PhD. Thesis
Tesis Doctoral

2020

EMOTIONAL HUMAN-ROBOT INTERACTION USING PHYSIOLOGICAL SIGNALS

Mikel Val Calvo

Doctoral program in Intelligent Systems
Programa de doctorado en Sistemas Inteligentes

Universidad Nacional de Educación a Distancia

José Ramón Álvarez Sánchez (UNED)
José Manuel Ferrández Vicente (UPCT)

Universidad Nacional de Educación a Distancia

Doctoral program in Intelligent Systems
Programa de doctorado en Sistemas Inteligentes

EMOTIONAL HUMAN-ROBOT INTERACTION USING PHYSIOLOGICAL SIGNALS

Author / *Autor*: Mikel Val Calvo

Thesis advisors / *Directores de la tesis*:
Dr. D. José Ramón Álvarez Sánchez
Dr. D. José Manuel Ferrández Vicente

PhD. Thesis
Tesis Doctoral

EMOTIONAL HUMAN-ROBOT INTERACTION USING PHYSIOLOGICAL SIGNALS

Author / *Autor*: Mikel Val Calvo

Thesis advisors / *Directores de la tesis*:
Dr. D. José Ramón Álvarez Sánchez
Dr. D. José Manuel Ferrández Vicente

Evaluation board / *Tribunal calificador*

Chairperson
Presidente

Secretary
Secretaria/o

Member
Vocal

Madrid, de septiembre de 2020

Abstract

The field of Emotional Robots begins to be perceived by many as a tangible reality that seems to develop robustly in the field of research, promoting interdisciplinary approaches. Emotional or socially intelligent robots try to emulate human social intelligence, to be integrated into our society naturally. To this end, robots must be able to learn both to detect and reproduce the way humans communicate, since in our relationships we use emotional signs that are expressed in non-verbal communication patterns, differences in the tone of voice, or nuances in the semantics of the message used. Therefore, the main aim of this thesis is to improve the human-robot interaction, to produce a more natural interaction for the user, under the hypothesis that the robot can improve its interaction if it can reduce its uncertainty when detecting the user's emotional states. Therefore, our goal is to find how to assess the user's emotional state by adding information related to physiological signals such as pulse, skin conductance, electroencephalography, and facial expressions.

Emotion estimation systems based on brain cortical activity, among other physiological signals are gaining special attention in recent years because of the possibilities they offer. The field of human-robot interactions could benefit from a broader understanding of the coding of the brain and physiological properties during emotion processing, together with the use of lightweight software and cheap wearable devices, and thus enhance the capabilities of robots to fully engage with the user's emotional reactions. On the other hand, facial expression recognition has been extensively researched over the past two decades because of its direct impact on the field of computer vision and affective robotics, but it has yet to address several drawbacks such as posture variations, occlusions, lighting, etc. To break down the complexity of the problem taking into account real-time constraints, the process is performed in two stages, automatic face detection, and facial expression recognition.

The affective interaction between humans and robots requires lightweight software and cheap wearable devices, that could boost this field. However, the estimation of emotions in real-time poses a problem that has not yet been fully addressed, namely the challenge of filtering artifacts and extracting features, while reducing processing time and maintaining high accuracy results. Thus, optimization processes must face several stages, such as artifact removal, feature extraction, feature smoothing, and pattern classification, while maintaining real-time constraints.

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. Standard experimental designs face emotion estimation using constant stimuli to have control over the variables under experimentation, allowing for the development of the research, but staying far from real scenarios where emotions are dynamically evoked. The experimental design is by far the most important issue which must be faced, therefore, to properly evaluate our methodologies,

firstly, standard databases using constant stimuli have been used to design a well-validated methodology. Secondly, a dramatic film has been proposed to dynamically evoke emotions. This made it possible to demonstrate that our method was robust enough even in realistic scenarios. Finally, an experimental design of human-robot interaction implies a complex paradigm in which dynamic emotional feedback from users is required, which is far from the initial standard methodologies of emotional stimulation. To minimize the leap involved in the experimental paradigm shift, a strategy has been proposed that combines the two previous approaches in which the robot conducts a dramatic story under a dynamic emotion stimulation strategy that, in addition, uses constant stimuli to evoke specific emotions after which it requests emotional feedback from the user.

In conclusion, emotion estimation is achieved by the use of three different sources, brain patterns, signals from the autonomous neural system, and facial expressions, which allowed us to measure emotions within the developed methodologies. From physiological signals, a combination of “negative-positive” and “relax-intense” emotions can be achieved, while for the facial expressions seven discrete emotions (“happy”, ”surprise”, ”neutral”, ”sad”, ”fear”, ”disgust”, ”angry”) can be measured with significant confidence.

This exploratory research study proposes realistic experimental designs, using dramatic films and story-telling robots, to evoke emotions in the users, and assessing previously self-designed methodologies, to be able to make estimates of the users’ emotional states in real-time. Regardless of the multiple restrictions, and all the aspects that could still be improved, this research can outline the feasibility of the proposed methodology in realistic scenarios.

Resumen

El campo de la robótica emocional comienza a ser percibido por muchos como una realidad tangible que parece desarrollarse con fuerza en el campo de la investigación promoviendo enfoques multidisciplinarios. Los robots emocionales o socialmente inteligentes intentan emular la inteligencia social humana para integrarse en nuestra sociedad de forma natural. Para ello, los robots deben ser capaces de aprender tanto a detectar como a reproducir la forma en que los humanos se comunican, ya que en nuestras relaciones personales utilizamos signos emocionales que se expresan en patrones de comunicación no verbales, diferencias en el tono de voz o matices en la semántica del mensaje utilizado. Por lo tanto, el objetivo principal de esta tesis es mejorar la interacción humano-robot para producir una interacción más natural para el usuario, bajo la hipótesis de que el robot puede mejorar su interacción si es capaz de reducir su propia incertidumbre a la hora de detectar los estados emocionales del usuario. Es por ello, que el objetivo se centra en investigar cómo evaluar el estado emocional del usuario añadiendo información relacionada con señales fisiológicas, tales como el pulso, la conductancia de la piel, la electroencefalografía o las expresiones faciales.

Los sistemas de estimación de emociones basados en actividad neural cortical, entre otras señales fisiológicas, están recibiendo una atención especial en los últimos años debido a las posibilidades que ofrecen. El campo de las interacciones entre humanos y robots podría beneficiarse de una comprensión más amplia de la codificación de las propiedades de la actividad neural cortical y señales fisiológicas, durante el procesamiento de las emociones, junto con el uso de programas informáticos ligeros y dispositivos baratos que sean portátiles y orientados al usuario, y así mejorar las capacidades de los robots para participar plenamente en las reacciones emocionales del usuario. Por otra parte, el reconocimiento de las expresiones faciales se ha investigado ampliamente en las dos últimas décadas debido a su impacto directo en el campo de la visión por computadora y la robótica afectiva, pero aún debe lidiar con inconvenientes como las variaciones en la postura, las oclusiones, la iluminación, etc. Para desglosar la complejidad del problema teniendo en cuenta las limitaciones en tiempo real, el proceso se lleva a cabo en dos etapas, la detección automática de caras y el reconocimiento de las expresiones faciales.

La interacción afectiva entre los seres humanos y los robots requiere un software ligero y dispositivos baratos que sean portátiles que puedan potenciar este campo. Sin embargo, la estimación de las emociones en tiempo real plantea un problema que aún no se ha abordado plenamente, como es el reto de filtrar los artefactos de la actividad neuronal cortical y la extracción de características, al tiempo que se reduce el tiempo de procesamiento y se mantienen los resultados en el reconocimiento. Así pues, los procesos de optimización deben hacer frente a varias etapas, como la eliminación de artefactos, la extracción de características, el suavizado de características y la clasificación de patrones, manteniendo al mismo tiempo las limitaciones en tiempo real.

La interacción afectiva entre el hombre y el robot sigue siendo un área de investigación activa, en parte debido a los grandes avances de la inteligencia artificial. Actualmente, el diseño de dispositivos autónomos que funcionan en entornos terapéuticos reales se ha convertido en una realidad plausible. La interacción afectiva humano-robot requiere que un robot analice el estado emocional del interlocutor humano e interprete las respuestas emocionales que pueden utilizarse, no sólo en la interacción sino, por ejemplo, para provocar las respuestas terapéuticas deseadas. Por lo tanto, es necesario ampliar las técnicas

experimentales a paradigmas más realistas, en los que se pueda explorar completamente la capacidad de estimación de las emociones. Los diseños experimentales estándar se enfrentan a la estimación de las emociones utilizando estímulos constantes para facilitar el análisis de las respuestas evocadas, permaneciendo lejos de escenarios realistas, donde las emociones se evocan dinámicamente. El diseño experimental es, con mucho, la cuestión más importante a la que hay que enfrentarse, por lo que, para evaluar adecuadamente nuestras metodologías, se ha seguido una estrategia incremental. En primer lugar, se han usado bases de datos estándar que utilizan estímulos constantes para diseñar una metodología bien validada. En segundo lugar, se ha propuesto una película dramática para evocar dinámicamente las emociones. Esto permitió demostrar que nuestro método era lo suficientemente robusto incluso en escenarios realistas. Por último, un diseño experimental de la interacción entre el hombre y el robot implica un paradigma complejo en el que se requiere una retro-alimentación emocional dinámica por parte de los usuarios, lo que dista mucho de las metodologías estándar iniciales de estimulación emocional. Para minimizar el salto que supone el cambio de paradigma experimental, se ha propuesto una estrategia que combina los dos enfoques anteriores en la que el robot lleva a cabo una historia dramática bajo una estrategia de estimulación emocional dinámica que, además, utiliza estímulos constantes para evocar emociones específicas, tras lo cual solicita la retro-alimentación emocional del usuario.

A modo de conclusión, la estimación de las emociones se consigue mediante el uso de tres fuentes diferentes, patrones cerebrales, balance entre sistemas simpático y parasimpático, y expresiones faciales, que nos permiten medir las emociones tanto en el modelo emocional continuo como en el discreto. A partir de señales fisiológicas se puede lograr estimar una combinación de emociones “negativa-positiva” y “relax-intenso”, mientras que para las expresiones faciales se pueden reconocer siete emociones discretas (“felicidad”, ”sorpresa”, ”neutral”, ”tristeza”, ”disgusto”, ”miedo”, ”ira”) con una confianza significativa.

Esta investigación exploratoria propone diseños experimentales realistas utilizando estímulos dinámicos, como las películas dramáticas o el uso de robots que cuentan historias, para evocar emociones en los usuarios, a la vez que se evalúan metodologías para poder hacer estimaciones de los estados emocionales de los usuarios en tiempo real. Independientemente de las múltiples restricciones, y de todos los aspectos que aún podrían mejorarse, esta investigación puede esbozar la viabilidad de la metodología propuesta en los diferentes escenarios realistas propuestos.

Acknowledgments

My last 11 years have been dedicated exclusively to academic training at the UNED University. This is a huge amount of time for a young person, which can only be explained by the ability of that university to improve my skills as a student. After having been a regular student almost everywhere, I finally found my best personal motivation to learn under the UNED methodology, which consists mainly of personal hard work reading non-smoking books in solitude. The way I enrolled in this university was totally by accident. I forgot to register on time at the University of Granada, so in September 2009 I discovered that my future for the following year was dark black. This was almost my last and biggest mistake of my life. However, fortune was on my side "the UNED allows you to register until November".

A few years later I was deciding which department I would apply to for my final project. I found the project "A study of the computational capabilities of in vivo cell cultures and their application to robotic control, Director: Felix de la Paz". After a short talk, I found myself packing my things in the van with my faithful friend Ira, on a one-way trip to Elche. The bucolic atmosphere that the palm trees imprint on the landscape, their devastating heat in summer, the dust rains, but above all, the people I met, have since made this place a home that I hope will be temporary since life makes ephemeral things beautiful and unforgettable.

The research in the Neuroprosthesis and Visual Rehabilitation Unit of Eduardo Fernandez Jover in Elche suffers from a severe arrhythmia that causes scientific artistic spasms of the most genuine. Like the rabbit hole that leads to the world of psychodelic ideas, I set out to pursue my white rabbit, the unpromising path to a doctorate.

With the invaluable help of Félix de la Paz, José Manuel Ferrández considered me to carry out the project of the present thesis. José Ramón Álvarez-Sánchez led me around the chessboard, preventing me from coming up with extravagant ideas that would be worthy of Humpty Dumpty himself. Eduardo Fernández Jover lent me his lab and advised me. My parents and my brother, again, supported me during this hard travel.

Apart from the cast of genuine characters that shook my conscience, I keep a faithful and loving memory of the wanderings made together with Javier Alegre Cortés, Antonio Manuel Lozano, Alfonso Rodil and Juan Sebastian Suarez. I will never forget the political-philosophical discussions with Alvaro García, Manuel Bayonas, and Lawrence Humphreys. The amazing English lessons with Zaida Molins. Nor the classes on electrophysiology by Cristina Soto and Gema Martínez, as well as the bad jokes of Antonio Alarcón.

Finally, in a very different, parallel, and no less important story, the biggest fortuitous mistake in this story happened:

"Once upon a time there was a little house on the edge of a small village called Valverde, where lived a beautiful woman of a strong character whose name I will not give any reason for the moment. On a beautiful summer afternoon, a handsome young man dared to put his arms around her, which caused her to cough and laugh very hard. The young man, astonished, tried to help her with a small pat on the back, from which a beautiful child named Jan was born".

The story of this child's wanderings coincides with the end of this short story and the beginning of the next stage of my life, and it is necessary to finish because I will tell you in another time a very different story that has nothing to do with this one.

We want to acknowledge to Programa de Ayudas a Grupos de Excelencia de la Región de Murcia, from Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia and RTI2018-098969-B-100 from the Ministerio de Ciencia, Innovación y Universidades, y PROMETEO/2019/119 from the Generalitat Valenciana.

To my lovely and crazy family,

Ana and Jan.

Contents

List of Figures

Chapter 1

Introduction

The aim of this thesis is to demonstrate that it is possible to improve affective human-robot interaction by reducing the uncertainty of the robot when detecting the user's emotional state. To this end, the use of physiological signals and facial expressions is proposed which implies a multidisciplinary approach. Therefore, it is indispensable to justify the theoretical background that allows us to make such an approach.

Consequently, the next section 1.1 defines the starting point of this research and the field of study is contextualized in section 1.2. Next, in sections 1.3 and 1.4, the biological context in which emotions are produced is explained, and in section 1.5, the main emotional models proposed by the field of study of psychology are presented. Finally, the current state of the art in the scientific field in which this thesis is being developed is reviewed, section 1.6.

The next chapter 2 establises the hypothesis and the objectives to be developed. Chapter 3 defines the methods used, both hardware and software developed, sections 3.1 and 3.2, which in our case is provided for future use by the scientific community. Chapter 4, presents the articles that make up the body of the thesis. These articles have been carried out in an incremental manner according to the hypotheses and objectives set out, in such a way that the methodologies developed for the first article are used in the following articles, and the same is true of the third article with respect to the first and second articles. Section 4.4 contains a series of collaborative research studies of relevance for the consecution of this thesis objectives. The fourth chapter, also includes the general discussion of the results obtained, section 4.5, although each article has its own detailed discussion of results. Finally, chapter 5 describes the conclusions.

1.1 Scientific context

The era of robotics has just started to jump into the real world thanks to the last advances in the fields of control theory and deep and reinforcement learning [Liu and Theodorou, 2019, Recht, 2019]. The general desire that robots perform complex tasks requiring a high level of expertise in complex environments is driving their evolution fast in areas like perception [Burschka, 2019, Stroessner, 2020], sensing and interaction [Reig et al., 2020, Engwall et al., 2020], and in fact, on some tasks they are outperforming humans [Silver et al., 2016]. Nevertheless, there is still a big challenge in the field of robotics directly related to how humans perceive and sense the world [Yang et al., 2018]. Humans are not only capable of perceiving and sensing the world, but they are also conscious about it and they can respond emotionally. Both properties are qualitatively superior from the previous ones and have become one of the next science barriers in the actual era.

Related to consciousness and autonomous learning, several approaches have been developed that reveal the endeavor of building artificial cognitive architectures [Cialdella et al., 2020, Samsonovich, 2020, Kelley and Twyman, 2020]. Concerning emotions, a recent branch of studies based on human emotion recognition is being developed to both give robots the ability to measure human emotions and to understand how emotions are produced [Suguitan et al., 2020, Hasegawa and Takahashi, 2020, Vásquez and Matía, 2020, Martínez Albán et al., 2019, Lee et al., 2020, Todd et al., 2020].

1.2 The scope of the study

There are four main capacities of the human brain in processing the environment:

- First, the capability of sensing the world: In the field of robotics, sensing has been one of the first concerns and, in fact, several types of sensors have been developed.
- Second, the capability of perceiving: This is a qualitatively superior abstraction over the sensing result, where the sensed data is related to meaning.
- Third, emotions: Human perceptions are biased by emotional states at different degrees, that is the reason why emotions play a fundamental role in social interactions.
- Four, the faculty of reasoning: This is the highest abstraction over all prior capabilities, which allow us not only to dynamically accommodate a changing world but additionally to modify it.

In figure 1.1 these capabilities are represented in a graphical sense where qualitative interactions can be abstracted. The main idea of this abstraction is to isolate the possible subsets of interactions that will allow us to clearly define our field of study. Currently, the field of robotics is making huge leaps in the highlighted subareas. Our research area is focused on subarea 3, that merges the act of, sensing and perceiving the emotional states of robot users but still not reasoning.

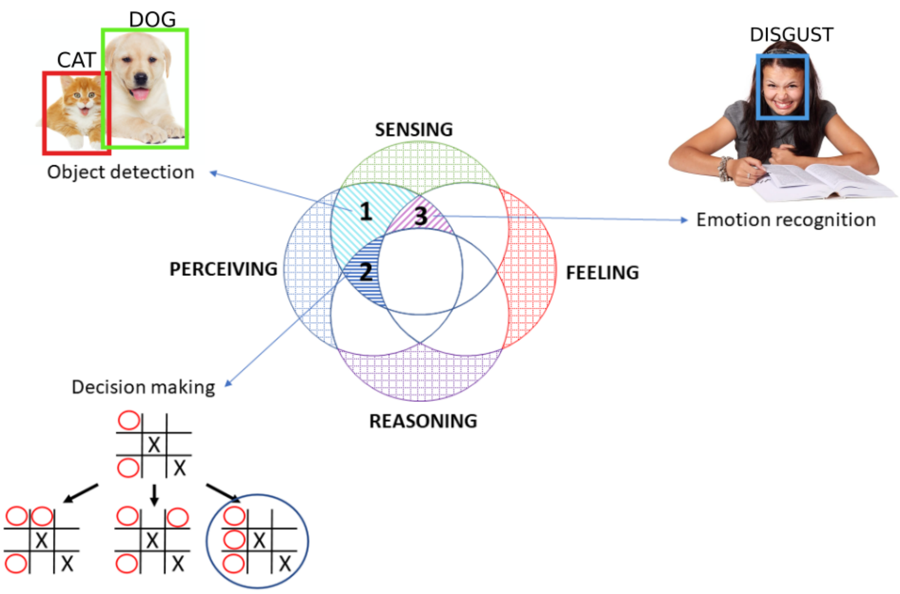

Figure 1.1: Our research reference point.

1.3 The biological sense

The central nervous system (CNS) has a series of systems that make it possible to obtain information about both the environment, the body, and itself. Therefore, the interaction with the environment is carried out through effectors that alter it. Although indeed, it can also be carried out through language in social interaction, it is necessary to point out that these interactions in turn require motor actions to be able to articulate the contents of our thoughts. This is important for the analysis of the functional organization of the CNS since in some way it must reflect this importance in its structure. For instance, main regions of the cortex and their associated high-level functions together with the cerebellum are shown in figure 1.2. Human high level capabilities are distributed over the brain structure leading to a highly emerging system of parallel dynamic neural behaviors.

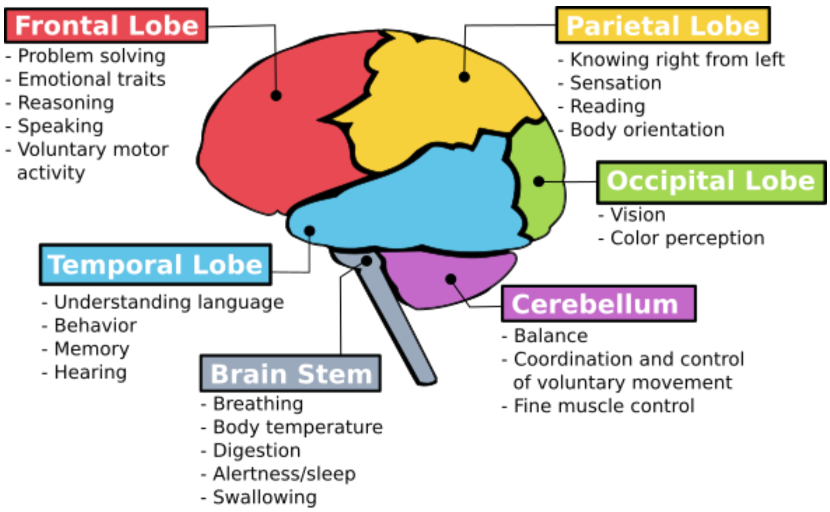

Figure 1.2: The autonomous nervous system.

From the architectural perspective of the brain, figure 1.3 shows how different structures have evolved to cope with tasks that must be performed autonomously or consciously.

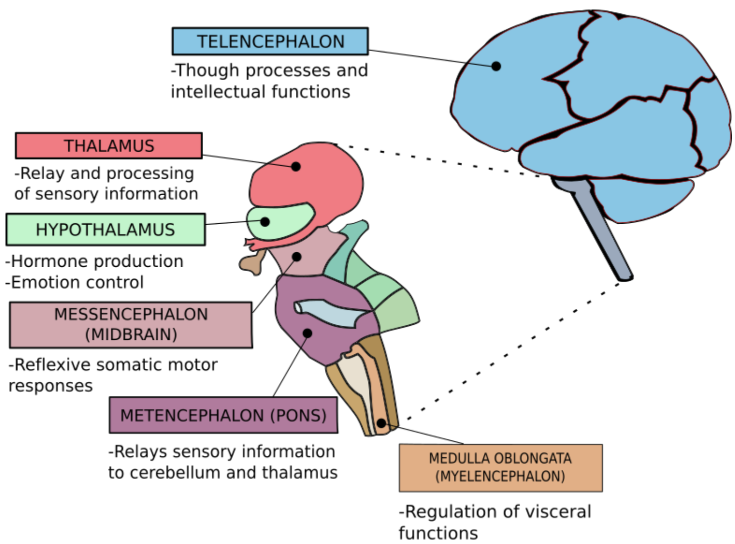

Figure 1.3: Major regions of the mammalian brains.

Hence, the brain stem is in charge of most of the autonomous functions such as relay and processing of sensory information, hormone production, emotion control, generation of reflexive somatic motor responses and regulation of visceral functions such as cardiovascular activity, while the telencephalon plays the role of producing the consciousness, thought processes and intellectual functions such as memory storage, multi-modal processing, or emotion management.

Finally, in 1.4 a topological view of interactions between parasympathetic and sympathetic systems is shown. In 1994, Dr. Stephen Porges introduced the Polyvagal Theory [Porges, 1995]. Porges described an interaction between the central nervous system and source nuclei in the brain stem. This interaction involves pathways in the autonomic nervous system that connect somatomotor components to visceral efferent pathways (that regulate muscles of the face) and to visceromotor components (that regulate the heart, among others). A model is shown in figure 1.4 describing how the limbic structure interacts balancing suppressions and activations of both, the parasympathetic and sympathetic circuits. As a result, the output of the sympathetic and parasympathetic autonomous systems produce variations in these systems.

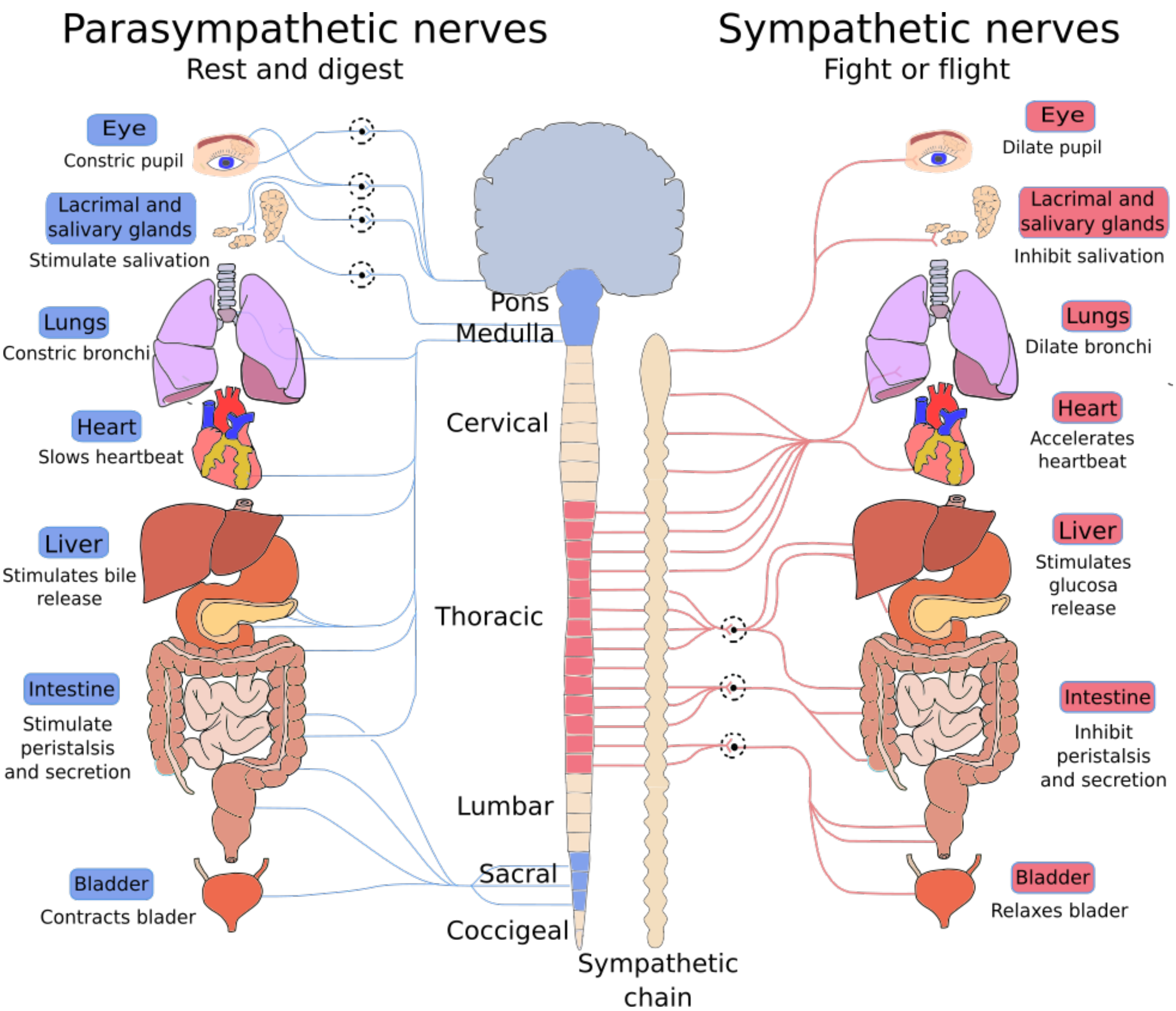

Figure 1.4: The sympathetic and parasympathetic nervous system.

1.4 The biological model abstraction

It is necessary for the study to justify an abstract biological model from which the mathematical model will emerge. As shown above, there is a highly interconnected and structured central nervous system. The abstract interactions are shown in figure 1.5. From this biological model, our assumption is based on the idea that emotional cues can be measured by signals processed both in the cortex and through physiological signals directly influenced by the parasympathetic and sympathetic systems.

