IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 4, 2020, accepted June 18, 2020, date of publication July 3, 2020, date of current version July 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3007109

Affective Robot Story-Telling Human-Robot Interaction: Exploratory Real-Time Emotion Estimation Analysis Using Facial Expressions and Physiological Signals

MIKEL VAL-CALVO[1,2], JOSÉ RAMÓN ÁLVAREZ-SÁNCHEZ[1], JOSÉ MANUEL FERRÁNDEZ-VICENTE[2], AND EDUARDO FERNÁNDEZ[3,4]

[1]Departamento de Inteligencia Artificial, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain
[2]Departamento de Electrónica, Tecnología de Computadoras y Proyectos, Universidad Politécnica de Cartagena, 30202 Cartagena, Spain
[3]Instituto de Bioingeniería, Universidad Miguel Hernández, 03202 Elche, Spain
[4]CIBER-BBN, 28029 Madrid, Spain

Corresponding authors: Mikel Val-Calvo (mikel1982mail@gmail.com) and José Ramón Álvarez-Sánchez (jras@dia.uned.es)

This work was supported in part by the Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia under Project Programa de Ayudas a Grupos de Excelencia de la Región de Murcia, in part by the Ministerio de Ciencia, Innovación y Universidades under Project RTI2018-098969-B-100, and in part by the Generalitat Valenciana under Project PROMETEO/2019/119.

ABSTRACT Affective human-robot interaction is still an active area of research in part due to the great advances in artificial intelligence. Now, the design of autonomous devices that work in real therapeutic environments has become a plausible reality. Affective human-robot interaction requires a robot to analyze the emotional state of the human interlocutor and interpret emotional responses that can be used, not merely in the interaction but, for example, to provoke desired therapeutic responses. It is, therefore, necessary to broaden experimental techniques into more realistic paradigms, where the capacity of emotion estimation can be completely explored. This exploratory paper proposes a realistic experimental paradigm in which the robot employs a dramatic story to evoke emotions in the users, and tests previously self-designed methodologies to be able to make estimates of the users' emotional state in real-time. Regardless of the multiple impediments and restrictions, and all the aspects that could still be improved, this paper can outline the feasibility of the proposed methodology in realistic scenarios.

INDEX TERMS Affective state, blood volume pressure, EEG, emotion estimation, face emotion recognition, galvanic skin response, human-robot interaction, real-time.

I. INTRODUCTION

Affective HRI (Human-Robot Interaction) is one of the most challenging tasks the research community is facing, but recent technological advances allow for the development of new attempts. The main objective of affective HRI is to build intelligent systems that can adapt to the changing mood of users, in order to enhance communication in real-time [1]. To cope with the lack of emotional connection between humans and machines, emotion detection must meet some requirements such as being automatic, reliable and adaptable.

Several social groups could benefit from the development of affective HRI. This is the case for lonely elders, children with an autism spectrum disorder or people with limited capabilities of communication. For many of them, communicating emotions is a problem that could be solved by the help of affective computing. For instance, using wearable sensors for measuring their physiological responses with the addition of an analysis of their behavioral responses, such as facial expressions or body gestures, could improve the attention given to the users by having a closer insight of their feelings, and therefore, improve their quality of life and happiness. For the case of autism spectrum disorder or children with difficulties to express their emotions, affective HRI can be used to allow them to express emotions through story-telling strategies by remotely controlling a robot. As an example, the use of puppets to help children learn how to express emotions has been widely studied [2]–[4]. In that way, children could

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqing Zhang.

improve their expressiveness and therefore allow them to integrate better in society. Regarding elders, a robot that could understand their feelings could be an appropriate tool to mitigate their loneliness. Therefore, it would be desirable to develop a deep understanding of how robots can effectively influence users' emotions, both evoking and detecting them, in order to be able to adapt to interactions dynamically and effectively.

Emotion recognition is an interdisciplinary field that requires knowledge from different domains such as psychology, neuroscience, signal processing electronics, and artificial intelligence among others. It can be addressed with the use of different types of signals. On one hand, physiological signals such as electroencephalography (EEG), galvanic skin response (GSR), or heart-rate variations by measuring blood volume pressure (BVP) or by electrocardiogram, can be used. These are internal signals which reflect the balance between sympathetic and parasympathetic systems, as is the case for BVP and GSR, while EEG manifests changes in the cortical areas of the brain. On the other hand, there are externally observable cues such as facial expressions, body gestures, or speech. While the internal signals are considered to be more objective due to the intrinsic properties of several functional areas of the central nervous system, the external ones remain as subjective measures of the expressed emotions, which can be intentionally modulated or manifested as very subtle changes, such as for facial expressions [5]. Taking all of this into account, recent approaches tend to exploit multiple sources in parallel [6].

Emotional models are needed to give users a homogeneous reference system for self-assessment of emotions, both involved in the learning process and affective HRI. Historically, two main models have been developed which remain controversial: the discrete emotional model, which assumes that emotions are qualitatively differentiated neurophysiological responses [7] that produce independent emotional experiences; and the dimensional model, which assumes continuous quantified relationships among emotions [8]. For the present paper, the dimensional model has been chosen, but the assessment space is discrete, to simplify the number of labels for the users to choose from [9], [10].

Research in emotion recognition involves a series of tasks to be developed. It requires the definition of the set of signals to be chosen as sources of information. The correlation between signals must be studied to better understand the expression of emotions and, therefore, requires the development of an adequate selection of stimuli and, more generally, the causal model underlying the experimental design that allows the generation of emotions. Finally, as detection is one of the main objectives, feature extraction methods must be developed and tested according to a set of algorithms for statistical inference.

Regarding the selection of sources, EEG signals are considered a useful source, as they measure the brain responses, reflected on the cerebral cortex, during emotion processing [11], [12], both in perception and expression, and is sensible to valence in the dimensional emotional model. GSR and BVP signals, on the other hand, reflect the balance between sympathetic and parasympathetic systems of the autonomic nervous system. While GSR is mainly driven by the sympathetic subsystem, BVP reflects the balance of both subsystems. Both are sensitive to arousal [13] in the dimensional emotional model. Finally, changes in facial expressions can easily be measured using cameras and sometimes reflect spontaneous changes in users' emotions.

Regulation of emotions occurs as a result of close interactions between various subsystems of the central nervous system under behavioral demands to dynamically adapt to changes in the environment in order to produce complex behaviors. The interaction involves the autonomic system, which alters the balance of the sympathetic and parasympathetic systems which can be measured by BVP and GSR signals. It also affects the prefrontal cortex and temporal lobes which can be measured by EEG. Finally, facial expressions are directly modulated by the amygdala's innervations while also guided by high-level behavioral intentions [14], [15]. As a result of those synergic interactions across the central nervous system, respiratory and electrodermal activity in conjunction with electroencephalographic and facial expression measurements may thus provide the necessary information on emotion processing [16]–[21], Fig. 1.

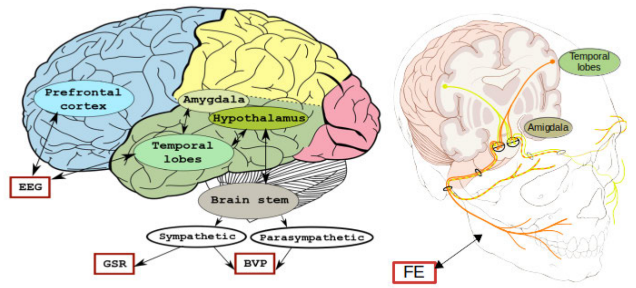


FIGURE 1. Brain areas. Graphical depiction of the inter-relations between brain areas involved in emotion processing. (partially modified [22]).

In recent years, deep learning has attracted interest in this field, as it has proven to have great results in the fields of computer vision and natural language processing, mainly due to the ability to learn high-level hierarchical representations [23]. As for the case at hand, several research studies have tried to attack the problem by the use of multi-modal sources which require the fusion of the information at hand. This fusion strategy can be done at two different levels. The first and easiest involves training a single model for every single source and finally perform a score-level fusion. The second and the hardest requires feature-level fusion in order to allow the models to take advantage of the intrinsic correlations among different sources but this is typically more difficult as their representations are not always directly compatible. So the problem becomes to find a proper representation of the set of sources to exploit the information presented. Moreover, such models must treat properly both temporal

FIGURE 2. **General electrophysiological and camera acquisition system for human-behavior recordings and processing in real-time. Middle-Left: long-term view of EEG raw data acquired with the OpenBCI system. Middle-to-bottom-Left: BVP, GSR and TMP signals acquired with the Empatica E4 device. Middle-Right: WebCam signal acquired using a self-customized driver. Bottom-Right: short-term view of EEG signals for the selected temporal window and frequency/spectrogram plot.**

and spatial representations in addition to the integration of different types of data streams [24].

Previous studies have been carried out in dramatic environments, such as comedy or theater performances, measuring the empathic responses of a group of volunteers, but the employed HRI systems still lack the capability of dynamically measure and adapt to users' emotional responses. The present paper involves a realistic scenario where a robot dynamically drives users' emotional responses by a story-telling affective HRI. The robot sequentially presents a series of stimuli, which are connected by a dramatic thread. A dramatic story was created in order to allow the robot to induce emotional changes on users, trying to compensate for the lack of a simple implementation of a convincing android facial expression. It is a matter of an existential story about the nature of the human being, as a guide for the robotic existence, to induce the users to reflect emotionally. The robot's story strategy covers fundamental existence dilemmas such as love, nature-human relation, and war, among others. The aim of this approach involves three main questions:

- Whether the effect of such an experimental emotional driving paradigm can be measured over users' physiological responses, in a population-based exploratory data analysis.
- At what extent each users' emotional estimation can be performed, based on the evoked properties of physiological signals or facial expressions.
- Assess whether the affective HRI has produced an emotional engagement, based on the subjective experience of the users.

II. MATERIALS AND METHODS

The present paper aims to answer a set of questions. First, to analyze the effect of affective HRI on the users' physiological responses by collecting data from physiological signals. Second, to explore the plausibility of such an approximation for the case of a real-time emotion estimation methodology in terms of accuracy reports. Finally, to evaluate the subjective experience of users regarding their emotional engagement towards the affective HRI.

