2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Assessing Arousal Through Multimodal Biosignals: A Preliminary Approach

Rita Correia, Daniel Agostinho, Isabel Catarina Duarte, Daniela Sousa, Ana Pina Rodrigues, Miguel Castelo-Branco and Marco Simões

Abstract— **The increase in Autism Spectrum Disorder (ASD) prevalence estimates over the last decades has driven a quest to develop new forms of rehabilitation that can be accessible to a larger part of this population. These rehabilitation approaches often take the form of computer games that are blind to the user's emotional state, which compromises their efficacy. In this study, a set of physiological signals were acquired in simultaneous with functional Magnetic Resonance Imaging (fMRI) with the future prospect of combining both kinds of data to create models capable of assessing the true emotional state of their users based on physiological response as a measure of autonomic nervous system, having as ground truth the activity of targeted brain regions. This paper describes an initial approach, focusing on the information contained on the physiological signals alone. A total of 35 features were extracted from biosignals' segments and subsequently used for automatic classification of arousal state (High Arousal vs. Low Arousal). The suboptimal results, although some extracted features present statistically significant differences, underline the challenging nature of our proposal and the added obstacles of recording physiological signals in the magnetic resonance environment. Further exploration of the measured signals is needed to gather a bigger number of discriminative features that can improve classification outcomes.**

I. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects social and communication skills, as well as normal patterns of behavior, interests and activities [1]. As of 2014, 16.8 per 1000 children aged 8 years were diagnosed with ASD in the United States of America (USA), which represents an increase of 150% when compared with 2000 estimates [2]. Hence, the number of people that can benefit from new and improved rehabilitation approaches is enormous and continues to rise.

Over the last decades, there has been an increasing interest in serious gaming as an alternative or complement approach to the traditional therapeutical interventions. A serious game is a game with an educational purpose, going beyond the sole purpose of entertainment. This rehabilitation tool represents a low-cost option that allows for the repeated practice of different skills that are usually impaired in the ASD population. Autistic individuals, however, generally present an increased sensory sensitivity, which may compromise the full potential and efficacy of the serious games, the presentation of the wrong type and number of stimuli may lead to the disengagement of the user from the game, or even to a complete rejection of the intervention [3]. Therefore, the next step must be to optimize the serious games based on the emotional state of its users.

In this sense, this project aims to develop biofeedback-based models, particularly designed for the ASD population, that, based on autonomic nervous system (ANS) physiological signals, can infer the state of the user, having as ground truth the neuronal activation evoked by different emotion eliciting stimuli.

While physiology based automatic emotion assessment has been substantially considered for typically developed (TD) individuals [4]–[8], it is underexplored for ASD. To the best of our knowledge, there are only three papers describing automatic emotion classification in autistic subjects. By measuring EDA, PPG, skin temperature, EMG and ECG on children with ASD while they performed computerized cognitive tasks, Liu et al. (2008) successfully attempted to classify emotional states of liking, anxiety and engagement in this population, achieving accuracies of 82.9 % with Support Vector Machines (SVM) [9]. Kushki et al. (2015) classified anxiety-related arousal using metrics derived from the ECG and a modified Kalman filter obtaining an average specificity of 92% and sensitivity of 99% [10]. More recently, Sarabadani et al. (2020) automatically discriminated positive from negative valence during high and low arousal in ASD obtaining accuracies of 78.1% and 84.5% for high arousal and low arousal, respectively, using K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA) and SVM, and combining the outputs of all the classifiers using a Majority Vote to enhance the performance [11].

While these positive outcomes suggest that emotion recognition is a viable approach in ASD, the evidence of emotion dysregulation in this population [12] seems to be overlooked. This evidence means that the use of self-assessment questionnaire responses or labels based on the general population's emotional perception of a stimulus as ground truth is of limited accuracy. For this reason, we believe that, due to its spatial resolution that allows for the precise mapping of brain regions or networks of interest, functional Magnetic Resonance Imaging (fMRI) is the ideal true state indicator. Sessions involving this imaging technique, however,

This work was supported in part by the Santander/University of Coimbra seed project BioHab."

All authors are with the Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT), from the Institute of Nuclear Sciences applied to Health (ICNAS) of the University of Coimbra. D. A., R. C. and M. S. (email: msimoes@dei.uc.pt) are also with the Center for Informatics and Systems of the University of Coimbra (CISUC).

