

Emotion recognition through physiological sensors using supervised learning reinforced with facial expressions

Sebastián González¹², Matias Alonso¹², Fernando Elkfury¹², Jorge Ierache¹²³

¹Instituto de Sistemas Inteligentes y Enseñanza Experimental de la Robótica.

²ESIICA Universidad de Morón (1708) Morón Argentina.

³Laboratorio de Sistemas Información Avanzados Universidad de Buenos Aires(C1063) Ciudad Autónoma de Buenos Aires, Argentina.

{sebastianlgonzalez, matialonso,
felkfury, jierache}@unimoron.edu.ar

Abstract. A great deal of information is transmitted through facial expressions, skin conductance and heart rate. One of the most interesting characteristics of these physiological values is that they contribute to the determination of emotions. This work objective is to predict the emotional state of a subject through heart rate, galvanic response and face capturing. The different steps of the development are described, the experimental results with several classifiers, the management and registration through a multimodal framework and the storage of the stimulus information (images and videos), face and physiological sensor data (skin conductance and heart rate) in a unified database, to be processed by supervised learning models, which will seek to predict arousal values. Ending with a predicted emotional profile for the test subject, representable in the arousal/valence plane for a given time.

Keywords: Affective computing, heart rate, skin conductance, facial expressions, supervised learning.

1 Introduction

Not so many years ago, the inclusion of emotional processes in software and hardware design was purely utopian and, for some, certainly absurd. However, for some people like Rosalind Picard, a researcher at the Massachusetts Institute of Technology (MIT), this field of research held the potential to be a great, hence the term affective computing arose in 1997 with the publication of her book “Affective Computing”. [1]. Picard [2] states that in order to design a device that can process information imitating the human mind, it must be endowed with both the ability to think and to feel. And for a machine to acquire these skills, it must have the ability to perceive a large set of stimuli that can come from the environment as well as from the subject with whom it interacts. A feature that today is not found in many everyday devices. Unimodal and multimodal solutions for eliciting emotions and their representation are presented considering dimensional and categorical approaches [3]. We are working with different combinations of

physiological parameters: EEG [4] [5], Galvanic Skin Response (GSR) [6] [7], HR [8], Temperature, blood pressure or combinations of these [9] [10]. Skin conductance along with parameters associated with heart rate prove to be indicators of tension or arousal. While facial expressions are more effective valence values calculations. This work objective is to predict the emotional state of a subject through heart rate, galvanic response and face capturing. The developed model will receive, as input signal, data from the selected biosignals from which the excitation/relaxation state will be inferred. The performance of multiple supervised machine learning algorithms will be evaluated: Support Vector Machines (SVM), k-nearest neighbors (KNN), Adaboost, Random Forest (RF), and Random Forest (RF). Section two presents a dimensional approach and introduces the stimulus dataset, section three presents a categorical approach for emotion representation, section four presents the sensors used and the associated emotions, section five presents the experimental design, section six details the face capturing post processing and supervised learning, presenting results obtained using several classifiers. Section seven presents the conclusions and future lines of research.

2 Dimensional approach - Image and videos database

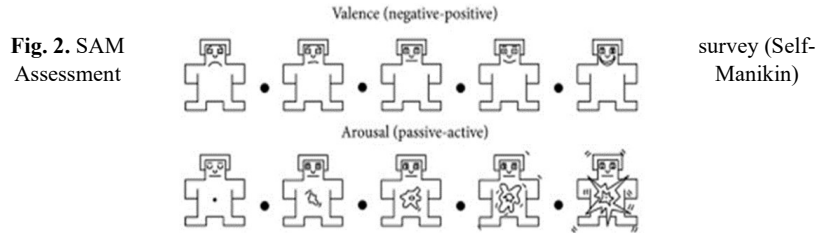
The dimensional approach implies that affective states are distributed in a continuous space which dimensional axes indicate the quantification of a feature [11]. One of the most widely accepted models is James Russell's circumplex model [12], also known as the Arousal-Valence model. This model is a two-dimensional one, with the axes being Excitement, or Arousal, (relaxed vs. excited) and Valence, or Valence, (pleasure vs. displeasure). In this model, emotions are located in a continuous space defined by two axes: arousal and valence (see Fig. 1). From now on, we will refer to this model as the Arousal-Valence model. The arousal axis measures the degree of activation (i.e., the level of excitement or relaxation), while the valence axis measures the pleasantness toward emotional experiences (from unpleasant to pleasant). Subsequently, a third axis known as dominance, which indicates the control a person has over an emotion, was also considered. For this work, the IAPS image set [13] was used, which is a collection of more than 1000 photographs depicting objects, people, landscapes, and situations of everyday human life. Each of these images has been evaluated by more than a hundred people -men and women- in the affective dimensions of valence (level of liking/dislike of the image), arousal or activation (level of activation/calmness provoked by the image) and dominance (level of control of the subject over the image), using a pictographic scale.