A. ACQUISITION SOFTWARE: GePHYCAM

To record and collect the data, a self-produced software, GePHYCAM [25], is developed. This application looks forward to being accessible to the whole scientific community, providing a resourceful tool for human-behavior experimental paradigms, covering the following functionalities (see Fig. 2 and Fig. 3):

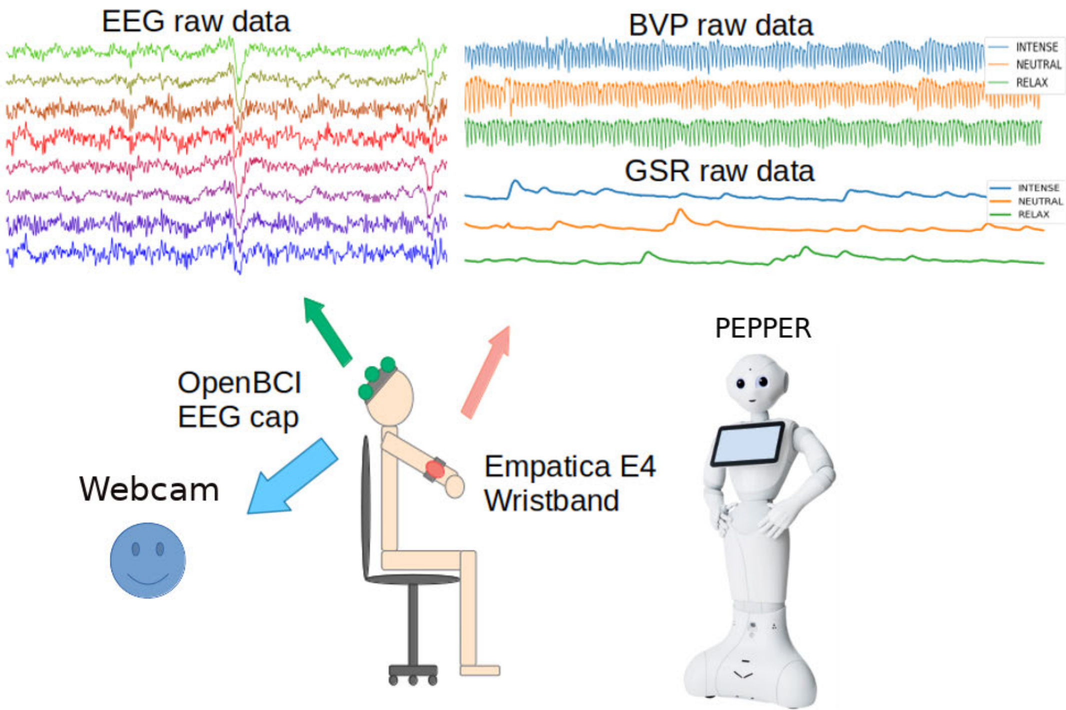


FIGURE 3. Experimental design picture. Top-Left: EEG raw data acquired with the OpenBCI system. Top-Right: GSR and BVP signals acquired with the Empatica E4 device.

1) Real-time acquisition and visualization of EEG, BVP, GSR, TMP and WEBCAM signals.
2) Trigger synchronization by a TCP/IP interface which allows start/stop recordings remotely.
3) Data recording on EDF (European Data Format) files for electrophysiological signals and MP4 file format for the audio-visual signals.
4) Online behavior labeling interface in which labels are synchronized and stored on EDF files.

B. DATABASE

A total of 16 volunteers (5 male, 11 female) aged between 19 and 40, participated in the present study. Participants were required to rate each scene of the dramatic story using the Self-Assessment Manikin (SAM) on two discrete 3-point scales, {NEGATIVE, NEUTRAL, POSITIVE} for valence and {RELAXED, NEUTRAL, INTENSE} for arousal. During each of the key scenes of the experiment, a set of iphysiological (EEG, BVP, GSR) and facial expression measurements were performed with the use of the Empatica E4 wristband, an OpenBCI system, and a standard webcam. For the OpenBCI cap, four prefrontal and four temporal electrodes, {F3, T7, P7, F7, F4, T8, P8, F8}, were used as they proved to be the best areas for emotion estimation [9], [10], [26], [27]. The Empatica E4 wristband was placed on the non-dominant hand to avoid artifacts when users perform self-assessment ratings.

Twelve different scenes, connected by a dramatic thread, were created from audio-visual resources such as documentaries and films, which were edited to accomplish a series of requirements. Each scene must be longer than one minute, to allow proper heart rate measurements, and each scene must drive a constant emotion. The duration (in seconds) of the scenes and the content are further explained in Table 1. After each scene users must perform self-assessment based on the two discrete valence and arousal dimensions and, also, are required to express their current emotions. The experiment is approximately 60 minutes long as it depends on the time spent by each user to explain their emotional responses after each scene.

TABLE 1. Story scenes specification. Time column is duration in seconds.

ID	Time	Description
Scene 1	164	Love and birth of a baby
Scene 2	73	Love between transsexuals
Scene 3	176	Happy video-clip by Pharrel Williams
Scene 4	60	Repressive law on homosexuality in the US 70s
Scene 5	98	Consequences of toxic discharges into water in Minamata, Japan, in the 70s
Scene 6	150	Animal mistreating
Scene 7	103	O'Barry talks about flippers emotional intelligence
Scene 8	130	Flipper killed himself at the hands of his caretaker
Scene 9	119	Interdependence between nature
Scene 10	179	The end of the Rapanui on Easter island
Scene 11	153	Nuclear bombs and war related apocalypse
Scene 12	177	Measures taken by humanity to overcome the problems of climate change and poverty

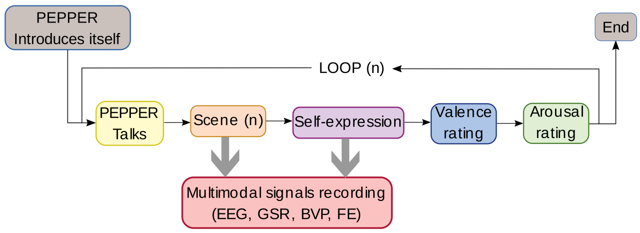


FIGURE 4. **Experimental design diagram.**

The experiment is conducted entirely by a Pepper robot, developed by SoftBank Robotics, after a prior preparation of the volunteer with the acquisition hardware (see Fig. 4):

1) Pepper introduces itself to the volunteer and after a short talk, asks the volunteer to rest with their eyes closed, for one minute.
   a) An explanatory message is shown in the chest tablet. The volunteer has to activate that interaction by pressing on the screen of the tablet. During this interaction, physiological signals are acquired, EEG, BVP, and GSR, using the OpenBCI and the Empatica E4 wristband.
   b) For the next interaction, the robot asks volunteers if they know about Plato's allegory of the cave. A message is shown on the interactive screen to allow them to specify "yes" or "no". Regardless of the response of each volunteer, Pepper explains Plato's allegory and after that, it asks them if they have had any experience where they have felt misunderstood, relating to the protagonist of the myth. An interactive screen is shown to allow the volunteer to activate the recording and tell Pepper of a similar experience of their own. The volunteer must click on the screen when finished.
   c) After that two consecutive interactive screens are shown, each of them allows the volunteer to perform the quantitative self-assessment using the SAM mannequins. This whole process is performed in order to allow the volunteers to learn the interactive process with the robot.
2) From that point until the end of the story, the robot acts by iteratively telling the story. First, develop the drama. Second, show a scene in the tablet while physiological signals are acquired. Third, the volunteer explains the evoked emotions while the robot records the volunteers with his front camera. Fourth, self-assessment on the valence-arousal discrete dimensions.
3) Finally, Pepper asks volunteers to tell their thoughts, both positive and negative, about life.

III. DATA ANALYSIS

Each of the following sections addresses the methodology applied when processing the aforementioned physiological signals and facial expressions.

A. PHYSIOLOGICAL SIGNALS PREPROCESSING

1) BVP PEAK DETECTION PREPROCESSING

The E4 wristband uses a photoplethysmogram sensor which allows BVP signal measurements. Processing steps involve a series of stages to obtain noise-free inter-beat intervals to properly code the signal properties. First, the moving average is computed over the raw data, where regions of interest are selected as the amplitude of the signal is larger than the moving average. R-peaks are marked at the maximum of each region of interest, which allows the computation of the interbeat intervals (time interval between two successive R-peaks of heartbeats) time series. Finally, detection and rejection of outliers are performed.

2) GSR SIGNAL PREPROCESSING

The E4 wristband captures the conductance, in microsiemens (μS), of the skin by measuring the potential difference between two electrodes while a tiny amount of current is applied between them. Due to the low sampling rate, 4Hz, of the E4 wristband, only tonic components were analyzed. The tonic component, called the skin conductance level, was obtained using a Savitzky-Golay filter [28] (window length=31, order=2).

3) EEG PREPROCESSING

EEG signals are arranged in a three-dimensional matrix containing n trials, c channels, and s samples at a sampling rate of 250 Hz. First, given that each signal has its own scaling factor values, signals are standardized using a z-score method. Second, a filter bank, based on sixth-order Butterworth filters, is applied for all n, c, and s, within a set of 5 non-overlapping bandwidths: 1-4 Hz, 4-8 Hz, 8-16 Hz, 16-30 Hz, and 30-50 Hz. An EEG oriented artifact removal technique (EAWICA) was used in this methodology. It was analyzed and validated with EEG brain patterns by Val-Calvo *et al.* [27] under real-time conditions.

B. FEATURE EXTRACTION

To answer the first two questions which address this paper, two different analyses were performed. First, population-based exploratory data analysis is carried out to analyze the statistical correlation between experienced emotions and the properties of the set of features computed for the EEG, BVP and GSR signals. Second, subject dependent classification is performed to check the feasibility of the emotion recognition methodologies, proposed for the experimental paradigm in question.