The authors thank the participants and their families for joining the study, as well as the patients' associations of APPDA-Viseu and Coimbra that supported the recruitment.

are quite expensive and nonportable, which limits their applicability. With this study, we intend to find ANS physiological patterns that are representative of the targeted brain regions modulation, so that it can be inferred outside the MR scanner.

To this end, respiration, photoplethysmography (PPG), electrodermal activity (EDA), electroencephalography (EEG), pulse oximetry (SpO2) and pupil size were recorded simultaneously with fMRI in ASD patients and a matched TD group, while watching short videos, chosen specifically to induce different kinds of emotional response.

This paper describes an initial approach to the experiment, which includes feature extraction from the physiological signals and subsequent binary classification into high or low arousal states. Given the early nature of this study, the data acquired from the fMRI are not yet considered. Instead, the ratings of arousal from the database where the videos were taken from were used as classification labels.

II. Methods

A. Participants

Fourteen individuals with ASD (1 female), and 13 typically developed (TD) individuals (2 females) took part in this study. Participants (or their legal representatives) signed an informed consent to participate in the study. Every subject completed the entire task. Table I provides a detailed description of the participants.

TABLE I. DEMOGRAPHIC DESCRIPTION OF THE ASD AND TD GROUPS, INCLUDING AGE, FULL-SCALE INTELLIGENCE QUOTIENT (FSIQ), EMPATHY QUOTIENT (EQ), AUTISM SPECTRUM QUOTIENT (AQ) AND THE AUTISM DIAGNOSTIC OBSERVATION SCHEDULE (ADOS-II) TOTAL SCORE. EACH SCORE IS PRESENTED IN TERMS OF GROUP AVERAGE AND STANDARD ERROR, IN BRACKETS. GROUP DIFFERENCES WERE ASSESSED WITH A TWO-SAMPLE T-TEST, WITH P-VALUES ON THE LAST COLUMN. GROUPS ARE MATCHED BY AGE AND EQ.

	ASD	TD	*P*
N	14	13	
AGE	21.58 (1.36)	23.15 (0.91)	**0.34**
FSIQ	94.50 (2.97)	111.23(4.30)	<0.01
EQ	38.45 (4.54)	45.62 (2.89)	**0.18**
AQ	24.17 (1.49)	15.38 (1.54)	**<0.01**
ADOS-II	11.17 (0.72)	-	-

B. Experimental Task

The task follows a block design. Each block consists of a 15 second video presentation trailed by a self-assessment period and is preceded by a rest period of approximately the same duration. The protocol is composed by 3 task runs, and each run is made up of 10 video trials.

The 30 videos (10 videos x 3 runs) were taken from the Chieti Affective Action Videos (CAAV) database and represent different actions, examples include hugging someone, stealing from another or simply hanging a jacket. For our experiment we chose to use videos recorded in the 1st person perspective and to coincide the gender of the participant to the main actor [7]. Each video in the database is accompanied by the mean rating of valence and arousal given by an evaluation group using a 9-point Self-Assessment Manikin (SAM). Consequently, each video falls into one of the following 3 categories: low valence and high arousal (LVHA); high valence and high arousal (HVHA); no valence and low arousal (NVLA). Thus, 10 videos of each category were selected to integrate the task.

For the self-assessment, the subjects were asked to rate the video they just watched also in the 9-point SAM scale. For this purpose, participants used a joystick.

Before each session, the task was explained and participants were asked to rate some training videos, to guarantee that both concepts of valence and arousal were understood, and that they knew how to operate the joystick.

C. Data Acquisition

EEG, EDA and SpO2 were acquired using the MP150 system and AcqKnowledge 4.2 software (BIOPAC Systems, Inc.). Respiration and PPG were recorded using the Physiological Measurement Unit of the MRI scanner (Siemens Healthcare) and pupil size was registered using the EyeLink 1000 Plus Eye Tracker with the long-range mount (SR Research Ltd.). Due to the hypersensitivity of the ASD population, we tried to simplify and reduce preparation time as much as possible, thus, EEG was acquired using only 3 electrodes, placed on the forehead, and either the right or left earlobe and temporal area. EDA was measured using 2 Ag/AgCl electrodes taped to the proximal phalanges of the index and middle fingers of the participant's nondominant hand. SpO2 and PPG were measured using a pulse oximetry and pulse finger sensors, respectively. Respiration was measured with a respiratory cushion attached to the participant using a respiratory belt. EEG, EDA and SpO2 were recorded with a sampling rate of 5000Hz, PPG and Respiration were acquired at 400Hz and pupil size at 500Hz.