The regulation of emotions involves a series of central nervous subsystems that interact with each other to produce complex behaviors. In this context, behavioral demands induce their coordination to produce changes that allow dynamic adaptation to the latter. Several subsystems intervene in these processes, from high to low levels of nervous activity, which involve close interactions between the central and autonomic nervous systems in different ways. Although the hypothalamus regulates part of the autonomic subsystems, many of the activities of the hypothalamus are, in turn, regulated by certain cortical areas, as well as by the central nucleus of the amygdala, which processes inputs from the external environment. The amygdala comprises several nuclei on the medial aspect of the temporal lobe, mainly the anterior hippocampus and the tip of the temporal horn [Téllez-Zenteno and Hernández-Ronquillo, 2012]. The amygdala receives inputs from the association cortex and its main projections are towards the septal area and the prefrontal cortex, mediating from emotional responses to sensory stimuli [Strominger et al., 2012].

Emotions may provide rapid and reliable responses to recurrent life challenges and therefore, as a result of these synergic interactions throughout the CNS, see figure 1.5, respiratory and electrodermal activity together with electroencephalographic measurements and facial expressions can provide the necessary information on the processing of emotions [Damasio, 1998, Hagemann et al., 2003].

1.4.1 Electroencephalography

Electroencephalography (EEG) measures the electrical changes produced by the neural populations under the scalp and skull, related to the brain activity that occurs in the upper layers of the cerebral cortex. The synchronization and desynchronization of these neural populations produce variations that can be measured by electrodes located on the scalp. What the EEG measures are the oscillation of the slow fields of the neuron populations [Buzsáki et al., 2012]. When a single neuron fires, transmitting a signal from the presynaptic to the postsynaptic neuron, the action potential of the former produces the release of glutamate which, in turn, causes glutamate binding in the latter, producing a change in voltage called excitatory post-synaptic potential (EPSP). For a large population of neurons, under synchronized activity, these EPSPs add up to stronger signal measurements. An electrode located in the occipital lobe, which belongs to the area of the visual cortex, is capable of measuring alpha-synchronized brain activity when the eyes are closed while complex information processing is not being performed. On the contrary, when the eyes are opened, a complex and unsynchronized signal is produced. Therefore, brain activity can be considered with Claude Shannon's information theory [Shannon, 1948], where the highly synchronized activity is related to low information processing, so that the more synchronized the neurons in the brain are, the fewer data processing takes place Singer [1993]. Such a property has made it possible to measure several synchronized events, such as epileptic seizures [Saab and Gotman, 2005, Ocak, 2009, Mirzaei

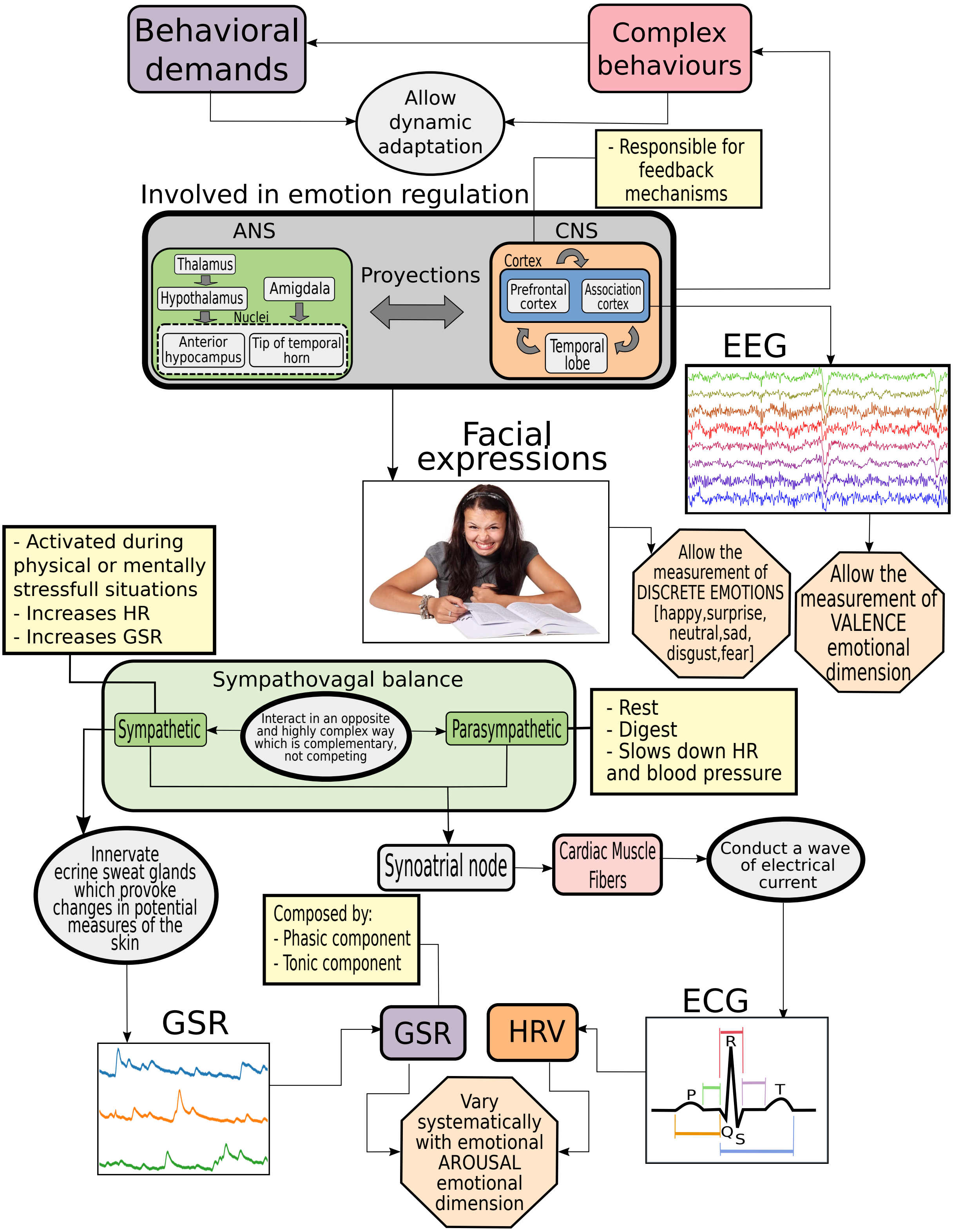

Figure 1.5: The nervous system integration model.

et al., 2010, Kumar et al., 2014, Al Ghayab et al., 2019], which caused large populations of neurons to start firing rapidly and at the same time. In that context, the brain uses a set of oscillatory rhythms [Berger, 1929, Dement and Kleitman, 1957, Foster et al., 2017] to allow parallel communication across neuronal structures. These rhythms can generally be characterized in the following set of frequency bands, from slowest to fastest: delta $0.1 - 4Hz$, theta $4 - 8Hz$, alpha $8 - 15Hz$, beta $15 - 30Hz$, and gamma $> 30Hz$.

1.4.2 Blood volume pressure

The blood volume pressure (BVP) is obtained through a photoplethysmogram that obtains the volumetric changes in the superficial blood vessels by illuminating the skin and measuring the changes in light absorption. When a cardiac cycle occurs, the heart pumps blood through the cardiovascular system, distending the arteries and arterioles just behind the skin to measure the pressure pulse. Signal analysis of the BVP requires the introduction of some brief concepts regarding the properties of signal dynamics and their related biological processes. In 1.6, several key points can be seen, each related to significant biological events:

- The diastolic points are the pressure in the blood vessels between heartbeats when the heart is at rest. It is the part of the heart's cycle during which the heart fills with blood. The diastolic points are the key points used for estimating the interval between heartbeats (IBI), also called the RR intervals.
- The systolic points measure the maximum pressure when the ventricles contract, during the emptying part of the heart cycle.
- Dichrotic notch points are related to the transient increase in aortic pressure when the aortic valve is closed, called a dichrotic wave. It can be used as a marker for the end of the systole period.
- The point of the second-wave marks the highest pressure point related to the dichrotic wave.

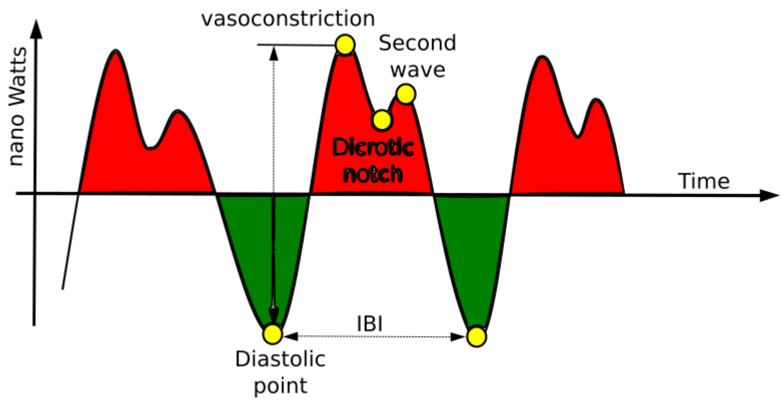

Figure 1.6: An example of a BVP signal. The output may vary depending on user and conditions.

The BVP signal is often used to calculate the heart rate variability (HRV), so the variability of the corresponding RR intervals is calculated. The shape and time dynamics of the BVP waveform reflect arterial changes correlated with interlocking changes in the sympathetic and parasympathetic nervous systems. Heart rate variability has been associated with changes in emotional arousal [Peper et al., 2007, Jönsson, 2007]. On the one hand, excessive emotions such as fear, worry, panic or stress, emotional tension, or high state anxiety, can produce a decrease in HRV [Nickel and Nachreiner, 2003], presumably related to attention focus and motor inhibition [Jönsson, 2007]. Conversely, emotions related to appreciation or love may increase HRV [Petrocchi et al., 2017]. Changes in the range of HRV may reflect, to some extent, the presence of an emotional state. The stronger the emotional engagement, the longer it takes for the signal to return to the baseline. These signals can provide insight into the level of activation of an emotional state, i.e. a trauma or a state of vigilance. Besides, BVP can be measured with the use of wearable devices that have some advantages over electrocardiogram(ECG) systems because they are user friendly and offer easy experimental set-ups.

1.4.3 Galvanic skin response

Galvanic skin response (GSR) measures the differential potential of the skin produced by changes in the activity of the sympathetic system, which in turn innervate the eccrine, apocrine, and apoecrine sweat glands [Donadio et al., 2019]. The first ones participate in the process of regulation of emotions and reflect changes in emotional arousal [Caruelle et al., 2019], producing changes in the potential measurements of the skin. The sweat glands on the palmar and plantar surfaces may have evolved to increase sensitivity and grip in some circumstances and are therefore related to psychological stimuli [Edelberg, 1972, Boucsein, 2012], thus allowing the quantification of the level of arousal during cognitive processes and the processing of emotions.

The conductance of the skin is composed of the phasic and tonic components. While the former is related to rapid changes (aproximately $1-3per/min$ during rest and $20per/min$ in high arousal states) in the differential potential of the skin, the latter represents changes on a slower time scale (from seconds to minutes).

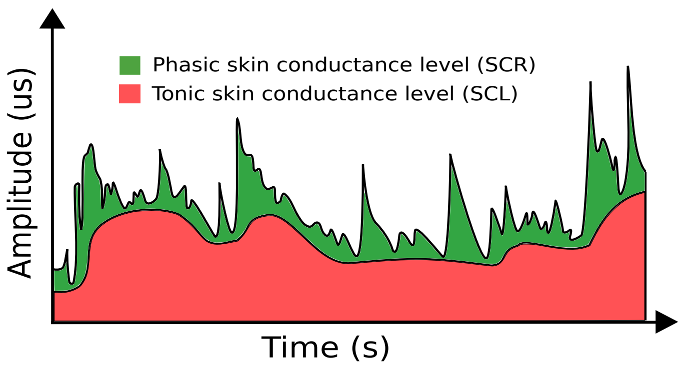

Figure 1.7: Electrodermal activity. Phasic and tonic components visualization.

The phase component is therefore the reflection of the sympathetic system for environmental stimuli in the short term, while the tonic component shows slow changes. Both are directly related to the level of arousal.

1.4.4 Facial expressions

Charles Darwin wrote at the end of the 19th century "The expression of the emotions in man and animals" Darwin [1872] where he stated that some basic emotions remain universal across some species. This original research opened a field that has been extensively studied in various scientific disciplines. Then, the study of facial expressions received little attention because of the lack of a proper tool to measure facial expressions and, as a result, the prevailing view was that facial expressions could not provide accurate information regarding emotion [Landis, 1924, Frois-Wittman, 1930, Bruner and Tagiuri, 1954]. Later, in the 20th century, Paul Ekman followed this research [Ekman and Friesen, 1971, Ekman et al., 1988, Ekman, 1997] and stated that among other emotional expressions, there are seven universal expressions of discrete emotions shared across cultures: anger, disgust, fear, happiness, sadness, and surprise; the neutral emotion is defined as a baseline.

The main issues regarding facial expressions are related to the ability of human beings to inhibit emotional expressions in contrast with other affective phenomena but still raise emotion physiology [Ekman, 1992] because of the contribution to emotion processing and facial muscles motor control of the amygdala [Livneh et al., 2012, Gothard, 2014].

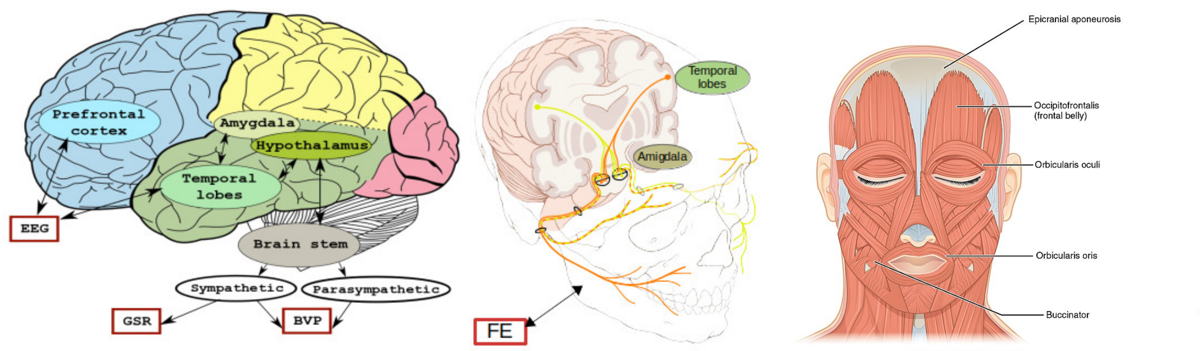

Figure 1.8: Facial muscles.

1.5 Emotional models

Different factors make the recognition of emotions a difficult task [Mi et al., 2019] and despite the variability of emotional responses and their dependence on culture [Jack et al., 2012], it is necessary to narrow down the reference system for experimental research by creating an emotional model. On the one hand, there is no basic truth for self-evaluation, since the assessment of experienced emotions is guided by emotional models developed in the field of psychology. These can generally be grouped as discrete and dimensional models. The former assumes that emotions are qualitatively differentiated neurophysiological responses that produce independent emotional experiences [Roseman et al., 1990]. Following Ekman's [Ekman, 1976] research, emotions are generally divided into six prototypical categories: anger, disgust, fear, happiness, sadness, and surprise; plus the neutral emotion is defined as a baseline. In contrast, the dimensional approach consisting on a dimensional model with two orthogonal dimensions, valence and arousal, captures the continuous quantified relationships between emotions [Russell, 1980]. The valence dimension reflects the emotional reactions related to pleasantness or unpleasantness of the evoked reaction while the arousal dimension measures the level of excitement. Furthermore, there are significant variations for each individual in the correlation between the properties of the measured physiological signals and the respective emotion, therefore, studies differ in methodology to attempt to better bridge this gap. Figure 1.9 shows graphically the relation between both models.

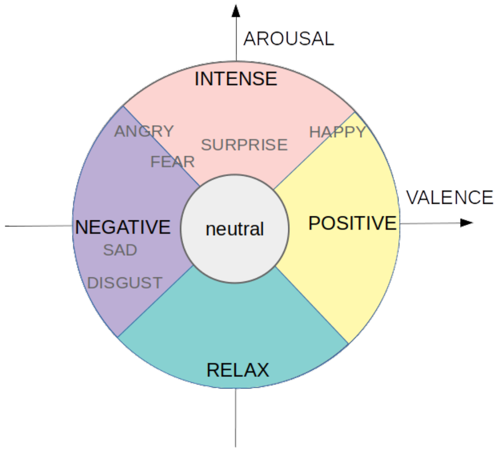

Figure 1.9: The valence-arousal dimensional model.

1.6 State of the art in emotion estimation

The estimation of emotions can be achieved through different strategies and sources of information. On the one hand, the hand-crafted features can be computed using a standard machine learning approach, in which the goal is to find a set of relevant invariant features that allow for robust estimation. On the other hand, deep learning strategies have received special attention because of the capacity for generalization they achieve when fed with sufficient data. Besides, the use of an appropriate stimulus must be carried out since it is an important factor concerning the strategy to be performed. Standard automatic learning algorithms do not require a large amount of data as deep learning techniques do, and the computation requirements on each strategy must be considered when real-time constraints are imposed. Finally, it is also necessary to take into account the properties of the stimulus, static or dynamic, and the physiological effect it generates in humans.

The latest advances in the field of emotion recognition have been made about new techniques developed in the field of deep learning. Their use has outperformed conventional approaches using traditional feature extraction techniques. These new advances include the use of temporal and spatial information of brain activity models based on deep learning [Zheng and Lu, 2015, Zheng et al., 2019, Yin et al., 2017, Tripathi et al., 2017, Nguyen et al., 2018, Zheng et al., 2018, Song et al., 2018], deep convolutional neural networks [Santamaria-Granados et al., 2018, Chen et al., 2019] and Long-Short-Term-Memory (LSTM) based architectures [Chen and Jin, 2015]. Another approach, that is often used, is based on network ensembles [Gjoreski et al., 2018], in which information is fed through a set of deep architectures that must converge at some point using a score fusion strategy. Although they generally achieve better performance, since each network extracts different relevant features and may lead to minimizing generalization error, they are computationally heavy. Other architectures take advantage of splitting the feature extraction step for each information source while having a multi-modal feature representation layer just before classification takes place. Also, novel techniques such as domain adversarial learning [Chai et al., 2016] or attention mechanisms have been proposed recently. If the main objective was to find a methodology for emotion estimation, researchers have already developed methods that optimize that process and even achieve estimation in a subject-independent paradigm [Lan et al., 2018, Li et al., 2019] that allows the discovery of latent and shared components among subjects.

Although deep learning has generally outgrown most standard machine learning methods, it cannot still be a general approach to users, and such networks are very time-consuming in both hyper-parameter and parameter tuning. Consequently, deep learning methods remain user-dependent and require a huge training and tuning effort that is not desirable for an affective human-robot interaction (affective-HRI) system to be adopted in practical situations. Furthermore, most previous research studies do not take into account the real-time limitations, except for [Liu et al., 2017a]. This requirement imposes constraints on computing resources that generally cannot be supported by lightweight or wearable devices when processing deep learning models, but which can be better adapted using standard machine learning methods.

With respect to HRI experimental designs, most studies rely on speech analysis [James et al., 2018, Lakomkin et al., 2018, Pan, 2018, Tsiourti et al., 2019] and facial expressions [Liu et al., 2017b, Chen et al., 2017, Faria et al., 2017b,a, Chen et al., 2018b,a, Benamara et al., 2019, 2020] as the sole source. Other novel approaches consider multi-modal sources of information [De Silva and Ng, 2000, Jung et al., 2004, Castellano et al., 2013, Cid et al., 2015, Perez-Gaspar et al., 2016] to better accomplish the task at hand, based on the assumption that correlations between signals related to the triggered emotions would help the algorithms to improve their accuracy. Finally, only a few have considered the use of physiological signals Kulic and Croft [2003], Liu et al. [2006], Kulic and Croft [2007], Maaoui and Pruski [2010].

Bearing this in mind and taking into account that this research has to cover a large set of items, a decision has been made to use standard machine learning methods with the use of carefully and meaningfully chosen features to address the task of emotion recognition through physiological signals, while an ensemble deep learning method for emotion estimation in facial expressions has been developed through a research collaboration with Benamara et al. [2019, 2020]. Therefore, the development of such powerful deep learning techniques for physiological signal processing is beyond the scope of this research study. Hence, the current research aims to demonstrate that affective-HRI is possible with the use of inexpensive, wearable devices, in collaboration with automatic and easy-to-tune signal processing techniques.

Chapter 2

Objectives

The main objective of this thesis is to build a methodology for the estimation of emotions in real-time under an affective human-robot interaction (affective-HRI) paradigm. To develop such a methodology, several sources of information have been taken into account concerning their physiological relationships on the processing of emotions in the CNS and the autonomous nervous system. Thus, this process must address several issues regarding the properties of each signal, taking into account not only real-time limitations but also experimental design decisions for human-robot interaction. A decision was made based on the assumptions that EEG, GSR, and BVP are physiological signals closely related to the processing of emotions. As explained above, these signals reflect measurable involuntary emotional reactions that, as a result, are very difficult to socially mask during the emotional experience [Kim et al., 2004, Kim and André, 2008]. Although not as reliable, facial expressions are used for non-verbal and affective communication between human beings, and therefore, have been considered for this research study. Such an approach has been addressed by considering a set of items:

1. EEG signals are considered a reliable source for the estimation of valence emotion but are widely affected by artifacts. Therefore, an artifact removal technique that processes EEG signals in real-time must be developed.

2. Valence emotion estimation must be done by EEG signals using a light methodology in terms of computing resources. Besides, it must be robust in terms of classification accuracy using a low density of electrodes, so the location over the brain areas is crucial. Also, the features must be carefully selected to reflect the differences in brain activity through the perceived emotions.

3. GSR and BVP signals are considered a reliable source for the estimation of arousal emotional dimension. The methodology should address noisy artifacts, in particular BVP signals, which are highly sensitive to wrist movements. The acquisition of these signals can be done with a low cost and wearable device.

4. For the estimation of facial expressions, several incomes have to be faced. Firstly, facial detection must be carried out taking into account different scenarios such as variations in brightness, occlusions, or body postures, which may affect the information available to recognize them. Also, most databases have been labeled by humans, so they contain an inherent bias due to the complexity of emotion perception.

5. A realistic scenario, or at least close to it, should be designed to adequately evaluate the proposed methodology under an affective-HRI paradigm.

6. An appropriate statistical evaluation method must be performed, standard cross-validation is not suitable for time series, instead, "leave-one-out" validation strategies must be used.

The first article that constitutes this thesis [Val-Calvo et al., 2019a] covers items one and two, an optimized methodology covering EEG signal acquisition, artifact removal, and real-time processing with the desired objective of valence emotion estimation. This research study was conducted using a well-defined and validated database for the field of EEG emotion estimation, the SEED database [Duan et al., 2013, Zheng and Lu, 2015]. The third item was achieved in the second article [Val-Calvo et al., 2020a], which merges three different sources of information, EEG, GSR, and BVP, to include arousal emotional dimension estimation into the previously designed methodology. Meanwhile, the experimental design went beyond a more realistic scenario by using a dramatic film that is argued to modulate emotional changes dynamically. The fourth item has been achieved in a collaboration with Benamara et al. [2019, 2020] by developing a real-time face detection and facial expression recognition, based on an ensemble deep learning approach and evaluated in several databases of facial expressions, FER-2013 [Goodfellow et al., 2013], SFEW 2.0 [Dhall et al., 2011] and ExpW [Zhang et al., 2017]. Finally, the fifth item has been achieved in the last article in this thesis [Val-Calvo et al., 2020b], which merges both sources, physiological signals and facial expressions, for the task of emotion estimation. A customized experimental design has been made based on the standard strategies of emotion stimulation used in the research of the understanding and measurement of emotions, but combined in an interesting way to provide an almost realistic scenario in the field of affective-HRI. As for the use of appropriate statistical evaluation, the sixth item has been taken into account as an essential requirement for all the articles that make up this thesis.

As part of the development necessary to achieve this objectives and within the work required for the articles that make up the thesis, three software applications have also been implemented, Biosignals [Val-Calvo, 2020a], GEERT [Val-Calvo, 2020b], and GePHY-CAM [Val-Calvo, 2020c]. These are further explained in more detail in the material and methods part in the following chapter.

Chapter 3

Material and methods

3.1 Hardware

The following devices have been used for the affective HRI:

1. Empatica E4 wristband by a Massachusetts Institute of Technology (MIT) spin-off.

2. OpenBCI hardware for the EEG acquisition system.

3. Pepper robot from Aldebaran robotics.

3.1.1 Empatica E4

The E4 is a medical-grade wearable device that offers real-time physiological data acquisition. It measures blood volume pulse, from which heart rate variability can be derived, captures motion-based activity with a three-axis accelerometer, measures the constantly fluctuating changes in certain electrical properties of the skin (GSR), and reads peripheral skin temperature. It allows the synchronization of such signals by the use of an internal clock. The GSR and temperature sensors have a sampling rate of $4Hz$ while the accelerometers and the BVP signal are measured at $64Hz$. The combination of the GSR and BVP sensors enables to simultaneously measure the balance of sympathetic nervous system activity and heart rate. More detailed information can be found at https://www.empatica.com/en-eu/research/e4/.

3.1.2 OpenBCI hardware for the EEG acquisition system

The OpenBCI system is a collection of headsets, boards, sensors, and electrodes that allow the research on brain-computer interfacing to purchase high-quality equipment at affordable prices. In this research the OpenBCI Cyton Board, an Arduino-compatible 8-channel neural interface with a 32-bit processor, is used. At its core, the OpenBCI Cyton Board implements the PIC32MX250F128B microcontroller, giving it lots of local memory and fast processing speeds. The board comes pre-flashed with the chipKIT™ bootloader, and the latest OpenBCI firmware. Data is sampled at 250Hz on each of the eight channels.

The OpenBCI Cyton Board can be used to sample brain activity (EEG), muscle activity (EMG), and heart activity (ECG). The board communicates wirelessly to a computer via the OpenBCI USB dongle using RFDuino radio modules. It can also communicate wirelessly to any mobile device or tablet compatible with Bluetooth Low Energy. More detailed information can be found at https://openbci.com/.

3.1.3 Pepper robot from Aldebaran robotics

Pepper is the world's first social humanoid robot able to recognize faces and basic human emotions. Pepper was optimized for human interaction and can engage with people through conversation and his touch screen. It has 20 degrees of freedom for natural and expressive movements, speech recognition and dialogue available in 15 languages, perception modules to recognize and interact with the person talking to him, touch sensors, LEDs and microphones for multimodal interactions, infrared sensors, bumpers, an inertial unit, 2D and 3D cameras, and sonars for omnidirectional and autonomous navigation while being an open and fully programmable platform. More detailed information can be found at https://www.softbankrobotics.com/emea/en/pepper.

3.2 Software

Wearable devices are tools of particular interest because of the possibilities they offer for brain-computer interface (BCI) related research, such as motor imagery, emotion estimation, or attention-related studies, which could benefit from open source applications. Pragmatically, a project of this type requires technical solutions to be provided. Therefore, a set of applications has been gradually developed to meet the requirements of experimental designs. All the code has been developed with pure python libraries in an object-oriented paradigm. The software engineering process was carried out taking into account cohesion and coupling for proper modularization. This allows the applications to be modular, scalable, and easy to maintain; in fact, this is a key aspect of any scientific tool to allow researchers to make modifications and to fulfill the specific requirements of each experimental scenario.

The philosophy of these applications is based on a supervised machine learning approach and therefore they offer two modes of interaction:

- The first allows for real-time acquisition and processing to generate a database and build models.
- The second provides online signal processing using pre-trained models to classify brain patterns.

The proposed applications are versatile and easily adaptable to different experimental designs while maintaining high-performance real-time signal processing. Aware that pattern recognition is in constant development, these applications offer the option of importing python external scripts, which must have a predefined structure, to include self-developed machine learning methodologies. Besides, as experimental designs require event synchronization, a TCP/IP interface has been provided. These applications are expected to be accessible to the entire scientific community, providing a resourceful tool for experimental paradigms of human behavior, which encompasses the following functionalities:

- Real-time acquisition and visualization of physiological signals.
- Trigger synchronization by a TCP/IP interface that allows start/stop the recordings remotely.

- Recording of data on European Data Format (EDF) for physiological signals.
- Online behavior labeling interface whose labels are synchronized and stored in the EDF files.
- Online signal processing with self-developed methodologies through the possibility of importing external python scripts.