For population-based exploratory data analysis, the set of features was computed taking into account the full-time series corresponding to each scene for each signal type, and then z-scored relative to the baseline measurements. On the other hand, subject dependent classification consists in splitting the signals corresponding to each scene in sliding windows to compute a set of features. Thus, three independent classification processes were done to test the feasibility of affective

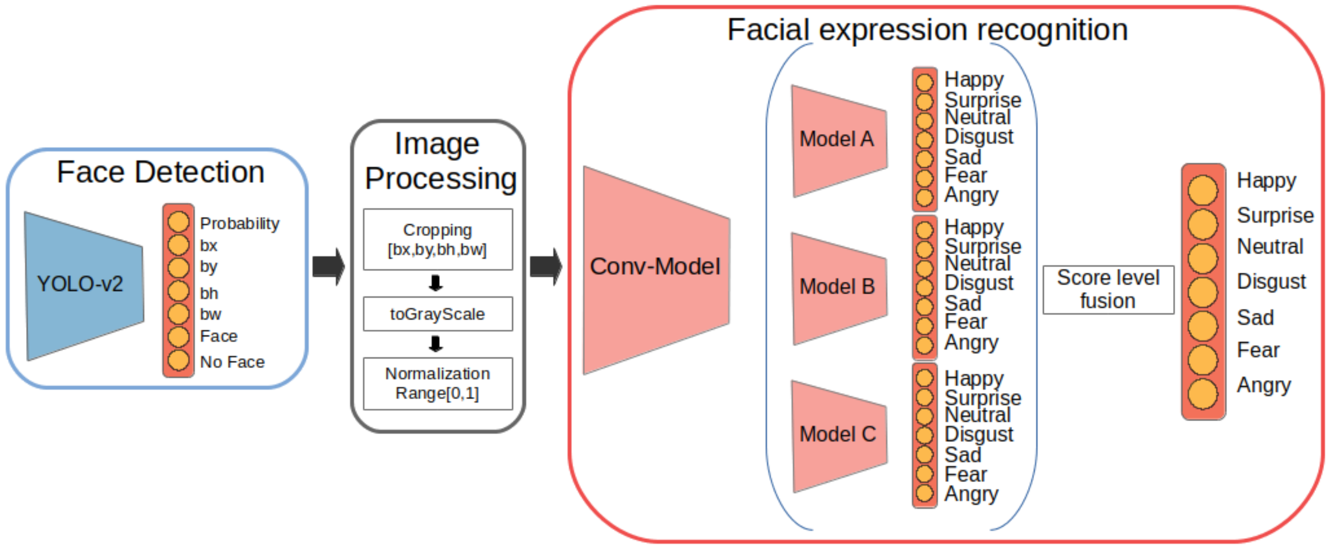


FIGURE 5. Facial expression estimation. First stage, face detection by convolutional deep learning model and preprocessing of detected face. Second stage, features extraction and classification by an ensemble of convolutional deep learning models.

state estimation. Therefore, the valence emotional dimension can be estimated by the use of EEG signals and the arousal by the use of GSR and BVP signals.

The following set of EEG features were computed based on the oscillatory properties of brain signals:

- Differential Entropy: Computed as a metric for measuring the predictability of a signal, whose values have a probability density function similar to a Gaussian distribution, N(μ, σ^2), as is the case for EEG signals. It can be defined for a signal X as $h(X) = \frac{1}{2}\log\left(2\pi e\sigma^2\right)$.
- Amplitude Envelope [29]: Computed through the Hilbert transform with the Neuro Digital Signal Processing Toolbox [30] python library developed at Voytek's Lab.
- Petrosian Fractal Dimension [31]: Defined as $PFD = \log(N)/(\log(N) + \log(N/(N + 0.4N_\delta)))$, where N is the series length, and N_δ is the number of sign changes in the signal derivative.
- Higuchi Fractal Dimension [32]: Higuchi's algorithm can be used to quantify the complexity and self-similarity of a signal.
- Fisher Information [33]: Fisher information is a way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ of a distribution that models X.

The last three EEG features mentioned have been computed with the PyEEG python library [34].

The set of GSR features computed based on the properties of skin conductance level time series were:

- Average of the series of amplitude values (offset).
- Average slope of the series of amplitude values.
- Standard deviation of the series of amplitude values.

The set of BVP features computed based on the properties of interbeat intervals (IBI) time series were:

- Average heart rate, computed as the inverse of inter-beat intervals.
- Standard Deviation of a IBI interval series.
- Root Mean Square of the successive differences of IBI.
- Standard deviation of IBI differences.
- Number of IBI differences greater than 20 milliseconds (NN20).
- Ratio between NN20 and the total number of IBI intervals.
- Number of NN interval differences greater than 50 milliseconds (NN50).
- Ratio between NN50 and the total number of IBI intervals.
- Triangular index: The ratio between the total number of IBI and the maximum of the IBI histogram distribution.
- Low Frequency: The power density estimation for the frequency band in the range [0.04, 0.15] Hz.
- High Frequency: The power density estimation for the frequency band in the range [0.15, 0.40] Hz.
- Sample Entropy: Used for assessing the complexity of the IBI interval series.

i Heart rate variability measurements and features were computed with the `pyHRV` python library [35].

C. FACIAL EXPRESSION RECOGNITION

Facial expression estimation is achieved by a combination of steps in two stages (see Fig. 5). In the first stage, facial detection is performed to simplify emotion estimation inference. This is achieved with the use of a convolutional deep learning model [36] that can work with real-time constraints. The detected face is then preprocessed: the image is cropped to extract the region of interest, converted from RGB to grayscale, resized to a resolution of 48 × 48 pixels, and finally normalized into a [0,1] range. In the second stage, the preprocessed image is fed into a low level feature extraction layer and a deep convolutional ensemble of neural networks to obtain the emotion classification [37], [38]. This ensemble model was trained on the FER-2013 database [39] achieving a 72.47% accuracy on the test set.

The FER-2013 database consists of 3 subsets containing 48×48 pixels images: 28709 images dedicated to training, 3589 images for validation and 3589 images for testing. All images include the following labeling: 0 angry, 1 disgust, 2 afraid, 3 happy, 4 sad, 5 surprised and 6 for neutral.

In the approach presented, a model is trained with a database of images of static facial expressions, however, it is evaluated on dynamic facial expressions, while volunteers explain their emotions.

Since the database in this paper cannot be made public, and in order to allow the research community to compare the results, the outcome of our approach on the public RAVDESS database [40] has also been evaluated.

The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) is a validated multi-modal database of emotional speech and song. The database is gender-balanced consisting of 24 professional actors, vocalizing lexically-matched statements in a neutral North American accent. The speech includes calm, happy, sad, angry, fearful, surprise, and disgust expressions; and the song contains calm, happy, sad, angry, and fearful emotions. Each expression is produced at two levels of emotional intensity, with an additional neutral expression.

D. CLASSIFICATION

After computing a set of meaningful features, the feature space must be carefully transformed in order to allow machine learning algorithms to exploit statistical inferences. In that way, smoothing the feature space deals with both, the amount of variability that emerges due to subtle changes in emotional states across trials, and with the lack of stability over time of the computed features. Therefore, a Savitzky-Golay filtering method [28] is used. Also, Quantile-Transform method (histogram equalization to uniform distribution) followed by the Min/Max scaling method is performed to deal with outliers, which can severely damage the performance of the classifiers, and state the range of values according to the input requirements of classifiers.

Then, the classification process is performed using a set of 8 standard classifiers: K-nearest neighbors, Support Vector Machine with linear and radial basis function kernels, Decision Trees, Random Forests, Ada-Boost, Gaussian Naive-Bayes, and Quadratic Discriminant Analysis. Results have been obtained using default hyper-parameter values in the Scikit-learn python library [41]. Also, it has been used to ensure that samples in the validation set are reasonably independent of the samples in the training set [42], [43]. In that context, the Leave-One-Out strategy was used to asses the correct performance of the methodology and the macro-average F1-score metric was used.

IV. RESULTS

In the following sections, offline analysis results are presented. First, population-based exploratory data analysis is performed. The current experimental paradigm is validated regarding the subjective self-assessment of volunteers. Also, causal and correlation effects are presented, which both help to answer the first question formulated in this paper. Second, subject dependent classification is analyzed for each source of signals to prove the feasibility and reliability of the accuracy results. Finally, the subjective assessment of the experiment, carried out by volunteers, is analyzed.

A. EXPLORATORY POPULATION-BASED DATA ANALYSIS

The first attempt at this exploratory analysis is to validate the experimental design. The dramatic story was necessary to boost emotions dynamically, so the effect on the subjective self-evaluation (scene labeling) of the volunteers must be pointed out. Fig. 6 shows the distribution of the frequencies of each discrete set of emotions in the scenes labeling for the emotional dimensions of valence and arousal. It can be noted that the dramatic story is balanced in terms of the properties of positive and negative stimuli, as neutral self-assessments were not often used by the volunteers. It is therefore clear that the dramatic story and the chosen scenes evoked very different emotional states in relation to the valence dimension. On the other hand, in the excitement dimension, the balance was produced by intense and neutral emotions since almost no stimulus was qualified as relaxing.

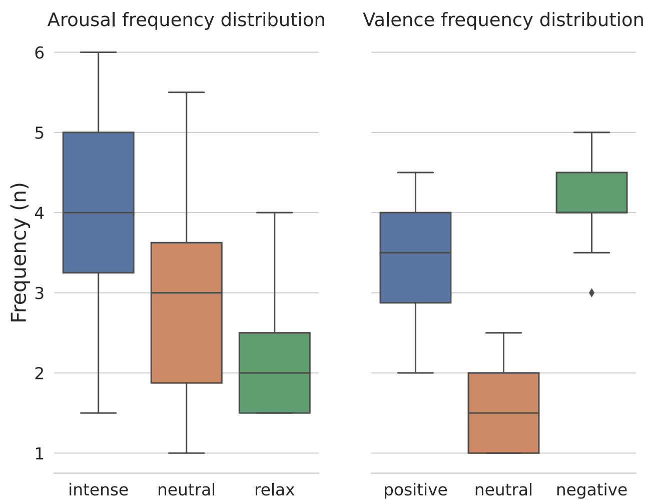


FIGURE 6. Distribution of the frequency of scenes labeled with each emotion in the whole story by all participants. Left: box-plots for frequency of labels for the arousal dimension. Right: box-plots for frequency of labels for the valence dimension.

