REAL WORLD GROUP EMOTIONAL ANALYTICS USING ELECTRODERMAL ACTIVITY SIGNALS

Gonçalo Salvador

ABSTRACT

Emotion Recognition is a task easily performed by humans, however, a great challenge to replicate in computers. The current work develops several tasks required to perform emotion recognition. Such as: 1) the development of novel self-assessment annotations tool; 2) the benchmarking of a new device to perform physiological data acquisition in group settings; 3) evaluation of collective group emotions while watching a long-duration uncalibrated audiovisual content in the wild.

Being the last enumerated task the most relevant one for the current work. In the latter task the aim was to analyze the similarities in simultaneous annotations across different participants and develop a new approach to identify the time regions where the audience reacted with higher intensity based on the EDA data and machine learning techniques.

The annotations performed by the participants do not follow the expected, revealing some limitations in the annotations process. Regarding the application of clustering algorithms, hierarchical clustering with average linkage outperformed the remaining, providing the areas in which the audience had high/low emotional reaction.

Index Terms— Emotion recognition; Electrodermal activity; Emotional self-assessment; Valence-Arousal scale; Clustering algorithms.

1. INTRODUCTION

Humans can express emotions to each other based on body language, facial expressions and speech, therefore emotions can be recognized based on these traits which occur naturally. Nevertheless, these cues do not reliably represent the individuals inner emotional states, since these can be influenced by factors such as environment, cultural background, personality or mood, and they can also be effortlessly faked. The same is not applied to physiological manifestations.

The Autonomous Nervous System mediates the body response to internal or external stimulus, thus, modulating the physiological manifestations of the emotion felt [1]. The physiological manifestations are expressed in physiological signals such as the Electrocardiogram (ECG), Electroencephalogram (EEG), EDA, and others. The link between emotions and physiological responses can be used for emotion recognition. These responses are detectable, easily collected using wearables in a non obtrusive way, and they can reflect human emotions more reliably, since they can not be controlled or faked [2, 3]. With this in mind, it was considered that physiological signals are the preferable method to perform emotion recognition [1].

To study emotions, a first step should be to understand the concept of emotion. Over the years, several definitions have been proposed, although no consensus has yet been reached. Generally, the most accepted concept is that emotions can be described according to two different models: discrete model and continuous (or affective) model. In the discrete model, the emotional experiences are described based on a list of words to label emotions into categories. However, the list of words used for this description is still widely debated. This discretization of emotions can be difficult, since the distinction boundary between two emotions is often blurred, and the meaning of the chosen words are culturally dependent [1]. On the other hand, continuous models aim to describe emotions based on continuous scales addressing two factors: the correlation between distinct emotions, and the quantification of a certain emotion. Russel et al. [4] proposed a 2D Valence-Arousal space, in which valence describes how pleasant an emotion is, and arousal which describes the intensity level [5]. Although other models such as the valence-arousal-dominance have been suggested [6], the valence-arousal model is the most widely accepted [7].

A factor which greatly influences emotions is the group effect. Humans are highly social beings that tend to live in complex social structures. Thus, many emotions are experienced in social contexts where there can be several interactions between group members [8]. Previous work on emotion recognition focuses mainly on the analysis of emotion in an individual setting and in controlled environments, ignoring important dimensions. Hence, to evaluate emotions experienced by the subjects in a group setting, it is necessary to collect the data simultaneously from all participants.

The present work aims to fill a gap in emotion analysis by evaluating the emotional states in a group setting using long-duration uncalibrated elicitation content, i.e. movies. Although most studies focus on analysing emotions in an individual setting, being in a group environment can have different effects on the experienced emotions. The analysis of the emotional states in group setting was carried out based on EDA data acquired simultaneously from all participants. In addition, this work also seeks to develop an emotion self-assessment tool for smartphones. This tool is developed based on the Valence-Arousal model, enabling the users to perform their emotional self-assessment in a group setting with any sort of elicitation material, with minimal distraction.

The remainder of this work is organized as follows: Section 2 displays experimental data acquisition steps in a group setting, presenting the data acquisition and emotional annotation tools, along with the setup used to show the elicitation content and store the acquired data. Sections 3 and 4 summarize the methodologies and results&discussion, respectively, of the emotional analysis performed on the collective data. These sections contains 2 different analysis: the first consists in evaluating the annotations given by the participants and the second suggests a new approach to identify time regions where the audience reacted with higher intensity, based on EDA data and unsupervised machine learning techniques. Section 5 presents the benchmarking of a new device, the BITalino R-IoT, capable of acquiring EDA and Photoplethysmogram (PPG) data simultaneously across several participants. This section is subdivided into 2 parts: the first one presents the methodologies used and the second part displays the results&discussion of the benchmarking process. Finally, Section 6 draws the main conclusions obtained throughout the current work.

2. COLLECTIVE EMOTION ASSESSMENT

The device chosen to acquire physiological data during the visualization of audiovisual elicitation contents was the Xinhua Net Future Media Convergence Institute (FMCI) device. This device was selected due to its capability to wirelessly acquire EDA data from up to 20 devices simultaneously. The performance of this system was evaluated in [9], showing the feasibility of signal acquisition with no significant data loss, in collective settings. The FMCI device consists of a small wrist bracelet with two electrodes connected; two electrodes are attached to the palm or finger area, making it a highly mobile data acquisition device. The device collects EDA data through an embedded sensor designed to acquire EDA signals with a bandwidth between 0 and 5 Hz with a sampling frequency of 1 Hz.