3.2.1 Biosignals application

Title: General BVP, GSR, temperature (TMP) and accelerometers (ACC) experimentation in real-time, an open source software for physiological real-time experimental designs

Authors: Mikel Val Calvo

URL: `https://github.com/mikelval82/Biosignals`

License: GNU General Public License v3.0

DOI: 10.5281/zenodo.3759262

This application has been built to further the research on the study of physiological signals, such as BVP, GSR, temperature (TMP) and accelerometers (ACC), related to human behavior, with the use of the Empatica E4 wristband. Figure 3.1 shows the layout of the BIOSIGNALS interface. Each section has been numbered to properly explain the role of each of the components:

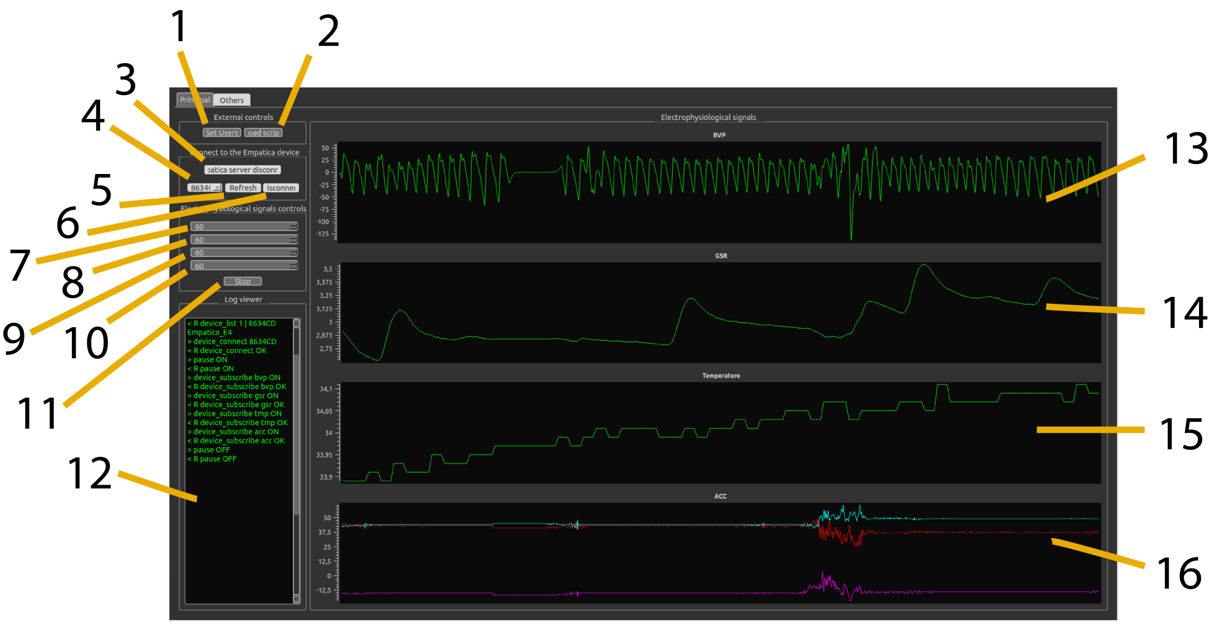

Figure 3.1: BIOSIGNALS: General physiological experimentation in real-time, an open source software for real-time physiological signals acquisition and processing with the Empatica E4 wristband.

1. Set Users: Allows for setting the name of the user file where acquired data will be stored.
2. Load Script: For loading external python scripts developed by users for specific experiments.
3. Empatica Server Connect/Disconnect: Enables the connection and disconnection between the Empatica Server and the Biosignals.
4. Empatica ID: A spin box to specify to which Empatica E4 device to connect.
5. Refresh: Requests to the Empatica Server the ID of Empatica E4 devices available.
6. Connect: Enables the connection and disconnection between the specified device through the Empatica Server and the Biosignals driver.
7. BVP Window size (seconds): A Spin Box to set the window size of the BVP signal.
8. GSR Window size (seconds): A Spin Box to set the window size of the GSR signal.
9. TMP Window size (seconds): A Spin Box to set the window size of the TMP signal.
10. ACC Window size (seconds): A Spin Box to set the window size of the ACC signals.
11. Start/Stop: A button to start/stop real-time visualization.
12. Log viewer: Shows information regarding the internal state of the application.
13. BVP Long Term View: Allows the visualization in real-time of acquired BVP signals.
14. GSR Long Term View: Allows the visualization in real-time of acquired GSR signals.
15. TMP Long Term View: Allows the visualization in real-time of acquired TMP signals.
16. ACC Long Term View: Allows the visualization in real-time of acquired ACC signals.

3.2.2 GEERT application

Title: General EEG experimentation in real-time, an open source software for BCI real-time experimental designs

Authors: Mikel Val Calvo

URL: `https://github.com/mikelval82/GEERT`

License: GNU General Public License v3.0

DOI: 10.5281/zenodo.3759306

This application has been built to further the research on the study of EEG signals related to human behavior, with the use of the OpenBCI system. Figure 3.2 shows the layout of the GEERT interface. Each section has been numbered to properly explain the role of each of the components:

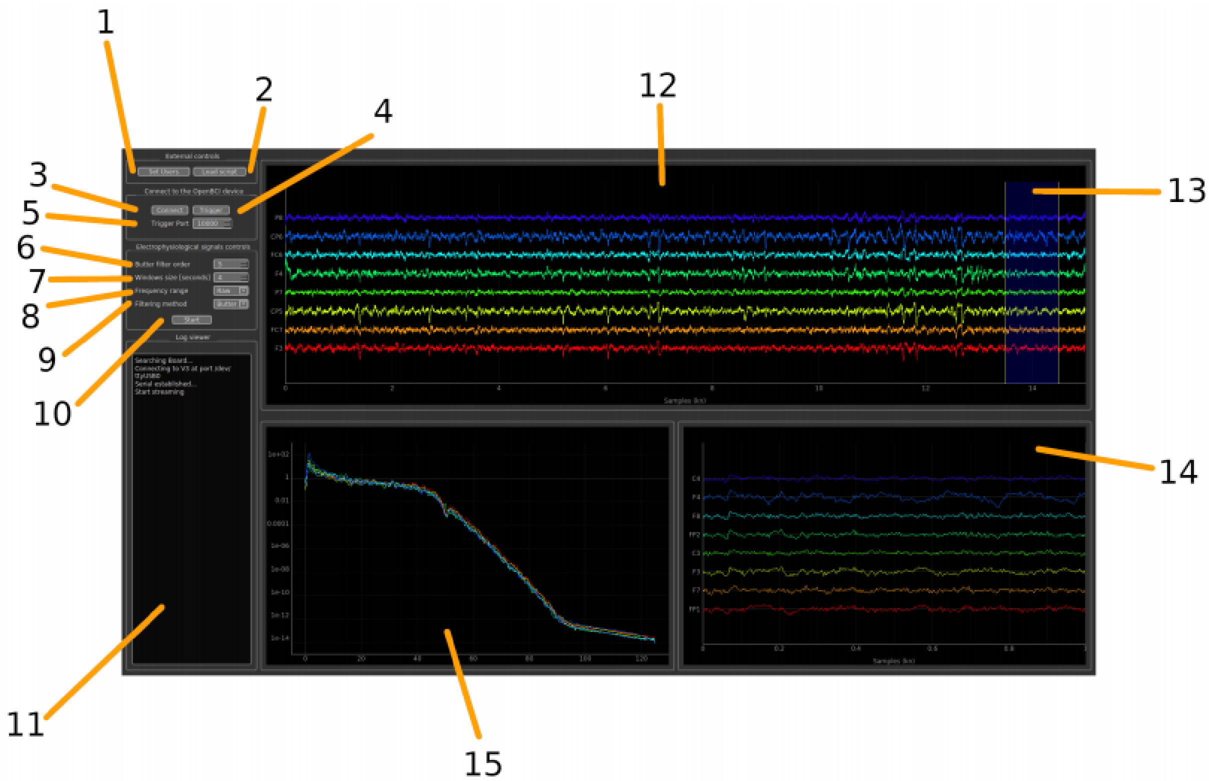

Figure 3.2: GEERT: General EEG experimentation in real-time, an open source software for BCI real-time experimental designs. A novel open source software application for real-time physiological signals acquisition and processing with the OpenBCI.

1. Set Users: Allows for setting the name of the user file where acquired data will be stored.
2. Load Script: For loading external python scripts developed by users for specific experiments.
3. Connect: Enables the connection and disconnection between the OpenBCI driver and GEERT through the corresponding serial device.
4. Trigger: Starts the trigger server for event synchronization.
5. Trigger port: A Spin Box to specify the TCP/IP port.
6. Butter filter order: A Spin Box to set the order of the Butterworth filter.
7. Windows size (seconds): A Spin Box to set the window size of the Short Term View.
8. Frequency range: Sets the frequency range of the Butterworth filter applied to signals.
9. Filtering method: Determines the online artifact removal (OAR) method to be applied.
10. Start: Starts and stops real-time visualization.
11. Log viewer: Shows information regarding the internal state of the application.

12. Long Term View: Allows the visualization in real-time of acquired EEG signals. By default 60 seconds time window is shown.

13. Adjustable Sliding Window: Defines the EEG temporal length and location to be shown in the Short Term View. Size and location can be interactively changed in real-time even during the acquisition.

14. Short Term View: Allows the visualization in real-time of acquired EEG signals falling inside the length and temporal location of the Adjustable Sliding Window.

15. Spectrogram Term View: Allows the visualization of each EEG signals spectrogram in real-time.

3.2.3 GePHYCAM application

Title: GePHYCAM: General electrophysiological and camera acquisition system for human behaviour recordings in real-time

Authors: Mikel Val Calvo

URL: https://github.com/mikelval82/GePHYCAM

License: GNU General Public License v3.0

DOI: 10.5281/zenodo.3727503

This application looks forward to providing a resourceful tool for human-behavior experimental paradigms using several information sources such as EEG, GSR, BVP, TMP, ACC, and VIDEO. Figure 3.3 shows the layout of the GePHYCAM interface. Each section has been numbered to properly explain the role of each of the components:

1. Save: Allows for setting the name of the user files where acquired data will be stored.

2. Script: For loading external python scripts developed by users for specific experiments.

3. Trigger: Starts the trigger server for event synchronization.

4. Trigger PORT: A Spin Box to specify the PORT to be used.

5. Trigger IP: A Spin Box to specify the IP direction to be used.

6. Start: Starts and stops real-time visualization.

7. Empatica Server Connect/Disconnect: Enables the connection and disconnection between the Empatica Server and the GePHYCAM.

8. Refresh: Requests to the Empatica Server the ID of Empatica E4 devices available.

9. Empatica ID: A spin box to specify to which Empatica E4 device to connect.

10. E4: Enables the connection and disconnection between the specified device through the Empatica Server and the GePHYCAM E4 driver.

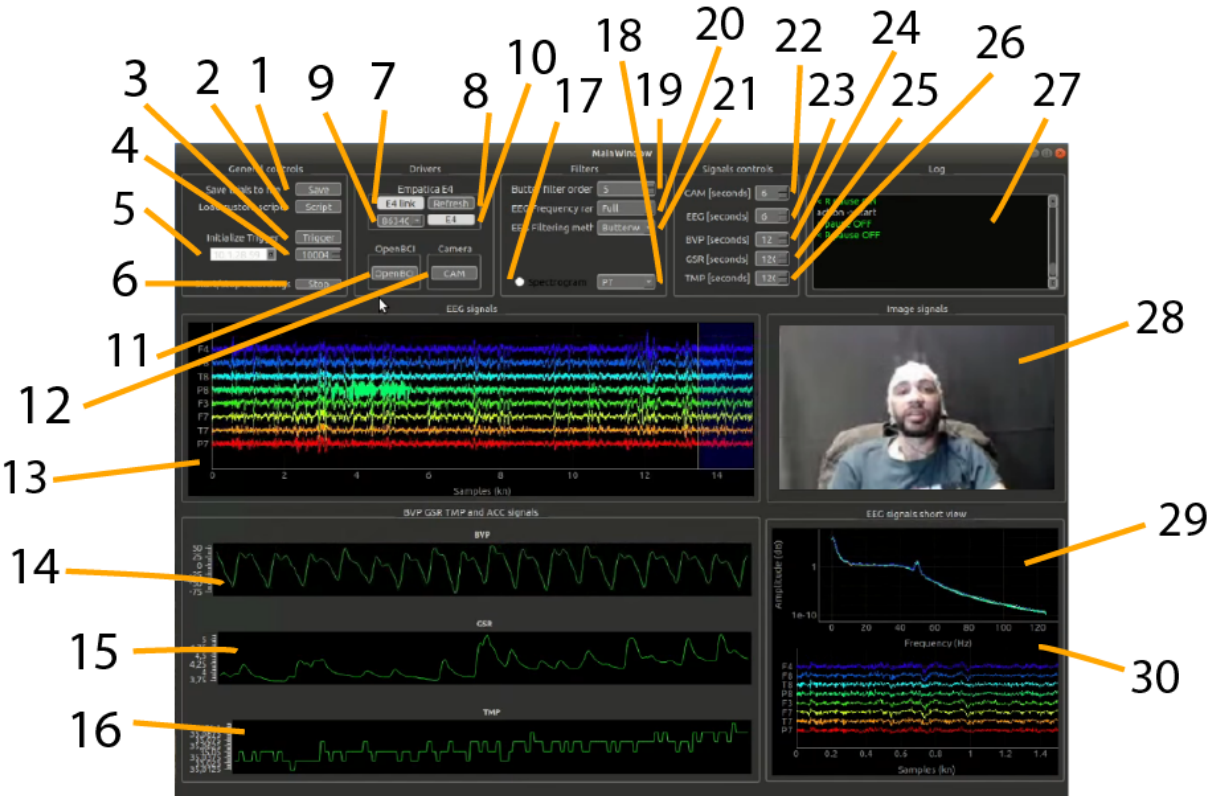

Figure 3.3: GePHYCAM: General electrophysiological and camera acquisition system for human behaviour recordings in real-time.

11. OpenBCI: Enables the connection and disconnection between the OpenBCI and the GePHYCAM BCI driver through the corresponding serial device.

12. CAM: Enables the connection and disconnection between the webcam and the GePHYCAM CAM driver.

13. Long Term View: Allows the visualization in real-time of acquired EEG signals.

14. BVP Long Term View: Allows the visualization in real-time of acquired BVP signals.

15. GSR Long Term View: Allows the visualization in real-time of acquired GSR signals.

16. TMP Long Term View: Allows the visualization in real-time of acquired TMP signals.

17. Spectrogram: Activates/deactivates the visualization of the spectrogram in the spectrogram term view. When it is activated the spectrogram is shown, otherwise, power spectral density function is shown for each EEG channel.

18. Channel_ID: A spin box that selects the EEG channel to be shown for the spectrogram.

19. Butter filter order: A Spin Box to set the order of the Butterworth filter.

20. Frequency range: Sets the frequency range of the Butterworth filter applied to signals.

21. Filtering method: Determines the EEG online artifact removal (OAR) method to be applied.

22. CAM Window size (seconds): A Spin Box to set the window size of the webcam video signal.

23. EEG Window size (seconds): A Spin Box to set the window size of the EEG signals.

24. BVP Window size (seconds): A Spin Box to set the window size of the BVP signal.

25. GSR Window size (seconds): A Spin Box to set the window size of the GSR signal.

26. TMP Window size (seconds): A Spin Box to set the window size of the TMP signal.

27. Log viewer: Shows information regarding the internal state of the application.

28. Webcam viewer: Allows the visualization in real-time of acquired webcam signal.

29. Spectrogram Term View: Allows the visualization of each EEG channel spectrogram in real-time.

30. Short Term View: Allows the visualization in real-time of acquired EEG signals falling inside the length and temporal location of the adjustable sliding window.

3.2.4 Software design principles

As mentioned before, the design principles have been followed taking into account cohesion and coupling for proper modularization. All of the aforementioned applications follow the same structural design approach and share the core structure so it would be redundant a specific explanation for each of them. Therefore, a high-level description is provided that allows us to understand the design principles of each application. As can be noted in figure 3.4 where the high-level flow diagram scheme is detailed, one of the key points of the design is focused on drivers and the data managers. Hence, each device has its driver specification which in turn is assigned a specific data manager. That means, BIOSIGNALS and GEERT have only one driver each but for GePHYCAM three drivers are needed running in parallel with their corresponding data managers. As a result, each application requires a specific graphical interface definition but overall the logic is versatile, easy to maintain, and scalable.

- APP: This class serves as the main software module that builds the application. It has been designed using PyQt5, a Python binding of the cross-platform graphical user interface (GUI) toolkit Qt, implemented as a Python plug-in. It launches all the other modules and forks several processes. First, the APP main fork. Second, a Driver is forked for each device, using the multiprocessing python library. Finally, it forks a thread for each Data Manager, using the multithreading python library.

- Constants: Sets the whole constants used along with the application. Thus facilitating the specification of such general variables in a unique module. All adjustable parameters are set in this object.

- Queue: This Queue follows the FIFO rule.

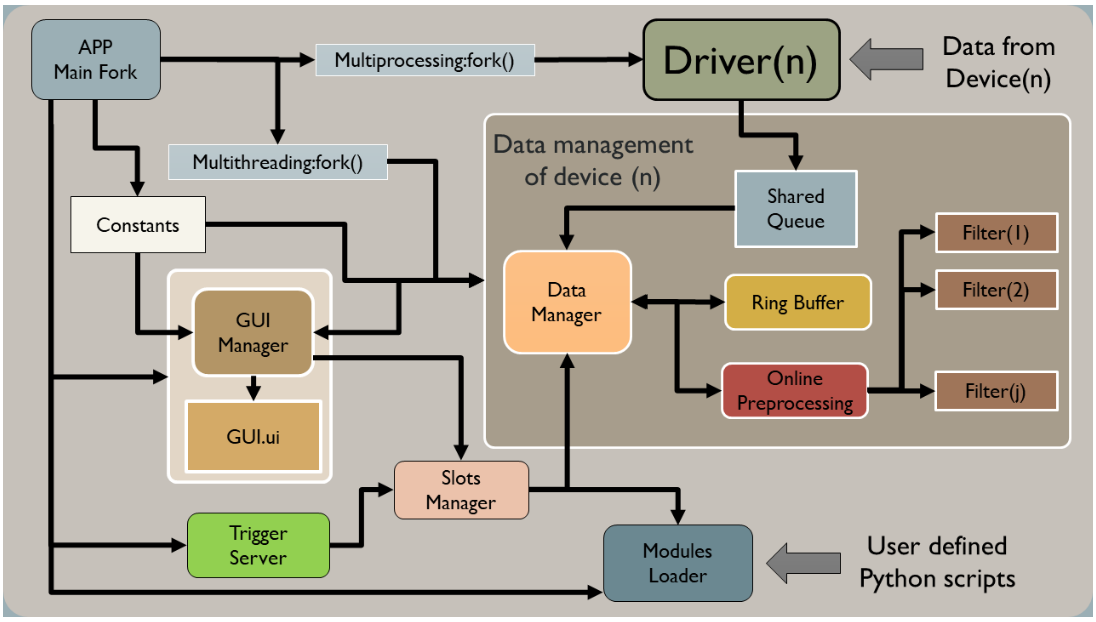

Figure 3.4: Flow diagram.

- Ring Buffer: Stores the last values for each sampled source of information. The fixed size is defined by default in the Constants but it can also be dynamically modified during the GUI interaction through the 'Window size' Spin Box. Each time the fixed size is modified, the Ring Buffer is emptied.

- Driver: Meets the requirements of the real-time acquisition. It inherits from multiprocessing python library. It offers the interface between the APP and the hardware. Inter-process communication is facilitated by Queue. Once the fork starts running, iterates acquiring samples that are indefinitely queued.

- Data Manager: It is defined as an interface between the GUI and the Driver to properly separate the management of the acquired data. It inherits from a multithreading python library. It has access to the shared Queue so its role consist on extracting iteratively samples from the queue to insert them on a Ring Buffer.

- GUI Manager: Offers the management of each of the components of the GUI.

- GUI.ui: Implements the APP graphics. It has been designed using QtDesigner, the Qt tool for designing and building GUIs.

- Trigger Server: A TCP/IP server for event synchronization. It can receive client connections and handles each request by notifying the Slots Manager using a callback function.

- Slots Manager: The purpose of this module is to bring the application to the versatility of adding a set of callbacks that allow synchronizing the overall logic.

- Online Filtering: This module has been implemented as an interface for a set of real-time signal processing methods.

- Modules Loader: The main purpose of this module is to expand the versatility of the APP for a wide range of research scenarios. It is required that users code the scripts with a predefined structure as shown in 3.1. The module will receive a reference to the APP instance so total access to the underlying object instances is offered.

Algorithm 3.1 Imported python scripts required structure.

```
# -- Add imports --
# import numpy as np

# -- Class definition --
class pipeline():
  ''' Predefined structure.'''
  def __init__(self, app):
      # -- initialize variables --
      self.log = app.log

  def run(self):
      # -- code desired behaviour --
      self.log.update_text('Start␣computing␣the␣user␣module.')
```

Main workflow scheme: Once an app is launched it can work as follows; Data Manager is threaded and starts iteratively trying to extract values from the queue. The thread runs a loop that is controlled by a logical condition. The iteration process cannot acquire data if the queue is empty. Then, the Driver is forked and automatically detects the corresponding device to connect with. Following, the driver starts iteratively stacking data into the Queue from the hardware. Both the thread and the fork form a consumer-producer system. The Ring Buffer acts as a short term memory. The Data Manager fills the Ring Buffer each time a new value is obtained from the Queue. Once the Ring Buffer is full, it starts overwriting the oldest data and so on. At this point, two stages are provided:

- Stage 1: Real-time visualization can be performed by pushing the start button. It does not imply user data storage. The start button has been designed to test the quality of the signal acquisition. The start button state will be changed, so pushing it again will stop visualization. Parameter settings (user filename, filter order, window size, online artifact removal method, and so on) should be done on this stage.

- Stage 2: Once the signals are properly acquired, the Trigger button can be pushed that initializes a TCP/IP server. The application only stores data in between a pair of client received requests. It is expected that one request indicates the start instant for event synchronization while the second, the stop instant. For the following, two pairs of events form a trial. During the trial duration, the Ring Buffer slots a call-back to the Data Manager whenever an amount of data equal to the window size is fulfilled, that is, a sample is generated. At the end of the event, the Data Manager permanently stores acquired samples in the users' file with its corresponding meta-data. Data Manager counts the number of samples generated and the number of trials, other metadata values are obtained from the Constants module.

Versatile workflow scheme Modules Loader allows the definition of own produced scripts. Each imported script must follow a predefined structure. The versatility comes due to the reference given to the imported script to the main object instance. Thus, different scenarios could be provided through this component.

- Offline scenario: The user could experiment with the main workflow scheme. After collecting the required data, a script could be loaded containing the coding for an offline analysis, i.e., training a machine learning model and store it on disk for future use.

- Online scenario: The user, could perform online computation thanks to the reference given, having access to all the underlying object instances. Access to the Data Manager to acquire EEG samples online is therefore provided. Moreover, applying OAR methods or even using pre-trained machine-learning models for online predictions can be also easily implemented.

The main highlights of the applications consist of its versatility: First, modularizing the interaction between the GUI manager and the Driver, which allows the embedding of drivers for other acquisition systems. Second, the separation between the GUI and the logic, where the former allows future modifications and performance of the GUI components, without altering the logic, and vice-versa; Finally, the Modules Loader, further expands the versatility, by offering the flexibility of incorporating self-developed scripts for specific scenarios and workflow schemes. The proposed applications are versatile and easily adaptable to different experimental scenarios while maintaining high-performance signal processing in real-time.

Chapter 4

Results

4.1 Article 1: Real-Time EEG Emotion Estimation

Title: Optimization of Real-Time EEG Artifact Removal and Emotion Estimation for Human-Robot Interaction Applications

Authors: Mikel Val-Calvo, José R. Álvarez-Sánchez, Jose M. Ferrández-Vicente and Eduardo Fernández

Journal: Frontiers in Computational Neuroscience, volume 13 article 80.

Impact factor: 2.535, JCR2019-Q2, 16/59 Mathematical & Computational Biology

Published: 26 November 2019

DOI: 10.3389/fncom.2019.00080

This paper covers items 1, 2, and 6 of the objectives. An optimized methodology covering EEG signal acquisition, artifact removal, and real-time processing with the desired objective of valence emotion estimation. In this experiment, an optimization process has been performed for real-time processing requirements which allow us to partially denoise EEG signals while also being acute in the estimation of the emotion, using standard machine-learning algorithms and a set of meaningful features. To develop such a methodology, the SEED database [Zheng and Lu, 2015, Duan et al., 2013] has been used which evoked three discrete emotions with regard to the valence dimension. First, both methods, EAW-ICA [Mammone and Morabito, 2014] and ICAW [Mahajan and Morshed, 2015] have been compared and modified for online removal of ocular artifacts. Second, a classification approach has been proposed using a set of features related to the complexity of the signals, and Savitzky-Golay (SG) filtering method is proposed as an alternative to Linear Dynamic Systems (LDS). SG smoothing is significantly faster than LDS and even improves the accuracy reports.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 26 November 2019
doi: 10.3389/fncom.2019.00080

Optimization of Real-Time EEG Artifact Removal and Emotion Estimation for Human-Robot Interaction Applications

Mikel Val-Calvo [1,2], José R. Álvarez-Sánchez [2]*, Jose M. Ferrández-Vicente [1] and Eduardo Fernández [3,4]*

[1] Departamento Electrónica, Tecnología de Computadoras y Proyectos, Universidad Politécnica de Cartagena, Cartagena, Spain, [2] Departamento de Inteligencia Artificial, UNED, Madrid, Spain, [3] CIBER-BBN, Madrid, Spain, [4] Instituto de Bioingeniería, Universidad Miguel Hernández, Alicante, Spain

Affective human-robot interaction requires lightweight software and cheap wearable devices that could further this field. However, the estimation of emotions in real-time poses a problem that has not yet been optimized. An optimization is proposed for the emotion estimation methodology including artifact removal, feature extraction, feature smoothing, and brain pattern classification. The challenge of filtering artifacts and extracting features, while reducing processing time and maintaining high accuracy results, is attempted in this work. First, two different approaches for real-time electro-oculographic artifact removal techniques are tested and compared in terms of loss of information and processing time. Second, an emotion estimation methodology is proposed based on a set of stable and meaningful features, a carefully chosen set of electrodes, and the smoothing of the feature space. The methodology has proved to perform on real-time constraints while maintaining high accuracy on emotion estimation on the SEED database, both under subject dependent and subject independent paradigms, to test the methodology on a discrete emotional model with three affective states.

Keywords: real-time, EEG, artifact removal, emotion estimation, HRI

OPEN ACCESS

Edited by:
German Castellanos-Dominguez, National University of Colombia at Manizales, Colombia

Reviewed by:
Pablo Varona, Autonomous University of Madrid, Spain
Xavier Vrijdag, The University of Auckland, New Zealand

****Correspondence:***
Mikel Val-Calvo
mikel1982mail@gmail.com
José R. Álvarez-Sánchez
jras@dia.uned.es

Received: *30 June 2019*
Accepted: *08 November 2019*
Published: *26 November 2019*

Citation:
Val-Calvo M, Álvarez-Sánchez JR, Ferrández-Vicente JM and Fernández E (2019) Optimization of Real-Time EEG Artifact Removal and Emotion Estimation for Human-Robot Interaction Applications. Front. Comput. Neurosci. 13:80. doi: 10.3389/fncom.2019.00080

1. INTRODUCTION

The use of Electroencephalography (EEG) signals for emotion estimation has been in the point of view of the field for the last decades. The future use of systems that could perform real-time emotion estimations in subjects under different health conditions would improve the application of therapies in different scenarios. One of the most promising fields for the application of such methodologies is affective human-robot interaction (HRI).

Under the paradigm of emotion recognition, robots will allow the development of automatic systems for the treatment and evaluation of the brain patterns of patients, taking into account the emotional content and, furthermore, to have the ability to adapt their behavior as the mood of the patient changes dynamically.