The architecture of the dramatic script imprints a fingerprint of evoked emotions and specifies intrinsic relationships between them. Fig. 7 shows the correlation of the volunteers' self-evaluations on the scenes. Cross-interactions between the emotional dimensions indicate that POSITIVE and RELAXED are highly correlated while POSITIVE and INTENSE are highly anti-correlated. Also, POSITIVE and

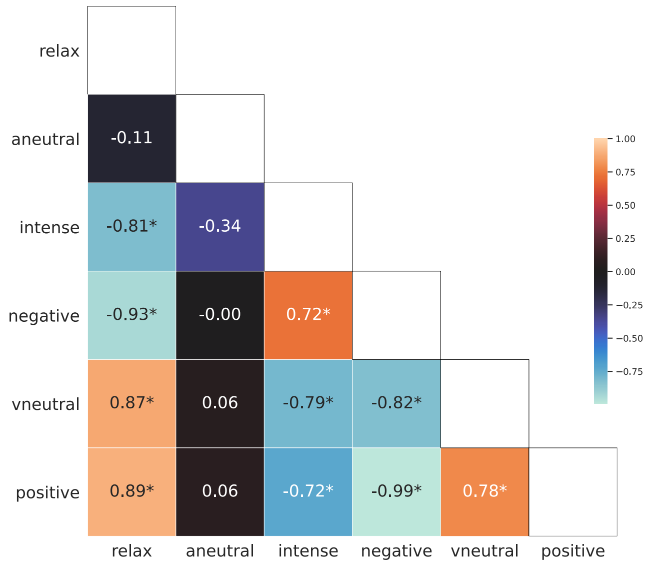


FIGURE 7. **Correlation matrix (corrgram) of emotions ratings over scenes for each pair of emotions. The valence and arousal dimensions are taken into account. Correlation values labeled with (*) are statistically significant $P_{value} < 0.05$.**

NEUTRAL-VALENCE are highly correlated. Two main conclusions can be drawn from these results. First, the majority of volunteers agree to rate some scenes as NEGATIVE while the ratings of POSITIVE and NEUTRAL-VALENCE indicate that some scenes are not as clearly defined as those rated primarily as NEGATIVE. On the other hand, it seems that for most volunteers the POSITIVE emotions can also be felt as RELAX regarding the arousal dimension. In contrast, NEGATIVE and INTENSE are highly correlated, suggesting that videos rated mainly as NEGATIVE cause a dramatically stronger impact than those rated mainly as POSITIVE. Finally, for NEUTRAL-VALENCE and RELAX, there seems to be a high interrelationship, leading to the conclusion that scenes rated mainly as NEUTRAL-VALENCE have a close relation with those rated as POSITIVE. Taking that into account, a pragmatical decision has been taken for the classification analysis, the NEUTRAL-VALENCE and RELAX labels have been considered to be equal.

Another perspective on the evolution of the emotions evoked can be seen in Fig. 8, where the majority of votes for each scene are shown. From this point of view, it can be extrapolated that the valence dimension is mostly balanced into the extremes, while in the arousal dimension, and relative to the valence dimension, the RELAX and NEUTRAL majority ratings seem to correlate highly with the scenes mainly rated as POSITIVE, and only the scenes mainly rated as INTENSE are clearly defined, as they have a significant coherence among the volunteers. Therefore, highlighting more evidence that the aforementioned pragmatic decision of merging both NEUTRAL-VALENCE and RELAX.

So far, subjective ratings have been analyzed, however, an important aspect is the objective effect of the designed emotional drive, over the physiological responses, which allow us to understand and prove that subjective feelings reflect unbalanced sympathetic and parasympathetic subsystems. In the present case and considering the way the features have been computed, some statistically significant correlations have been found in the BVP and EEG features, which are shown graphically in Fig. 9 and Fig. 10. Concerning the dimension of excitation and for the case of BVP measurements, the standard deviation of IBI differences characteristic is highly correlated with the INTENSE excitation emotion. On the other hand, no GSR features showed statistically significant correlations. As for the valence dimension, for the case of EEG measurements, Fisher's information on the T7 temporary electrode for the gamma band is highly correlated with the NEGATIVE valence emotion, while the same Fisher's information but on the P8 parietal electrode for the beta band, is highly correlated with the NEUTRAL valence emotion.

B. CLASSIFICATION OF EMOTIONS USING A SUBJECT-DEPENDENT PARADIGM

As the future aim is to build automatic systems for emotion recognition on affective HRI scenarios, valence and arousal emotion estimation was performed regarding physiological signals. For facial expression recognition, although the model is able to estimate seven discrete emotions, the estimation process has been simplified in order to map from seven discrete emotions {Neutral, Surprise, Happy, Sad, Angry, Fear, Disgust}, detected by the aforementioned ensemble of convolutional models, into three discrete valence emotions {NEUTRAL, POSITIVE, NEGATIVE}. It seems that the context plays a fundamental role in deciding the surprise valence towards positive or negative. That is still an open question, so we decided to categorize surprise as NEUTRAL, see Table 2. This simplification allowed the homogenization of the estimation outputs of the whole system and to validate the results obtained regarding the self-assessment of volunteers.

TABLE 2. **Discrete emotion mapping.**

Emotional dimension	Label
Neutral, Surprise	NEUTRAL
Happy	POSITIVE
Sad, Fear, Angry, Disgust	NEGATIVE

For the present paper, the task was faced independently for each emotional dimension, as the focus is on exploring the plausibility of emotion estimation under this novel paradigm. Regardless that GSR has not shown any statistically meaningful correlations on the population-based EDA, arousal is estimated using GSR and BVP signals.

As mentioned before, in order to properly validate the performance of the models, a Leave-One-Out validations strategy has been used. In fact, in order to exhaustively test the performance, 20 classification iterations were carried out. In each one, a set of two random scenes were

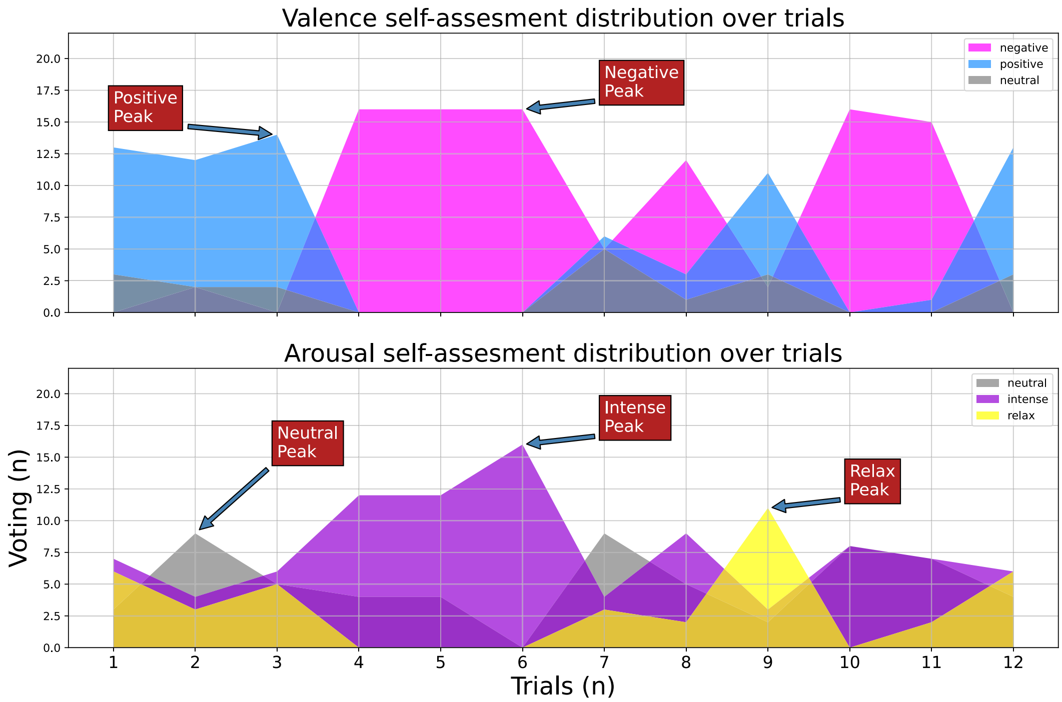


FIGURE 8. Rating distribution over selected scenes for all participants. Top: rating over time for the arousal dimension. Bottom: rating over time for the valence dimension.

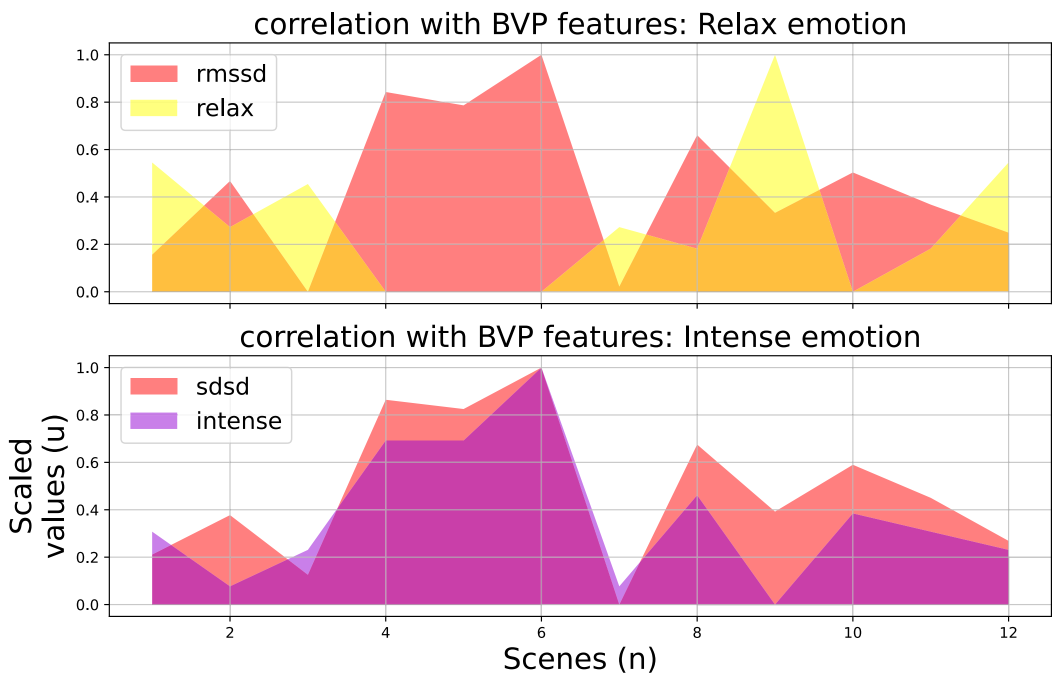


FIGURE 9. Correlation between two features of the BVP signals with the arousal majority ratings over time. RMMSD is Root Mean Square of successive differences in IBI. SDD is Standard Deviation of IBI differences.

selected as trials, each of them belonging to differentiated labels (POSITIVE and NEGATIVE), and the final f1-score value represents the mean and standard deviation of these iterations.