The volunteers' emotional annotation was performed retrospectively using the EmotiphAI annotation tool developed by Bota et al. [10]. With EmotiphAI the users are able to see the video clips which they are currently annotating to recollect the emotions experienced, and afterwards annotate the emotional state according to the Valence-Arousal emotional model.

Each experiment was realized with a group of between 9 and 4 volunteers (μ=5.1, σ=1.6), older than 18 years old, without any known pathology. An assistant was present to ensure the protocol was properly followed; due to the COVID-19 pandemic, all experiments were performed following the sanitary guidelines from the national health authority (Direção Geral da Saúde - DGS). The trials were conducted in a familiar and comfortable environment for the participants to eliminate any bias, such as the stress of being in a new environment, and simulate as much as possible a real-world unconstrained scenario. The documents to be sign by the participants (e.g. informed consent document) as well as the experimental procedure were authorized by the ethics committee of the University under the process #1005890. To ensure the data privacy of each participant, a pseudonym was assigned and the data collected was disassociated from the participant's private information (e.g. name).

The audiovisual elicitation content consisted of recent uncalibrated movies that premiered in the last 3 years, thus approaching current topics. Furthermore, these contents covered 7 different genres to elicit a broad range of emotions.

A *Raspberry Pi 4 Model B* was used as a set-top media center to display the elicitation content, run the EmotiphAI data acquisition and annotation software, and store the physiological data. This media center was connected to a LCD monitor (*SAMSUNG UE65TU7025* with 65" - 165 cm) to exhibit the elicitation content. Furthermore, this device also receives the data sent in real-time by the EDA acquisition devices, and stores it locally. For the data acquisition, each participant had one Xinhua Net FMCI device connected to their wrist or forearm, this device has two EDA electrodes, which were placed in the palm of the hand with Silver–Silver Chloride (Ag/AgCl) electrodes. The Raspberry Pi embedded EmotiphAI software ensured the synchronization between the data acquisition and the video, by starting the two simultaneously.

After, the viewing of the movie, participants were asked to fill their self-assessment annotations regarding the content watched. The subjects' emotional annotation was performed using EmotiphAI's annotation tool using their mobile phones or a PC provided by the research assistant. Both the EDA signals and the emotional annotations are stored in the same HDF5 file in a hierarchical format. For each user a HDF5 dataset is created, containing all the information acquired from this user i.e. the EDA signal and the user's annotations.

3. PROPOSED METHODOLOGY

3.1. Signal Preprocessing

Data processing was performed on a Python 3 environment, with the support of the BioSPPy (version 2) toolbox [11]. The first step in the pre-processing of the EDA data was outlier removal and manual selection of which signals/participants to use. The exclusion criteria for the manual selection was based on the overall quality of the signals, i.e., saturated signals, interruptions amidst the acquisition, and signals with a constant amplitude were removed.

The EDA signal was interpolated to 10 Hz using a cubic spline interpolation. Afterwards, the signal was filtered with a 4th order low pass Butterworth filter with a 1 Hz cutoff frequency. Following the filter, the signal was smoothed using a 10 point moving average following the approach described in [12]. After these procedures, the signals were normalized per subject so that its range is between zero and one [1]. To decompose the EDA into its Electrodermal Response (EDR) and Electrodermal Level (EDL) components the *cvxEDA* algorithm was applied [13]. Finally, the identification of the fiducial points was achieved based on the method proposed by [14], which has the advantage of not requiring any type of threshold. The algorithm returns the onset, peak and end of each event.

3.2. Analysis of the Synchrony between Annotations

A first analysis was performed on the self-reporting given by the volunteers upon watching a movie. Namely, the potential synchrony between time regions in which each participant performed an annotation, and the total number of annotations across all the participants. Figure 1 gives an example on how these metrics were achieved. The red and blue lines represent the number of annotations throughout the duration of the movie and the annotations from both participants have an timewise overlap, thus they are considered to be simultaneous.

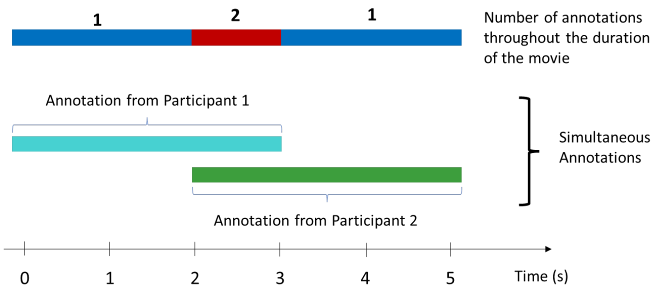

1
2
1
Number of annotations throughout the duration of the movie
Annotation from Participant 1
Simultaneous Annotations
Annotation from Participant 2
0
1
2
3
4
5
Time (s)

Fig. 1: Example of the number of annotation in each time instant and illustration of a simultaneous annotation across two participants

Afterwards, a histogram was plotted with the density of each annotation value for the Valence and Arousal dimensions. Each histogram had 5 columns, one per each annotation value from 1 to 5, with the total number of annotations performed with those values.

A further step was to analyse simultaneous annotations. The EDA data was concatenated for the entire duration of the time window considered synchronous. In the example demonstrated in Figure 1 EDA data from Participants 1 and 2 would be joined from the start of the annotation of Participant 1 until the end of the annotation of Participant 2. The EDA signal from the remaining participants (those who did not annotate in that time region) was also concatenated into a different group, for the entire duration of the time window considered synchronous. Furthermore, the mean EDA signal was calculated, along with the Standard Deviation (STD) for simultaneous annotations periods. Likewise, the procedure was implemented for the participants with no annotations during the timestamp. With the goal of comparing different simultaneous annotations, the annotations values, along with the number of EDR events and the Pearson Correlation Coefficient (PCC) between the EDA signals of each participant involved in the simultaneous annotations were acquired.