From the perspective of the field of robotics, emotions estimation can be performed by evaluating the dynamical changes over facial expressions, body language, voice tone, EEG patterns, and physiological signals, related to the equilibrium between the parasympathetic and sympathetic autonomous systems. The EEG is a non-invasive method of high temporal resolution that could allow real-time recognition of emotional responses. Also, it can provide a better understanding of the user's behavior and emotional responses which involve facial expression, tone of voice, or body gestures, which may remain hidden as is the case for patients with expression and mobility problems. Therefore, in this article, EEG patterns will be analyzed and related to emotional responses, as they may provide a different perspective on patients' emotional responses.

Most research studies using EEG have presented methodologies that used offline and supervised artifact removal obtaining high accuracy results, however, often involving the use of complex deep learning machines that require hyper-parameter tuning (Khosrowabadi et al., 2014; Zheng and Lu, 2015; Zheng et al., 2017; Song et al., 2018). Both processes could take up to several days or even weeks of preparation which are not affordable for domains of study where real-time constraints are involved. On the other hand, the problem of real-time recognition has been already addressed by Liu et al. (2010, 2017) using the IADS database and own-produced video database, respectively, using Fractal Dimensions as the main feature for emotion recognition.

As EEG emotion estimation has proved to be affordable in different ways, the next barrier is to perform such task under real-time constraints. This process faces two main problems: online artifact removal and classification with high accuracy results. The former is usually performed in two steps. Firstly decomposing the signals using independent component analysis (ICA) and recompose the signals for the next step. Secondly, visualizing the signals to manually remove the parts which are related to artifacts. The latter involves the following procedures:

- Feature extraction, to represent the information as a set of features.
- Feature smoothing, to remove variability over time.
- Scaling the training samples taking into account the underlying data distribution.
- Dimensional reduction by means of feature selection techniques.
- Model selection and hyper parameterization for optimal generalization.

Finally, the development of such a methodology that could work under real-time constraints must deal with two main obstacles: artifact removal and accurate classification across sessions and subjects.

1.1. Online Artifact Removal

The most common artifacts presented in EEG signals are electro-oculographic (EOG) artifacts, muscle artifacts, and 50 Hz background noise. Artifact removal is necessary, as it reduces possible classification errors and reduces the amount of processed information. On the other hand, care must be taken while carrying out such a process, since valuable information in the signals could be damaged.

Taking into account these assumptions, an automatic artifact removal method can be developed using the following approach. Firstly, 50 Hz background noise can be easily removed by a notch filter based on IIR filters. Secondly, EOG artifacts, such as blinking, are often presented within slow frequency bands, below 5 Hz (Rani and Mansor, 2009), while muscle artifacts are usually presented within medium to high-frequency bands 20–300 Hz (Muthukumaraswamy, 2013). Therefore, muscle artifacts are partially removed outside the range of 1–50 Hz when filtering the signals, since this range includes the best frequency bands for emotion estimation: delta (1–4 Hz), theta (4–8 Hz), alpha (8–16 Hz), beta (16–30 Hz), and gamma (30–50 Hz). As several studies report (Zheng and Lu, 2015; Zheng et al., 2015, 2017), the most effective band ranges for emotion estimation, are beta and gamma bands. Finally, EOG artifacts can be effectively removed with real-time constraints by using independent component analysis (ICA) methods combined with wavelet analysis. Although several real-time EOG artifact removal methods have been developed, only two methodologies (Mahajan and Morshed, 2014; Mammone and Morabito, 2014) based on these approaches were tested.

1.2. Emotion Estimation

EEG emotion estimation is considered a challenging task due to different factors. Self-evaluation is needed as there is no basic truth about emotion classification and thus, the assessment performed over experienced emotions is a subjective task. Therefore, a series of emotional models developed in the field of psychology must be used to guide the self-evaluation process. The most used, in the field of EEG emotion estimation, are the discrete and dimensional models (Russell, 1980; Roseman et al., 1990). The former is based on the assumption that emotions produce differentiated and independent emotional responses. The latter assumes that emotions are manifested dynamically with subtle inter-relations among them. The use of discrete labels for emotions is based on the affective-defensive emotional model following Davidson and Fox (1982) and Davidson (1992). There is still no clear evidence of whether emotions affect the brain patterns across specific regions or spread over cortical and sub-cortical areas (Kragel and LaBar, 2016). Due to the variability in brain patterns, it is difficult to find invariant patterns across sessions and subjects. In this paper, in total eight electrodes were used, six temporal electrodes and two prefrontal (AF3, T7, TP7, P7, AF4, T8, TP8, P8), since they have proved to be the best brain areas for emotion estimation (Zheng and Lu, 2015; Zheng et al., 2017).

EEG signals can be approached through several domains: time, frequency, time-frequency, information theory, and signal complexity. Features should be stable across sessions to avoid the greatest amount of variability while carrying as much valuable information as possible. To work in that direction, feature smoothing is an effective technique that helps to erase such variability over time. Linear Dynamic Systems (LDS), moving average or Savitzky-Golay (SG) among other techniques can be used (Zheng and Lu, 2015; Zheng et al., 2015, 2017). Regarding scaling, outliers must be taken into account, by choosing an appropriate methodology, since some methods such as standardizing or min-max scale approaches can damage the feature space for the classification step.

One key step in such machine learning strategies is the dimensional reduction stage for the selection of relevant and stable features over time, which faces two main problems: First, in real scenarios there is no access to the underlying distribution related to the target, this makes it difficult to find relevant features in a way that is closely related to bias-variance trade-off (Kohavi and John, 1997). Second, finding the optimal set of features often involves NP-hard search spaces and the selected model must take into account time constraints in real-time scenarios.

Moreover, in the EEG emotion estimation, time series corresponding to trials are split into a series of samples. The main assumption is that time series related to trials are independent of each other but related to the evoked emotion, so the time series is homogeneous within the trial and heterogeneous between trials (Tashman, 2000). This makes the EEG time series be a special case. While in the time series prediction paradigms, as is the case for regression models, past is used to predict the future, supervised learning models assume independence of samples and do not care about the time order of the samples. Therefore, predefined cross-validation schemes for supervised learning algorithms are not suitable for model performance evaluations.

For the dimensionality reduction step, different approaches differ in the way they exploit the relation between features and target (Kohavi and John, 1997). In general, they are defined as filter-based, wrapper-based and embedded methods. On one hand, filter-based methods perform feature selection independently from the learning process and are based on the assumption that a feature that has higher variance may contain more useful information. On the other hand, wrapper-based methods combine feature selection and the learning process to select an optimal feature subset. This is also the case for embedded methods, which perform a penalty against complexity during the learning process to reduce the degree of overfitting or variance of a model by adding more bias.

Wrapper and embedded methods involve the use of nested cross-validation procedures which may lead to increased computational cost and possible overfitting, especially when a small number of observations are available. Also, as mentioned earlier, these processes, when applied with predefined algorithms, do not take into account the particularities of the EEG time series, so that the feature subset estimates are further biased.

Regarding the classifier to be chosen for the methodology, in recent years, very powerful deep learning approaches have been developed and tested in the emotion estimation field (Tripathi et al., 2017; Yin et al., 2017; Song et al., 2018). Although they have proved to be promising tools, they usually require a very large amount of time for hyper-parameter tuning, so there exists a need to find an approach that could yield automatically and in a short time, while still achieving high accuracy performances.

1.3. State of the Art

In this work, the proposed methodology is compared with the latest performances in the field of emotion estimation. A set of research studies are used for comparison as they show the best results in terms of accuracy of the results, albeit the experimental conditions which are not completely equivalent due to the real-time constraints imposed on the present study.

First, Khosrowabadi et al. (2014) developed a Biologically Inspired Feedforward Neural Network called ERNN. They produced a database containing 57 subjects using emotionally tagged audio-visual stimuli, achieving an average performance of 70.83 % for arousal and 71.43 % for valence dimensional spaces, using the 5-fold cross-validation method.

Second, Zheng and Lu (2015) produced their own database, SEED, for the estimation of three affective states. Deep Belief Neural Networks (DBNs) were used to analyzing critical frequency bands and channels through the weight distributions of the trained DBNs. With a selection of 12 channels, the best accuracy result obtained was a mean accuracy of 86.65 % using the first 9 trials as the training set and remaining 6 ones as the testing set, Inter-trial (IT), for each subject.

Third, Zheng et al. (2017) explored a set of popular features used in the domain of EEG emotion estimation. The SEED database was used. Differential entropy and together with the Graph regularized Extreme Learning Machine (GELM) classifier outperformed state of the art results. Mean accuracy of 60.93 % was obtained using the Leave-one-out validation scheme (LOO). For the inter-sessions validation scheme (IS) a mean accuracy of 79.28% was obtained.

Fourth, Tripathi et al. (2017) compared the use of both Deep and Convolutional neural networks (DNN, CNN) using the DEAP database. The valence and arousal dimensions were split into three categories. The DNN model achieved 58.44 and 55.70 %, while the CNN model achieved 66.79 and 57.58 %, respectively using the Leave-one-out validation scheme.

Later, Song et al. (2018) developed a novel Dynamical Graph Convolutional Neural Network (DGCNN) tested over the SEED database. Differential entropy features of five frequency bands were combined resulting in average recognition accuracy of 90.40 % using the first 9 trials as the training set and the remaining 6 as the testing set.

Finally, a comparison is made taking into account those experiments where real-time constraints were faced. As mentioned above, Liu et al. (2017) developed a real-time emotion recognition system which uses a three-level classification approach and a real-time artifact removal algorithm. Regarding the classifying strategy, in the first level, high-arousal and valence emotions versus neutral emotions were estimated with an average accuracy of 92.26. For the second level, positive vs. negative emotions, with an average accuracy of 86.63 were estimated. For the last level, joy, amusement, and tenderness were classified at an average accuracy of 86.43. The training and validation scheme was done by using 8 trials to elicit 7 discrete emotions and one neutral state and the same amount of stimuli as a test set in a real-time emotion estimation scenario.

2. METHODOLOGY

The objective of the present paper is to perform the whole process involved in EEG emotion estimation under real-time constraints.

To prove its feasibility, it will be tested on the SEED database. The process comprises six main steps:

1. Online artifact removal: EAWICA.
2. Feature extraction: Differential Entropy, Amplitude Envelope, Petrosian Fractal Dimension, Higuchi Fractal Dimension, and Fisher Info.
3. Feature smoothing: Savitzky-Golay filter.
4. Feature scaling: Quantile transform followed by min-max scaler.
5. Feature selection: Based on the chi-squared statistic.
6. Classification: Nearest Neighbors (KNN), Support Vector Machines (SVM) with linear and radial basis function kernels, decision trees, random forest, AdaBoost, naive Bayes, and Quadratic Discriminant Analysis (QDA).

2.1. SEED Database

The SEED database (Zheng and Lu, 2015) has 15 subjects but the experiment was performed three times each, with a time interval of one week. Emotions were quantified in terms of three discrete categories: POSITIVE, NEGATIVE, and NEUTRAL. A set of 15 emotional-tagged videos were employed, each approximately 180 s long. The international 10–20 system for EEG acquisition was used with a set of 62 channels.

2.2. Online Artifact Removal

Two main approaches that are affordable in real-time constraints are used: EAWICA (Mammone and Morabito, 2014), and ICA-W (Mahajan and Morshed, 2014) methods. The performance over artificial artifactual data is first analyzed to compare them under controlled conditions but finally, are compared over real EEG samples obtained from the objective SEED database.

Both approaches use the same underlying philosophy, that is, they employ a divide and conquer strategy to isolate the artifacts as much as possible, both in time-frequency domain through the wavelets transform decomposition and by analyzing the independent components sources. Based on the assumption that artifactual and EEG signals are linearly mixed but statistically independent, and that propagation delays through the mixing medium are negligible, ICA seems to be an optimal tool for decomposing an identifying the source of artifactual signals effectively. In order to properly take into account either sub-Gaussian and super-Gaussian signals, the Extended-Infomax ICA algorithm (Bell and Sejnowski, 1995) is used in both approaches, which allows the computation of the unmixing matrix, so that the components are as independent as possible.

Spurious isolated oscillations are then automatically detected by a means of entropy and Kurtosis measurements. On one hand, the entropy value for EOG artifacts is expected to be low due to the regular shape so they are more predictable in comparison to neural oscillations. On the other hand, peak distributions with highly positive Kurtosis values are expected for the same type of artifacts (Mammone and Morabito, 2014). Both approaches have been compared using an analysis over all the frequency range bands (delta, theta, alpha, beta, gamma) as well as over the delta band only.

2.2.1. ICA-W

EEG signals are decomposed in a series of independent components (ICs), where is expected that independent sources are separated from each other. Artifactual ICs are identified by analyzing the statistical properties in terms of Kurtosis and Multi-Scale Sample Entropy measurements. To remove as little information as possible, ICs identified as artifactual are further selected for bandpass decomposition with wavelet analysis. Decomposed wavelet independent components (WICs) require a second identification stage with the aim of zeroing only those wavelet components carrying artifactual information. Finally, the original signals are reconstructed with the inverse transforms of wavelets and ICA decompositions (Mahajan and Morshed, 2014).

2.2.2. EAWICA

The original EAWICA method proposes the isolation approach of the artifactual signal component by first computing the wavelet components over the EEG signals within the frequency ranges associated with the emotion estimation task. Thus, once the information is bandpass filtered, ICA decomposition is applied to isolate artifactual data in a series of WICS. In order to automatically detect artifactual WICS, Kurtosis, and Renyi entropy measurements are used. Those marked as artifactual are further split into a series of time windows with a temporal interval of one second defined as epochs, which in case of being marked as artifactual, are zeroed to remove as little information as possible. Finally, ICA reconstruction followed by wavelet components addition is performed to reconstruct the original signals (Mammone and Morabito, 2014).

2.2.3. Differences and Modifications

One of the main differences between both algorithms is that EAWICA methods improve the localization of the artifactual data by first band-passing the signals, which also helps the ICA algorithms to properly identify the sources as it takes the advantage of the redundancy by having more data. Another key difference is the way both methods apply the threshold steps to identify the artifactual data. While the ICA-W performs an automatic threshold method that works in the frequency domain (wavelet components), the EAWICA method performs in the time domain. Regarding the EAWICA threshold was restrictive, in order to improve upon this, a design decision has been taken to allow the variation of the thresholds by manually adjusting them in terms of the quartiles over the distribution values.

2.2.4. Metrics

To properly compare both methods, EEG signals have been artificially contaminated with a set of artificially generated artifacts as is done by Mammone and Morabito (2014). A series of measurements are computed to compare both methods: root-mean-square error (RMSE), correlation (CORR), mutual information (MI), and coherence (C) together with timing measurements, will allow the best method to be chosen.

2.3. Emotion Estimation Methodology

The proposed methodology has been designed for its future use on subject dependent paradigms in the domain of HRI. This implies that the time consumption of each of the following processes must accomplish real-time constraints. Therefore, this philosophy guides the decision-making process taking into account an optimal balance between fast computation and accuracy.

2.3.1. Preprocessing

EEG signals are arranged in a three-dimensional matrix containing n trials, c channels, and s samples at a sampling frequency, f_s. First, given that each signal has its own scaling factor values, signals are standardized using the 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, 4–8, 8–16, 16–30, and 30–50 Hz.

2.3.2. Feature Extraction Methodology

Once the data-set has been preprocessed, a set of features are computed based on the oscillatory properties of brain signals:

- Differential Entropy (DE): computed as a metric for measuring the predictability of signal X, whose values have a probability density function similar to a Gaussian distribution, $N(\mu, \sigma^2)$, as is the case for EEG signals. It can be defined as $h(X) = \frac{1}{2}\log\left(2\pi e\sigma^2\right)$.
- Amplitude Envelope (AE): computed using the Hilbert transform (Boashash, 1992).
- Petrosian Fractal Dimension (PFD): defined as $PFD = \log(N)/\left(\log(N) + \log(N/(N + 0.4N_\delta))\right)$, where N is the series length, and N_δ is the number of sign changes in the signal derivative (Petrosian, 1995).
- Higuchi Fractal Dimension (HFD): Higuchi's algorithm can be used to quantify the complexity and self-similarity of a signal (Accardo et al., 1997).
- Fisher Information (FI): 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 (Fisher, 1925).

All features have been computed using a sliding window of 6 seconds as suggested by Candra et al. (2015), without overlapping. Each training sample represents the computed features for each time window. Features are computed for each band/channel and later concatenated for each training sample. Thus, resulting in a feature set of 435 samples with 200 features. AE has been computed with the Neuro Digital Signal Processing Toolbox (NeuroDSP) python library (Cole et al., 2019) developed at Voytek's Lab. PFD, HFD, and FI have been computed with the PyEEG python library (Bao et al., 2011).

2.3.3. Feature Smoothing

Emotions are often considered static in the field of EEG to simplify the data processing for the classifiers, albeit continuous and subtle changes should be considered in the time domain. It has been noticed previously (Zheng and Lu, 2015; Zheng et al., 2015, 2017), that considering the temporal dependence and variation of emotions during the stimuli improves the performance of the training step. To do that, smoothing the feature space allows us to filter out those components that are unrelated to emotional states, thus becoming a key step for the design of an optimal methodology. In other words, smoothing the feature space deals with 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.

Savitzky-Golay (SG) filtering method is proposed as an alternative to Linear Dynamic Systems (LDS). Both approaches have the property of outperforming the classification accuracy reports above the results obtained without smoothing the feature space, but also SG smoothing is significantly faster than LDS and even improves the accuracy reports.

2.3.4. Feature Scaling

Feature scaling is a key step in preprocessing data. Outliers can severely damage the performance of the classifiers while looking for statistical differences. Moreover, some machine-learning algorithms for the dimensionality reduction and classification processes require data to have a predefined range of values. The process of scaling data must be performed taken into account both constraints to properly feed the algorithms in the next steps. In this paper, the Quantile-Transform method (histogram equalization to uniform distribution) followed by the Min/Max scaling method is performed. The former is a non-linear method for scaling data distributions which is robust to outliers. The later allows re-scaling in a positive range of values $[0 - 1]$ as the dimensional reduction proposed method requires positive values as input.

2.3.5. Dimensionality Reduction

As mentioned in the introduction, wrapper and embedded methods combine feature selection and the learning process by the use of nested cross-validation schemes but this leads to biased results when taking into account the particularities of EEG time series. Therefore, χ^2 feature selection technique has been chosen as it is a filter-based method where the selection of features is based on the chi-squared statistic which measures the lack of independence between a feature and the target, without involving any cross-validation biased scheme nor combining the selection and the learning process.

2.3.6. Set of Classifiers

Classification process has been performed using a set of eight 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 with default hyper-parameter values. The Scikit-learn python library (Pedregosa et al., 2011) has been used.

2.3.7. Performance Evaluation

The crucial point is to ensure that samples in the validation set are reasonably independent of the samples in the training set. Therefore, three different validation methods are reported in this paper.

- Validation across trials: Using the first nine trials as the training set and remaining six ones as the testing set for each subject and session.

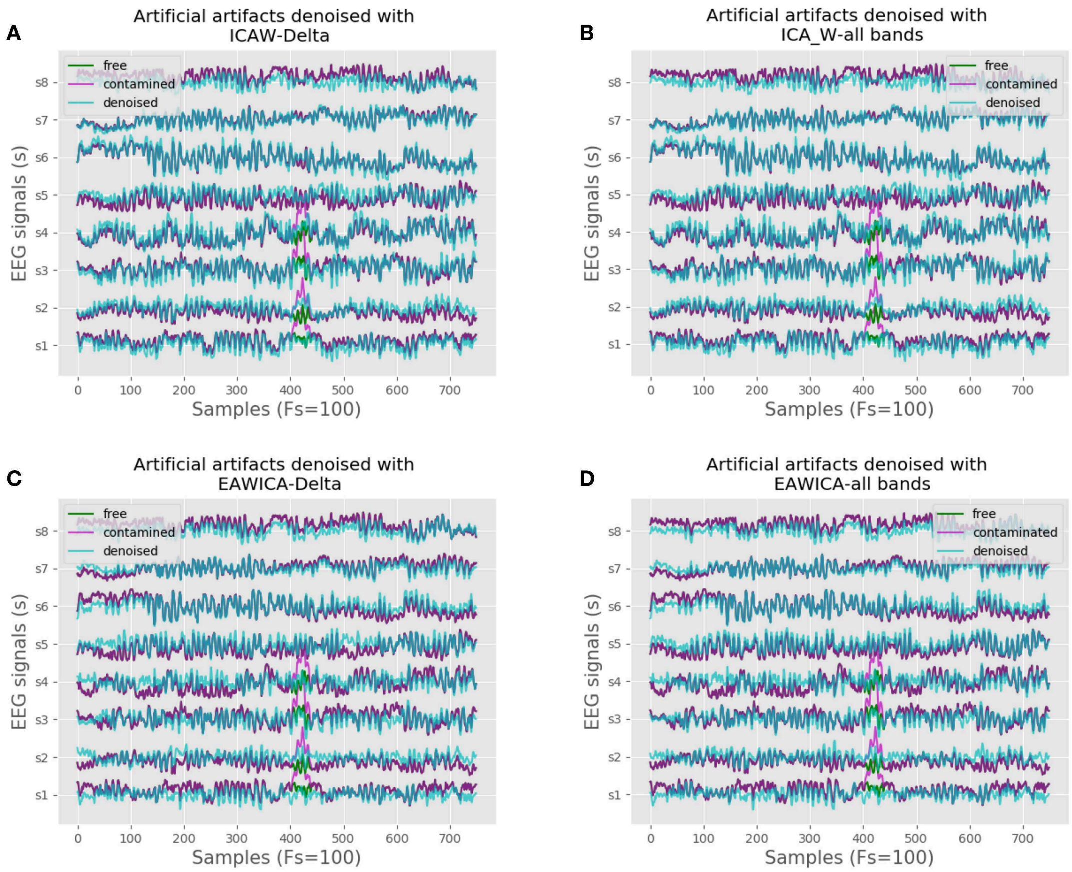

FIGURE 1 | EAWICA and ICA-W methods performance comparison over artificial artifactual data. **(A)** ICA-W using only the delta band component of artifactual signals. **(B)** ICA-W using all bands components of artifactual signals. **(C)** EAWICA using only the delta band component of artifactual signals. **(D)** EAWICA using all bands components of artifactual signals.

TABLE 1 | Metrics comparison for the four cases.

	RMSE	Correlation	Mutual information	Time (s)
EAWICA - ALL	0.27	0.87	0.66	0.19
ICA-W - ALL	0.16	0.92	0.87	0.60
EAWICA - DELTA	0.29	0.86	0.64	0.36
ICA-W DELTA	0.14	0.93	0.93	0.68

- Validation across sessions: Train and test are performed over the whole set of sessions pairs for each subject.
- Validation across subjects: Leave-one-subject-out validation scheme is used.

3. RESULTS

3.1. Online Performance Over Artificial and Real Artifactual Data

Figure 1 shows the performance of both EAWICA and ICA-W methods applied over artifactual data. Both methods have been applied, on one hand, using all the frequency ranges of interest (delta, theta, alpha, beta, gamma) and, on the other hand, over the slowest frequency range (delta). It can be noted that ICA-W focuses on the artifactual data better than EAWICA, moreover, the latter seems to affect the signal in all the frequency ranges. As a comparison, a set of metrics are shown in **Table 1**. Taking into account these metrics, ICA-W outperforms the obtained results in terms of CORR, MI, and RMSE, with regard to the artifactual-free signals. Concerning time consumption, EAWICA performs the filtering process in 0.19 s while still having low RMSE and high CORR and MI results, and ICA-W takes 0.6 s.

Signal filtering performance can also be evaluated in terms of the coherence lost relative to the original artifact-free signal. Coherence is referred to as the cross-frequency spectrum of two signals, it is a measurement of the loss of the filtered signals relative to the original, in the frequency domain. **Figure 2A** shows the coherence estimates between both the EAWICA and ICA-W filtering methods relative to the artifact-free signals, where it can be noted that EAWICA applied over all frequency ranges, severely damages the original signal in all the frequency spectrum. On the contrary, EAWICA applied over the delta band

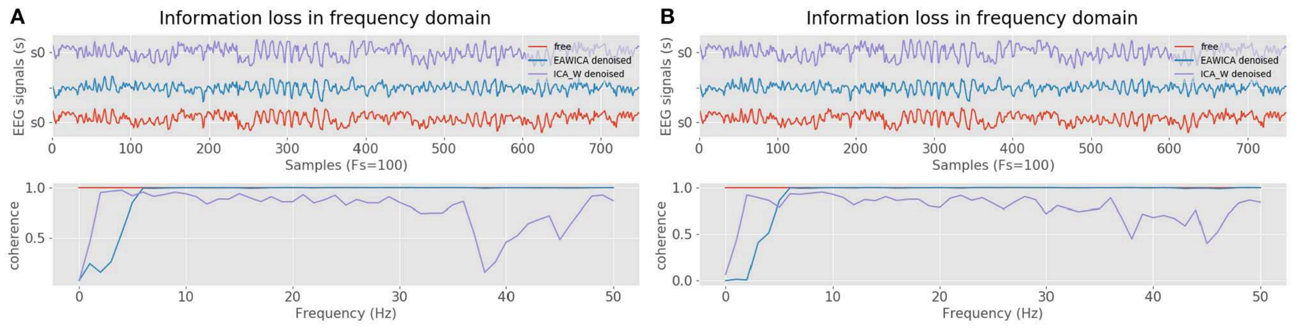

FIGURE 2 | Coherence estimates between filtered (EAWICA and ICA-W) signal and artifact-free signal. **(A)** Coherence estimation after artifact removal using all frequency bands. **(B)** Coherence estimation after artifact removal using only the delta band.

outperforms the signal cross-frequency measurements against ICA-W as it can be shown in **Figure 2B**.

Artificial artifactual data is useful in order to compare the performance of a set of algorithms taking into account the basic truth, but real domains often involve more complex signals. Therefore, both methods need to be compared against real artifactual EEG data, which in this case is obtained from the SEED database. **Figure 3A** shows the performance of both algorithms and for the real case EAWICA seems to outperform ICA-W. **Figure 3B** better shows how the EOG artifacts are removed by EAWICA while still maintaining all the information in the frequency spectrum. Regarding time consumption, EAWICA required 0.4 s and ICA-W 2.2 s both for filtering 6 s of 8 signals with a sampling rate of 200 Hz. Taking all these results into account, EAWICA has been selected for the methodology.

3.2. Emotion Estimation

In machine learning applied to emotion estimation, the standard K-fold cross-validation is often applied. At that point, there is a key issue that arises when the performance of the models is evaluated. These methods cannot be directly used with time series data as they assume that there is no relationship between the observations, that is, each observation must be independent while in fact, they are not. The EEG time series data in emotion estimation strongly correlated along the time axis. The randomization performed with cross-validation methods make it likely that for each sample in the validation set, numerous strongly correlated samples exist in the training set but this defeats the very purpose of having a validation set: the model has prior information about the validation set, leading to optimistic performance reports on it. Such an analysis could provide an insight into how the selected model works, or if there exists a statistical difference between samples, but not if a correlation between these statistical differences and the task at hand is really present. Furthermore, any estimate of the performance will be optimistic and any conclusion based on this performance will be biased and could be completely wrong. This problem is not only related to the classification step, as a result, it also arises in the dimensionality reduction step if predefined algorithms are used, which do not take into account these assumptions.

Therefore, to properly evaluate the performance, three different validation schemes were used. The evaluation has been performed taking into account different subsets of features, ranging from 1 to 200.

3.2.1. Feature Smoothing

Figure 4 shows the comparison of applying LDS or SG methods on the feature space. It can be clearly noted that smoothing the features space makes it possible to clearly observe the correlation with the corresponding targets. Moreover, the SG method makes the features space even less noisy. Therefore, although the LDS smoothing method works well eliminating the variability, the SG method outperforms the obtained results, both in terms of removing such variability, as well as with regard to the time consumption needed. While LDS needs roughly 200 s for smoothing a feature space of 435 samples $\times$ 200 features, SG takes approximately 73 ms.