D. Signal Processing

Photoplethysmography

To reduce noise contamination, the PPG signal was bandpass filtered using a 6th order Butterworth filter with a lower cut-off frequency of 0.5Hz and a higher cut-off frequency of 20Hz.

The clean PPG signal was then used to compute the Heart Rate (HR) by identification of the PPG pulse peak. HR is affected by both, the sympathetic and parasympathetic nervous systems, and is one of the most popular measures when it comes to emotion assessment.

Electrodermal Activity

EDA data were high-pass filtered with a 0.5Hz cut-off frequency as it was being collected, and it was later low-pass filtered using a 5th order Butterworth filter with a 1Hz cut-off frequency.

Electroencephalography

As expected, the EEG recordings were considerably contaminated by MR gradient switch artifacts. To correct them, the FMRI Artifact Slice Template Removal (FASTR) algorithm from the FMRIB plug-in for EEGLAB (version 1.21) was used. Feeding the algorithm with the corrupted signal and the events for each slice acquired, it computes an

average template for the artifact and subtracts it from the signal, locked to each slice trigger.

Signal Segmentation

Lastly, signals were divided into 30 second segments, time-locked to the beginning of each video, this way each segment includes the 15 seconds of the video, the self-assessment period, and some seconds after.

E. *Feature Extraction*

For each video trial, a total of 35 features was extracted from the different biosignals. To account for possible carryover effects, the value of the feature for the last 5 seconds of the previous rest period was subtracted after extraction.

For each signal, a brief description of the extracted features is given in Table II.

F. *Statistical Analysis*

In a first approach, with the intention of inspecting the extracted features for significant differences among conditions and groups, the means of every feature for each subject were computed, for the conditions of High Arousal (HA) and Low Arousal (LA). Given its' subjective nature, for this preliminary study, the valence dimension was not considered. The HA and LA conditions were then obtained by condensing the 3 original ones (LVHA, HVHA, NVLA) and were defined considering both the database ratings and the self-assessed arousal values given by each participant. Using the database ratings, a trial was labeled as LA if the arousal rating for the corresponding video was lower than 4 and labeled as HA otherwise. For the self-assessment, k-means clustering was performed on each participant's answers individually to partition them into 2 clusters. Trials were then classified as HA or LA based on the cluster they fell into. Wilcoxon signed rank tests were then performed to look for statistically significant differences in feature values between HA and LA, for each group, and Wilcoxon rank sum tests were applied to look for differences between groups for the two conditions.

G. *Classification*

In order to explore the accuracy of automatic emotion assessment in the data acquired with our experimental protocol, 4 classification algorithms were applied. The ratings of the CAAV database for arousal were used to label the data.

The considered classifiers were a Euclidean Minimum Distance Classifier (MDC – Euclidean), a K-Nearest Neighbors (KNN) and SVM using a Radial Basis Function (SVM RBF) kernel and a linear (SVM Linear) kernel. The optimal parameters for the KNN (number of neighbors, K) and SVM (cost, C and Kernel Parameter, γ) were determined by applying a 50/50 partition on the training set 5 times and choosing the parameters that resulted in the smallest classification error. After the hyperparameters were selected, the classifiers were retrained with all training data for the chosen parameters.

The classifiers were then tested for both intraparticipant and interparticipant classification. For the intraparticipant

approach, data from each subject were randomly split using the 70:30 ratio, where 70% of the data were used to train the classifier and the remaining data were used for testing. This process was repeated 30 times to avoid outlier results. As for the interparticipant classification, we employed the Leave One Subject Out (LOSO) method where the data from each participant are used once for testing, while the rest of the participants' data are used to train the classifier.

Finally, to ascertain if the accuracies of the classifiers were significantly greater than chance level (50%), permutation tests were used. For each partition, after testing, the true labels of the test set were iteratively shuffled, and accuracies were calculated using the random labels as the predicted classes. The number of times that these accuracies were greater than the one obtained with the classes predicted by the classifier, were then counted.

TABLE II. LIST AND DESCRIPTION OF THE FEATURES EXTRACTED FROM THE PHYSIOLOGICAL SIGNALS.