3.3. Collective Intelligence Analysis

A second analysis focuses on determining the time regions where the audience reacted with higher intensity. To this end, the first step is to extract representative features from the EDA signal. A wide range of features are observed in the literature [1], from which 13 were selected based on the information provided by each feature for emotion recognition. The features extracted were: Number of EDR onsets; Number of EDL onsets; Area under EDR events; Sum of the startle magnitudes; Sum of EDR event amplitudes; Sum of EDL event amplitudes; Sum of the response duration; Sum of the onset-peak times; Mean EDR; STD EDR; Dynamic range; Mean of the EDA derivative; Surface area between the EDA and its linear regression. These features were selected based on a trade-off between the number of features and the information provided by each individual feature.

These features were extracted from a group EDA signal performing a group analysis to see where the audience as a whole reacted. This group EDA signal was calculated by determining the mean EDA across all participants using a moving window with a length of 3 seconds and an overlap of 2 seconds. Note that the participant's EDA signals used to calculate the group EDA signal were previously normalized across all participants. For feature extraction the signal was divided into windows, in the literature, it was observed that windows should be between 10 to 300s [15]. In the current work, a window size of 20s with an overlap of 5s was used. Given that the average movie scene duration has between 1 and 3 minutes, a window smaller than this would be too granular and may not contain sufficient information for emotion recognition, while a longer window could encompass very different reactions, compromising the results.

The feature vectors were given as input for clustering algorithms. The clustering algorithms were applied to group the periods of the movie where the audience reacted similarly. The number of clusters in the hierarchical clustering was determined using the life-time criteria, while for the K-means several number of clusters were tested as input to the algorithm. In particular, the number of clusters was increased from 2 to 8 until there was a some distinction in the movie

[1] *minmax _scale* function of the scikit-learn tool

scenes in each clusters. From the resulting clusters, the corresponding movie clips were extracted to analyse the scenes which triggered such reactions. This analysis consisted in counting the number of clips in each cluster, their total duration along with a visualization of the clips to see if there were any similarities between them, and check if there was an emotional context behind such scenes. Assuming the premise that each emotional scene triggers an emotional response and that neutral scenes do not trigger any emotional response. This would mean that, ideally, every scene in clusters that only contain strong or emotional clips, triggered an emotional reaction, and all the emotional reactions elicited by the clips in that cluster were similar, thus being in the same cluster.

Note that this method is correlated with the intensity of the users' emotional reaction (Arousal), and less correlated to how positive or negative an emotion is (Valence). The current work relies on the EDA, which is strongly related to the Arousal dimension. The Mean Affective Profile (MAP) was determined using the methods described in the work of Fleureau et al.[16]. The MAP is a validated methodology in the literature that reflects the arousal variations of a global audience during a movie. This measurement was used as a ground truth to evaluate the performance of the clustering methods.

4. RESULTS & DISCUSSION

The current work focuses on the analysis of the data collected using as elicitation content the movie "Spider-Man: Far From Home". Due to the pandemic situation in which this work was developed it was hard to gather volunteers for the experiment leading to several delays in the acquisition process, thus only one movie was analysed. The movie at hand had a total duration of 1 hour and 55 minutes (6900 seconds), and it is classified on IMDB as an Action, Adventure and Sci-Fi film[2]. Data was acquired from 7 volunteers, from whom 2 where male; the average age of the participants was 20 years old, with a STD of 0.7. Participants 0 and 5 were excluded from the analysis, due to poor EDA signal quality.

4.1. Analysis of the Synchrony between Annotations

Figure 2 displays the total number of annotations performed by all participants throughout the movie. The parts of the movie without any annotation can be expected since the participants were only performing annotations of the most relevant parts of the content. In terms of the period in which there were annotations, the number of annotators for a given timestamp varies considerably, since participants react differently to the same elicitation. As such, a clip that triggers an emotional reaction in one participant may not trigger an emotional reaction in another participant. Even when clips trigger an emotional reaction across several participants there could be a delay between emotional responses.

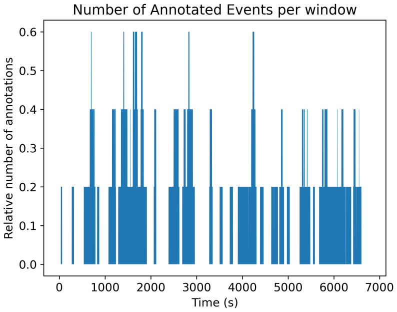


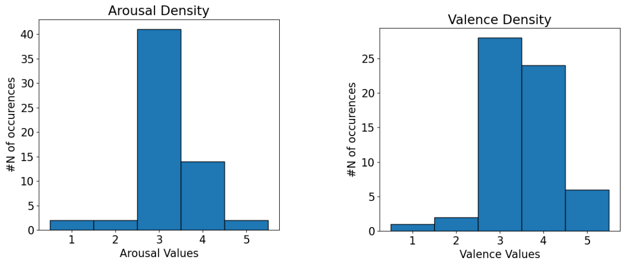
Fig. 2: Total number of annotations throughout the movie duration.