3.2.2. Trial Validation Tests

Figure 5 shows the $\mu \pm \sigma_M$ performance for inter-trial validation tests for all subjects in each session. A set of 9 trials for each subject have been selected for training while 6 are used as a test set, were 2 trials are present for each class (POSITIVE, NEGATIVE, NEUTRAL). For each training step, a set of features ranging from 1 to 200 have been used. This evaluation provides an insight into the robustness of the method in terms of generalization performance of the model in a more realistic scenario for the unseen. The best mean accuracy report for the best subset of features for each subject is (82.3 ± 4.4) % for session 1, (78.9 ± 5.7) % for session 2, and (80.5 ± 8.6) % for session 3.

3.2.3. Sessions Validation Tests

Figure 6 shows the $\mu \pm \sigma_M$ performance for inter-session validation tests for all subjects and for all session to session pairs. For each training step, a set of features ranging from 1 to 200 have been used. As the SEED database consists of the same set of stimuli for each session, these results prove the stability of the selected features over time. The best mean accuracy report for the best subset of features for each subject is (74.6 ± 8.8) % for session 1 to session 2, (74.9 ± 11.1) % for session 2 to session 1, (76.8 ± 8.1) % for session 3 to session 1, (77.0 ± 10.1) % for

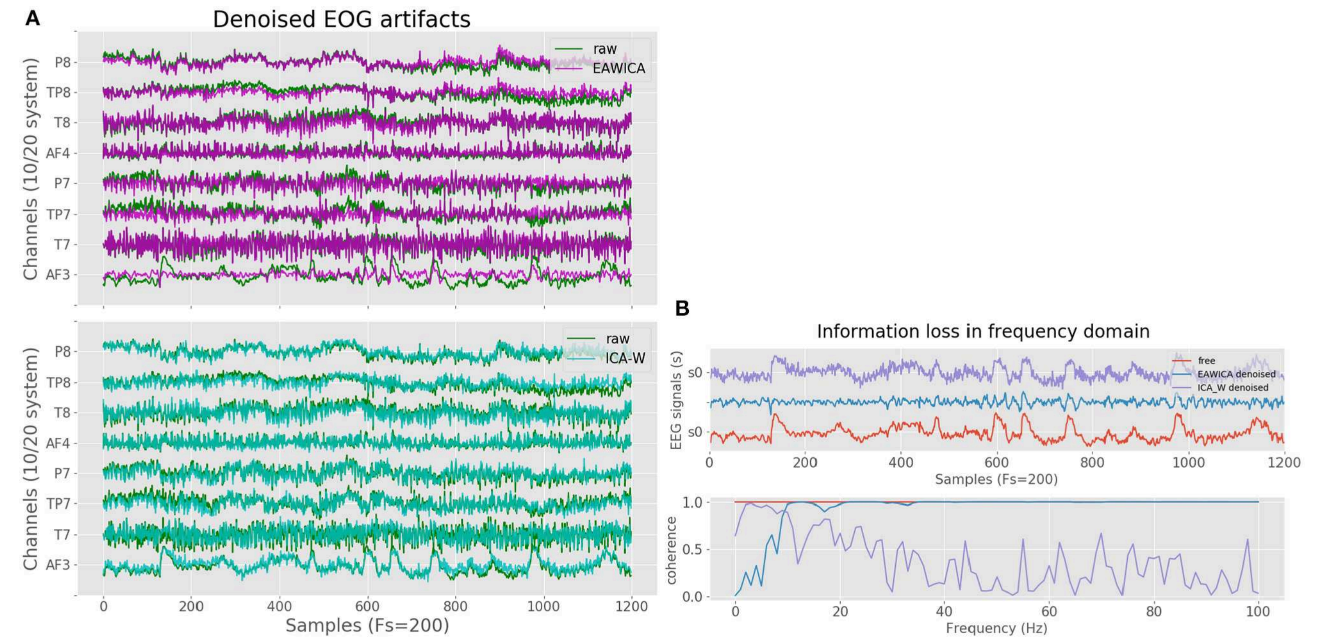

FIGURE 3 | EAWICA and ICA-W methods performance comparison over real EEG data. **(A)** Comparison between EAWICA and ICA-W artifact removal strategies in the time domain. **(B)** Coherence comparison between EAWICA and ICA-W artifact removal strategies in the frequency domain.

session 1 to session 3, (76.4 ± 8.4) % for session 2 to session 3, and (75.9 ± 6.7) % for session 3 to session 2.

3.2.4. One Subject Out Validation Tests

Figure 7 shows the $\mu \pm \sigma_M$ performance for inter-subject validation tests. A leave-one-out subject evaluation scheme has been used. For each training step, a set of features ranging from 1 to 200 have been used. This validation scheme provides an insight into the robustness of the selected set of features for the subject independent paradigm, confirming that underlying common processes exist across subjects and that the selected features are closely related to invariant properties of the brain oscillations dynamics for the evoked emotion experimentation. The best mean accuracy report for the best subset of features for each subject is (77.6 ± 5.6) % for session 1, (73.4 ± 6.7) % for session 2, and (77.1 ± 7.7) % for session 3.

3.2.5. Time Consumption

The main objective of this methodology is to perform under real-time constraints while having high accuracy results in terms of emotion estimation. Thus, time consumption analysis is needed as a comparison for the design decisions. Methodologies based on real-time constraints have to deal with the following key processing steps:

- Time consumption of the artifact filtering process and feature extraction steps.
- Time consumption of the feature smoothing and scaling steps.
- Time consumption of the classifying and fine-tuning step.

Feature smoothing has shown to be a key step to further improve the performance of models but those filtering processes cannot be performed over unique samples. Therefore, a set of samples for each trial should be present in the training set before smoothing the features. Thus, the experimental paradigm in a real scenario needs to be performed on three main stages. First, a real-time signal acquisition while the subject is stimuli evoked. In that stage, signals could be stored without any filtering process to alleviate the computation effort of the acquisition application. The second should consist of the model training step. Each EEG set of signals would be split on windowed samples for the processing of training sets. This stage should perform the following steps: artifact removal, feature extraction, feature smoothing, and scaling, and finally model training. For this stage, it is a requirement for the time consumption to be short, in order to reduce the amount of time the subject under study is waiting. Finally, the following steps for each acquired sample should be performed: online artifact removal, feature extraction, smoothing, and scaling taking into account training samples to properly transform the computed features, and the final prediction step.

For this experimental paradigm, several time metrics have been computed for each step. As a comparison, **Table 2** shows the details of each processing step for the case of offline training and online prediction stages. The software used was the Scikit-learn python library (Pedregosa et al., 2011) running in an Intel Core i7-8700 K (3.70 GHz).

Offline training is analyzed taking into account one subject session, which means using $435 * 200(samples/features)$. The process of online filtering performed in (151.89 ± 0.23) s. The feature extraction step required (79.27 ± 0.03) s. Concerning feature smoothing, LDS required (149.35 ± 0.07) s, while SG was able to perform the same task in approximately (67.31 ± 0.01) ms. The scaling process was performed in approximately (98.42 ± 0.01) ms. Feature selection was

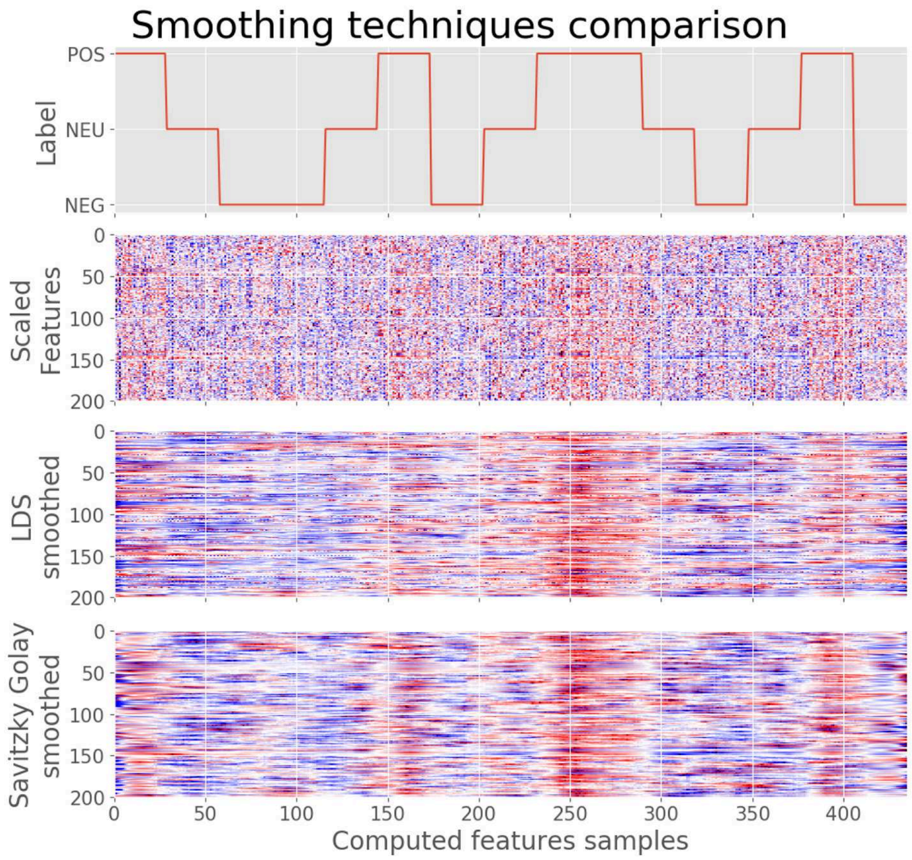

FIGURE 4 | Feature smoothing applied over the feature space. At the **(top)**, the label of each short film in the time domain. Below, the feature space after Quantile-Transform normalization. Then, feature space smoothed using LDS. At the **(bottom)**, feature space smoothed using Savitzky-Golay filtering. As it can be noted, label smoothing clarifies the correlation of the feature space regarding labels.

computed for the worst-case scenario, selecting only the best feature, which required (1.87 ± 0.05) ms. The classifying step was proposed for a set of eight classifiers. The proposed approach consists of selecting a range of features taking into account the aforementioned results to reduce the amount of time to find the number of features that best generalizes over the unseen. The amount of time needed for each classification of a number X of selected features for the set of eight classifiers is approximately (873.65 ± 0.62) ms. Furthermore, there is no need for fine-tuning, as the previous results show that the methodology is robust enough without hyper-parameter tuning.

On the other hand, the online prediction is analyzed regarding the time consumption for one unique sample. Online artifact removal is carried out in (200 ± 16) ms, while feature extraction is performed in (182.25 ± 0.08) ms. Based on previous results, SG smoothing is used instead of LDS, as the processing time has been shown to reduce. Furthermore, feature selection is not computed as the set of best features are previously determined during the offline training stage. As SG smoothing and normalization steps cannot be performed over one unique sample, the strategy must involve the set of features computed for the offline training stage, thus, the time required is the same as in the offline training. Stacking the incoming samples in this data structure allows for both processes to be applied, and once the incoming sample has been smoothed and normalized, taking into account previous training samples, it is unstacked for the final prediction step over the best selected model during the offline training, at the time cost of (38.8 ± 1.1) µs.

4. DISCUSSION

The proposed methodology accomplishes the main processing steps required for a real-time emotion estimation approach and overcomes the many difficulties presented in this field of study.

Artifact removal is the first step of the whole process and therefore plays a very important role in the outcome. As mentioned earlier, only two methods for online artifact removal were tested, EAWICA and ICA-W, since they show to be feasible for real-time constraints. EAWICA outperforms ICA-W when using real EEG data and was therefore chosen as part of this method. A modified version of EAWICA, constrained to the delta band, was used to reduce artifacts (EOG), and therefore reduce computation time.

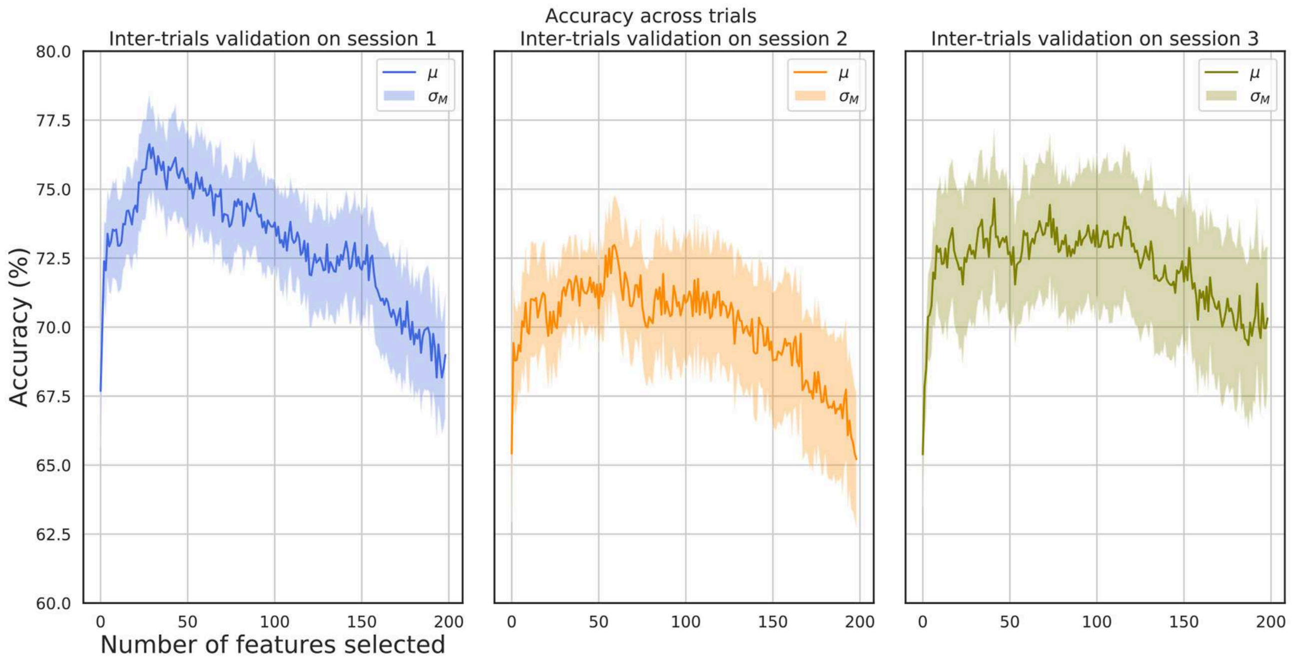

FIGURE 5 | Inter-trial validation accuracy. The figure on the left shows the accuracy results for validation between trials in session 1, the middle figure for session 2 and the right figure for session 3.

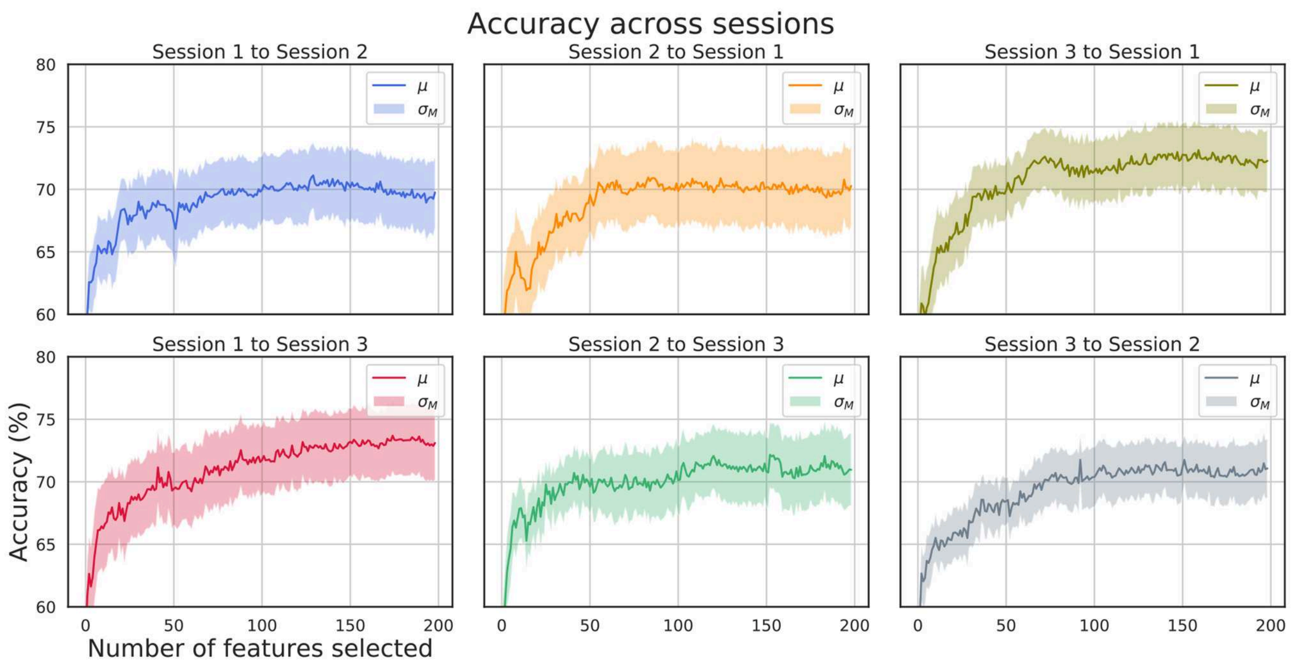

FIGURE 6 | Inter-session validation accuracy. Six different combinations of inter-session validation are shown.

In this paper, 8 electrodes were used, six temporal electrodes and two prefrontal, placed at AF3, T7, TP7, P7, AF4, T8, TP8, and P8, since these were previously shown in the literature to be the best brain locations to use for emotion estimation (Zheng and Lu, 2015; Zheng et al., 2017). The chosen set of electrodes showed to be a good choice since it offered a proper balance between the more informative electrodes and redundancy for the classifying step (Kohavi and John, 1997), and achieved good

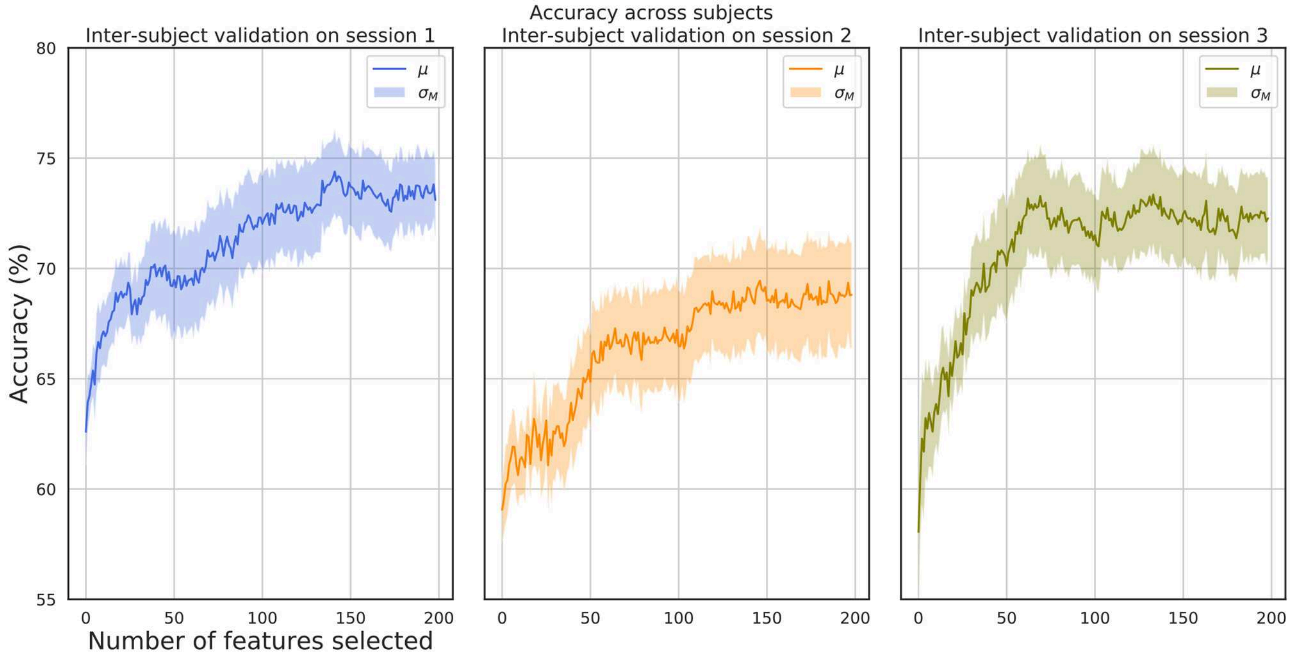

FIGURE 7 | One subject out validation accuracy with respect to selected features. Inter-subject validation for the three sessions.

TABLE 2 | Time consumption regarding offline training and online prediction stages.

	Offline training	Online prediction
(Samples, features)	(435, 200)	(1, 200)
Artifact removal	151.89(23) s	200(16) ms
Feature extraction	79.27(3) s	182.25(8) ms
LDS smoothing	149.35(7) s	None
SG smoothing	67.31(1) ms	67.31(1) ms
Normalize	98.42(1) ms	98.42(1) ms
Feature selection	1.87(5) ms	None
Training/Predicting	873.65(62) ms	38.8(11) μs
Total	381.55 s	548.02 ms

results. Due to the great number of electrodes employed, high dimensional space is often a limitation in the domain of EEG signal analysis. Such signals have a complicated structure since their intrinsic properties are non-linear and non-stationary, thus, the need for balance between a wide dimensional space of features that must be treated carefully with feature reduction techniques, often biased by the statistics in hand (Fan and Li, 2006), and a set of low components feature space, which would be desirable.

The robustness of the methodology is higher when only a subset of eight temporal and prefrontal electrodes is used, leading to a feature space with fewer dimensions. On the other hand, the results show that in general, a subset ranging from 50 to 100 features leads to the optimal accuracy results, probably because the redundancy on the features space enhances the performance of the classifiers.

In the HRI domain, it is crucial to ensure that real-time emotion estimation is a quick and versatile process. The set of selected features chosen for this methodology is easy to compute in any type of computer and can be easily implemented in any programming language, allowing the quick development of portable systems with high accuracy results, as is the case for the openBCI system. Also, these features allow the interpretation of the phenomenon under study, as they are direct measurements of the properties of brain patterns, being far from black-box techniques, which use deep-learning approaches such as auto-encoders (Chai et al., 2016), or very complex features with difficult interpretation in biological terms (Zheng et al., 2015).

A proper validation scheme is important for every novel methodology for it to be comparable to those in literature. As mentioned earlier, predefined cross-validation schemes for supervised learning algorithms are not suitable for model performance evaluations in EEG emotion estimation (Tashman, 2000). Liu et al. (2017) proposed a real-time methodology which recognized eight different brain affective states with a three-step level classification. This methodology is therefore comparable to the proposed method, however, the validation scheme which was used is not clearly stated, and could, therefore, be backtesting, IT or cross-validation. Without this knowledge, a proper comparison of the results can not be achieved. On the other hand, since different validation schemes were found to be used in state of the art, this method was validated using IT, IS, and LOO (see **Table 3**), in order to properly compare the results. As observed in **Table 3**, the proposed method is close to the results obtained by previous studies in terms of accuracy for the cases of IT and IS schemes,

TABLE 3 | State of the art comparison.

	Validation scheme	Database: Emotional model	Accuracy (%)
Khosrowabadi et al. (2014)	Cross-validation	DEAP: valence/arousal	70.83/71.43
Zheng and Lu (2015)	IT	SEED: [pos., neg., neu.]	86.65
Zheng et al. (2017)	LOO and IS	SEED: [pos., neg., neu.]	60.93 and 79.28
Tripathi et al. (2017)	LOO	DEAP: valence/arousal	66.79/57.58
Song et al. (2018)	IT	SEED: [pos., neg., neu.]	90.4
Liu et al. (2017)	Not specified	Own produced: [neu., non-neu.]/[pos., neg.]/[joy, amusement, tenderness]/[sad, angry, fear, disgust]	92.26/86.63/86.43/65.09
Proposed method	IT/IS/LOO	SEED: [pos., neg., neu.]	82.27/76.36/77.59

and outperforms the results obtained by previous studies in the LOO scheme, which is the most complex due to inter-subject variability.

5. CONCLUSION

Our method has proved to be robust and fast, reaching comparable results to state of the art in subject dependent and independent analysis for EEG emotion recognition. An accurate and computationally light EEG emotional estimation methodology could allow the use of portable and cheap devices in the domain of emotional HRI.

This method uses a three-categories emotional model; however, for more complex emotional models, more complex deep learning strategies must be implemented (Zheng et al., 2018; Zhao et al., 2019). Therefore, even maintaining the three-categories emotional model, this method could be improved by altering the filtering methods and with a better coding strategy, that is, with a set of features that better describe the invariant relationships of emotionally evoked brain patterns and their corresponding categories.

Since this method has proven to be fast, over 1 s total processing time, and reliable, 82.27, 76.36, 77.59% for {IT, IS, LOO} validation schemes respectively, it fulfills the proposed task. Therefore, it is an optimal methodology for HRI, that could further the research in this field.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://bcmi.sjtu.edu.cn/~seed/downloads.html.

AUTHOR CONTRIBUTIONS

MV-C and JÁ-S developed the methodology. JF-V and EF designed and supervised the theoretical approach and results.

FUNDING

We would like to acknowledge Programa de Ayudas a Grupos de Excelencia de la Región de Murcia, from Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia. Also, we would like to acknowledge project RTI2018-098969-B-100 from Ministerio de Ciencia, Innovación y Universidades of Spain, and project PROMETEO/2019/119 from Generalitat Valenciana.

REFERENCES

Accardo, A., Affinito, M., Carrozzi, M., and Bouquet, F. (1997). Use of the fractal dimension for the analysis of electroencephalographic time series. *Biol. Cybern.* 77, 339–350. doi: 10.1007/s004220050394

Bao, F. S., Liu, X., and Zhang, C. (2011). Pyeeg: an open source python module for eeg/meg feature extraction. *Comput. Intell. Neurosci.* 2011:406391. doi: 10.1155/2011/406391

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129

Boashash, B. (1992). Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals. *Proc. IEEE* 80, 520–538. doi: 10.1109/5.135376

Candra, H., Yuwono, M., Chai, R., Handojoseno, A., Elamvazuthi, I., Nguyen, H. T., et al. (2015). "Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (IEEE), 7250–7253.

Chai, X., Wang, Q., Zhao, Y., Liu, X., Bai, O., and Li, Y. (2016). Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. *Comput. Biol. Med.* 79, 205–214. doi: 10.1016/j.compbiomed.2016.10.019

Cole, S., Donoghue, T., Gao, R., and Voytek, B. (2019). Neurodsp: a package for neural digital signal processing. *J. Open Source Softw.* 4:1272. doi: 10.21105/joss.01272

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain Cogn.* 20, 125-151. doi: 10.1016/0278-2626(92)90065-T

Davidson, R. J., and Fox, N. A. (1982). Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants. *Science* 218, 1235–1237. doi: 10.1126/science.7146906

Fan, J., and Li, R. (2006). Statistical challenges with high dimensionality: feature selection in knowledge discovery. *arXiv preprint math/0602133*.

Fisher, R. A. (1925). "Theory of statistical estimation," in *Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 22* (Cambridge: Cambridge University Press), 700–725.

Khosrowabadi, R., Quek, C., Ang, K. K., and Wahab, A. (2014). ERNN: a biologically inspired feedforward neural network to discriminate emotion from EEG signal. *IEEE Trans. Neural Netw. Learn. Syst.* 25, 609–620. doi: 10.1109/TNNLS.2013.2280271

Kohavi, R., and John, G. H. (1997). Wrappers for feature subset selection. *Artif. Intell.* 97, 273–324. doi: 10.1016/S0004-3702(97)00043-X

Kragel, P. A., and LaBar, K. S. (2016). Decoding the nature of emotion in the brain. *Trends Cogn. Sci.* 20, 444–455. doi: 10.1016/j.tics.2016.03.011

Liu, Y., Sourina, O., and Nguyen, M. K. (2010). "Real-time EEG-based human emotion recognition and visualization," in *2010 International Conference on Cyberworlds* (Singapore: IEEE), 262–269.