Fig. 11 shows the overall tendency taking into account different sets of features, from 1 to 13 for arousal dimension and the following set of $\{1, 5, 10, 15, 20, 25, 30\}$ number of features, for valence dimension. Achieved accuracy results

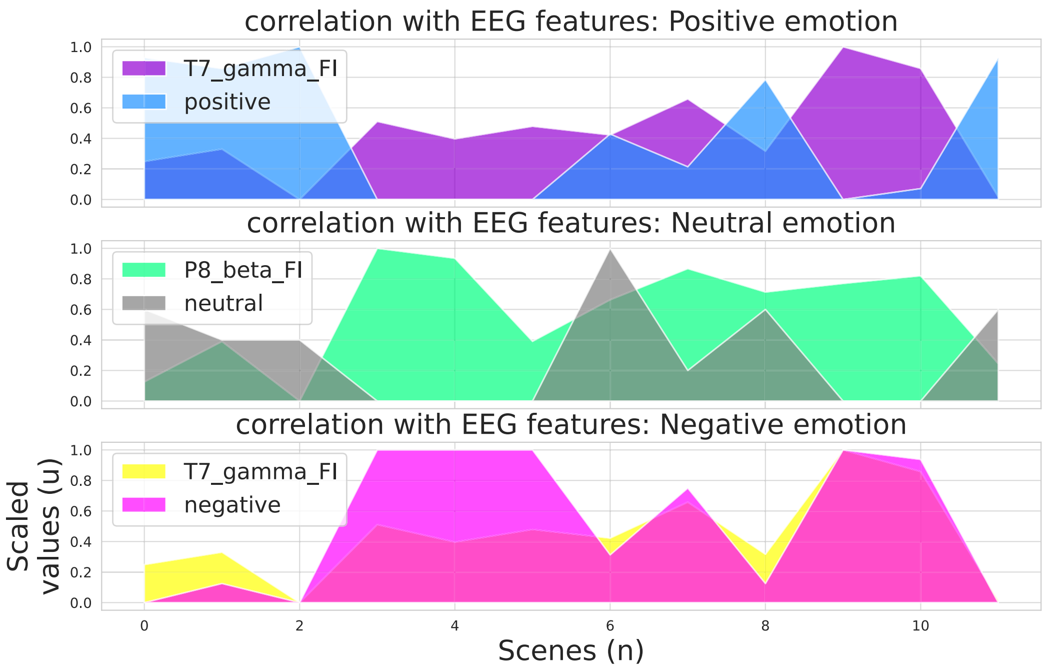


FIGURE 10. EEG most correlated or anti-correlated features across the scenes. Fisher's information of both, T7 temporary electrode for the gamma band (T7_gamma_FI) and P8 parietal electrode for the beta band (P8_beta_FI).

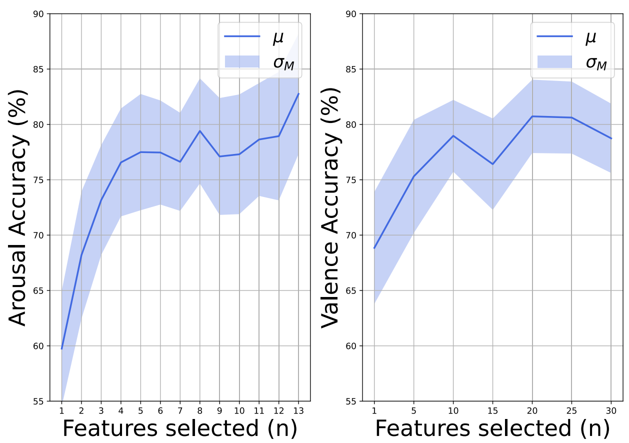


FIGURE 11. Arousal and Valence classification accuracy results. Macro-average F1-score validation metric used. The performance was evaluated with different subsets of features. Solid line is the mean value and shadow area represent de standard error of mean.

are higher than 80% on average for both emotional dimensions.

Fig. 12 shows the coherence of the facial expression recognition model on the RAVDESS database, taking into account that the real emotion expressed by the actors has been mapped to the set of three discrete emotions {NEUTRAL, POSITIVE, NEGATIVE}. The model seems to fail mainly on angry and surprise emotions but in general, is successful in the estimation, regardless of the lack of temporal information.

On the contrary, facing a real-world scenario where emotions are expressed sincerely and not merely acted, without any emphasis on expressing them and therefore creating no bias on the outcome of the recognition system, the model is only capable of having meaningful results for some subjects, as it can be noted in Fig. 13.

C. EXPERIMENT RATING QUESTIONNAIRE

In order to rate the affective HRI experience of users, a series of questions have been done after the experiment. Fig. 14 shows box-plots for the distribution of ratings assigned by all participants to the first 5 questions.

q1. When you started the experiment, were you in a good mood to interact with the robot? or did the robot make you nervous? Rate from 0 to 10, with 0 being fully uncomfortable, 5 neutral and 10 fully comfortable.

q2. Did you like the story Pepper told you, or did it seem like a series of unconnected videos with a meaningless thread? Rate from 0 to 10 the story, being 0 fully unconnected, 5 neutral and 10 fully connected.

q3. What level of emotional engagement, empathy, did you generate in the interaction with the robot? Evaluate from 0 to 10 the perceived empathy, being 0 without any empathy, 5 neutral and 10 full empathy.

q4. Do you consider yourself extroverted or introverted? Evaluate from 0 to 10, being 0 fully introverted, 5 in equal parts, 10 fully extroverted.

q5. Do you consider that the robot brings dramatic value to the story, or a tablet with the same audio and videos would cause the same level of involvement in the story? Rating from 0 to 10, with 0 being the robot that contributes nothing and 10 being the robot that makes me fully involved.

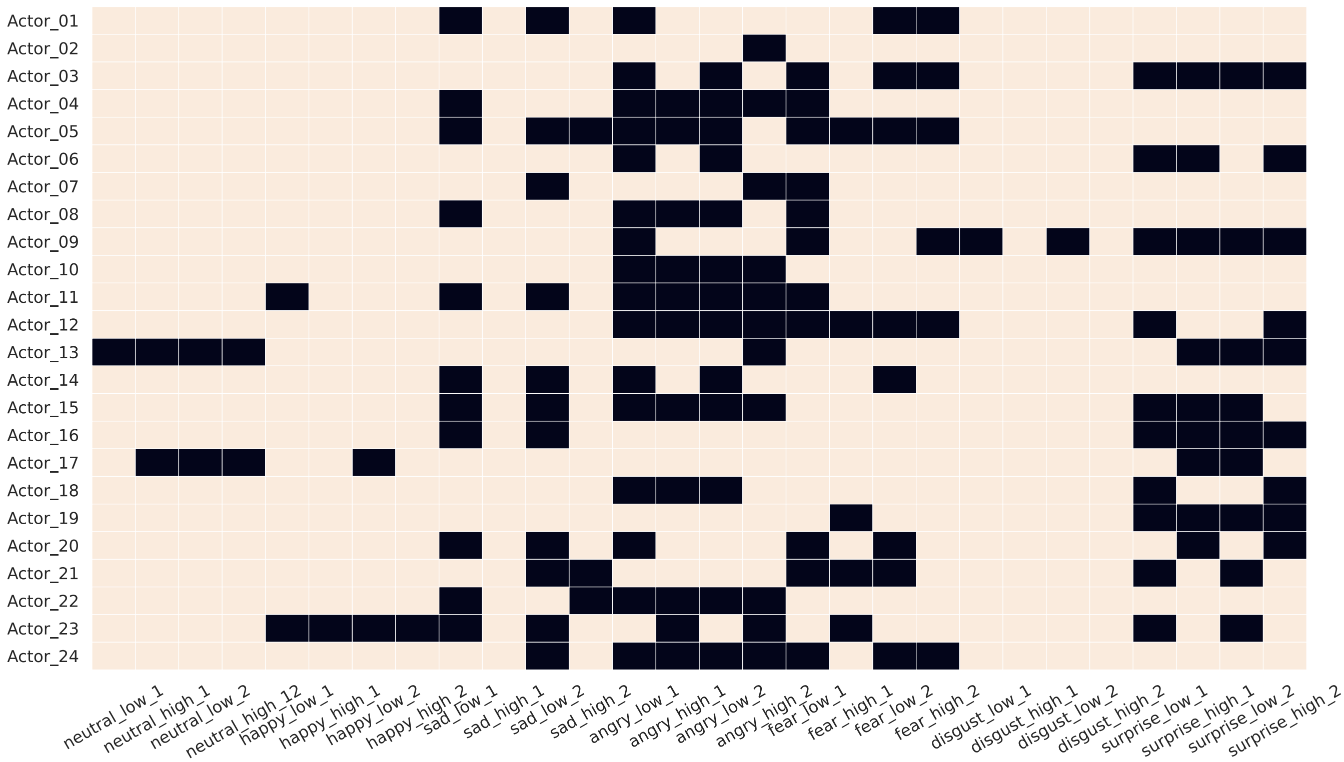


FIGURE 12. Facial expression recognition coherence on the acted out emotions of the RAVDESS database. Black cells show incorrect estimations.

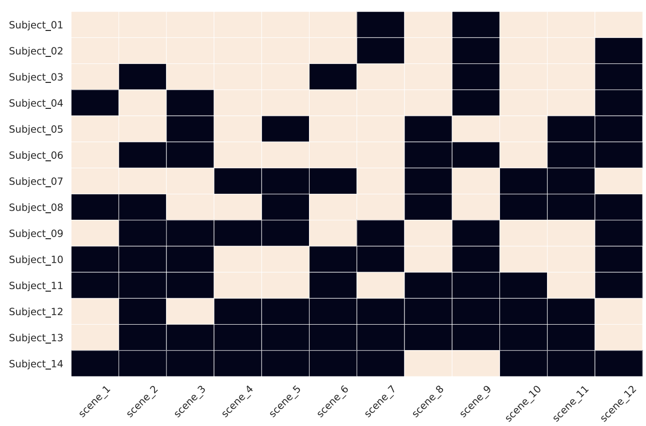


FIGURE 13. Facial expression recognition coherence for the self-expressed emotional reactions during the experiment. Black cells show incorrect estimations.