Figures 3a and 3b display a density histogram with the total number of annotations per value, for the Arousal and Valence dimensions, respectively. In terms of the values of the annotations, 39.3% corresponded to neutral states of Valence and Arousal and, as it is possible to see in these figures, the main annotation value in these dimensions was a 3 (with a slight tendency for higher values), which is a neutral emotional state. These values were not expected since the movie consisted of a high-pass superhero movie, with several fight scenes, along with some emotional and comical parts. One would expect that the annotations of the relevant parts of the film were mainly positive. Predominantly in the Arousal dimension, 65.6% of the annotations values were 3, although, since this dimension measures the intensity of emotions, in an action movie such as the Spider-Man, with very intense fight scenes, plot twists and comical scenes it was expected that the emotions elicited would be more intense, as the film suggests. These annotations suggest that a simple, unmeaningful conversation between two characters would elicit an emotion as intense as a fight scene where Spider-Man almost dies, or a comical scene when Peter gets caught by a friend in an awkward situation. The predominant Valence annotation value was also a 3, though in this case, it is clearly seen that these annotations tend to higher values, which describe a positive emotion as expected in these movies[3]. Nevertheless, in both dimensions, the number of annotations with values of 1 and 2 (extreme values) were very low (as expected), since these describe negative emotions, such as angry or scared, and inactive emotions, such as boredom or sadness, which are emotions that these kinds of movies do not aim to elicit in their audiences.

To establish a comparison between several simultaneous annotations, Table 1 displays 8 representative simultaneous annotations, out of the 20 total simultaneous annotations, along with some descriptive metrics. It is possible to see

[2] https://www.imdb.com/title/tt6320628/

[3] In contrast to a horror movie where it should be expected that the emotion elicited would be mainly negative



(a)

(b)

Fig. 3: Density histograms with the total number of annotations per value for the Arousal (a) and Valence (b) dimensions.

that the PCCs (ranged between -1 and 1) are all close to 0, suggesting that the EDA signals in simultaneous annotations have small correlation between them. In fact, the mean PCC between EDA signals of simultaneous annotations is 0.45 with a STD of 0.24, which implies a low average correlation in these signals. Furthermore, the number EDR events detected in these periods can be quite similar in some cases, although the results are not consistent across all simultaneous annotations, with some cases having a very different number of events being detected. Therefore, it is possible to conclude that the number of EDR events does not present a reliable correlation between the EDA data responses in simultaneous annotations

The EDA signals in simultaneous annotations displayed a tendency to increase in amplitude over the period of the annotations, while for the remaining participants, which did not annotated in the same period, the same was not observed, displaying a decreasing trend in the EDA signal. Furthermore, the signals during simultaneous annotations display few similarities, as observed by the PCC obtained across the participant who annotated in the same segment. Regarding the individuals' self-reports, the obtained values do not follow the expected by a review performed by an expert annotator, thus revealing a lack of comprehension of the annotations scales by the participants, a lack of engagement towards the content and/or the annotation task, leading them to perform the annotations carelessly, with minimal attention. The expert annotator consisted in a person with a background in emotional analysis and a vast knowledge of the Valence and Arousal scales. These difficulties in assessment methods have already been described in [16], since these methods can be strongly biased by the level of attention of the participant, by subjective factors in the perception of the scenes or their annotation. Furthermore, they can also be considered to be intrusive, thus leading the participants to be reticent in performing a genuine self-assessment. A possibility to overcome the annotation tools limitations is to perform an emotional analysis solely based on the acquired physiological signals, in this case the EDA, and the movie content.

4.2. Collective Intelligence Analysis

In Figure 4 it is possible to see the MAP, where each point represents the mean arousal of the audience at that instant. The mean MAP throughout the whole duration was 9.64×10^{-5} with a STD of 1.33×10^{-4}. This image also contains a representation of the most relevant scenes of the movie, the shaded light blue areas represent the periods in which such scenes occurred. Indeed, high correspondence is observed between the identified scenes by the annotator and the peaks of the MAP. Hence, verifying the correlation between higher arousal states given by the EDA data and stronger emotional scenes of the movie, i.e. intense scenes with strong emotions elicit high arousal emotional states in the audience.

Regarding the application of different clustering methods with features extracted from the group EDA signal, Table 2 presents a characterization of the resulting clusters. These results were obtained by extracting features from the group EDA signal, they represent the overall emotional response to the movie. Based on the analysis of the movie clips in each cluster, clusters 1 and 2 of the hierarchical average linkage, cluster 1 of the hierarchical single linkage and cluster 1 of the hierarchical complete linkage are all composed exclusively of intense movie clips that portray fight scenes, comical clips and emotional parts of the film, so these clusters successfully group the parts of the movie where the audience had a more intense emotional reaction (with higher arousal). It was observed that these clusters have the lowest duration and contain the most relevant scenes, very similar to each other in the elicitation emotional content. For the remaining clusters (with greater length), although they may also contain scenes that can be labelled as emotional, they mainly contain very long-lasting clips with "dead zones", i.e. filling parts of the movie where the history is developing without eliciting any relevant emotion.