Fig. 1. A graphical representation of the circumplex model of affect with the horizontal axis representing the valence dimension and the vertical axis representing the arousal or activation dimension.

This classification by quadrant can be achieved thanks to the dataset that comes included with the IAPS which, among other data, provides us with the arousal and valence values linked to the images.

Therefore, from our approach, we selected N of the most representative images from each quadrant, to do this we rely on the arousal mean and valence mean values provided by the IAPS dataset itself. Another source stimuli used was the Database of Emotional Videos from Ottawa or DEVO [14] (Database of Emotional Videos from Ottawa). This collection of emotional video clips can be used in a similar way to the IAPS images. The Ottawa Emotional Video Database (DEVO) includes 291 short video clips drawn from unfamiliar sources to reduce familiarity and avoid influencing participants' emotional responses. The quadrant classification can be achieved thanks to the dataset included with the DEVO which, among other data, provides us with the arousal and valence values linked to the image. Video (DEVO) and image (IAPS) stimuli were selected, following a similar pattern, so that the stimuli shown per phase try to lead the subject to the same emotional quadrant. In this way, their biometric values can be measured on the same emotional set. This selection was made according to the arousal and valence values of the stimuli, so we can evaluate the received stimulus by comparing it with a Self-Assessment Manikin survey (see fig. 2. SAM) completed by the subject after receiving the stimulus.



3 Categorical approach - Microsoft Face Cognitive Services

As explained in [15] the categorical approach was initially developed by psychologist Paul Ekman, who claimed that there was a set of six basic and universal emotions that are not determined by cultures. This set is composed of joy, fear, sadness, anger, disgust, and surprise. For face emotion recognition we will use the "Face" service offered by Microsoft Azure Cognitive Services. [16] This service is capable of infer the following emotions: anger/anger, contempt, disgust, fear, happiness, sadness and surprise from a static image provided. The Face service can perform emotion detection on a facial expression. However, it worth noting that facial expressions alone do not necessarily represent the internal states of people. Therefore, it will be added to the conductance and heart rate measurements as another factor in the interpretation of emotional state. The interface to communicate with this service (Face) will have the following structure:

Example response in JSON format of a Microsoft "Face" API request for photo Fig. 3:

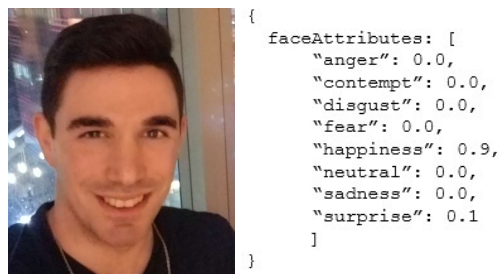


Fig. 3. Image sent to Face for emotional recognition and its associated response.

In which, each of the emotions will have a decimal value associated with the weighting of that emotion according to the image provided. Being the emotion with the highest value the dominant one.

4 Sensors and associated emotions

In this study, two types of biometric parameters will be taken into account. On the one hand, parameters associated with heart rate and on the other hand, skin conductance will be considered. The information concerning heart rate makes it possible to evaluate the changes that take place between cardiac cycles. Heart rate variability (HRV) measures how close or spaced one heartbeat is to the other. A large HRV implies that the beats are closely spaced, which is interpreted as a low heart rate or low arousal (activation of the parasympathetic nervous system). In contrary, a low HRV implies that the beats are close together, which is interpreted as a high heart rate or high arousal (sympathetic nervous system activation). During episodes of stress or emergency situations, the sympathetic system is activated resulting in fight or escape responses, including increased heart rate, causing heart rate variability. [17]. On the other hand, skin conductance depends on the activity of the skin sweat glands and reacts to the slightest, almost imperceptible changes in hand sweating. The stronger the activity of the sweat glands, the moister the skin becomes and the better the current is conducted. As a result, capillary conductance increases. The conductance is measured in microsiemens. The activity of the sweat glands in the skin is determined by the vegetative nervous system, which consists, in part, of the sympathetic and parasympathetic systems. The skin sweat glands are activated by the sympathetic, so it is a good indicator of internal tension. The sympathetic nervous system is activated after exposure to stress, mental activity, emotional arousal or a fright and prepares the body to act in borderline situations increasing conductance, pulse, blood pressure and blood glucose level to have an instant source of energy and a boost in attention. [18]