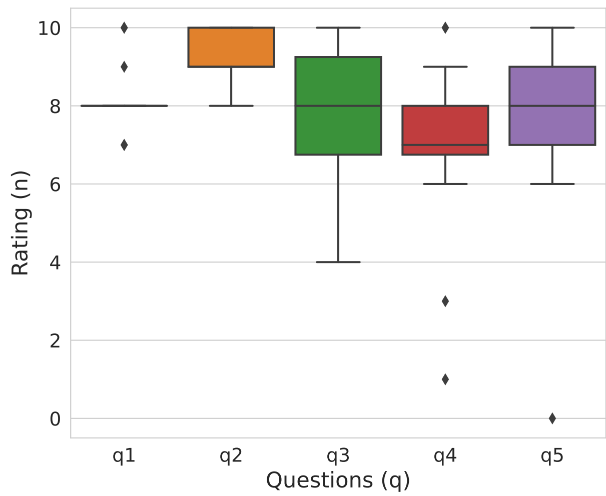


FIGURE 14. Distribution of ratings assigned by all participants to questions regarding the experiment experience (q1, q4, and q5 have some outliers). As q6 is a binary question, it is not represented in these box-plots.

q6. In the hypothetical case of having to express an emotional experience, if you had to choose between a fully unknown person or a robot, who would you choose?

Regarding the last question q6, 37.5% of users would choose a robot. Finally, as a measure of emotional engagement, the mean time spent by all volunteers after each scene is shown in Fig. 15. It can be noted that most users tend to spend more time as the affective HRI goes on.

V. DISCUSSION

To create a more realistic scenario, a dramatic story has been chosen as the emotional drive for the robot to engage emotionally with the volunteers. The dramatic story talks about some of the most important human philosophical questions such as love, the relationship between humans and nature, war and the future of humanity. Such a paradigm tries to evoke the emotions of volunteers, to make them think about them and express their deeper insights both verbally and emotionally. Bias is one of the main questions for any experimental design and it is directly related to the experimental design.

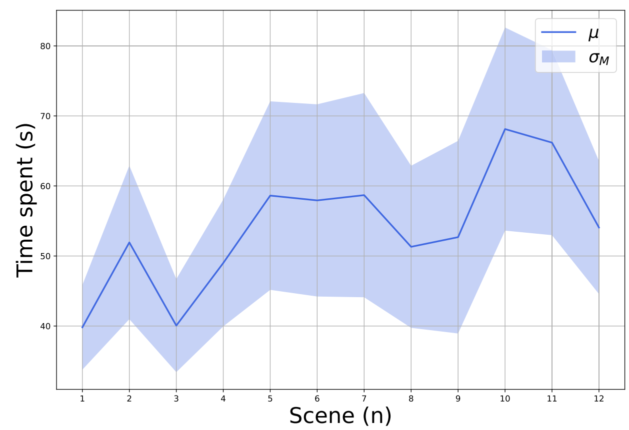


FIGURE 15. Mean and standard deviation of the time spent by all users during the self-expressions of emotions after each scene during the dramatic story. Solid line is the mean value and shadow area represent de standard error of mean.

For the experiment at hand, after each scene, volunteers were completely free to express their emotional thoughts to each of the questions the dramatic story proposes. Such a paradigm makes emotion prediction more complex, as observed when comparing the results on the RAVDESS database, which consists of a set of actors expressing emotions, with the results obtained from the self-developed recordings in a completely realistic scenario. Several papers have carried out an affective HRI approach following at some point the same paradigm as in the RAVDESS database, that is, asking volunteers to act a series of emotional reactions [44]–[48]. This causes volunteers to overreact their facial expressions. This is also the case for the FER-2013 database which has more than twenty thousand facial expressions that are overreacted, causing bias in any result obtained using this type of data. Therefore, this is an important issue that must be faced to properly validate the results obtained, since, on the one hand, these databases allow the development of research in the field, but on the other, they are still quite far from reality.

Facial expression recognition is still a challenging task due to several problems. Firstly, databases developed are usually carried out with actors, who overreact facial expressions, as is the case for FER-2013 and RAVDESS databases. Moreover, for a proper algorithm to be developed, the temporal and spatial connection of facial expressions should be taken into account. That means, training a model over static facial expressions is not enough to achieve accurate results. Indeed, to properly exploit the emotional information contained in the self-expressed emotional reactions, models should take into account the dynamics inherent to each expressed emotion which are, also, interviewed with facial movements related to the current speech. Finally, culture and personal differences arise between subjects, and therefore, algorithms should be fine-tuned for some volunteers to improve the accuracy on them as it can be noticed in the comparison between emotional reactions from two distinct volunteers shown on Fig. 16, Fig. 17, Fig. 18 and Fig. 19. Fig. 16 and Fig. 17 show the evolution of the predicted facial expressions corresponding to the real frames in order to properly evaluate the reliability of the model for an expressive subject. On the contrary, Fig. 18 and Fig. 19 show the same evolution of predictions and facial expressions for a non-expressive subject.

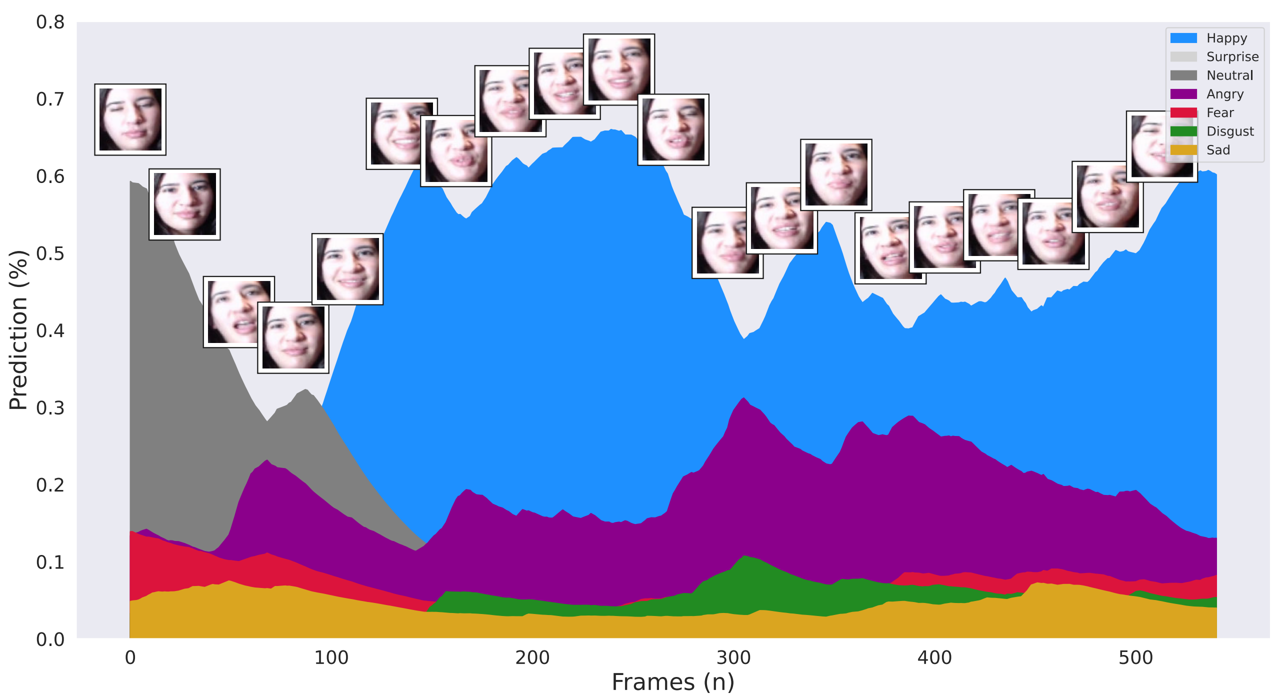


FIGURE 16. Expressive subject self-expressions. POSITIVE facial expression evolution.

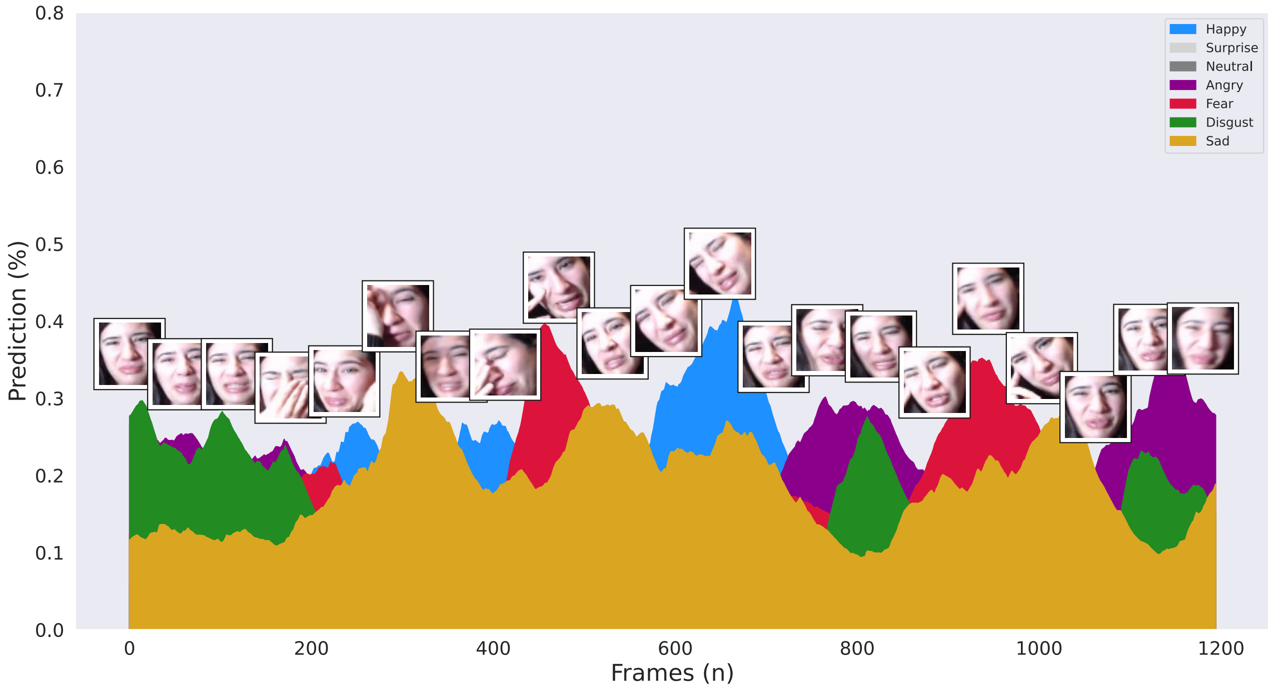


FIGURE 17. **Expressive subject self-expressions. NEGATIVE facial expression evolution.**