A further approach to verify the clusters containing "dead zones" (besides the Length) is to identify the cluster with the initial instant of the movie, i.e. the first few seconds of the movie that displays the production companies. These instants are considered "dead zones" and should be included in the clusters with low emotional-intensity response. Regarding the clusters obtained with the hierarchical ward linkage method, there was no clear distinction in the clips contained in each cluster, i.e. all clusters seemed to contain both intense emotional clips as well as "dead zones" (even though cluster 1 appeared to contain less "dead zones" and more emotional clips). Concerning the clusters obtained with the K-means algorithm, it is clearly seen that this method is the one that produced the greater amount of clusters, although it is also the most difficult method to evaluate. Despite the fact that the length of each cluster is smaller, and as said before for the hierarchical clustering method, smaller clusters meant that each

Table 1: Comparison of the annotations and the EDA signals of the participants involved in each simultaneous annotations, along with a description of the correspondent movie scene

Participants	Annotations (V,A)			# EDR events			PCC			Scene description
1,4,6	1	4	6	1	4	6	1-4	1-6	4-6	Spider man is stressed with complicated questions and funny scene with may
	2,2	4,4	4,3	16	16	15	0,0424	0,6203	0,4707	
1,2,4	1	2	4	1	2	4	1-2	1-4	2-4	Intense Fight Scene
	3,3	3,1	3,4	51	94	18	0,2027	0,4841	0,3826	
1,2,6	1	2	6	1	2	6	1-2	1-6	2-6	Funny scenes and jokes / talk between Spider man and Fury
	3,3	3,3	4,2	49	15	29	0,2679	0,0529	0,1913	
1,3	1	3		1	3		1-3			Peter almost mistakenly kills a friend
	3,3	4,5		42	22		0,154			
1,3,6	1	3	6	1	3	6	1-3	1-6	3-6	Peter is scared for his friends safety
	3,3	5,4	4,3	57	34	40	0,6982	0,6492	0,8703	
1,3,6	1	3	6	1	3	6	1-3	1-6	3-6	MJ finds out about Peter being Spider-Man
	3,3	3,3	4,3	34	23	21	0,5624	0,71	0,2504	
1,4	1	4		1	4		1-4			Fight scene with lava monster
	3,3	3,3		14	4		0,6265			
2,4,6	2	4	6	2	4	6	2-4	2-6	4-6	Intense Fight Scene with water monster
	3,1	3,4	4,3	56	26	35	0,3578	0,1482	0,468	

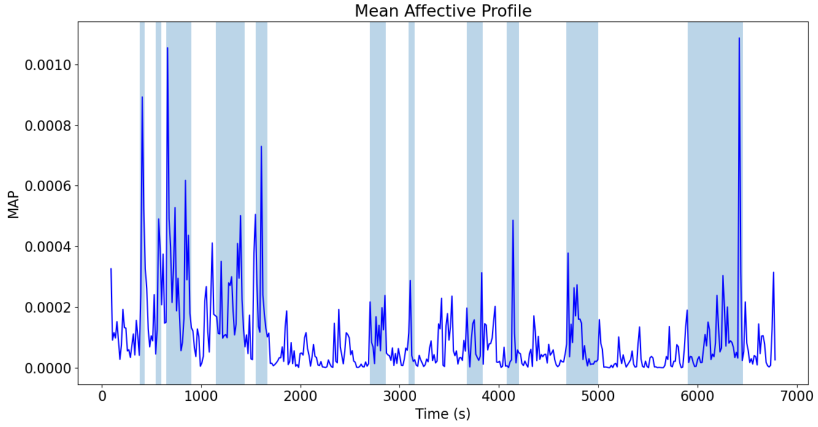
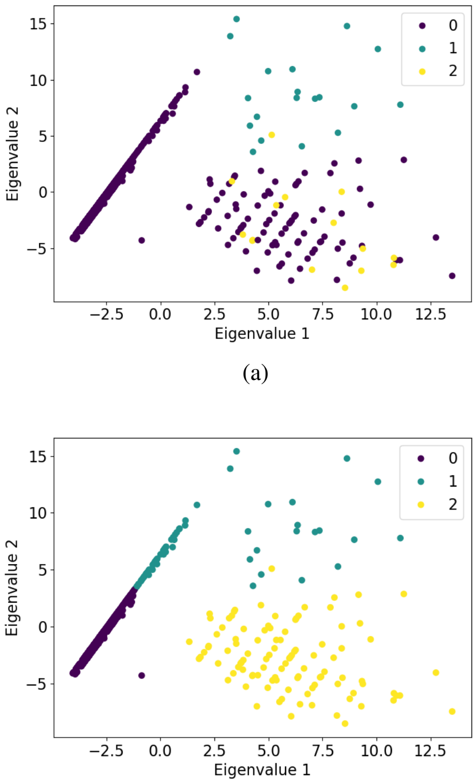


Fig. 4: MAP calculated throughout the duration of the movie, along the location of the most relevant scenes of the movie [17].



(a)

(b)

Fig. 5: Plot of the clusters obtained with different clustering methods using the group EDA signal, in a feature corresponding to the 2 eigenvectors obtained by a PCA of the 13 dimension feature space. (a) - Hierarchical clustering with average linkage and (b) - Hierarchical clustering with ward linkage.

cluster contained only relevant clips of the movie, in this case, some of the smaller clusters contain both relevant scenes as well as ”dead zones”. Even though two clusters seemed to stand out, clusters 3 and 5, they contained more intense emotional scenes than ”dead zones”, which suggests that the audience had a similar high arousal emotional response in these clusters.

Lastly, a comparison of the results achieved with the clustering methods with the MAP (considered as a validated method for emotional response profiling) is made based on the mean and STD of MAP for each cluster. So, if the clusters containing exclusively emotional movie clips have a high mean MAP, this means that these groups contain the part of the movie where the audience had a stronger emotional reaction. This would validate the correlation between intense scenes and strong emotional reactions, and the successful clustering of the more intense emotional reactions of the audience. In Table 2 it is possible to see that the clusters obtained with the hierarchical clustering with ward linkage have the lowest mean MAP as such these clusters did not group the parts of the movie where the audience had a stronger reaction. On the other hand, the smaller clusters obtained with the hierarchical clustering with average linkage have highest mean MAP, meaning that the clusters obtained through this method correspond to the areas where the audience had a more intense reaction, as expected from the analysis of the clips contained in these clusters.