Liu, Y.-J., Yu, M., Zhao, G., Song, J., Ge, Y., and Shi, Y. (2017). Real-time movie-induced discrete emotion recognition from EEG signals. *IEEE Trans. Affect. Comput.* 9, 550–562. doi: 10.1109/TAFFC.2017.2660485

Mahajan, R., and Morshed, B. I. (2014). Unsupervised eye blink artifact denoising of EEG data with modified multiscale sample entropy, kurtosis, and wavelet-ica. *IEEE J. Biomed. Health Inform.* 19, 158–165. doi: 10.1109/JBHI.2014.2333010

Mammone, N., and Morabito, F. (2014). Enhanced automatic wavelet independent component analysis for electroencephalographic artifact removal. *Entropy* 16, 6553–6572. doi: 10.3390/e16126553

Muthukumaraswamy, S. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Front. Hum. Neurosci.* 7:138. doi: 10.3389/fnhum.2013.00138

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. Available online at: https://arxiv.org/abs/1201.0490

Petrosian, A. (1995). "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," in *Proceedings Eighth IEEE Symposium on Computer-Based Medical Systems* (Texas: IEEE), 212–217.

Rani, M. S. B. A., and Mansor, W. (2009). "Detection of eye blinks from EEG signals for home lighting system activation," in *2009 6th International Symposium on Mechatronics and Its Applications, ISMA 2009* (Sharjah), 5164828.

Roseman, I. J., Spindel, M. S., and Jose, P. E. (1990). Appraisals of emotion-eliciting events: testing a theory of discrete emotions. *J. Pers. Soc. Psychol.* 59, 899–915. doi: 10.1037/0022-3514.59.5.899

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Song, T., Zheng, W., Song, P., and Cui, Z. (2018). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comput.* 1–1. doi: 10.1109/TAFFC.2018.2817622

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *Int. J. Forecast.* 16, 437–450. doi: 10.1016/S0169-2070(00)00065-0

Tripathi, S., Acharya, S., Sharma, R. D., Mittal, S., and Bhattacharya, S. (2017). "Using deep and convolutional neural networks for accurate emotion classification on deap dataset," in *AAAI* (New York, NY), 4746–4752.

Yin, Z., Zhao, M., Wang, Y., Yang, J., and Zhang, J. (2017). Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Comput. Methods Prog. Biomed.* 140, 93–110. doi: 10.1016/j.cmpb.2016.12.005

Zhao, L.-M., Li, R., Zheng, W.-L., and Lu, B.-L. (2019). "Classification of five emotions from EEG and eye movement signals: complementary representation properties," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)* (Shanghai: IEEE), 611–614.

Zheng, W.-L., Liu, W., Lu, Y., Lu, B.-L., and Cichocki, A. (2018). Emotionmeter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.* 49, 1110–1122. doi: 10.1109/TCYB.2018.2797176

Zheng, W.-L., and Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Mental Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

Zheng, W.-L., Zhang, Y.-Q., Zhu, J.-Y., and Lu, B.-L. (2015). "Transfer components between subjects for EEG-based emotion recognition," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (IEEE), 917–922.

Zheng, W.-L., Zhu, J.-Y., and Lu, B.-L. (2017). Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 10, 417–429. doi: 10.1109/TAFFC.2017.2712143

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.2 Article 2: Real-time multi-modal emotion estimation

Title: Real-time multi-modal estimation of dynamically evoked emotions using electroencephalography, heart rate and galvanic skin response

Authors: Mikel Val-Calvo, José R. Álvarez-Sánchez, Jose M. Ferrández-Vicente, Alejandro B. Diaz-Morcillo, and Eduardo Fernández

Journal: International Journal of Neural Systems 30(4)

Impact factor: 5.604, JCR2019-Q1, 18/136 Computer Science, Artificial Intelligence

Published: April 2020

DOI: 10.1142/S0129065720500136

This paper covers items 3 and 6 of the objectives. In this experiment, the set of estimated emotions has been increased to both dimensions, valence, and arousal, extending the information sources to the set of EEG, GSR, and BVP signals. Also, the methodology has been evaluated in the context of dynamically evoked emotions through a dramatic film. An in-house production database has been created to demonstrate the feasibility of such an approach under dynamic and more realistic conditions, as will be the case with an affective-HRI. For the experiments performed in this article both applications, BIOSIGNALS (subsection 3.2.1) and GEERT (subsection 3.2.2), were used for the acquisition of multimodal signals in real-time. First, an exploratory data analysis concerning the correlations, between the emotions themselves, and the correlation between emotions and features, has been carried out for the case of the selected dramatic film. Finally, the method for classifying emotions has been evaluated using the most relevant features, using well-suited evaluation methods based on Leave-One-Out strategies.

International Journal of Neural Systems, Vol. 0, No. 0 (2019) 1–17

Real-time multi-modal estimation of dynamically evoked emotions using electroencephalography, heart rate and galvanic skin response

Mikel Val-Calvo[1,2,‡], Jose Ramón Álvarez-Sánchez[1,‡*], Alejandro Díaz-Morcillo[3] Jose Manuel Ferrández-Vicente[2] and Eduardo Fernández-Jover[4]
[1] *Dpto. de Inteligencia Artificial, UNED, Madrid, Spain*
mikel1982mail@gmail.com, jras@dia.uned.es
[2] *Dpto. de Electrónica, Tecnología de Computadoras y Proyectos*
Univ. Politécnica de Cartagena, Cartagena, Spain; jmferrv@gmail.com
[3] *Dpto. Tecnologıas de la Información y las Comunicaciones, Univ. Politécnica de Cartagena, Cartagena, Spain*
[4] *Instituto de Bioingeniería, Univ. Miguel Hernández, Alicante and CIBER-BBN, Spain;e.fernandez@umh.esm*

Emotion estimation systems based on brain and physiological signals (EEG,BVP and GSR) are gaining special attention in recent years due to the possibilities they offer. The field of human-robot interactions (HRI) could benefit from a broadened understanding of brain and physiological emotion encoding together with the use of lightweight software and cheap wearable devices, and thus improve the capabilities of robots to fully engage with the user's emotional reactions. In this paper, a previously developed methodology for real-time emotion estimation aimed for its use in the field of HRI is tested under realistic circumstances using a self-generated database created using dynamically evoked emotions. Other real-time state of the art approaches face emotion estimation using constant stimuli in order to facilitate the analysis of the evoked responses, remaining far from real scenarios, since emotions are dynamically evoked. Therefore, the proposed approach studies the feasibility of the emotion estimation methodology previously developed, under an experimentation paradigm which imitates a more realistic scenario involving dynamically evoked emotions by using a dramatic film as the experimental paradigm. The emotion estimation methodology has proved to perform on real-time constraints while maintaining high accuracy on emotion estimation when using the own-produced dynamically evoked emotions multi-signal (EEG,BVP and GSR) database.

Keywords: real-time; EEG; BVP; GSR; artifact removal; emotion estimation; HRI

1. Introduction

Different factors make recognition of emotions a challenging task. On one hand, there is no basic truth for self-evaluation, as the assessment of experienced emotions is guided by emotional models developed in the field of psychology. These can be grouped generally as discrete and dimensional models. The former assumes that emotions are qualitatively differentiated neuro-physiological responses which produce independent emotional experiences,[1] while the dimensional approach captures continuous quantified relationships among emotions.[2] However, qualitative differences arise when moving across fuzzy boundaries, between valence and arousal. In addition, there are significant variations for each individual in the correlation between the properties of the measured physiological signals and the respective emotion, therefore, the studies differ in methodology to attempt to better bridge this gap. In order to find the invariant features in the above across individuals, data exploratory analysis (EDA) is performed. However, user-adapted HRI techniques are required to face inter-subject variability, therefore, focus on subject-dependent analysis is also performed.

Studies based on off-line emotion recognition were developed,[3–6] amongst other reasons, to measure and understand how emotions are produced. This feat was accomplished, however it remained somehow unrealistic if it was not further developed into a real-time process. The development of methodologies that work under real-time constraints must deal with two main obstacles: artifact removal and accurate classification across sessions and subjects. Off-line methodologies addressed this problem by performing offline analysis with the use of supervised artifact removal techniques, obtaining high accuracy results, however, often involving the use of complex deep learning machines that require hyper-parameter tuning, taking up to several days or even weeks of preparation which are not affordable for domains of study where real-time constraints are involved. The increasing use of robots that can interact with humans is generating a greater interest on the application of machine learning techniques for the recognition of human emotions, since they offer a more complex analysis. However, simplified and affordable systems must be kept in mind when developing such methodologies. Further research was carried out and on-line or real-time approaches were successfully carried out.[7,8] This research will use a previously self-developed real-time methodology to carry out emotion estimation.

Both off-line and real-time methodologies were accomplished using a plethora of stimuli, music, images, short audio-visual clips,[8–13] however, all these remain incomplete with regards to everyday human-human interactions, as these stimuli evoked coherent and predefined emotions. These approaches use emotionally constant stimuli in order to facilitate the analysis of the evoked responses. It is therefore necessary to further evolve the technique and achieve dynamically evoked emotions, using a different type of stimulus, since real-life scenarios involve dynamic changes in the mood of the subjects. Therefore, the proposed approach is to study the feasibility of the emotion estimation methodology under an experimentation paradigm which could imitate a more realistic scenario. On this topic, films have long been developed to emotionally engage the audience with alternating emotional stimuli as the different scenes progress and tell a story. Thus, a dramatic film which had a balance between negative and positive emotions was chosen as the stimuli for the experiment.

Emotion estimation can be performed by evaluating the dynamical changes over facial expressions, body language, voice tone, EEG patterns and physiological signals, related to the equilibrium between the parasympathetic and sympathetic autonomous systems. Emotion regulation involves a series of central nervous subsystems which interact with each other producing complex behaviours. In that context, behavioural demands induce their coordination to produce changes which allow dynamic adaptation to the later. In these processes several subsystems are involved, from high to low levels of nervous activity, involving close interactions between the central and autonomic nervous systems in different ways. While the hypothalamus regulates part of the autonomic subsystems, many of the activities of the hypothalamus are, in turn, governed by certain cortical areas, as well as the central nucleus of the amygdala, which processes inputs from the external environment. The amygdala comprises several nuclei on the medial aspect of the temporal lobe, mostly anterior hippocampus and the tip of the temporal horn.[14] The amygdala receives inputs from the association cortex and its major projections are to the septal area and prefrontal cortex, mediating from emotional responses to sensory stimuli. Within this high degree of specificity in the organisation of the central nervous system (CNS), Figure 1, emotions may provide quick and reliable responses to recurrent life challenges and therefore, as a result of those synergic interactions across the CNS, respiratory and electro-dermal activity in conjunction with electroencephalographic measurements may thus provide the necessary information on emotion processing.[?,15–20] Therefore, this research will be based on a simple (NEGATIVE-NEUTRAL-POSITIVE) discrete emotional model for the case of valence dimension estimation, and (RELAXED-NEUTRAL-INTENSE) for the arousal dimension estimation, as previously carried out,[21] to further simplify the problem at hand.

The combined use of Electroencephalography (EEG), blood-volume pressure (BVP) and galvanic skin response (GSR) signals for emotion estimation have been therefore in the view point of the field for the last decades.[9,11,22,23]

Whether brain patterns evoked by emotions can be mapped onto specific brain regions still remains unresolved. In fact, current studies suggest that information encoded during emotional experiences spread over cortical and subcortical areas.[24]

There is still no clear evidence on which of the local and global distributions of brain patterns are consistent among subjects, both in dimensional and discrete emotion models. Therefore, there is not yet a consensus on the relevant brain pattern features and brain regions suitable for emotion detection, invariant across subjects. In this paper, in total 8 electrodes were used, four temporal electrodes and four prefrontal, due to the previously explained relationships between the temporal lobes, the prefrontal cortex and emotion processing.

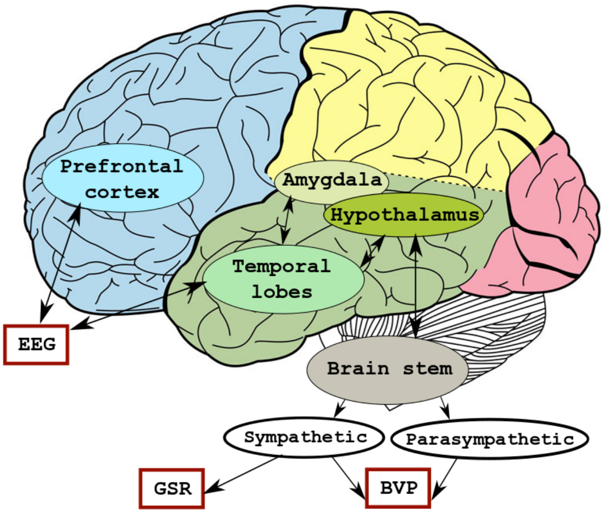

Figure 1: Graphical depiction of the inter-relations between brain areas involved in emotion processing.

In conclusion, the main aim is to focus on the feasibility of real-time emotion estimation for the case of a scenario involving dynamically evoked emotions. The developed methodology for real-time emotion recognition[25] will be used in the domain of HRI with the openBCI system for EEG signal acquisition, and the Empatica-E4 wristband for the BVP and GSR signals acquisition, taking that into account the pre-processing and feature extraction techniques must be performed with feasible real-time techniques. For validation of the methodology, the self-produced database using a dramatic film has been used.

2. Materials and Methods

The aim of the present paper is to perform the whole process involved in emotion estimation under real-time constraints for the case of dynamically evoked emotions. In order to test its feasibility, it will be tested on an own-produced multi-signal database using a dramatic film.

2.1. *Database*

In order to create a database for the experimental procedure a total of 18 volunteers, (50% female - 50% male) aged between 18 and 35, participated on the validation of such film. Participants rated arousal and valence dimensions using SAM mannequins on two discrete 3-point scales, (NEGATIVE-NEUTRAL-POSITIVE) for valence and (RELAXED-NEUTRAL-INTENSE) for arousal. Finally, for the acquisition of EEG; BVP and GSR signal respectively, in total 8 electrodes were used for the OpenBCI EEG cap, four prefrontal and four temporal, (F3, T7, P7, F7, F4, T8, P8, F8), since they have proved to be the best brain areas for emotion estimation.[3,10] Empatica E4 wristband was placed on the right hand as all of the volunteers were right-handed.

10 volunteers were selected for posterior analysis (3 females and 7 males aged between 18 and 23) as the signals obtained with 8 volunteers were severely corrupted due to noise artifacts or connection problems since the Empatica wristband was placed on their dominant hand. The participants provided their written consent and filled a health questionnaire, supervised by the Ethics Committee of the University Miguel Hernandez. All volunteers were physically and mentally healthy and were not taking any medication.

A commercial film was chosen in order to induce dynamic changes in the emotional responses of volunteers. The name of the film is "Cien Metros"(2016), which is a story of the self-acceptance and personal growth of character after overcoming a series of obstacles. "Ramón", who is detected a degenerative disease, amyotrophic lateral sclerosis, and although doctors assure him that he would not be able to walk a hundred meters, he decides to face life and train for an Iron-man, the toughest sporting event on the planet. With the help of his wife and his father-in-law, "Ramón" will undergo special training to show the world that surrender is not an option.

The film is 105 minutes long and was split into 27 different scenes based on the script structure. Scenes must be longer than one minute(when possible) in order to reduce variability in heart rate measurements, and each scene must contain a unified emotional drive. Scene length is specified in table 1.

Table 1: Film scenes specification

ID	*duration(s)*	ID	*duration(s)*	ID	*duration(s)*
1	583	10	135	19	264
2	307	11	126	20	138
3	339	12	522	21	163
4	139	13	115	22	205
5	402	14	114	23	416
6	108	15	168	24	55
7	247	16	47	25	225
8	160	17	146	26	586
9	92	18	332	27	207

Prior to film visualisation, a black screen was shown during 5 minutes in order to acquire EEG, BVP and GSR signals for a resting state, using the OpenBCI system and the Empatica E4 wristband, respectfully. The volunteers then watched the whole film, however a grey screen was shown for 6 seconds between scenes, in order to allow volunteers to rate the watched scene using the labels POSITIVE, NEGATIVE or NEUTRAL for the valence dimension, and RELAXED, NEUTRAL or TENSE for the arousal. This decision was taken in order to minimise the interruption of the visualization of the film, to affect as less as possible the development of emotions. During the whole visualisation and rating process, the EEG, GSR and BVP signals of the volunteers were acquired for posterior analysis, figure 2.

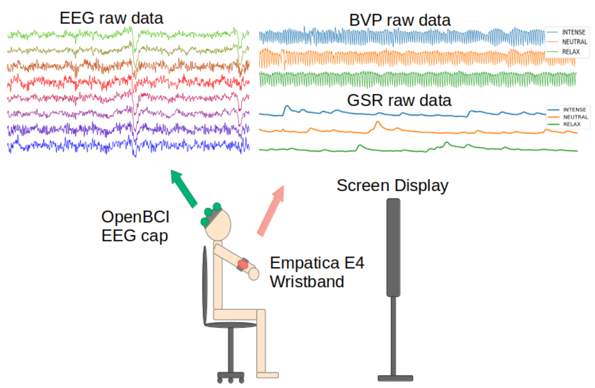

Figure 2: Experimental design picture. (**a**) Top-Left, EEG raw data acquired with the OpenBCI system. (**b**) Top-Right, GSR and BVP signals acquired with the Empatica E4 device.

2.2. *Experimental procedure*

In the following sections, processing, online artifact removal, feature extraction and performance evaluation methodologies are explained in detail.

2.2.1. *Signal processing*

BVP peak detection preprocessing The BVP signal is acquired with a photoplethysmogram sensor (PPG) to detect blood volume changes in the microvascular bed of tissue. The signal consists of the systolic peak, the dicrotic notch, and the diastolic peak. The systolic peaks are used in order to compute the inter-beat interval (IBI) time series which further allows the measurement of the balance between sympathetic and parasympathetic systems, with the use of the heart rate variability analysis.

As the morphological properties of the BVP signals vary across subjects, the process of peak detection must accomplish a series of steps. Standard methods involve the use of adaptive peak detection strategies based on the moving average computed over the raw data, where regions of interest (ROI) are selected as the signals amplitude is larger than the moving average. R-peaks are marked at the maximum of each ROI, which allows the computation of the IBI time series. Finally, detection and rejection of outliers is performed.

GSR signal preprocessing: The GSR signals is formed by two components: the skin conductance level (SCL) and the skin conductance response (SCR). The former represents slow changes in the sympathetic tone, varying from seconds to minutes. On the other hand, fast changes in the sympathetic tone are presented as the SCR component which is presented as a burst of GSR peaks. The last can be further split into two types, whether an stimuli evokes a response on the SCR, called an event-related SCR (ER-SCR), or whether it does not, called non-specific SCRs, which can happen spontaneously in the range of 1 to 3 per minute. For the present analysis only the SCL component was taken into account as a measure of the sympathetic tone during the film.

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, 64Hz, of the E4 wristband, only tonic components were analysed. The tonic component, called the skin conductance level (SCL), was obtained using a Savitzki-Golay filter (window length=31, order=2).

EEG preprocessing EEG signals are arranged in a three dimensional matrix containing n trials, c channels and s samples at a sample frequency, f_s. First, given that each signal has its own scaling factor values, signals are standardised 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 Hz to 4 Hz, 4 Hz to 8 Hz, 8 Hz to 16 Hz, 16 Hz to 30 Hz and 30 Hz to 50 Hz.

2.2.2. *Online artifact removal*

Artifact removal is necessary, as it reduces possible classification errors and reduces the amount of processed information. Thus, care must be taken while carrying out such process since valuable information in the signals could be damaged. An EEG oriented artifact removal technique (EAWICA) was used in this methodology. It was analysed and validated with EEG brain patterns in[25] under real-time conditions. On the other hand, BVP and GSR were filtered using standard techniques.

The most common artifacts presented in EEG signals are: electro-oculographic (EOG) artifacts, muscle artifacts and 50 Hz background noise. Firstly, 50 Hz background noise, which can be easily removed by a notch filter based on IIR filters. Secondly, EOG artifacts, such as blinking, are often presented within slow frequency bands, below 5 Hz,[26] while muscle artifacts are usually presented within medium to high frequency bands 20–300 Hz.[27] Therefore, muscle artifacts are partially removed outside the range of 1–50 Hz when filtering the signals, since this range includes the best frequency bands for emotion estimation: delta (1–4 Hz), theta (4–8 Hz), alpha (8–16 Hz), beta (16–30 Hz) and gamma (30–50 Hz). In fact, as several studies report,[3,10] the most effective band ranges for emotion estimation, are beta and gamma bands.

Finally, with regard to EOG artifacts, these can be effectively removed with EAWICA wich employs a divide and conquer strategy in order to isolate the artifacts both in time-frequency domain through the wavelets transform decomposition and by analysing the independent components sources (ICA). The original EAWICA method proposes the isolation approach of the artifactual signal component by first computing the wavelet components over the EEG signals within the frequency ranges associated to the emotion estimation task. Thus, once the information is band pass filtered, ICA decomposition is applied in order to isolate artifactual data in a series of WICS. In order to automatically detect artifactual WICS, Kurtosis and Renyi entropy measurements are used. Those marked as artifactual are further split into a series of time windows with a temporal interval of one second defined as epochs, which in case of being marked as artifactual, are zeroed with the aim of removing as less information as possible. Then, ICA reconstruction followed by wavelet components addition is performed to reconstruct the original signals.[28] EAWICA was modified to allow the variation of the thresholds by manually adjusting them in terms of the quartiles over the distribution values and restricted to perform over delta band[25] in order to remove effectively EOG artifacts while preserving as much information as possible for higher band ranges.

2.2.3. *Feature extraction*

The specifics of the feature selection and extraction processes for each type of signal are detailed amongst other necessary processes such as feature smoothing, feature scaling, dimensionality reduction and the set of selected classifiers.

Feature selection

(1) Population based EDA was done in order to find relevant correlations between the experienced emotions and a set of features computed. For this analysis, the features were computed taking into account the full time series corresponding to each scene for each signal type and zscored with respect to the baseline measure's. Therefore, for each scene, the window length used to compute each feature was equal to the length of

each scene, this was also done for the baseline recordings.

(2) Subject independent classification, consist on splitting information carried on each scene in sliding windows to compute a set of features. Thus, two independent classification processes were done in order to test the feasibility of the emotion estimation under such a dynamic emotion evoked context. The first process was performed with the aim of discriminating the emotions evoked in the valence dimension with the use of EEG signals, while the second one was performed for the arousal dimension with the use of both BVP and GSR signals.

   (a) Valence classification with EEG signals: All features have been computed using a sliding window of 6 seconds as suggested by[29] with three seconds overlapping. Features are computed for each band/channel and later concatenated for each training sample. Four consecutive samples were concatenated in order to provide temporal order information to the classifiers. Therefore, each training sample represents the computed features for four overlapped time windows.
   (b) Arousal classification with BVP and GSR signals: All features have been computed using a sliding window of 60 seconds, as most heart rate variability meassures cannot be computed reliable taking into account shorter durations, with thirty seconds overlapping. Features are concatenated for each training sample. Two consecutive samples were concatenated in order to provide temporal order information to the classifiers. Therefore, each training sample represents the computed features for two overlapped time windows. For the scenes with shorter lengths, the total scene length was taking into account for features computation and the resulting sample was concatenated with itself to be coherent with the rest of the training data shape.

BVP and GSR Feature extraction methodology for arousal estimation A set of features are computed based on the properties of IBI and SCL time series:

- GSR:
  - OFFSET: Average of the series of amplitude values.
  - SLOPE: Average slope of the series of amplitude values.
  - STD: Standard deviation of the series of amplitude values.
- BVP:
  - Mean Heart Rate: Average of heart rate.
  - SDNN: The Standard Deviation of a NN interval series.
  - RMSSD: Root Mean Square of the Successive Differences.
  - SDSD: Standard deviation of NN differences.
  - NN20: Number of NN interval differences greater 20 milliseconds.
  - PNN20: Ratio between NN20 and total number of NN intervals.
  - NN50: Number of NN interval differences greater 50 milliseconds.
  - PNN50: Ratio between NN50 and total number of NN intervals.
  - Triangular index (`TRI_INDEX`): The ratio between the total number of NNs and the maximum of the NN histogram distribution.
  - Low Frequency (LF): The power density estimation for the frequency band in the range [0.04, 0.15]Hz.
  - High Frequency (HF): The power density estimation for the frequency band in the range [0.15, 0.40]Hz.
  - Sample Entropy (SAMPEN): Used for assessing the complexity of NN interval series.

  Heart rate variability measurements were computed with the pyhrv python library.[30]

EEG Feature extraction methodology A set of features are computed based on the oscillatory properties of brain signals:

- Differential Entropy (DE): Is computed as a metric for measuring the predictability of signal X, whose values have a probability density function similar to a Gaussian distribution, $\mathrm{N}(\mu, \sigma^2)$, as is the case for EEG signals. It can be defined as $h(X) = \frac{1}{2}\log\left(2\pi e\sigma^2\right)$.
- Amplitude Envelope (AE): Computed by means of the Hilbert transform.[31]
- Petrosian Fractal Dimension (PFD): 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.[32]

- Higuchi Fractal Dimension (HFD): Higuchi's algorithm can be used to quantify the complexity and self-similarity of a signal.[33]
- Fisher Information (FI): 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.[34]

AE has been computed with the Neuro Digital Signal Processing Toolbox[35] python library developed at Voytek's Lab. PFD, HFD and FI have been computed with the PyEEG python library.[36]

Feature smoothing In the field of EEG, emotions are often considered as static, in order to simplify data processing for the classifiers, albeit continuous and subtle changes should be considered in the time domain, as it has been previously noticed that considering the temporal dependence and variation of emotions during the stimuli, improves the performance of the training step.[3,6,10] To carry this out, smoothing the feature space helps filter out those components which are unrelated with emotional states, thus becoming a key step for the design of an optimal methodology. In other words, 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. A Savitzky-Golay (SG) filtering method is used as it demonstrated to outperform in the classification accuracy reports above the results obtained without smoothing the feature space,[3] and in addition, SG smoothing is significantly faster while preserving high accuracy results.[25]

Feature scaling Feature scaling is a key step in preprocessing data. Outliers can severely damage the performance of the classifiers while looking for statistical differences. Moreover, some machine-learning algorithms used in both dimensionality reduction and classification processes, require data to have a predefined range of values. The process of scaling data must be performed taking into account both constraints, in order to properly feed the algorithms in the next steps. In this paper, a Quantile-Transform method (histogram equalisation to uniform distribution) followed by the Min/Max scaling method is performed. The former is a non-linear method for scaling data distributions which is robust with outliers. The later allows re-scaling in a positive range of values $[0-1]$, as the proposed dimensionality reduction method requires positive values as an input.

Dimensionality reduction Wrapper and embedded methods combine feature selection and the learning process by the use of nested cross validation schemes, however, this leads to biased results when taking into account the particularities of time series. Therefore, a χ^2 feature selection technique has been chosen as it is a filter-based method were the selection of features is based on the lack of independence between a feature and the target, without involving any biased cross-validation scheme, nor combining the selection and learning processes.

Set of classifiers Classification process have been performed using a set of 8 classifiers: K-nearest neighbours, 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 with default hyper-parameter values. The Scikit-learn python library[37] has been used.

2.2.4. *Performance evaluation*

In order to ensure a correct and unbiased performance, the most important point is to ensure that samples in the validation set are reasonably independent from the samples in the training set.[38,39] Therefore, Leave-One-Out (LOO) strategy was used in this paper to asses the correct performance of the methodology and the 'macro average $F1_{score}$ metric was used.

3. Results

As mentioned in the experimental procedures the analysis of the results is split into two main questions. First,EDA was carried out to explore whether the computed features correlate with the evoked emotions, in order to find the relevant relationships. Second, the feasibility of this real-time emotion estimation strategy under a subject dependent paradigm and realistic conditions.

3.1. *Exploratory population-based data analysis*

The first step, in order to study the viability of such an approximation using dynamic stimuli, requires the observation of the voting distribution over the previously selected scenes. This analysis has been performed taking into account the ratings of all 18 volunteers that watched the dramatic film, with the aim of validating the film, as a proper stimulus which evokes the desired set of emotions. Figure 11 shows the voting distribution over the selected scenes, from which the dynamics of the emotions evoked in the volunteers by the film can be deduced. This distribution can be taken as a finger print of the emotional engaging properties of this specific stimulus. As observed in the figure, several scenes were coherently rated by all the volunteers to correspond to a specific evoked emotion. Regarding the arousal dimension, it can be observed that the film was designed to induce intense emotions in the beginning, as the character in the movie is faced against a personal and crucial problem, and in the end, were he achieves a personal objective. On the other hand, the neutral-arousal is mainly dominant over relaxed-arousal, except in scene 19. Taking that into account, for the arousal model training process, neutral and relaxed labels were considered equally.