Signal	Feature Name	Extracted Features
EDA	meanEDA	Mean
	maxEDA	Maximum
	minEDA	Minimun
	mean_abs_fd_EDA	Mean Absolute First Difference
	mean_deriv_neg_EDA	Mean of Derivative for Negative Values
PPG	meanPPG	Mean
	maxPPG	Maximum
	minPPG	Minimum
	NNmean	Mean of the Normal-to-Normal (NN) time intervals
	SDNN	Standard Deviation of NN intervals
	SDSD	Standard Deviation of Successive Differences between NN intervals
	RMSSD	Root Mean Square of Successive Differences between NN intervals
	NN50	Number of Successive Differences greater than 50ms
	pNN50	Ratio between NN50 and total number of NN intervals
EEG	delta, theta, alpha, beta, gamma	Relative Power (delta, theta, alpha, beta and gamma bands)
Respiration	meanResp	Mean
	maxResp	Maximum
	minResp	Minimum
	mean_abs_fd_resp	Mean Absolute First Difference
Pupil Size	meanPupilSize	Mean
	maxPupilSize	Maximum
Heart Rate	meanHR	Mean
	maxHR	Maximum
	minHR	Minimum
	VLF	Relative Power (Very Low Frequency, Low Frequency, and High Frequency bands)
	LF	
	HF	
	RaLH	Ratio between Low and High Frequency Powers
SpO2	meanSpO2	Mean
	maxSpO2	Maximum
	minSpO2	Minimum

III. RESULTS

The significance levels that resulted from the statistical analysis are present in Tables I and II, for comparisons between conditions and groups, respectively. Only the features that suggest statistically significant differences for at least one scenario are shown in each table.

Apart from the MDC – Euclidean which returned poor accuracies, not significantly higher than chance level, all other

3 classifiers exhibited similar results on classifying HA vs LA. While accuracies for the intraparticipant classification for both groups are highly variable and for the most part, not significantly higher than chance, the interparticipant outcomes have a narrower distribution and exhibit accuracies higher than random chance more than 80% of the times. The median accuracy value for all 3 interparticipant modalities is of approximately 60%. To illustrate these findings, classification results for SVM Linear on intra and inter subject modalities on classifying HA vs LA are displayed in Fig.1.

TABLE III. P-VALUES OF PAIRWISE COMPARISONS FROM WILCOXON SIGNED RANK TEST (HIGH AROUSAL COMPARED TO LOW AROUSAL)

Feature	Database		Self-Assessment	
	Clinical (N = 14)	***Control (N = 13)***	***Clinical (N = 14)***	***Control (N = 13)***
meanPPG	0,71	0,31	**0,01**	0,19
minPPG	0,19	0,15	0,76	**0,02**
maxResp	0,67	**0,05**	0,67	0,17
meanPupilSize	0,67	**0,03**	0,76	1,00
meanHR	**0,01**	0,38	0,12	0,84
maxHR	**0,02**	0,45	0,24	0,84
minHR	**< 0,01**	0,59	**0,01**	0,95
RaLH	**0,04**	0,95	**0,01**	0,68

TABLE IV. P-VALUES OF PAIRWISE COMPARISONS FROM WILCOXON RANK SUM TEST (CLINICAL GROUP COMPARED TO CONTROL GROUP)

Feature	Database		Self-Assessment	
	High Arousal	***Low Arousal***	***High Arousal***	***Low Arousal***
minPPG	0,87	**0,04**	0,72	**0,03**
delta	**0,05**	**0,01**	**0,04**	**0,03**
beta	**0,03**	**0,01**	**0,03**	**0,03**
gamma	0,08	**0,04**	**0,05**	0,10
meanHR	**0,01**	0,68	**0,01**	0,17
maxHR	**0,01**	0,37	**0,01**	0,72
minHR	**0,01**	0,87	**0,01**	0,15

IV. DISCUSSION

The statistical analysis revealed that only a limited number of extracted features present statistically significant differences between HA and LA conditions as well as between groups. The time-domain features of the HR (meanHR, maxHR and minHR) seem to be significantly different between conditions for the clinical group, and also between groups for the HA condition, which is in accordance with the known relation of this signal with emotion discrimination [13].

These results help to explain the low classification accuracies and reveal the low discriminative power of the extracted features for distinguishing between HA and LA.

A possible explanation for the poor results of the intraparticipant modality is the low number of observations for each individual participant, which limits the generalization capacity of the models.