In Figure 5 it is possible to observe the categorization achieved with 2 distinct hierarchical clustering algorithms, with the average linkage (Figure 5a) and with the ward linkage (Figure 5b), since these were considered to be the best and worst results obtained, respectively. These images display a plot of the clusters in a two dimensional space achieved using a Principal Component Analysis (PCA) on the original 13 features extracted from the EDA data. Based on the data clustering with the ward linkage, it is possible to see, that each cluster is almost completely separated from the reaming ones; this happens since this algorithm aims to cluster data in order to minimal the variance within each group. On the other hand, the average linkage although it may suggest that the clusters are not correctly separated, especially clusters 0 and 2, this method determines the distance between an observation and a cluster based on the average distance between that observation and each element of the cluster, which can result in the clusters not appearing to be precisely separated in a reduced dimension case, such as this one.

In conclusion, it is possible to determine the areas of the movie where the audience experienced an emotional reaction with higher intensity. When comparing this clustering methodology with the literature MAP, the best performing methodology was hierarchical clustering with average linkage, since it provides a higher number of clusters with more areas in which the audience had a more intense emotional reaction and it also differentiates the areas in which the audience reacted based on the intensity of such emotional reaction, i.e. it ranks the already stronger emotional reactions based on their intensity level into different clusters. Nevertheless, these results only provide insight to when and how much an audience reacts; they are mainly related to the Arousal dimension of emotion since the only physiological signal acquired was the EDA. To have an insight into the valence level of the audience (how the audience reacted), other physiological signals related to the Valence, such as the PPG, should be analysed.

5. BENCHMARKING OF THE BITALINO R-IOT

5.1. Proposed Methodology

In previous researches the FMCI device has been benchmarked with the BITalino (r)evolution as a reference [9]. As such, the current work of benchmarking of the BITalino R-IoT, also uses the BITalino (r)evolution as a reference. The

Table 2: Table with the characteristics of the group video clips achieved using different clustering algorithms with the group EDA signal.

Clustering Algorithm	Cluster	Counts	Length (s)	Mean MAP (E-04)	STD MAP (E-04)
Hierarchical Average Linkage	0	22	6327	0,90	1,88
	1	7	327	4,84	2,27
	2	15	315	1,73	1,27
Hierarchical Single Linkage	0	16	6576	1,57	1,26
	1	15	315	1,87	2,00
Hierarchical Complete Linkage	0	9	6414	0.95	1,86
	1	8	393	4,40	2,41
Hierarchical Ward Linkage	0	84	4614	1,25	1,16
	1	22	1182	2,38	2,09
	2	79	2019	1,03	1,75
K-Means	0	14	324	1,41	1,93
	1	74	2364	1,96	2,09
	2	15	315	0,62	0,76
	3	36	801	1,12	0,75
	4	26	891	1,21	1,19
	5	54	2559	3,37	2,05
	6	33	753	1,40	1,51
	7	5	240	2,63	2,58

BITalino (r)evolution has been tested several times in the past, revealing to be a reliable physiological signal acquisition tool, obtaining high quality data [18, 19]. However, with BITalino (r)evolution it is only possible to simultaneously acquired data from 3 devices due to its communication via Bluetooth and the high data throughput[4,5]. So, it can not be used as an acquisition tool in group settings.

The device being evaluated in the current work, the BITalino R-IoT, communicates via Wi-Fi which enables it to collect information from several devices simultaneously, eliminating the limitation set with the Bluetooth communication in the BITalino (r)evolution. Furthermore, the BITalino R-IoT acquires data with 200 Hz Sampling Frequency (SF), which as been shown to be enough for most physiological signals, namely the ones herein foreseen (EDA and PPG) [20]. This device is composed of a rechargeable 3.7 V battery, it contains integrated accelerometer, gyroscope, magnetometer and Euler angles calculation with 3 degrees of freedom along with a temperature sensor [21][6]. Moreover, it is also possible to add two additional sensors with 12-bit resolution, used to collect EDA and PPG data in this case.

For the current work data was acquired from volunteers, older than 18 years old, without any known pathology. The data acquisition was carried out in an individual setting, using both devices simultaneously, with the electrodes placed on finger of the subjects non-dominant hand according. The acquired signals from each device were stored in different HDF5[7] files in a hierarchical format. For each user a two signal dataset were created containing all the information acquired from this user.

To synchronise the data collected by the two devices, accelerometer data was acquired by the BITalino (r)evolution using a sensor placed on the BITalino R-IoT shell. The synchronization was achieved by creating a prominent peak in the acceleration signals of both devices with a small stroke on the sensors. Given that the sensors remained still for the rest of the acquisition, the signals would be constant throughout the acquisition, except for this peak, thus marking the beginning of the acquisition in both devices. During the signal processing step, the acquired signals from each device were cropped on the acceleration peak location of their device.

Data processing was conducted on a Python 3 environment, with the support of BioSPPy(version 2) toolbox [11], a publicly available set of signal processing tools to analyse biosignals. Since the signals acquired with the BITalino R-IoT had a lower SF, these where interpolated to the same SF as the BITalino (r)evolution (1000Hz) using a cubic spline interpolation. The EDA signal was smoothed using the 15×SF point moving average following the approach described in [12].

The comparison between the signals acquired with both devices was achieved based on the detected peaks and onsets of the PPG and EDA signals, respectively. To compare the data morphology, the PCC was calculated between the signals of the two devices. For the PPG these values were calculated for a time period which corresponded to 0.25 s before the detected peaks and 0.5 s after. For the EDA these values were calculated between the onset and end point of each event. Beyond these metrics, the number of detected peaks was also evaluated (for both EDA and PPG signals), along with an extraction of the Heart Rate (HR) from the PPG signals.