5 Experiments design and results

The objective of this work is to predict the emotional state of a subject from a set of parameters obtained from various sources, including heart rate, galvanic response, and face capturing. To achieve this goal, we rely on a multimodal framework [19], a series of tests were designed in order to generate the necessary data to build a model as accurate as possible. These tests consist of inducing in the subject the different emotional states associated with each of the four quadrants of the arousal-valence plane, using Image (IAPS) and Video (DEVO) stimuli. While the subject is exposed to several stimuli, the data obtained from the various sources are persisted through the framework for further analysis and processing. Finally, a percentage of the data obtained will be used as test data for the system. This data set will be used to induce the emotional state and will not be used as training data. Six test subjects participated in the experimentation with an extension between six and nine minutes each, initially totaling nine sessions, with an average of 495 records per test. As can be seen in Fig. 4a, these stimuli are presented to the test subject in a planned and organized manner, simultaneously recording data from the heart rate sensor (eSense Pulse) [20], and from the sensor conductance sensor (eSense Skin) [21], in addition to the constant capture of the face image, to then send a set of particular images to the Cognitive Services

service of Microsoft (Face), also capturing the image of the screen with which the test subject interacts, the latter has the stimulus that the person is perceiving at that moment. These data will be dumped into a centralized database in order to later be able to exploit and analyze them. The test is made up of a total of five steps from the time the test subject gets ready with all the sensors, up to the session results presentation in a set of dynamic graphs. These tests consist of:

First step – Connectivity: This consists of connecting the sensors to the test subject and verifying the connection of the sensors. In addition, a general introduction to the test is presented.

Step Two – Initial SAM: In order to observe and consider the test subject's current emotional state, prior to the activity, the test subject fills a SAM (neutral without stimuli) survey where he/she can indicate on a scale of 1 to 9 his/her state of arousal and valence.

Step three - Stimulus(s) (per quadrant): A set of ten IAPS images and 5 DEVO videos were selected for each of the four quadrants of the arousal/valence plane. In the selection of images and videos from each set, the density of the arousal-valence values associated with them was taken into account, thus forming each set of images and videos with the most recurrent excitation-valence values for each one. The person is subjected to the image's projection for a period of 3 seconds per image and approximately 5 seconds per video.

Step four - SAM (per quadrant): At the end of each projection, the subject will be asked to answer a SAM survey. The objective of this phase is to verify that the intended emotional state has been successfully inferred. Fig 4b represents graphically what was explained during phase 3 and 4.

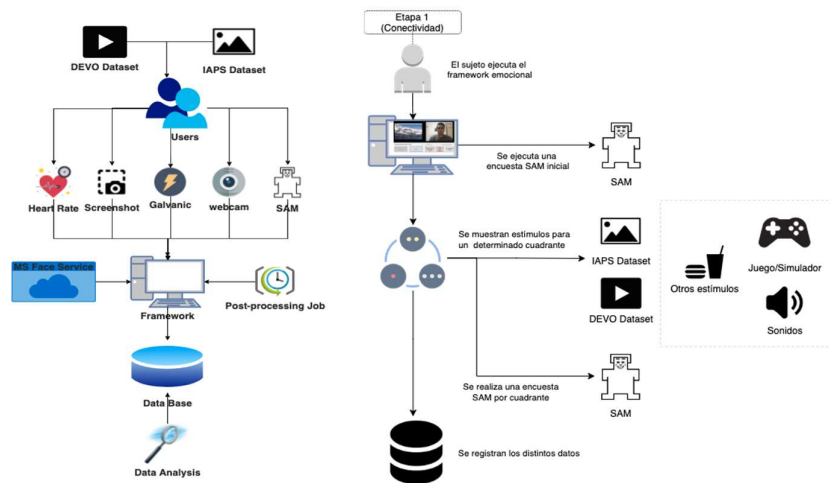


Fig. 4a. Conceptual system model. **Fig.4b** Test process.