The simultaneous acquisition of physiological signals and fMRI represents a great challenge. Besides the common noise sources, there is the added artifact caused by the gradient switch of the MRI scanner that severely contaminates most of the recordings and results in low signal-to-noise ratios (SNR). This interferes with the quality of the features derived from the signals and hinders the appearance of subtle differences between states, which could be valuable for their distinction.

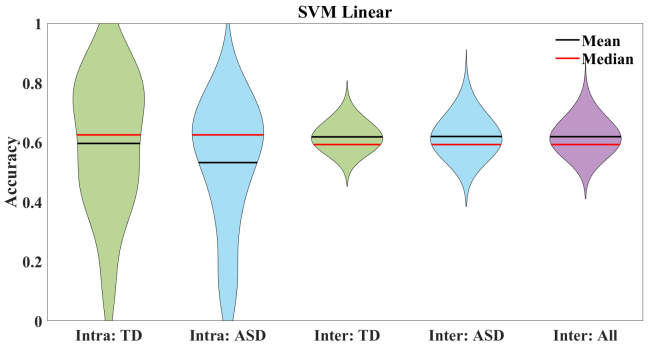

Figure 1. Distribution of accuracies achieved by SVM on classifying HA vs LA on intra and inter subject modalities.

This preliminary approach reiterated the challenging nature of this project and highlighted the need to further explore the biosignals in order to find meaningful features that can optimize the classification results.

REFERENCES

[1] C. P. Johnson and S. M. Myers, "Identification and Evaluation of Children With Autism Spectrum Disorders," *Pediatrics*, vol. 120, no. 5, 2007, doi: 10.1542/peds.2007-2361.

[2] J. Baio *et al.*, "Prevalence of autism spectrum disorder among children aged 8 Years - Autism and developmental disabilities monitoring network, 11 Sites, United States, 2014," *MMWR Surveill. Summ.*, vol. 67, no. 6, 2018, doi: 10.15585/mmwr.ss6706a1.

[3] S. H. Baum, R. A. Stevenson, and M. T. Wallace, "Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder," *Prog. Neurobiol.*, vol. 134, pp. 140–160, 2015, doi: 10.1016/j.pneurobio.2015.09.007.

[4] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, 2008, doi: 10.1109/TPAMI.2008.26.

[5] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Med. Biol. Eng. Comput.*, vol. 42, no. 3, pp. 419–427, 2004, doi: 10.1007/BF02344719.

[6] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, 2001, doi: 10.1109/34.954607.

[7] C. L. Lisetti and F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *EURASIP J. Appl. Signal Processing*, vol. 2004, no. 11, pp. 1672–1687, 2004, doi: 10.1155/S1110865704406192.

[8] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty," *IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans*, vol. 41, no. 6, pp. 1052–1063, Nov. 2011, doi: 10.1109/TSMCA.2011.2116000.

[9] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder," *Int. J. Hum. Comput. Stud.*, vol. 66, no. 9, pp. 662–677, 2008, doi: 10.1016/j.ijhsc.2008.04.003.

[10] A. Kushki, A. Khan, J. Brian, and E. Anagnostou, "A Kalman filtering framework for physiological detection of anxiety-related arousal in children with autism spectrum disorder," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 3, pp. 990–1000, 2015, doi: 10.1109/TBME.2014.2377555.

[11] S. Sarabadani, L. C. Schudlo, A. A. Samadani, and A. Kushski, "Physiological Detection of Affective States in Children with Autism Spectrum Disorder," *IEEE Trans. Affect. Comput.*, vol. 11, no. 4, pp. 588–600, 2020, doi: 10.1109/TAFFC.2018.2820049.

[12] A. C. Samson, J. M. Phillips, K. J. Parker, S. Shah, J. J. Gross, and A. Y. Hardan, "Emotion dysregulation and the core features of autism spectrum disorder," *J. Autism Dev. Disord.*, vol. 44, no. 7, pp. 1766–1772, 2014, doi: 10.1007/s10803-013-2022-5.

[13] P. Rainville, A. Bechara, N. Naqvi, and A. R. Damasio, "Basic emotions are associated with distinct patterns of cardiorespiratory activity," *Int. J. Psychophysiol.*, vol. 61, no. 1, pp. 5–18, 2006, doi: 10.1016/j.ijpsycho.2005.10.024.