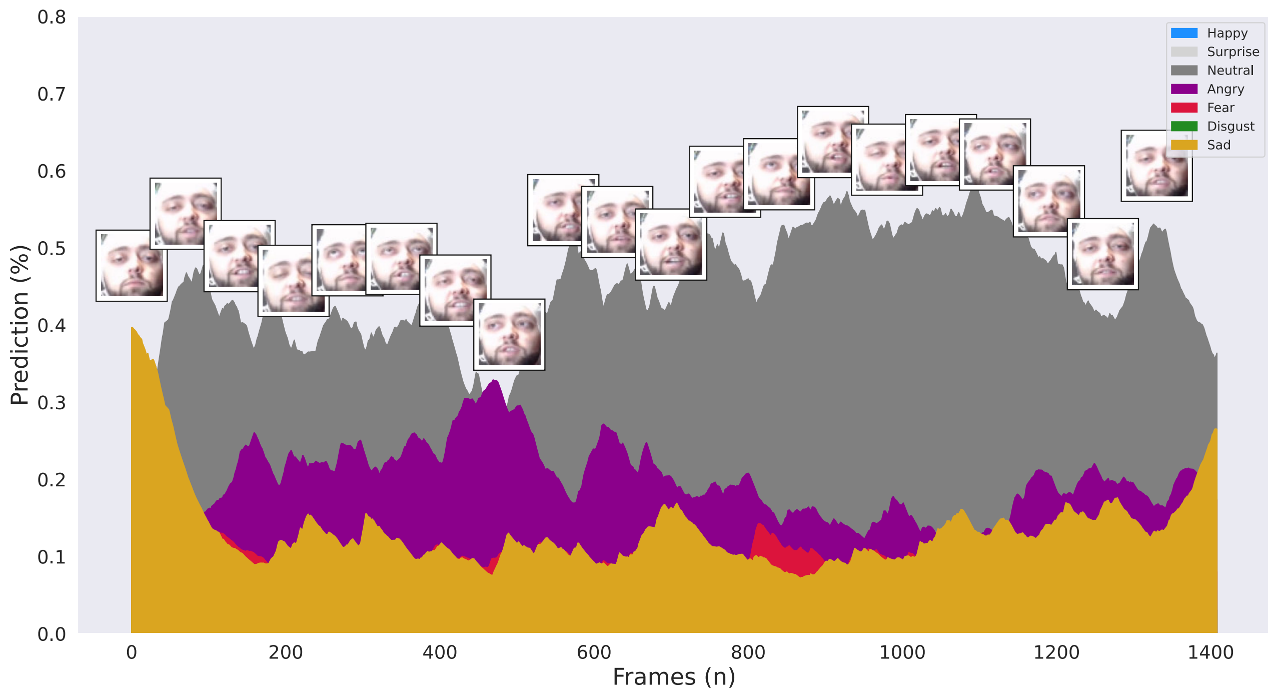


FIGURE 18. **Non-expressive subject self-expressions. POSITIVE facial expression evolution.**

As it can be noted, the model can capture the differences for each facial expressions and therefore the predicted label is close to reality, while this is not the case for a non-expressive subject which clearly shows a noticeable tendency towards a neutral facial expression where differences are too subtle for the model to be able to capture them.

Regarding the correlations between signals and emotions, the GSR signal did not show statistical results. This could be due to the activation of the sympathetic tone during the "self-expression" sections, which could be altering the balance of the autonomic system, and therefore, the correlation of this signal during the "watching" sections is disturbed.

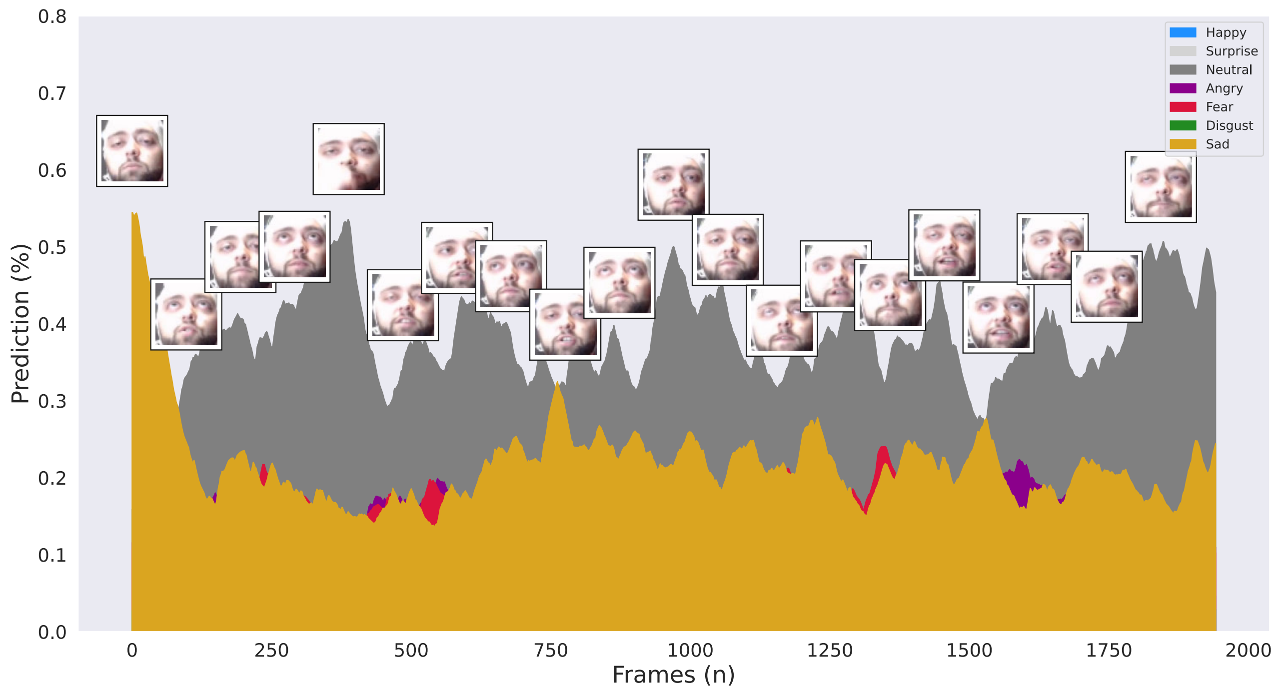


FIGURE 19. **Non-expressive subject self-expressions. NEGATIVE facial expression evolution.**

Concerning the physiological signals, both EEG, for valence prediction, and GSR with BVP, for the arousal prediction, showed to be robust even taking into account that our methodology has been developed without hyper-parameter tuning or the use of any powerful deep-learning model. Therefore, emotion estimation in such a paradigm is not only possible but has a wide margin for optimization. Regarding the validation methodology, the Leave-One-Out strategy was used to asses the correct performance. For the case of a set of samples computed from temporal signals, the temporal correlation must be taken into account and therefore reasonable independence must be maintained between training and test sets.

Taking into account that a population of size 16 is clearly not enough to extract any conclusion, it is noticeable that more than a third of the population would prefer a robot as an emotional companion instead of a human. In addition, Fig. 14 shows that most volunteers were in a positive mood which allowed them to empathize with the robot, moreover, they rated the story as engaging, and the robot was part of the effectiveness of such engagement, leading to the conclusion that robots are appropriate tools to develop affective HRI therapies.

VI. CONCLUSION

This paper has introduced a novel experimental paradigm that has proved to be pragmatically useful as a causal emotion generation mechanism. The use of computationally light emotional estimation methodologies plus wearable and cheap sensors could allow the development of affective HRI therapies in realistic scenarios. This method uses a three-category emotional model, however, for emotion estimation, regarding physiological signals, only two of them were used for each dimensional model. In addition, the Leave-One-Out validation scheme ensures the appropriateness of the proposed methodology in terms of the accuracy of the results, regardless of hyper-parameter tuning. While facial expression recognition provides a useful insight into which emotions are in process, for realistic scenarios, further research must be done in order to develop databases that are closer to reality.

REFERENCES

[1] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 2000.

[2] V. Vidler, "Use puppets to reach the emotionally disturbed," *Instructor*, vol. 81, pp. 9–68, May 1972.

[3] N. Currant, "The expansive educational value of puppets," *Acad. Therapy*, vol. 21, no. 1, pp. 55–60, Sep. 1985.

[4] S. R. Carter, "Use of puppets to treat traumatic grief: A case study," in *Proc. Elementary School Guid. Counseling*, 1987, pp. 210–215.

[5] Q. Yao, "Multi-sensory emotion recognition with speech and facial expression," Ph.D. dissertation, Dept. Comput. Sci. Comput. Eng., Univ. Southern Mississippi, Hattiesburg, MS, USA, 2014. [Online]. Available: https://aquila.usm.edu/dissertations/710/

[6] A. Konar, A. Halder, and A. Chakraborty, "Introduction to emotion recognition," in *Emotion Recognition: A Pattern Analysis Approach*. Hoboken, NJ, USA: Wiley 2015, ch. 1, pp. 1–45, doi: 10.1002/9781118910566.

[7] I. J. Roseman, M. S. Spindel, and P. E. Jose, "Appraisals of emotion-eliciting events: Testing a theory of discrete emotions," *J. Pers. Social Psychol.*, vol. 59, no. 5, p. 899, 1990.

[8] J. A. Russell, "A circumplex model of affect," *J. Pers. Social Psychol.*, vol. 39, no. 6, p. 1161, Dec. 1980.

[9] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 417–429, Jul. 2019.

[10] M. Val-Calvo, J. R. Álvarez-Sánchez, J. M. Ferrández-Vicente, A. Díaz-Morcillo, and E. Fernández-Jover, "Real-time multi-modal estimation of dynamically evoked emotions using EEG, heart rate and galvanic skin response," *Int. J. Neural Syst.*, vol. 30, no. 4, Apr. 2020, Art. no. 2050013.