Step 5 - Data Consolidation and Synchronization: As seen in Fig. 5, once the test is completed, the data in the open text format, comma separated values (CSV) from both

sensors are taken and placed in a predetermined location within the project path. To then extract the data from each file and persist them in the unified database in order to have a single, integrated data source with the values of the different sensors. Each set of data obtained from the sensors (heart rate and conductance) comes with a set of standardized data, both of which contain the **TIMESTAMP** value, that indicates the time when the pulse or conductance measurement was obtained. In addition, each stimulus, SAM survey and other events are recorded with a time stamp from the computer's system schedule, so that all data (both event and sensor data) can be dumped into the same timeline and related.

Once finished, the data post-processing phase begins. As seen in Figure 5, this stage can be divided in two: generation of the dataset and its exploitation. the generation of the dataset includes the processing of face captures and the generation of dataset for data mining.

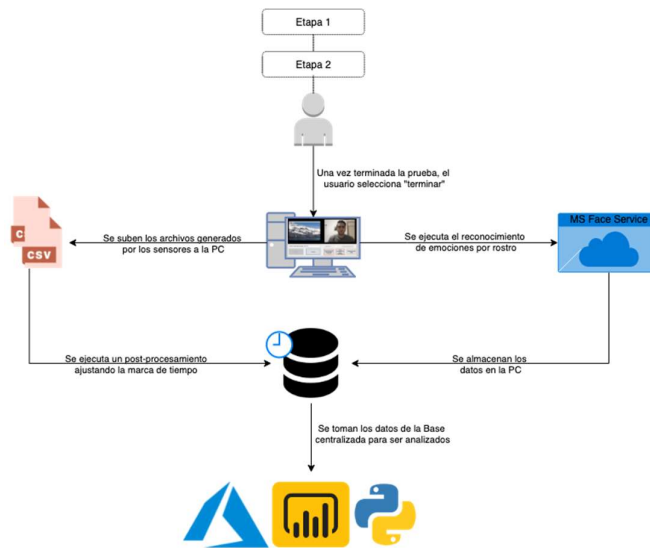


Fig. 5. Graph representing the last phases of the test.

Once the subject's test is finished, we will try to obtain the face's emotional features obtained during the test. The captures are sent to Microsoft's "Face" service. This service uses a categorical emotional model for classification, thus obtaining "discrete" emotions. In order to analyze the emotions of the face, the categorical values provided by the service must be converted to dimensional values. To perform this conversion, the study carried by FaceReader [22] was taken as a reference, in which, from the analysis of a face image, a valence value is inferred by subtracting the predominant negative emotion from the measured value of happiness, resulting in a value belonging to the interval [-1; 1]. The FaceReader calculation does not provide arousal values for

static figures (images), therefore, only the valence value will be considered and associated to the arousal value provided by the supervised learning model based on heart rate and skin conductance values.

Unlike FaceReader, where each of the emotions has independent values, i.e., the sum of them can exceed the value of 100%, the "Face" service implemented in the current work provides percentage values for each of the emotions, making the sum of all of them 100%. As the service does not provide independent values for each negative emotion, each of them must be considered, replacing the predominant negative emotion (value used by FaceReader) by the sum of all negative emotions obtained.

$$\text{Joy} - \Sigma (\text{negative emotions}) = \text{valence} \quad (1)$$

Considering as "negative emotions": sadness, anger, fear and disgust.

Once the sensor data is stored in the database, a process begins immediately to generate a new CSV text file that links the heart rate sensor, skin conductance sensor and recorded events (face and stimulus image capture) records via the timeline, this file will then be used for the supervised learning functionality. Such file will consist of a table with the following columns; TimeStamp, HR (heart rate), RR (time between beats), HRV (heart rate variability), MicroSiemens (conductance), SCR (conductance responses), SCR_MIN (conductance responses per minute), ArousalMean (average arousal), ValenceMean (average valence), ArousalSD (excitation standard deviation), ValenceSD (valence standard deviation), PhaseName (phase name), MatchesSam (Value indicating whether the SAM survey matches the presenting quadrant).

The arousal/valence value, from that table, will be taken from the average value published in the dataset corresponding to the stimulus ID that was displayed at that time. Once this process is completed, before the data is exploited by the supervised learning process, we checked if the SAM responses provided by the test subject are consistent (quadrant relevance) with the values provided by the stimuli, for example if the resulting quadrant was HA_PV (High Arousal - Positive Valence) it corresponds to the values completed in the SAM survey. If it does not match the expected one, the MatchesSam column is marked with the value false, so that it is not considered in the supervised learning process. Once this process is finished, these data will be analyzed by the supervised learning model, which is detailed below. In the supervised learning process we applied different algorithms KNN, Random Forest, SVM (RBF) , SVM (POLY) were applied in order to classify the test subject's arousal level by using supervised learning in order to obtain the arousal value of the subject during the test, more specifically, if it had a high or low arousal. The high arousal value will be assigned the value 1 and the low arousal the value 0, we consider a high arousal if it is higher than 5, on a scale of values ranging from 1 to 9 (9 being the maximum arousal). The supervised learning process, which takes place after the CSV of the previous point has been generated, consists of three well marked.