[4] https://plux.info/software/43-opensignals-revolution-000000000.html

[5] https://www.bluetooth.com/specifications/bluetooth-core-specification/

[6] https://bitalino.com/storage/uploads/media/manual-riot-v12.pdf

[7] https://www.hdfgroup.org/solutions/hdf5/

5.2. Results & Discussion

Based on Table 3 it is possible to see that the PPG results achieved with both devices are very similar. In terms of the number of detected peaks these are very close to each other, with about 95% of the detected peaks being correctly matched between the two signals. Furthermore, the analysis on the waveform similarities around each detected peak show a mean PCC value very close to 1, indicating an almost perfect similarity in waveform of the PPG signals acquired with both devices. Even though, the minimum PCC is close to 0 in the first 2 participants and smaller than 0 in the last participant, which demonstrates almost no correlation in these areas, the STD is very small in all participants which indicates a very low deviation from the mean value. All time difference values across the table are considered to be very small, with the maximum and minimum time differences being 0.1 and -0.1 s, which expresses a negligible time deviation in peak detection in the signals from the 2 devices. Lastly, the differences in the PCC extracted from each PPG signal are in the order of magnitude of 10^{-1}, which are also negligible considering the range of values for this metric.

Regarding the EDA signal comparison, Table 4 encompasses the results obtained. In this table it is possible to see that there are some differences in the number of detected onsets and % of matched onsets in participant 0, although in participants 1 and 2 the number of onsets are very close to each other, with about 90% matching onsets. Overall, the number of onsets detected with each device is very similar, with the great majority of the detected onsets being matched between the two devices. In terms of the waveform similarity, the mean PCC are above 0.9 for all participants, with an overall value of 0.94, which indicates an almost perfect similarity in waveform of the acquired EDA signals. The overall minimum PCC is observed in participant 0, even though these value still demonstrates some similarities in waveform. The observed mean EDA time differences are all considerably close to 0, showing a good data correspondence across devices.

6. CONCLUSIONS

Throughout the current work, several problems related with real-world group emotional analysis were addressed. Regarding the annotations performed by the participants, they were reported, mainly, as neutral states, which was not expected given that the movie consisted of a high-pass superhero movie, containing several fight scenes, with some emotional and comical parts. Thus revealing a lack of comprehension of the annotation's scales by the participants, a lack of engagement towards the content and/or the annotation task. Nevertheless, the signals during simultaneous annotations displayed few waveform similarities.

To overcome the annotation tool limitations, an emotional analysis solely based on the acquired physiological signals and movie content was performed. This analysis was conducted by extracting features from the mean EDA signal of the group and applying clustering algorithms to group the areas of the movie where the audience experienced a similar emotional reaction. The clustering results were then compared with the literature, namely the MAP. Based on this analysis it was possible to conclude that best performing methodology was hierarchical clustering with average linkage. This clustering methodology provides a higher number of areas in which the audience had a more intense emotional reaction, divided into two distinct clusters. Furthermore, within the areas in which the audience had a more intense emotional reaction, this method also provides a differentiation in the intensity of the reaction with one cluster having a mean MAP of 4.84E-04 and the other cluster having a mean MAP of 1.73E-04.

To address existing limitations in the real-world collective data acquisition, the BITalino R-IoT was evaluated; this device revealed to be a great asset for the acquisition of physiological data in collective environments, being able to collect two physiological signals simultaneously across several member of an audience. With the use of this device it would be possible to collect EDA and PPG data, thus providing a window of information to the Valence dimension of emotions. The benchmarking of a new device able of acquiring one additional physiological signal in a group setting, establishes a path to future works in the area of group emotion recognition.

Throughout the present work, other topics were also addressed, namely the development of a real-time emotional annotation tool, in the form of a smartphone application. The developed tool performs an unbounded emotional annotation which is promising approach for annotation of previously uncalibrated and unseen content with higher reliability. This tool was also evaluated in terms of its usability and mental workload and the results displayed a high usability and a low mental workload, thus providing a reliable emotional annotation with minimal distraction. Further details about this annotation tool can be seen in [22].

Overall, the current work fulfilled the objectives drawn at the beginning, expanding the state-of-the-art by developing a new self-assessment tool and implementing machine learning methods to emotional assessment in a collective setting, namely of the audience EDA signal. Future work will focus on expanding the database using the protocol developed with the BITalino R-IoT acquiring EDA and PPG data; further validate the annotation tool developed and applying the developed emotional analysis method (application of clustering algorithms to EDA data) on other movies.

7. ACKNOWLEDGEMENTS

This document was written and made publically available as an institutional academic requirement and as a part of the evaluation of the MSc thesis in Biomedical Engineering of the

Table 3: Comparison between the number of detected peaks, PCC, temporal difference and extracted HR for the PPG signal extracted with the BITalino (r)evolution and BITalino R-IoT.

	Nb Values			Waveform Similarity			Time difference (s)				HR (bpm)	
Participant	#N peaks R-IoT	#N peaks (r)evolution	% matched peaks	Mean PCC	STD PCC	Min PCC	Mean	STD	Max	Min	R-IoT	(r)evolution
0	3633	3747	89,78	0,91	0,15	0,02	-0,02	0,02	0,10	-0,10	77,8	77,7
1	3287	3228	96,59	0,96	0,08	-0,02	-0,01	0,02	0,10	-0,10	62,29	62,29
2	3961	3979	96,33	0,97	0,06	-0,47	-0,04	0,02	0,09	-0,10	68,65	68,66
Overall	10881	10954	94,17	0,94	0,11	-0,47	-0,03	0,03	0,10	-0,10		

Table 4: Comparison between number of detected onsets, PCC, temporal difference and event duration for the EDA signal extracted with the BITalino (r)evolution and BITalino R-IoT.