[11] Z. Yin, Y. Wang, L. Liu, W. Zhang, and J. Zhang, "Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination," *Frontiers Neurorobot.*, vol. 11, p. 19, Apr. 2017.

[12] R. Adolphs, "Recognizing emotion from facial expressions: Psychological and neurological mechanisms," *Behav. Cognit. Neurosci. Rev.*, vol. 1, no. 1, pp. 21–62, Mar. 2002.

[13] E.-H. Jang, B.-J. Park, M.-S. Park, S.-H. Kim, and J.-H. Sohn, "Analysis of physiological signals for recognition of boredom, pain, and surprise emotions," *J. Physiol. Anthropol.*, vol. 34, no. 1, p. 25, Dec. 2015.

[14] M. Diano, M. Tamietto, A. Celeghin, L. Weiskrantz, M.-K. Tatu, A. Bagnis, S. Duca, G. Geminiani, F. Cauda, and T. Costa, "Dynamic changes in amygdala psychophysiological connectivity reveal distinct neural networks for facial expressions of basic emotions," *Sci. Rep.*, vol. 7, no. 1, p. 45260, May 2017.

[15] S. Wang, R. Yu, J. M. Tyszka, S. Zhen, C. Kovach, S. Sun, Y. Huang, R. Hurlemann, I. B. Ross, J. M. Chung, A. N. Mamelak, R. Adolphs, and U. Rutishauser, "The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity," *Nature Commun.*, vol. 8, no. 1, pp. 1–13, Apr. 2017.

[16] D. Hagemann, S. R. Waldstein, and J. F. Thayer, "Central and autonomic nervous system integration in emotion," *Brain Cognition*, vol. 52, no. 1, pp. 79–87, Jun. 2003.

[17] J. Zheng, R. F. Stevenson, B. A. Mander, L. Mnatsakanyan, F. P. Hsu, S. Vadera, R. T. Knight, M. A. Yassa, and J. J. Lin, "Multiplexing of theta and alpha rhythms in the amygdala-hippocampal circuit supports pattern separation of emotional information," *Neuron*, vol. 102, no. 4, pp. 887–898, 2019.

[18] G. Girardeau, I. Inema, and G. Buzsáki, "Reactivations of emotional memory in the hippocampus–amygdala system during sleep," *Nature Neurosci.*, vol. 20, no. 11, p. 1634, 2017.

[19] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, May 1993.

[20] J. L. Andreassi, *Psychophysiology: Human Behavior and Physiological Response*. East Sussex, U.K.: Psychology Press, 2010.

[21] A. R. Damasio, "Emotion in the perspective of an integrated nervous system," *Brain Res. Rev.*, vol. 26, nos. 2–3, pp. 83–86, 1998.

[22] P. J. Lynch. *Cranial Nerve VII*. Accessed: Dec. 23, 2006. [Online]. Available: https://commons.wikimedia.org/wiki/File:Cranial_nerve_VII.svg

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[24] D. Nguyen, K. Nguyen, S. Sridharan, D. Dean, and C. Fookes, "Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition," *Comput. Vis. Image Understand.*, vol. 174, pp. 33–42, Sep. 2018.

[25] M. Val-Calvo, "General physiological and camera acquisition system for human behaviour recordings and analysis in real-time," Zenodo, Mar. 2020, doi: 10.5281/zenodo.3727503.

[26] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[27] M. Val-Calvo, J. R. Álvarez-Sánchez, J. M. Ferrández-Vicente, and E. Fernández, "Optimization of real-time EEG artifact removal and emotion estimation for human-robot interaction applications," *Frontiers Comput. Neurosci.*, vol. 13, p. 80, Nov. 2019.

[28] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares Procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.

[29] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals," *Proc. IEEE*, vol. 80, no. 4, pp. 520–538, Apr. 1992.

[30] VoytekLab. (2018). *Neuro Digital Signal Processing Toolbox*. [Online]. Available: https://github.com/neurodsp-tools/neurodsp

[31] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," in *Proc. 8th IEEE Symp. Computer-Based Med. Syst.*, Jun. 1995, pp. 212–217.

[32] M. Affinito, M. Carrozzi, A. Accardo, and F. Bouquet, "Use of the fractal dimension for the analysis of electroencephalographic time series," *Biol. Cybern.*, vol. 77, no. 5, pp. 339–350, Nov. 1997.

[33] R. A. Fisher, "Theory of statistical estimation," *Math. Proc. Cambridge Philos. Soc.*, vol. 22, no. 5, pp. 700–725, Jul. 1925.

[34] F. S. Bao, X. Liu, and C. Zhang, "PyEEG: An open source Python module for EEG/MEG feature extraction," *Comput. Intell. Neurosci.*, vol. 2011, Oct. 2011, Art. no. 406391.

[35] P. M. Pedro Gomes and H. Silva. (2018). *pyHRV—Open-Source Python Toolbox for Heart Rate Variability*. [Online]. Available: https://github.com/PGomes92/hrv-toolkit/

[36] I. Itzcovich. (2018). *Yolo-Face-Detection*. [Online]. Available: https://github.com/iitzco/faced

[37] N. K. Benamara, M. Val-Calvo, J. R. Álvarez-Sánchez, A. Díaz-Morcillo, J. M. Ferrández-Vicente, E. Fernández, and T. B. Stambouli, "Real-time emotional recognition for sociable robotics based on deep neural networks ensemble," in *Proc. Int. Work-Conf. Interplay Between Natural Artif. Comput.*, in LNCS, Almería, Spain, vol. 11486. Springer, Jun. 2019, pp. 171–180, doi: 10.1007/978-3-030-19591-5_18.

[38] N. K. Benamara, M. Val-Calvo, J. R. Álvarez-Sánchez, A. Díaz-Morcillo, J. M. Ferrández-Vicente, E. Fernández, and T. B. Stambouli, "Real-time facial expression recognition for affective robotics using smoothed deep neural network ensemble," *Integr. Comput.-Aided Eng.*, to be published.

[39] Facial Expression Recognition Dataset. (Feb. 2013). *Challenges in Representation Learning: Facial Expression Recognition Challenge*. [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge

[40] S. R. Livingstone and F. A. Russo, "The ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in north American English," *PLoS ONE*, vol. 13, no. 5, May 2018, Art. no. e0196391. [Online]. Available: https://zenodo.org/record/1188976

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[42] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: An analysis and review," *Int. J. Forecasting*, vol. 16, no. 4, pp. 437–450, Oct. 2000.

[43] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Inf. Sci.*, vol. 191, pp. 192–213, May 2012.

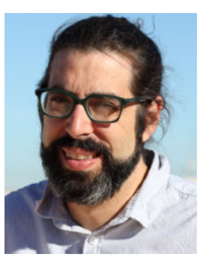
[44] C.-C. Tsai, Y.-Z. Chen, and C.-W. Liao, "Interactive emotion recognition using support vector machine for human-robot interaction," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2009, pp. 407–412.

[45] F. Cid, L. J. Manso, and P. Núnez, "A novel multimodal emotion recognition approach for affective human robot interaction," in *Proc. FinE*, 2015, pp. 1–9.

[46] L.-A. Perez-Gaspar, S.-O. Caballero-Morales, and F. Trujillo-Romero, "Multimodal emotion recognition with evolutionary computation for human-robot interaction," *Expert Syst. Appl.*, vol. 66, pp. 42–61, Dec. 2016.

[47] Z. Liu, M. Wu, W. Cao, L. Chen, J. Xu, R. Zhang, M. Zhou, and J. Mao, "A facial expression emotion recognition based human-robot interaction system," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 4, pp. 668–676, 2017.

[48] L. Chen, M. Zhou, W. Su, M. Wu, J. She, and K. Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction," *Inf. Sci.*, vol. 428, pp. 49–61, Feb. 2018.



MIKEL VAL-CALVO received the B.S. degree in software engineering and the M.S. degree in intelligent diagnostic, planning and control systems from the Universidad Nacional de Educación a Distancia, Madrid, Spain, in 2016 and 2017, respectively, where he is currently pursuing the Ph.D. degree in intelligent systems. His research interest includes the development of methodologies for real-time emotion recognition under the human–robot interaction paradigm.

JOSÉ RAMÓN ÁLVAREZ-SÁNCHEZ received the B.Sc. and M.Sc. degrees in physics from the Universidad Complutense de Madrid, Spain, in 1988, and the Ph.D. degree in physics from the Universidad Nacional de Educación a Distancia, Madrid, Spain, in 1997. He is currently an Associate Professor with the Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia. His main research interests include artificial neural networks and mobile robotics.

JOSÉ MANUEL FERRÁNDEZ-VICENTE was born in Elche, Spain. He received the M.Sc. and Ph.D. degrees in computer science from the Universidad Politécnica de Madrid, Spain. He is currently Full Professor with the Department of Electronics, Computer Technology and Projects, Universidad Politécnica de Cartagena, and the Head of the Electronic Design and Signal Processing Research Group, Universidad Politécnica de Cartagena. Nowadays, he is also the Vice President for Internationalization, Research, and Innovation, and the General Chairman at the International Work conference on the Interplay between Natural and Artificial Computation (IWINAC). His research interests include bioinspired processing, neuromorphic engineering, and emotional technologies.

EDUARDO FERNÁNDEZ received the M.D. degree from the University of Alicante, in 1986, and the Ph.D. degree in neuroscience, in 1990. He has been a Visiting Professor with the University of Utah, USA, Oldenburg, Germany, the Beth Israel and Harvard Medical School, USA, and the University of Vienna, Austria. He is currently a Full Professor of cellular biology, the Director of the Research Chair in Retinitis Pigmentosa Bidons Egara, the Director of the Neural Engineering Division, Bioengineering Institute, Universidad Miguel Hernández, and CIBER BBN, Spain, and an Adjunct Professor with the John Moran Eye Center, University of Utah, USA. He is a qualified M.D. who combines biomedicine (molecular and cellular biology, biochemistry, anatomy, physiology, and regenerative medicine) with the physical sciences (mathematics, physics, and applied chemistry) and engineering to develop new treatments and devices that can be applied to enhance the life of people that are affected by visual impairments more effectively. He is also trying to better understand brain plasticity in blind subjects and working on the development of new therapeutic approaches for retinal degenerative diseases.