Step 1 - Construction of the training dataset: the first, in which the CSV data is subjected to a standardization process, which rescales the data (all measurements are now in the same interval and with the same standard deviation) in order to better detect

variations and help the classifier algorithm not to lose accuracy due to the diversity of the numerical values of each record of the generated dataset.

The "Aroused" column is added to the already known columns, whose purpose is to identify for each given measurement whether the subject was aroused (1) or not (0), thus facilitating the training of the classifier model. The value (1) is assigned to a measurement if the average arousal value belonging to the stimulus that was being presented at that moment is greater than or equal to 5 and the value (0) if it is less. In order for the model to consider the variations of values as temporal slopes in its training, the values of the HR, HRV and Microsiemens measurements of the four measurements immediately preceding in time (t-1, t-2, t-3 and t-4) are added to each of the records in Table 3. Giving the training dataset the final format represented in Table 1. Finally, this dataset will be persisted in a CSV file "1_standardized_biometrics.csv".

HR	RR	HRV	MicroSiemens	SCR	SCR_MIN	PhaseName	
-0.6879460933785926	0.6603042394011167	0.24219508152680996	-0.12182417440038537	0	0	HA_PV	
-0.07971408073825882	0.015297318165465725	0.2504763787513362	-0.12182417440038537	0	0	HA_PV	
ArousalMean	Aroused	TimeStamp	HR-4	HR-3	HR-2	HR-1	
7.27	1	2020-10-25 19:04:03	-0.905171812178712	-0.5358880902185115	-0.8400040965386768	-0.7313912371386171	
7.27	1	2020-10-25 19:04:04	-0.5358880902185115	-0.8400040965386768	-0.7313912371386171	-0.6879460933785926	
HRV-4	HRV-3	HRV-2	HRV-1	MicroSiemens-4	MicroSiemens-3	MicroSiemens-2	MicroSiemens-1
0.08485043428081729	0.15110081205702486	0.19250729817965442	0.22287205466958296	0.12475989667508678	0.12475989667508678	-0.12182417440038537	-0.12182417440038537
0.15110081205702486	0.19250729817965442	0.22287205466958296	0.24219508152680996	0.12475989667508678	-0.12182417440038537	-0.12182417440038537	-0.12182417440038537

Table 1. standardized data consolidation

Step 2 - Model creation: the second step consists of iterating the dataset obtained in the previous step through all the classification algorithms that apply to the use case of the dataset and have a model created for each selected classifier, this process is carried out in the file. Each model will give us its respective performance values at the time of being created, in table 4 we can see the summation of these metrics for each classifier.

Looking at Table 2, the models with the best Accuracy / F1-Score / Cross Validation Avg Accuracy ratio are: KNN Default, Random Forest - Grid Search and Adaboost. These models score an average Accuracy in cross validation higher than 65%. KNN and Random Forest had an average F1-Score higher than 50%, which is acceptable, but indicates that the model still has opportunities for improvement in terms of false positive and false negative detection. In the case of Adaboost, the scores were about a 45%, which means that this algorithm was discarded from the evaluation, Random Forest produced better results with the test data and KNN for train data. However, KNN proved to be more efficient when making predictions with test datasets, which is why it was the model selected to continue the research.

			Accuracy	F1-Score		CV (Avg)
				0	1	Accuracy
KNN	Default	Train	0,87	0,85	0,89	0,61
		Test	0,57	0,47	0,64	0,66
	Grid-Search	Train	0,83	0,79	0,85	0,61
		Test	0,56	0,46	0,63	0,67
Random Forest	Default (overfitted)	Train	1,00	1	1	0,59
		Test	0,56	0,44	0,63	0,65
	Grid-Search	Train	0,70	0,64	0,75	0,60
		Test	0,64	0,53	0,71	0,65
SVM (RBF)	Default	Train	0,57	0	0,73	0,54
		Test	0,61	0	0,76	0,66
	Grid-Search (overfitted)	Train	0,99	0,99	1	0,56
		Test	0,55	0,43	0,63	0,62
SVM (Poly)	Default	Train	0,62	0,3	0,74	0,55
		Test	0,58	0,09	0,73	0,66
	Grid-Search	Train	0,82	0,8	0,8	0,54
		Test	0,47	0,43	0,63	0,57
Adaboost	Default	Train	0,76	0,71	0,79	0,57
		Test	0,58	0,45	0,45	0,69

Table 2. Main metrics summary.