	Nb Values			Waveform Similarity			Time difference (s)			
Participant	#N onsets R-IoT	#N onsets (r)evolution	% matched onsets	Mean PCC	STD PCC	Min PCC	Mean	STD	Max	Min
0	57	46	45,65	0,91	0,18	0,26	-0,17	0,62	1,04	-1,94
1	62	63	88,89	0,98	0,05	0,66	-0,17	0,47	1,88	-1,24
2	104	97	93,81	0,98	0,05	0,64	-0,13	0,46	0,87	-1,64
Overall	249	235	72,34	0,96	0,15	0,26	-0,15	0,51	1,88	-1,94

author at Instituto Superior Técnico. The work previously described was performed at Instituto de Telecomunicações (IT) and at the Department of Bioengineering of Instituto Superior Técnico (Lisbon, Portugal), in collaboration with the British Broadcasting Corporation Research and Development (BBC R&D), during the period February-October 2021, under the supervision of Prof. Ana Fred and Prof. Hugo Silva.

References

[1] P. Bota, C. Wang, A. Fred, and H. Silva, "A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals," *IEEE Access*, vol. 7, pp. 140 990–141 020, 2019.

[2] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: a review," in *Proc. of the IEEE Int'l Colloq. on Signal Processing and its Applications*, 2011, pp. 410–415.

[3] J. Domínguez-Jiménez, K. Campo-Landines, J. Martínez-Santos, E. Delahoz, and S. Contreras-Ortiz, "A machine learning model for emotion recognition from physiological signals," *Biomedical Signal Processing and Control*, vol. 55, p. 101646, 2020.

[4] J. Russell, "Affective space is bipolar." *Journal of Personality and Social Psychology*, vol. 37, pp. 345–356, 1979.

[5] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, 2018.

[6] A. Mehrabian, "Framework for a comprehensive description and measurement of emotional states." *Genetic, social, and general psychology monographs*, 1995.

[7] M. Bradley and P. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[8] S. Barsade and D. Gibson, "Group affect: Its influence on individual and group outcomes," *Current Directions in Psychological Science*, vol. 21, no. 2, pp. 119–123, 2012.

[9] P. Bota, C. Wang, A. Fred, and H. Silva, "A wearable system for electrodermal activity data acquisition in collective experience assessment." in *Proc. of the Int'l Conf. on Enterprise Information Systems: ICEIS*, 2020, pp. 606–613.

[10] P. Bota, P. Cesar, H. Silva, and A. Fred, "Unveiling the potential of retrospective ground-truth collection for affective computing."

[11] C. Carreiras, A. Alves, A. Lourenço, F. Canento, A. Silva, H. and. Fred *et al.*, "BioSPPy: Biosignal processing in Python," 2015–, [Online; accessed 14/10/2021]. [Online]. Available: https://github.com/PIA-Group/BioSPPy/

[12] S. Smith, *The scientist and engineer's guide to digital signal processing*. California Technical Publishing, 1997.

[13] A. Greco, G. Valenza, A. Lanata, E. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2015.

[14] K. Kim, S. Bang, and S. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and biological engineering and computing*, vol. 42, no. 3, pp. 419–427, 2004.

[15] R. Martinez, A. Salazar-Ramirez, A. Arruti, E. Irigoyen, J. Martin, and J. Muguerza, "A self-paced relaxation response detection system based on galvanic skin response analysis," *IEEE Access*, vol. 7, pp. 43 730–43 741, 2019.

[16] J. Fleureau, P. Guillotel, and I. Orlac, "Affective benchmarking of movies based on the physiological responses of a real audience," in *Humaine Association Conf. on Affective Computing and Intelligent Interaction*. IEEE, 2013, pp. 73–78.

[17] C. Wang and P. Cesar, "Measuring audience responses of video advertisements using physiological sensors," in *Proc. of the Int'l Workshop on Immersive Media Experiences*, 2015, p. 37–40.

[18] D. Batista, H. Silva, A. Fred, C. Moreira, M. Reis, and H. Ferreira, "Benchmarking of the BITalino biomedical toolkit against an established gold standard," *Healthcare technology letters*, vol. 6, no. 2, pp. 32–36, 2019.

[19] D. Batista, H. Silva, and A. Fred, "Experimental characterization and analysis of the BITalino platforms against a reference device," in *Int'l Conf. of the IEEE Engineering in Medicine and Biology Society*, 2017, pp. 2418–2421.

[20] S. Béres and L. Hejjel, "The minimal sampling frequency of the photoplethysmogram for accurate pulse rate variability parameters in healthy volunteers," *Biomedical Signal Processing and Control*, vol. 68, p. 102589, 2021.

[21] E. Ramos, H. Silva, B. Olstad, J. Cabri, and P. Lobato, "SwimBIT: A novel approach to stroke analysis during swim training based on attitude and heading reference system (AHRS)," *Sports (Basel, Switzerland)*, vol. 7, no. 11, 2019.

[22] G. Salvador, P. Bota, V. Vinayagamoorthy, H. Silva, and A. Fred, "Smartphone-based content annotation for ground truth collection in affective computing," in *ACM Int'l Conf. on Interactive Media Experiences*, 2021, p. 199–204.