In order to better describe the film properties regarding the evoked emotions, figure 3 shows the frequency distribution for all volunteers. In the arousal dimension, neutral dominates over relaxed. In the valence dimension, the distribution of the negative and positive emotions leans mainly towards positive, however, considering the previous figure 11, there are enough negative and coherent scenes in order to test the feasibility of statistical model.

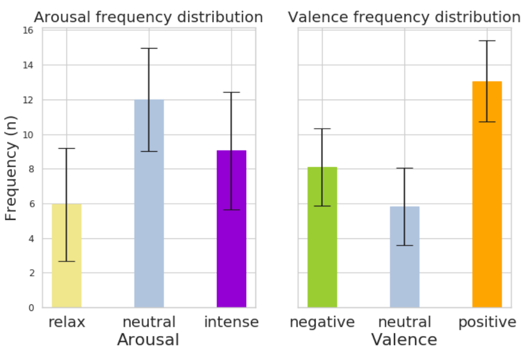

Figure 3: Frequency distribution in the whole film for all participants. (**a**) Left, shows the frequency rating for the arousal dimension. (**b**) Right, shows the frequency rating for the valence dimension.

An interesting question regarding the dynamics of emotions is to analyse whether the target emotions are correlated between them along the chosen emotional model. Of course, this must be considered as a finger print for the selected film, as different stimuli could develop different correlations. In order to properly check the correlation for the time series given as the rating of all participants over scenes, for each pair of emotions, taking into account both the arousal and valence dimensions, the Spearman correlation has been computed. As expected, figure 4 shows that positive and negative emotions in the valence dimension, are anti-correlated with a value of $Correlation = -0.82$, however, as shown in figure 11, neutral-arousal is the dominant emotion against relaxed-arousal, thus the anti-correlation is higher for the case of neutral-arousal versus intense-arousal $correlation = -0.86$ than relaxed-arousal versus intense-arousal $correlation = -0.54$. Positive-valence and relaxed-arousal are correlated $correlation = 0.66$ while relaxed-arousal and negative-valence are anti-correlated $correlation = -0.77$. On the other hand, intense-arousal is correlated with negative-valence $correlation = 0.46$ while it is not correlated to the positive-valence $correlation = -0.08$. Finally, both neutral-arousal and neutral-valence have a high correlation $correlation = 0.71$.

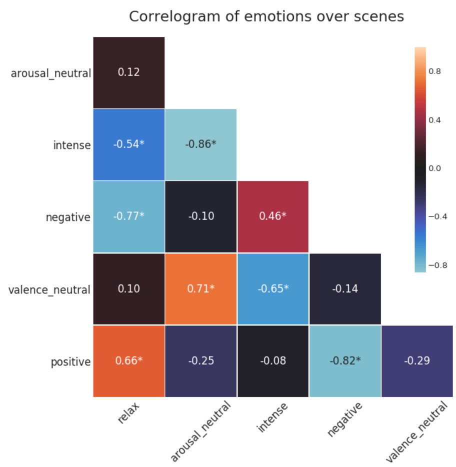

Figure 4: Correlogram of emotions ratings over scenes for each pair of emotions, taking into account both the valence and arousal dimensions. Correlation values labelled with (*) are statistically significant $P_{value} < 0.05$.

Following EDA analysis, the correlation of GSR, BVP and EEG computed features over the scenes, with regards to the majority rating of emotions is shown. This analysis has been performed taking into account the signals of the set of 10 selected volunteers, as their physiological signals were not severely corrupted by noisy artifacts or connection problems.

Figure 5 shows the correlation between OFFSET, SLOPE and STD features for the GSR signals. OFFSET is mainly correlated with intense-arousal $correlation = 0.49$, SLOPE is mainly anti-correlated with neutral-arousal $correlation = -0.40$ and OFFSET is mainly anti-correlated with relaxed-arousal $correlation = -0.40$. Std does not show any statistically significant correlation with any of the arousal emotions. From this figure, each emotion seems to be described by the selected set of features as they provide different correlations for each feature, meaning that the emotional responses could be modelled regarding the tendencies reflected on the GSR signal.

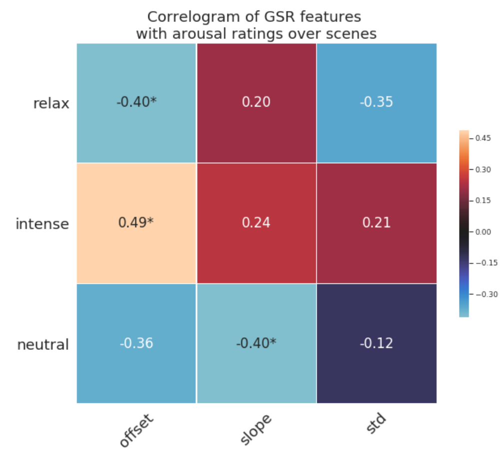

Figure 5: Correlogram between the set of features over scenes computed with the GSR signals regarding the arousal evoked emotions. Correlation values labelled with (*) are statistically significant $P_{value} <$ 0.05.

Figure 12 shows the correlation between OFFSET and SLOPE features of the GSR signals with the valence and arousal majority ratings over time.

Figure 8 shows the correlation between the whole set of selected features for the BVP signals regarding the majority arousal ratings over scenes. As it can be noted, the three arousal emotions could be distinguishable taking into account different subsets of features for each. As a case example, neutral-arousal mainly correlates with HF $correlation = 0.42$ while being mainly anti-correlated for intense-arousal $correlation = -0.45$, finally, relaxed-arousal is mainly anti-correlated with RMSSD $correlation = -0.38$.

Figure 13 shows the correlation between HF and RMMSD of the BVP signals with the arousal majority emotions ratings over time.

Regarding EEG features, Figures 9 and 10 show the correlation between the whole set of selected features for EEG signals with respect to the arousal ratings over scenes. Three main insights can be easily deduced from these figures. First, high correlated features are mostly accumulated in the beta and gamma frequency ranges. Positive-valence and negative-valence for all used electrodes, HFD feature for the beta and gamma bands, is correlated and anti-correlated respectively. Alternatively, neutral-valence does not show any relevant correlation with HFD but achieves its highest correlation values for the F7, F3, F4 and P8 electrodes for the FI feature, in the beta frequency range. Secondly, the correlation of features over scenes with positive-valence and negative-valence are mostly anti-correlated. Third, only negative-valence shows a significant correlation with the selected features for the slowest delta, theta and alpha frequency ranges as it can be noted for the several features, in all electrodes and in both hemispheres.

Figure 14 shows only one of those features that maximised the correlation or anti-correlation, on each electrode and frequency range, for each emotion.

3.2. *Classification of emotions using subject-dependent paradigm*

Valence emotion estimation was performed using EEG signals while for arousal estimation, BVP and GSR signals were used. As mentioned before in the methods section, LOO validation methodology was used to test the performance of the model, by selecting 2 trials corresponding to divergent emotions. A feature selection algorithm was used in order to evaluate the stability of the classification process using different subsets of features. The $F1_{score}$ (macro average) metric was used as it shows the highest penalisation when the performance of the trained model on the test set is unbalanced in favour of one of the targeted emotions. The classification process was performed by doing 20 iterations, on each iteration two randomly selected trials were chosen, belonging to divergent emotions in each case, from valence and arousal.

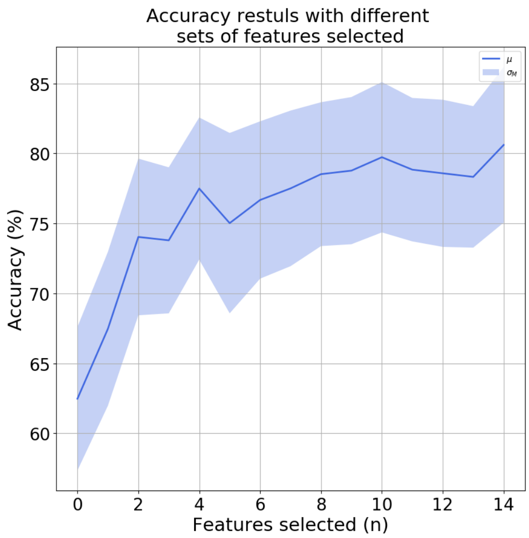

Figure 6: Arousal classification accuracy results using macro average $F1_{score}$ validation metric. The performance was evaluated with different subsets of features.

Figure 6 shows the $\mu \pm \sigma_M$ accuracy results for the classification of relaxed-arousal and neutral-arousal versus intense-arousal using a set of features ranging from a minimum of 1 to all the computed features, for all the volunteers selected for analysis.

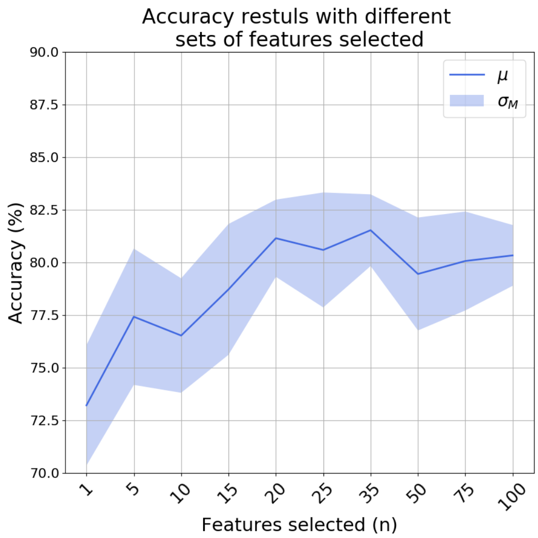

Figure 7: Valence classification accuracy results using $F1_{score}$ validation metric. The performance was evaluated with different subsets of features.

Figure 7 shows the $\mu \pm \sigma_M$ accuracy results for the classification of positive-valence versus negative-valence using a set of features ranging from a minimum of 1 to 500 computed features, using all the volunteers selected for analysis.

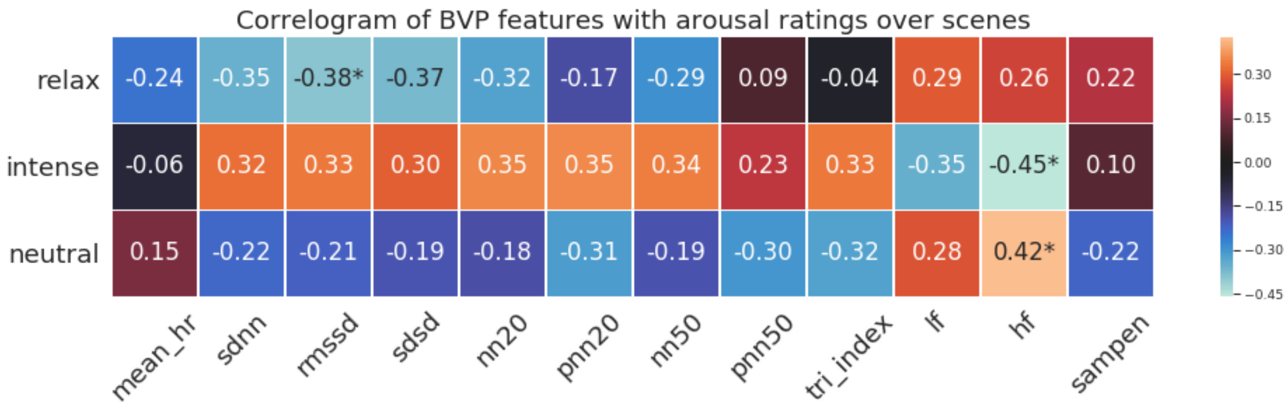

Figure 8: Correlogram between the set of features over scenes computed with the BVP signals regarding the arousal evoked emotions. Correlation values labelled with (*) are statistically significant $P_{value} < 0.05$.

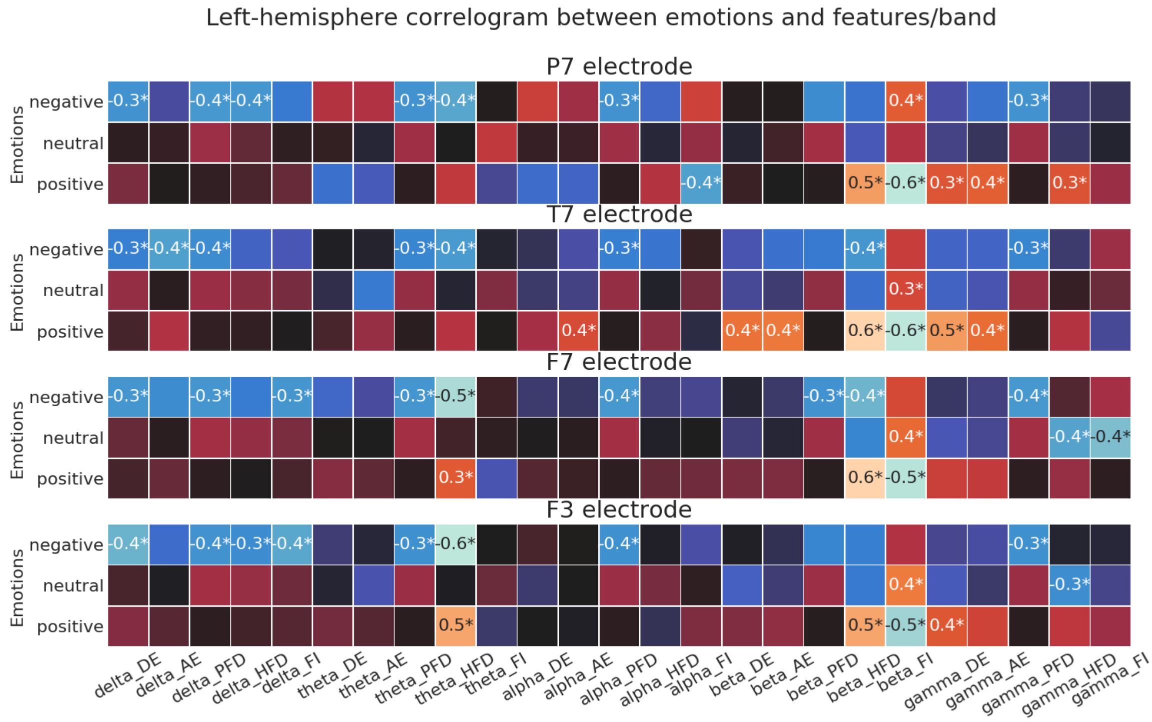

Figure 9: Left hemisphere correlogram between emotions and EEG computed features for each electrode and frequency range. Squares labelled with (*) are statistically significant $P_{value} < 0.05$.

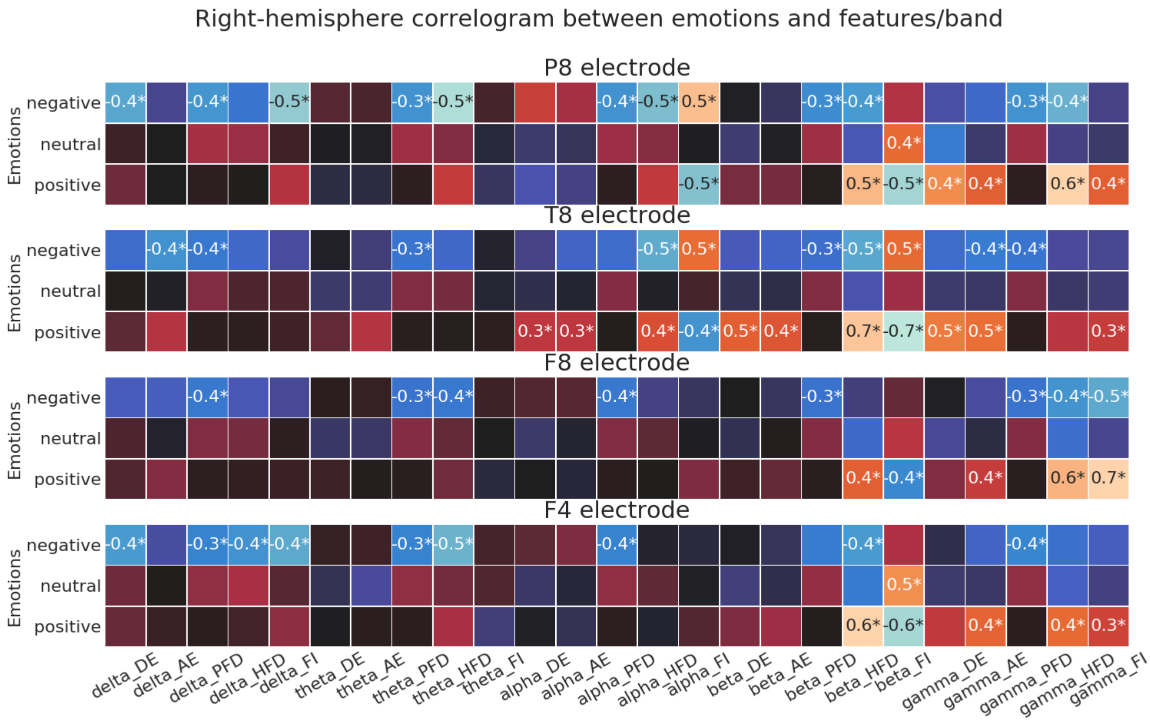

Figure 10: Right hemisphere correlogram between emotions and EEG computed features for each electrode and frequency range. Squares labelled with (*) are statistically significant $P_{value} < 0.05$.

4. Discussion

As expected, emotion estimation is not a straight forward task, it requires the implementation of several disciplines which have to be gracefully coordinated to achieve a successful outcome. The first of these barriers is the proper definition of an emotional model, which must include a range of emotions. As explained previously, this work has used a discrete emotion model including both the arousal and valence dimensions which have been previously used in literature.[10] The oversimplification of the emotional model at hand, was required in order to simplify the analysis under such a novel experiment.

The second barrier is a more physical one, signal acquisition and the apparatus. In order to obtain accurate results the acquisition apparatus must be reliable and have a constant performance. Out of the 18 volunteers, 10 recordings were of sufficient quality to perform any type of analysis on them. After a thorough study, both OpenBCI and Empatica E4 systems showed to be stable and consistent across subjects, and the erroneous signals were due to movements caused by the volunteers as the Empatica E4 wristband was placed on the dominant hand which users used to annotate emotional labels. Further and more detailed descriptions of the experimental procedure, together with a clear statement of the importance of the procedure steps could reduce the problem in the future.

Overall, using the reliable acquired data, the experimental procedure showed some expected results. Figure 3 shows the film has a varied distribution of all types of arousal and valence according to ratings carried out by the volunteers across all scenes, and was therefore a proper choice as a stimuli for the desired experiment. Figure 11, shows expected arousal and valence responses according to the film contents, since at the start the film exposes the main problem for the protagonist to encounter, respectively coinciding with both intense-arousal and negative-valence peaks; as the film progresses and the plot develops, both neutral-arousal and neutral-valence prevail; as the plot shows some resolve a relaxed-arousal peak and a positive-valence peak coincide; finishing, as the plot finds its resolve, with mostly intense-arousal and a mixture of positive and negative valence, coinciding with the mixture of resolve and sadness intended by the film director. These outcomes together with Figure 4, imply both expected and interesting results. Taking into account the contents and context of the film, which in a simplified manner is a dramatic film which offers a bad or negative starting situation, followed by an intense overcoming process, and ending with a relieving accomplishment of the protagonists goals with a constant reminder of the constant presence of the obstacle. Following the expected sensations that the film director intends, Figure 4 shows, as expected, that positive and negative valence are highly anti-correlated, and negative-valence was mostly related to intense-arousal while positive-valence was mostly related to relaxed-arousal, meaning positive and relaxed feelings appear after an intense and negative phase while as the protagonist faces and overcomes the obstacles.

Figure 5 and Figure 12 shows that for GSR features, each arousal state has a different correlation sequence. Neutral-arousal is anti-correlated to all features, intense-arousal is correlated with all features, and relax-arousal is correlated with one and anti-correlated with two features. This could hint at a specific response of the GSR features for each arousal state as previous studies suggest.[40] However, higher statistical significance and further experiments would be required before drawing any definitive conclusions.

Figure 8 and Figure 13 shows that for BVP features, neutral and intense arousal estates are correlated and anti-correlated respectively with HF, however the relaxed state is anti-correlated with RMSSD. Intense and relaxed states were expected to have opposite correlations with the same feature, but this was not the case. Further analysis and in-depth experiments are required in order to draw a proper conclusion, however, these findings could be the beginning of an more complete understanding of emotions.

Figures 9 and Figure 10 allow us to draw the following conclusions: First, highly correlated features are mostly accumulated in the beta and gamma frequency ranges for positive-valence and negative-valence while neutral-valence achieves its highest correlation values for the beta frequency range, hinting at a possibly different behaviour of the neural flow. Secondly, positive-valence and negative-valence opposite correlations for all features. Finally, only negative-valence shows a significant correlation with the selected features for the slowest delta, theta and alpha frequency ranges, in all electrodes and in both hemispheres. All those findings are mostly coherent with previous results,[3] regardless of the experimental paradigm and culture differences. In addition, there is no evidence to draw any conclusion about lateralization of emotions from the results obtained by this experiment.

Regarding each type of emotion, EDA analysis provided an insight of the relevant set of features for the description of physiological changes. However, those correlations are highlighted by the use of the population of volunteers, but for the case of a subject dependent paradigm, the inter-subject variability makes emotion estimation a harder task. A feature selection algorithm was therefore used in order to evaluate the stability of the classification process using different subsets of features. Furthermore, the main purpose in the analysis is to evaluate the feasibility of the emotion estimation methodology under the paradigms of dynamic emotions but not to optimise the classification accuracy, as this would require further research into the models and into hyper-parameter optimisation. Thus, the aforementioned machine learning models were built without hyper-parameter optimisation to perform emotion estimation, both for valence and arousal dimensions.

Valence emotion estimation was performed using EEG signals, while for arousal estimation BVP and GSR signals were used. As mentioned before in the methods section, LOO validation methodology was used for testing the performance of the model, selecting 2 trials corresponding to divergent emotions. However, such a process would be optimistic if only one iteration is performed, as the process of selecting those 2 trials could be statistically biased in favour of the trained model. Thus, the classification process is performed by doing 20 iterations, where on each iteration two randomly selected trials were chosen, ensuring they belong to divergent emotions on each case, from the valence and arousal classification processes. Therefore, on each iteration, the performance of the best classifier from the set of eight was chosen for performance evaluation, as done in.[25] The $F1_{score}$ is chosen since, when the performance of the test set is biased towards a specific result, it shows the highest penalisation, and as it is desirable to build models that properly generalise over the objective task, this was the best option.

EEG artifact removal and successive feature processing steps were previously probed[25] to perform in real-time constraints. In addition, the processes involved for both, BVP peak detection and GSR preprocessing and feature processing, does not require heavy computations. Therefore, the proposed methodology can be applied in real-time scenarios with meaningful results taking into account a near realistic scenario.

Several studies provide useful insights when measuring the properties of functional connectivity under different scenarios.[41–45] Although the proposed approach has not taken into account a broader study of functional connectivity properties during the evolution of emotions. Future work will include such an analysis to evaluate the connectivity properties during perceived emotions in order to better understand the dynamics and correlations between functional connectivity and physiological properties over time.

Finally, the next research step will make use of the proposed methodology under a HRI paradigm to study whether the robot will adapt its behaviour based on the estimated emotion.

5. Conclusions

This paper has introduced a method which has proved to be feasible under the assumption of dynamically evoked emotions in the subject dependent analysis for EEG, BVP and GSR emotion recognition. An accurate and computationally light emotional estimation methodology could allow the use of portable and cheap devices in domain of emotional HRI where more realistic scenarios are involved.

This method uses three categories emotional model, however, for the emotion estimation task only two of them were used for each dimensional model. Nevertheless, LOO validation scheme probed the feasibility of such approach although the classifiers were not tuned to optimise the accuracy results.

Finally, population based EDA analysis provide useful insights on which features are best for determining the recognition of emotions.

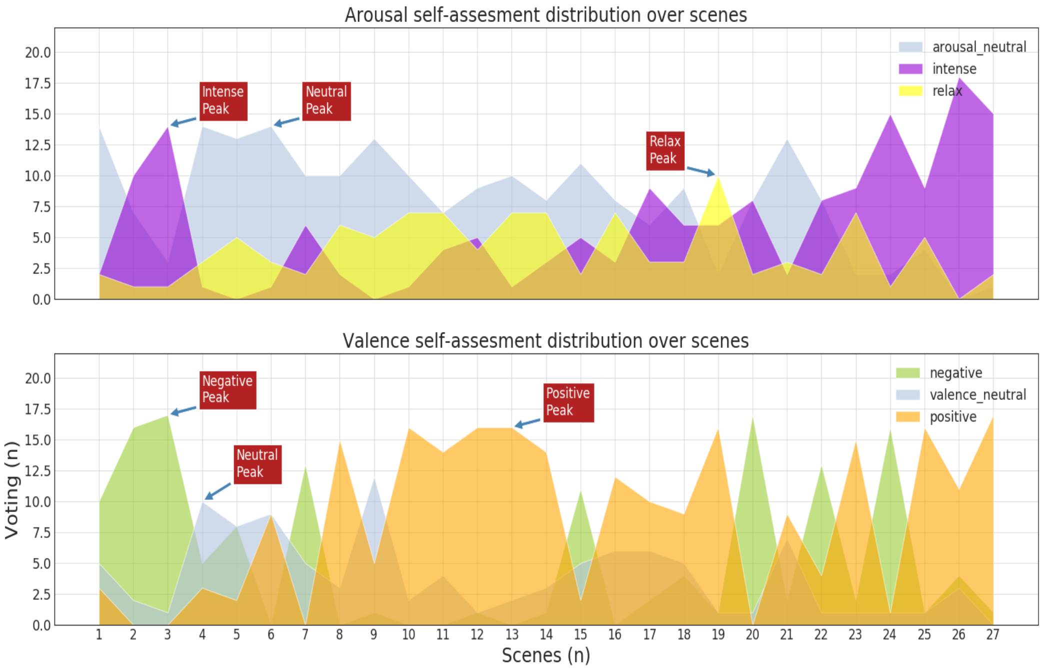

Figure 11: Rating distribution over selected scenes for all participants. (**a**) Top, shows the rating over time for the arousal dimension. (**b**) Bottom, shows the rating over time for the valence dimension.

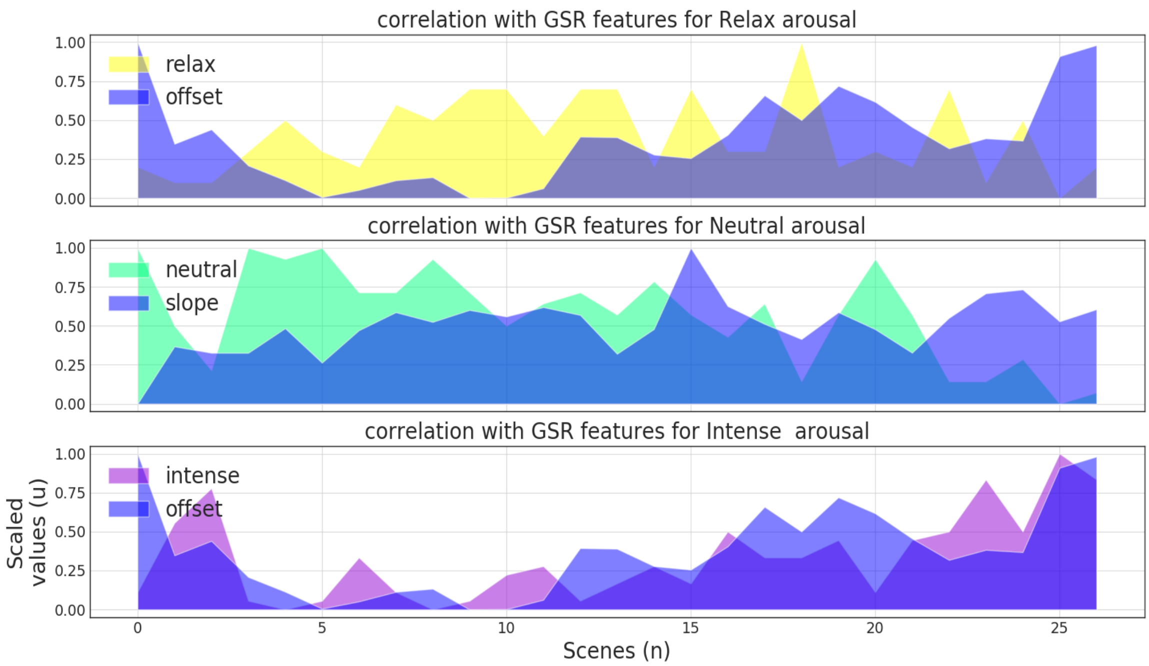

Figure 12: Correlation between OFFSET and SLOPE features of the GSR signals with the arousal majority ratings over time.