Step 3 - Prediction: In the last step of supervised learning, we proceed to perform the prediction of the arousal level of a subject. We take the selected model (in our case, KNN) and feed it with the data of the test session on which we want to perform the prediction. As output of this procedure, a CSV file is obtained, as shown in Table 5, which contains each measurement belonging to the test session associated to the expected (Arousal column) and predicted (Arousal predicted column) arousal level.

Once the test and the supervised learning process have been completed, it will be possible to visualize all the data that composed the test. A graph summarizing the prediction results is presented. Fig. 6 shows how the prediction output can be easily assigned to a timeline. In the first graph (top) we can see the training values, i.e., what the system is expected to predict, in the second (middle) the values predicted by the system and finally, in the third graph (bottom) we can see both graphs superposed

(predicted vs. expected). The “y” variable of this graph refers to the predicted arousal level (0 / 1) and the “x” variable to each time instant associated to the measurements performed in the test session. Thanks to the superposition we can easily identify when the prediction matches with the expected value.

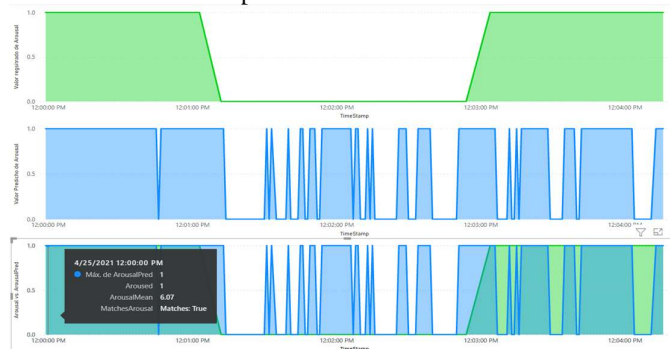


Fig. 6. Prediction of subject arousal in a test session (accuracy=75%).

Once the prediction of arousal through the supervised learning model and of valence from the Face service with selected images (according to their emotional state variation) has been performed, it will be possible to compose a visualization that presents the data of both predictions associated to the same timeline (Fig. 7), thus providing, for some time intervals, a complete prediction of the emotional profile applicable to the dimensional approach (arousal-valence plane).

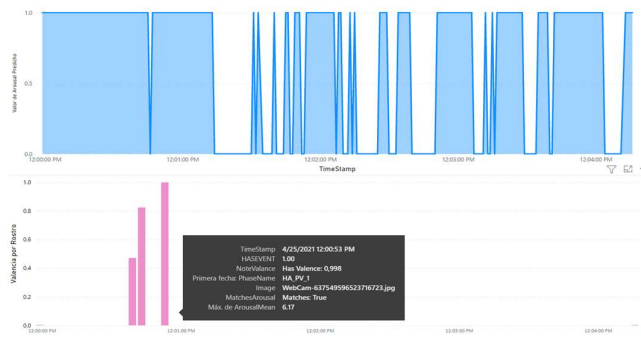


Fig. 7. Excitation Prediction (by classifier) and Valencia Prediction (by API Face)

Of the two graphs shown in Fig. 7, the first (top) shows the arousal value predicted by the algorithm (0 / 1); this is the same graph that can be seen in Fig. 6. The second graph (bottom) refers to the valence predicted by the Microsoft Face service based on the captures of the test subject's face. This valence level is represented in the interval [-1;1]. Each bar is associated with a photograph taken at the time instant associated with the X-axis at the point where the center of the bar is located, and its height indicates the predicted valence value belonging to the face capture taken at that instant. One more

point to take into consideration is that although the face captures are taken continuously at one-second intervals, only some of them are sent to the emotional prediction service when the subject shows a non-neutral expression on his face.

