% compared to SVM [9]). Moreover, majority of other works use laboratory equipment and bio-signal collection devices which are significantly more accurate and reliable than any smart watch. This factor on itself can affect the performance and success rate of the algorithms (in our case, negatively). Hence, given the considerably simpler structure of the proposed system and the small difference in performance, our system is overall more favourable.

[2]We remark that we did not have access to the data, codes, speed, or power consumption of other works to compare those quantitatively.

TABLE V

PERFORMANCE OF DIFFERENT ALGORITHMS.

Algorithm	Success Rate
Current work	64.66%
SVM [9]	66,95%
RSVM [9]	75,19%
SVM + GA [2]	90%
NN [7]	80,2%
DFA [1]	84,7%

Last but not least, we remark that other works do not report any measure of confidence which could be compared with the present work.

V. CONCLUSION

In this paper, we presented a new method of analysing physiological signals, by looking for peaks in the EDA signal, and showed a simpler solution to identify emotions. In our experiments, we use a statistical analysis of the peaks, level changes in the signal, and a weighted counter system for emotion recognition. Although the performance was very close to some others and not as good as some others such as a NN, it is always simpler and easier to implement. In the context of mobile and ESs, the small footprint and relaxed requirements on the resources gives the proposed system an advantage compared to others. Furthermore, we presented the concept of confidence in recognition and provided this measure as one of the outputs of the system. Confidence, which as shown in the paper, is not represented by success rate of the algorithm is an important factor which can help many systems that use recognized emotions as their inputs. This -often overlooked- parameter is of paramount importance in multi-modal emotion recognition systems, where different modes provide conflicting results.

REFERENCES

[1] E.-H. Jang *et al.*, "Analysis of physiological signals for recognition of boredom, pain, and surprise emotions," London, 2015.

[2] S.-N. Yu and S.-F. Chen, "Emotion state identification based on heart rate variability and genetic algorithm," in *EMBC*, 2015, pp. 538–541.

[3] J. J. Diehl and R. Paul, "Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders," *Research in autism spectrum disorders*, vol. 6, no. 1, pp. 123–134, 2012.

[4] S. D. Stagg *et al.*, "Does facial expressivity count? how typically developing children respond initially to children with autism," *Autism*, vol. 18, no. 6, pp. 704 – 711, 2013.

[5] E. Pattini and D. Rollo, "Response to stress in the parents of children with autism spectrum disorder," in *IEEE MeMeA*, 2016, pp. 1–7.

[6] D. Phan *et al.*, "Smartwatch: Performance evaluation for long-term heart rate monitoring," in *ISBB*, 2015, pp. 144–147.

[7] C. Lee, S. Yoo, Y. Park, N. Kim, K. Jeong, and B. Lee, "Using neural network to recognize human emotions from heart rate variability and skin resistance," in *27th EMBC*, 2005, pp. 5523–5525.

[8] E. H. Jang *et al.*, "Identification of the optimal emotion recognition algorithm using physiological signals," in *ICEI*, 2011, pp. 1–6.

[9] Y. Dai *et al.*, "Reputation-driven multimodal emotion recognition in wearable biosensor network," in *IEEE I2MTC*, 2015, pp. 1747–1752.

[10] M. T. Valderas *et al.*, "Human emotion recognition using heart rate variability analysis with spectral bands based on respiration," in *37th EMBC*, 2015, pp. 6134–6137.

[11] N. TaheriNejad and D. Pollreisz, "Assessment of physiological signals during happiness, sadness, pain or anger," in *6th MobiHealth*, 2016.

[12] N. TaheriNejad, A. Jantsch, and D. Pollreisz, "Comprehensive observation and its role in self-awareness; an emotion recognition system example," in *FedCSIS*, 2016, pp. 123–130.

[13] A. Landowska *et al.*, "Limitations of emotion recognition in software user experience evaluation context," in *FedCSIS*, 2016, pp. 1671–1680.