Fig. 8 shows, in the form of diamonds, the excitation and valence values of the stimuli presented at the times of the predictions. That is, what the system is expected to predict. As a circle, the three complete predictions (i.e., both excitation and valence values) given by the system. In the form of a square (colored orange) the SAM survey corresponding to the stimulus phase and at the bottom a table with all the traceability of the data obtained. This table shows for each prediction: the time stamp, the Stimulus Phase that was being presented at that moment, the meta-data of the stimulus presented (id and associated values), the predicted values and the prediction values are re-scaled to a scale of 1 to 9, the physiological values measured at that instant, the face capture taken at that instant and a value indicating whether the prediction was correct or not.

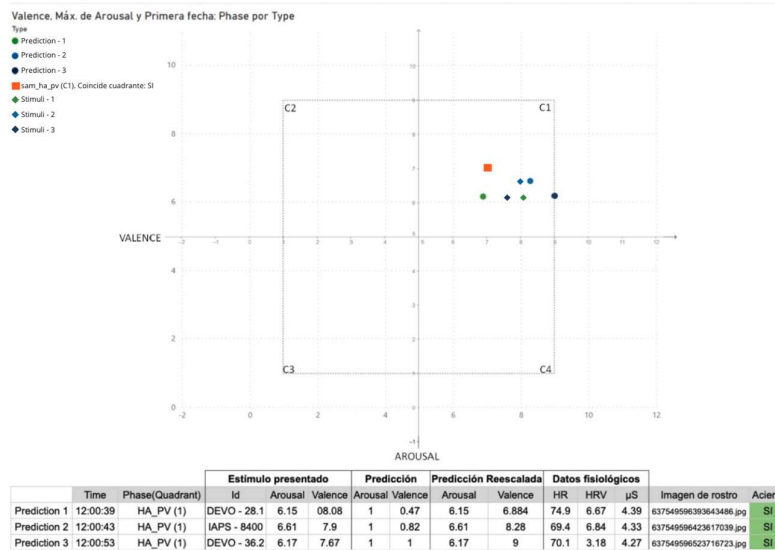


Fig. 8. Excitation and predicted valence in dimensional model.

As seen in the graph there are three pairs of diamonds (stimuli) and circles grouped by color (predictions), in the three cases both are located in the same quadrant, so we can say that the prediction of the emotional quadrant is consistent with the expected values (based on the stimuli presented), as well as the answer to the SAM survey of the test subject (represented by a square).

6 Conclusions and future lines of work

For this research, an emotional recognition framework [19] has been used as a basis, features have been added to provide greater multimodal recognition capabilities. By adding sensors that allow obtaining physiological information from people, it has been possible to include in the framework, information regarding heart rate, skin conductance and emotion recognition through facial expressions. Through multiple experiences with different participants, a basal data set was obtained in order to train the supervised machine learning model, with the ability to infer the state and arousal value of the test subject. As first approach to supervised learning, classifiers generated by algorithms such as adaboost, KNN, random-forest and SVM have been trained, obtaining results with accuracies between 65% and 80% regarding a binary classification of the arousal level of a person, being KNN the classifier with the best results. On the other hand, valence values were added at arbitrary time instants, measured from captures of the participants' faces through Microsoft's "Face" service, complementing the prediction and obtaining a result increasingly aligned to Russel's circumflex model, being able to represent both dimensions of the arousal/valence plane proposed in that model. Regarding future lines of work, it is contemplated to explore other algorithms for the supervised learning engine, with a larger number of samples in order to be able to use a Stochastic Gradient Descent (SGD) classifier, which specializes in datasets larger than one hundred thousand samples. Improve the selection process of face images based on the emotional variation, inferred by a proprietary classifier that uses logistic regression [23]. This will allow to have a continuous valence value, to associate each prediction of arousal to a valence level, achieving a completely two-dimensional result, where each time instant has a continuous representable value in the arousal/valence plane. Future lines of research are oriented to the multimodal integration with classifier models of other Excitation and Valence data associated to EEG, and Voice.

References

1. Baldasari, S. (2016). Computación afectiva: tecnología y emociones para mejorar la experiencia del usuario. Revista: Bit & Byte; año 2, no. 3, ISSN: 2468-9564 .Obtenido de <http://hdl.handle.net/10915/53441>
2. Picard, R. W. (1997). Affective Computing, Second Edition 1998 ed., Vol. 321.
3. H. Gunes, B. Schuller, M. Pantic et al., (2011) "Emotion representation, analysis and synthesis in continuous space: A survey", en Face and Gesture 2011, IEEE, 2011, págs. 827- 834, ISBN 978-1-4244-9140-7
4. X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network," in Proc. IEEE Int. Conf. Bioinform. Biomed., Shenzhen,China, Dec. 2016, pp.352-359
5. Y. Kwon, S. Shin, and S. Kim, "Electroencephalography Based Fusion Two-Dimensional (2D)-Convolution Neural Networks (CNN) Model for Emotion Recognition System," Sensors, vol. 18, no. 5, pp. 1383-1395,2018.

- 6 Howell, N., Devendorf, L., Tian, R. (Kevin), Vega Galvez, T., Gong, N.-W., Poupyrev, I., ... Ryokai, K. (2016). Biosignals as Social Cues. Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16. doi:10.1145/2901790.2901850
7. Fusar-Poli, P., Landi, P., & O'Connor, C. (2009). Neurophysiological response to emotional faces with increasing intensity of fear: A skin conductance response study. *Journal of Clinical Neuroscience*, 16(7), 981–982. doi:10.1016/j.jocn.2008.09.022
8. Azarbarzin, A., Ostrowski, M., Hanly, P., & Younes, M. (2014). Relationship between Arousal Intensity and Heart Rate Response to Arousal. *Sleep*, 37(4), 645–653. doi:10.5665/sleep.356
9. Ménard, Mickaël & Richard, Paul & Hamdi, Hamza & Daucé, Bruno & Yamaguchi, Takehiko. (2015). Emotion recognition based on heart rate and skin conductance. *PhyCS 2015 - 2nd International Conference on Physiological Computing Systems, Proceedings*. 26-32
10. Hariharan, A., & Adam, M. T. P. (2015). Blended Emotion Detection for Decision Support. *IEEE Transactions on Human-Machine Systems*, 45(4), 510–517. doi:10.1109/thms.2015.2418231.
11. H. Gunes, B. Schuller, M. Pantic et al., (2011) “Emotion representation, analysis and synthesis in continuous space: A survey”, en *Face and Gesture 2011*, IEEE, 2011, págs. 827-834, ISBN 978-1-4244-9140-7
12. J. Posner, J.A. Russell y B.S. Peterson, (2005) “The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology.”, *Development and psychopathology*, vol. 17, n° 3, págs. 715-34, ISSN: 0954-5794.
13. Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. In Technical Report A-8. Gainesville, FL: University of Florida
14. Ack Baraly, K. T., et al. (2020). Database of Emotional Videos from Ottawa (DEVO). *Collabra: Psychology*, 6(1): 10. DOI: <https://doi.org/10.1525/collabra.180>
15. P. Ekman, “Basic Emotions”, *Handbook of Cognition and Emotion*, T. Dalgleish y M.J. Power, eds., Chichester, UK: John Wiley & Sons, Ltd, 2005, cap. 3, págs. 45-60, ISBN: 978-0-47197-836-7
16. <https://azure.microsoft.com/es-mx/services/cognitive-services/face/#features> mayo 2021
17. J. A. Gallo Villegas, J. Farbiarz y D. L. Alvarez Montoya, «Análisis espectral de la variabilidad de la frecuencia cardíaca», *Iatreia*, vol. 12, n.º 2, págs. 61-71, 1999, issn: 2011-7965.
18. P Fusar-Poli, P Landi, C O'Connor - Neurophysiological response to emotional faces with increasing intensity of fear: A skin conductance response study, *Journal of Clinical Neuroscience*, Volume 16, Issue 7, July 2009, Pages 981-982 2009
19. Ierache Jorge, Sattolo Iris, Chapperón Gabriela, Framework multimodal emocional en el contexto de ambientes dinámicos, *RISTI*, N.º 40, 12/2020, DOI: 10.17013/risti.40.45–59
20. Mindfield® eSense Pulse | Biofeedback. <https://mindfield-shop.com/produkt/esense-pulse>
21. Mindfield® eSense Skin Response | Biofeedback. <https://www.mindfield.de/en/Biofeedback/Products/Mindfield%C2%AE-eSense-Skin-Response.html>
22. Loijens, L., & Krips, O. (2019). FaceReader methodology note. <https://info.noldus.com/free-white-paper-on-facereader-methodology>
23. Barrionuevo Carlos, Jorge Ierache, Iris Satollo, Emotion Recognition Through Facial Expressions Using Supervised Learning with Logistic Regression, 26th Argentine Congress, CACIC 2020, San Justo, Buenos Aires, Argentina, October 5–9, 2020, Revised Selected

Papers, Springer International Publishing Communications in Computer and Information Science , Vol 1409, pp 233-246, DOI 10.1007/978-3-030-75836-3.