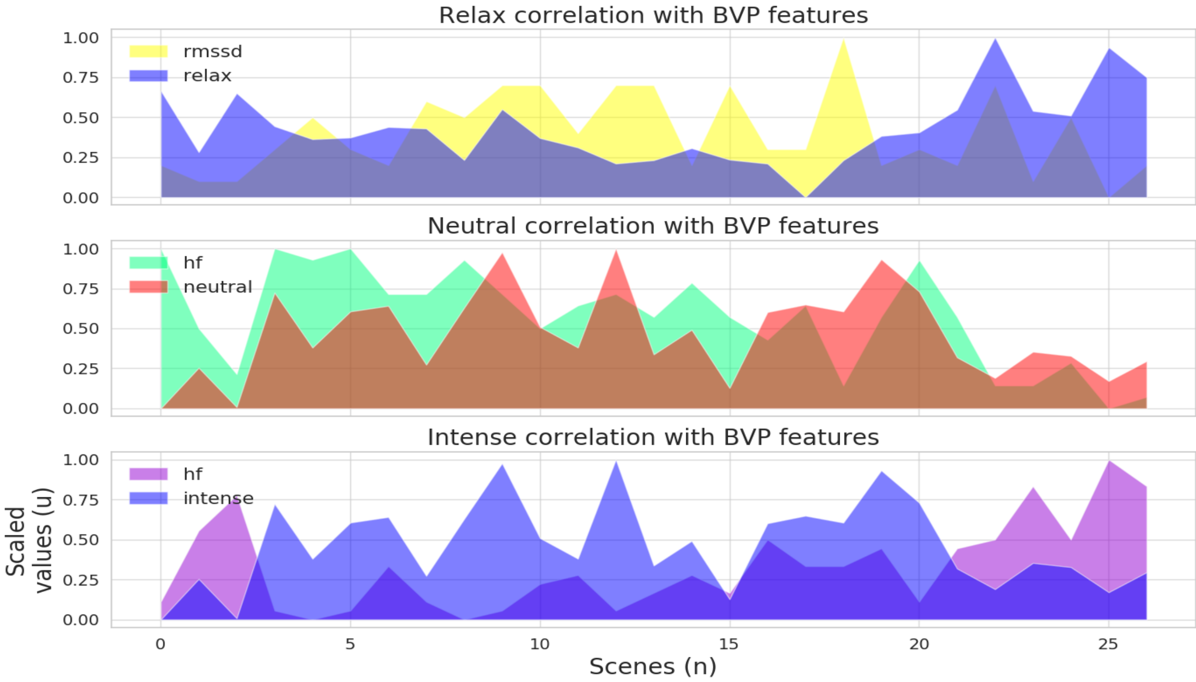

Figure 13: Correlation between HF and RMMSD features of the BVP signals with the arousal majority ratings over time.

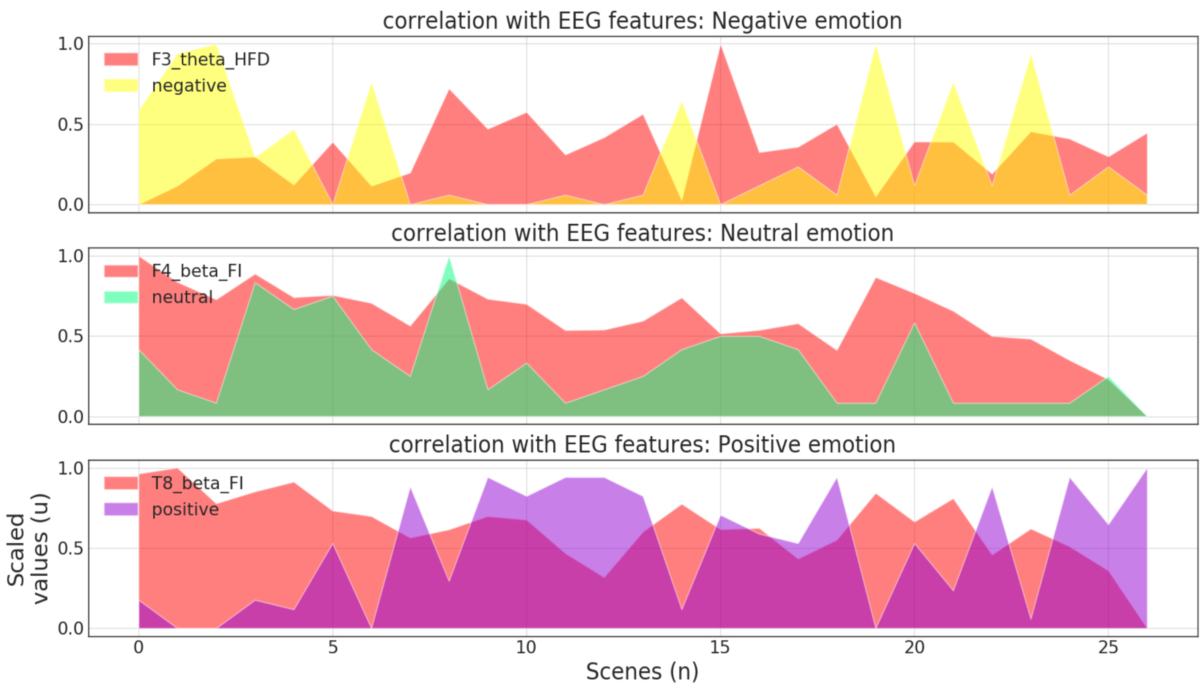

Figure 14: EEG most correlated or anti-correlated features across the scenes.

6. Acknowledgments

We want to acknowledge to Programa de Ayudas a Grupos de Excelencia de la Región de Murcia, from Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia and RTI2018-098969-B-100 from the Ministerio de Ciencia, Innovación y Universidades, y PROMETEO/2019/119 from the Generalitat Valenciana.

Bibliography

1. I. J. Roseman, M. S. Spindel and P. E. Jose, Appraisals of emotion-eliciting events: Testing a theory of discrete emotions., *Journal of personality and social psychology* **59**(5) (1990) p. 899.
2. J. A. Russell, A circumplex model of affect., *Journal of personality and social psychology* **39**(6) (1980) p. 1161.
3. W.-L. Zheng, J.-Y. Zhu and B.-L. Lu, Identifying stable patterns over time for emotion recognition from eeg, *IEEE Transactions on Affective Computing* (2017).
4. W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu and A. Cichocki, Emotionmeter: A multimodal framework for recognizing human emotions, *IEEE transactions on cybernetics* (99) (2018) 1–13.
5. S. Chen and Q. Jin, Multi-modal dimensional emotion recognition using recurrent neural networks, *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*, ACM2015, pp. 49–56.
6. W.-L. Zheng, Y.-Q. Zhang, J.-Y. Zhu and B.-L. Lu, Transfer components between subjects for eeg-based emotion recognition, *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, IEEE2015, pp. 917–922.
7. Y. Liu, O. Sourina and M. K. Nguyen, Real-time eeg-based human emotion recognition and visualization, *2010 international conference on cyberworlds*, IEEE2010, pp. 262–269.
8. Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge and Y. Shi, Real-time movie-induced discrete emotion recognition from eeg signals, *IEEE Transactions on Affective Computing* **9**(4) (2017) 550–562.
9. S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt and I. Patras, Deap: A database for emotion analysis; using physiological signals, *IEEE transactions on affective computing* **3**(1) (2011) 18–31.
10. W.-L. Zheng and B.-L. Lu, Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks, *IEEE Transactions on Autonomous Mental Development* **7**(3) (2015) 162–175.
11. M. Soleymani, J. Lichtenauer, T. Pun and M. Pantic, A multimodal database for affect recognition and implicit tagging, *IEEE Transactions on Affective Computing* **3**(1) (2011) 42–55.
12. R. Horlings, D. Datcu and L. J. Rothkrantz, Emotion recognition using brain activity, *Proceedings of the 9th international conference on computer systems and technologies and workshop for PhD students in computing*, ACM2008, p. 6.
13. S. Aydin, S. Demirtaş, K. Ateş and M. A. Tunga, Emotion recognition with eigen features of frequency band activities embedded in induced brain oscillations mediated by affective pictures, *International journal of neural systems* **26**(03) (2016) p. 1650013.
14. J. Kiernan, Anatomy of the temporal lobe, *Epilepsy research and treatment* **2012** (2012).
15. D. Hagemann, S. R. Waldstein and J. F. Thayer, Central and autonomic nervous system integration in emotion, *Brain and cognition* **52**(1) (2003) 79–87.
16. 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* **102**(4) (2019) 887–898.
17. G. Girardeau, I. Inema and G. Buzsáki, Reactivations of emotional memory in the hippocampus–amygdala system during sleep, *Nature Neuroscience* **20**(11) (2017) p. 1634.
18. P. J. Lang, M. K. Greenwald, M. M. Bradley and A. O. Hamm, Looking at pictures: Affective, facial, visceral, and behavioral reactions, *Psychophysiology* **30**(3) (1993) 261–273.
19. J. L. Andreassi, *Psychophysiology: Human behavior and physiological response* (Psychology Press, 2010).
20. A. R. Damasio, Emotion in the perspective of an integrated nervous system, *Brain research reviews* **26**(2-3) (1998) 83–86.
21. M. B. H. Wiem and Z. Lachiri, Emotion sensing from physiological signals using three defined areas in arousal-valence model, *2017 International Conference on Control, Automation and Diagnosis (ICCAD)*, IEEE2017, pp. 219–223.
22. S. D. Kreibig, Autonomic nervous system activity in emotion: A review, *Biological psychology* **84**(3) (2010) 394–421.
23. Z. Yin, M. Zhao, Y. Wang, J. Yang and J. Zhang, Recognition of emotions using multimodal physiological signals and an ensemble deep learning model, *Computer methods and programs in biomedicine* **140** (2017) 93–110.
24. P. A. Kragel and K. S. LaBar, Decoding the nature of emotion in the brain, *Trends in cognitive sciences* **20**(6) (2016) 444–455.
25. 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 in Computational Neuroscience* (2019).
26. M. S. B. A. Rani and W. Mansor, Detection of eye blinks from eeg signals for home lighting system activation, *2009 6th International Symposium on Mechatronics and its Applications, ISMA 2009*, 2009, p. 5164828.
27. S. Muthukumaraswamy, High-frequency brain activity and muscle artifacts in meg/eeg: a review and recommendations, *Frontiers in human neuroscience*

7 (2013) p. 138.

28. N. Mammone and F. Morabito, Enhanced automatic wavelet independent component analysis for electroencephalographic artifact removal, *Entropy* **16**(12) (2014) 6553–6572.
29. H. Candra, M. Yuwono, R. Chai, A. Handojoseno, I. Elamvazuthi, H. T. Nguyen and S. Su, Investigation of window size in classification of eeg-emotion signal with wavelet entropy and support vector machine, *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, IEEE2015, pp. 7250–7253.
30. P. M. Pedro Gomes, Hugo Silva, pyHRV - opensource python toolbox for heart rate variability (2018–), [Online; accessed ¡today¿].
31. B. Boashash, Estimating and interpreting the instantaneous frequency of a signal. i. fundamentals, *Proceedings of the IEEE* **80** (April 1992) 520–538.
32. A. Petrosian, Kolmogorov complexity of finite sequences and recognition of different preictal eeg patterns, *Proceedings Eighth IEEE Symposium on Computer-Based Medical Systems*, IEEE1995, pp. 212–217.
33. A. Accardo, M. Affinito, M. Carrozzi and F. Bouquet, Use of the fractal dimension for the analysis of electroencephalographic time series, *Biological cybernetics* **77**(5) (1997) 339–350.
34. R. A. Fisher, Theory of statistical estimation, *Mathematical Proceedings of the Cambridge Philosophical Society*, **22**(5), Cambridge University Press1925, pp. 700–725.
35. VoytekLab, Neuro digital signal processing toolbox *https://github.com/neurodsp-tools/neurodsp , (2018).*
36. *F. S. Bao, X. Liu and C. Zhang, Pyeeg: an open source python module for eeg/meg feature extraction,* Computational Intelligence and Neuroscience ***2011*** *(2011) Article ID 406391.*
37. *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,* Journal of Machine Learning Research ***12*** *(2011) 2825–2830.*
38. *L. J. Tashman, Out-of-sample tests of forecasting accuracy: an analysis and review,* International journal of forecasting ***16****(4) (2000) 437–450.*
39. *C. Bergmeir and J. M. Benítez, On the use of cross-validation for time series predictor evaluation,* Information Sciences ***191*** *(2012) 192–213.*
40. *A. M. Lal and T. Narula, Emotion recognition using skin conductance and sentiment analysis.,* International Journal of Advanced Research in Computer Science ***10****(4) (2019).*
41. *R. Yuvaraj, M. Murugappan, U. R. Acharya, H. Adeli, N. M. Ibrahim and E. Mesquita, Brain functional connectivity patterns for emotional state classification in parkinson's disease patients without dementia,* Behavioural brain research ***298*** *(2016) 248–260.*
42. *M. Ahmadlou and H. Adeli, Functional community analysis of brain: A new approach for eeg-based investigation of the brain pathology,* Neuroimage ***58****(2) (2011) 401–408.*
43. *Y. Li, D. Cao, L. Wei, Y. Tang and J. Wang, Abnormal functional connectivity of eeg gamma band in patients with depression during emotional face processing,* Clinical neurophysiology ***126****(11) (2015) 2078–2089.*
44. *M. Ahmadlou, A. Adeli, R. Bajo and H. Adeli, Complexity of functional connectivity networks in mild cognitive impairment subjects during a working memory task,* Clinical Neurophysiology ***125****(4) (2014) 694–702.*
45. *J. delEtoile and H. Adeli, Graph theory and brain connectivity in alzheimer's disease,* The Neuroscientist ***23****(6) (2017) 616–626.*

4.3 Article 3: Affective Robot Story-telling Human-Robot Interaction

Title: Affective Robot Story-telling Human-Robot Interaction: Exploratory Real-time Emotion Estimation Analysis using Facial Expressions and Physiological Signals

Authors: Mikel Val-Calvo, José R. Álvarez-Sánchez, Jose M. Ferrández-Vicente and Eduardo Fernández

Journal: IEEE Access, 8: 134051–134066

Impact factor: 3.745, JCR2019-Q1, 35/156 Computer Science, Information Systems

Published: July 2020

DOI: 10.1109/ACCESS.2020.3007109

This paper covers items 5 and 6 of the objectives. This last experiment covers the feasibility of such a methodology for real-time emotion recognition using various sources of information. It combines the previously developed methodologies to allow the robot to estimate emotions in real-time in both the valence and arousal dimensions, together with the discrete emotion model that corresponds to seven emotions in the case of facial expressions. An internal production database has been created to provide a realistic affective-HRI scenario. The experiment has been designed in such a way that the robot tells a story to the users using audiovisual stimuli, after which it requires the users' emotional feedback by asking them to express the emotions caused by these stimuli. Besides, exploratory analysis is performed with regards to the correlations of the emotions themselves and the emotions and the features of each source of signals. The performance of each emotion estimation subsystem is evaluated using Leave-One-Out strategies for the classification task. Finally, a short questionnaire is performed to evaluate the user's acceptance and empathy with regards to the affective-HRI.

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.

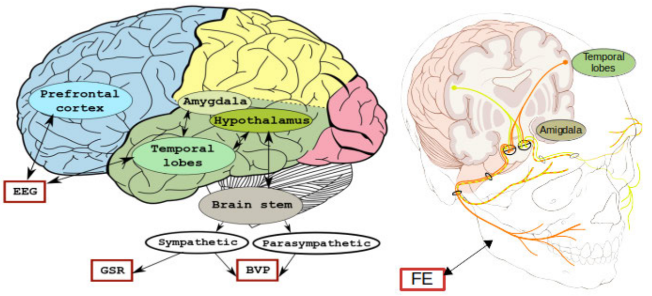

FIGURE 1. Brain areas. Graphical depiction of the inter-relations between brain areas involved in emotion processing. (partially modified [22]).

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):

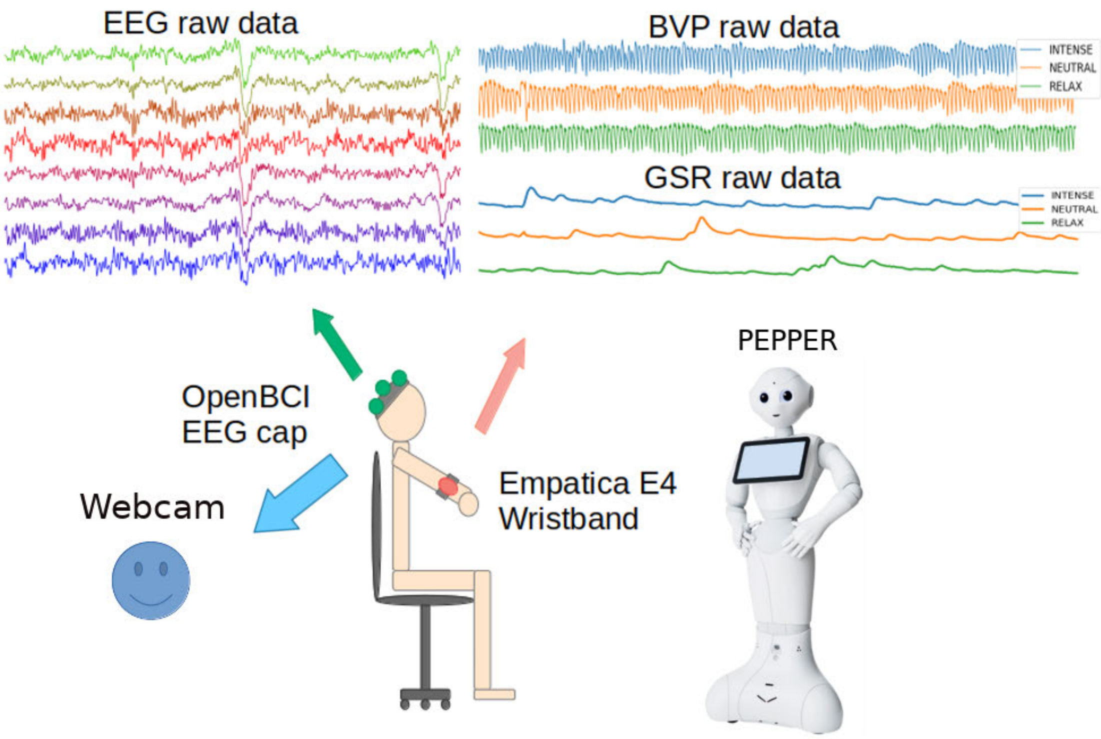

FIGURE 3. Experimental design picture. Top-Left: EEG raw data acquired with the OpenBCI system. Top-Right: GSR and BVP signals acquired with the Empatica E4 device.

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

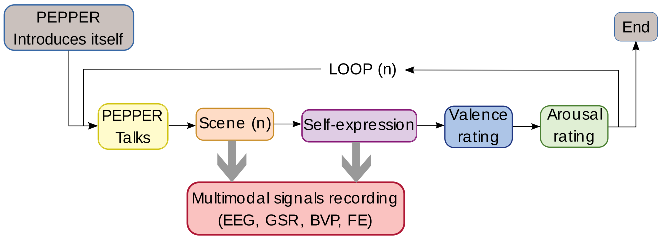

FIGURE 4. **Experimental design diagram.**

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

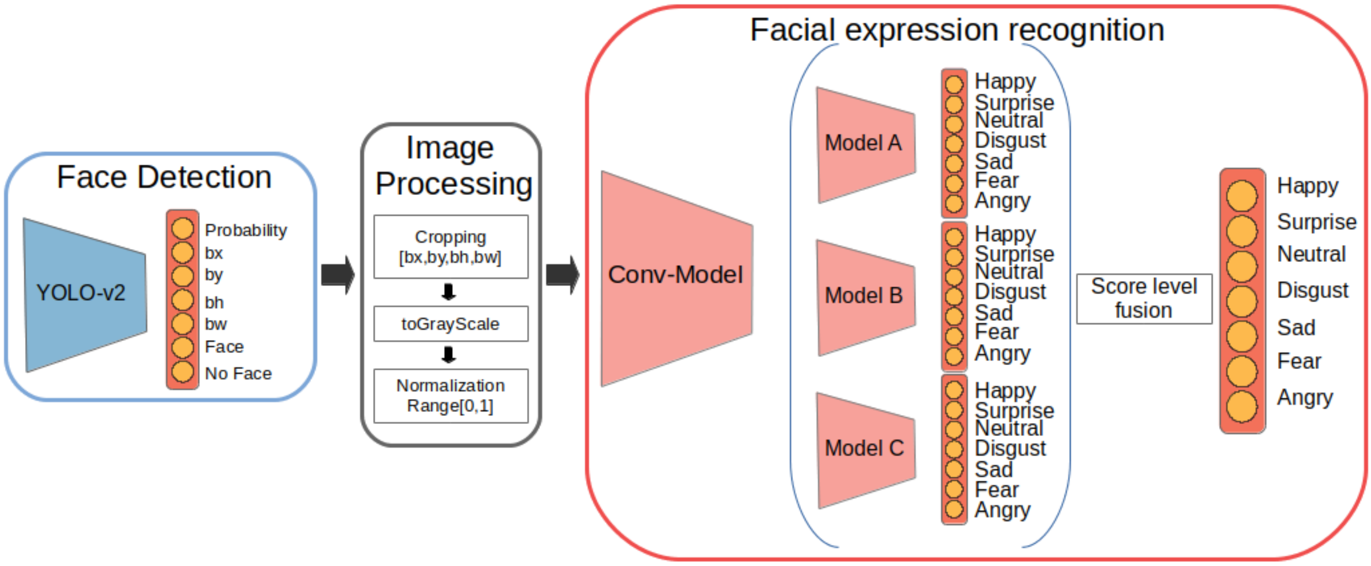

FIGURE 5. **Facial expression estimation. First stage, face detection by convolutional deep learning model and preprocessing of detected face. Second stage, features extraction and classification by an ensemble of convolutional deep learning models.**

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 $\times$ 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.

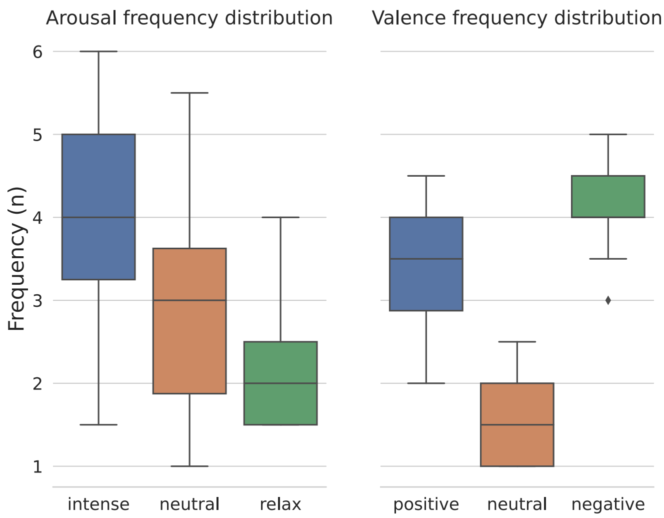

FIGURE 6. Distribution of the frequency of scenes labeled with each emotion in the whole story by all participants. Left: box-plots for frequency of labels for the arousal dimension. Right: box-plots for frequency of labels for the valence dimension.

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

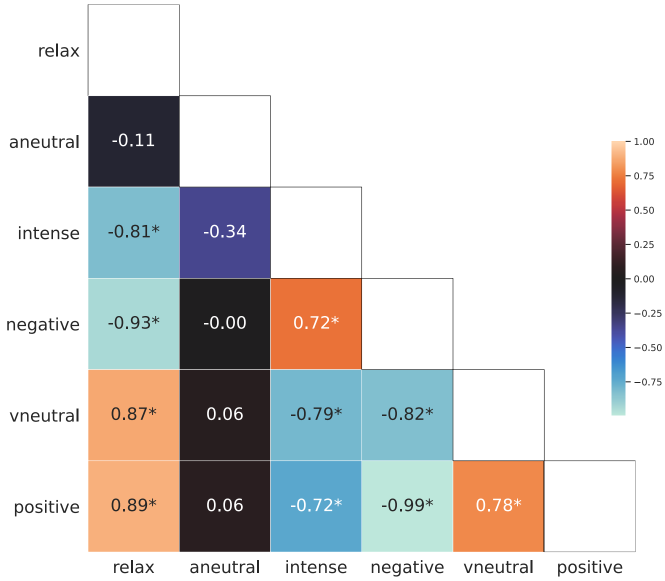

FIGURE 7. Correlation matrix (corrgram) of emotions ratings over scenes for each pair of emotions. The valence and arousal dimensions are taken into account. Correlation values labeled with (*) are statistically significant $P_{value} < 0.05$.

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

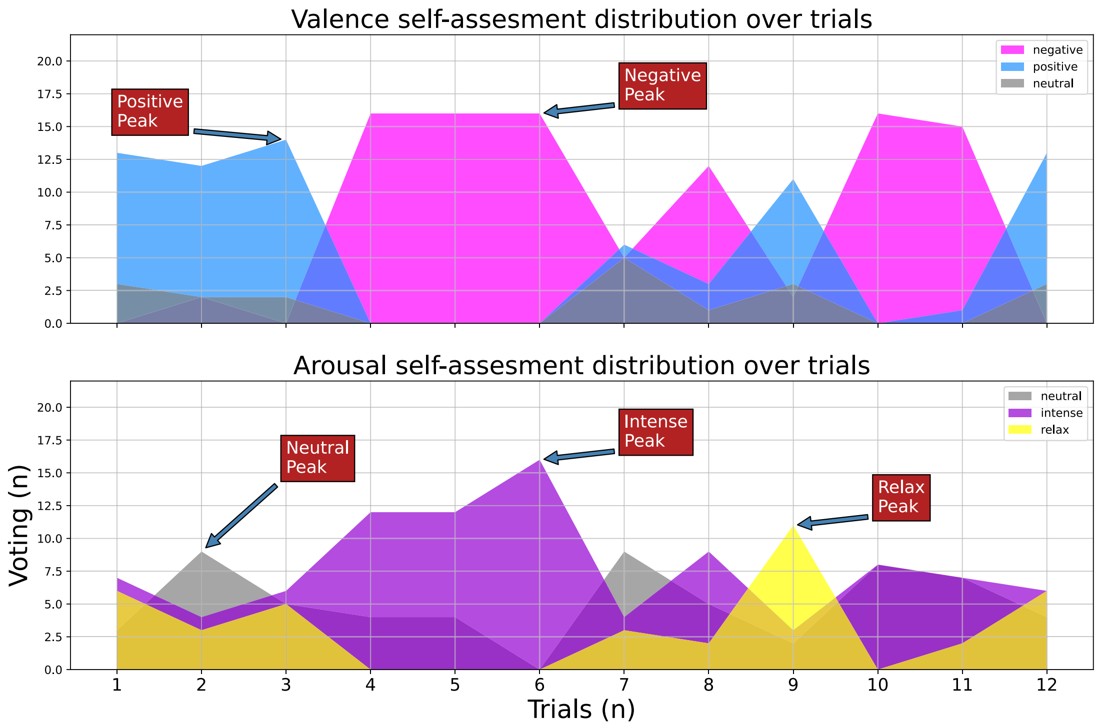

FIGURE 8. **Rating distribution over selected scenes for all participants. Top: rating over time for the arousal dimension. Bottom: rating over time for the valence dimension.**

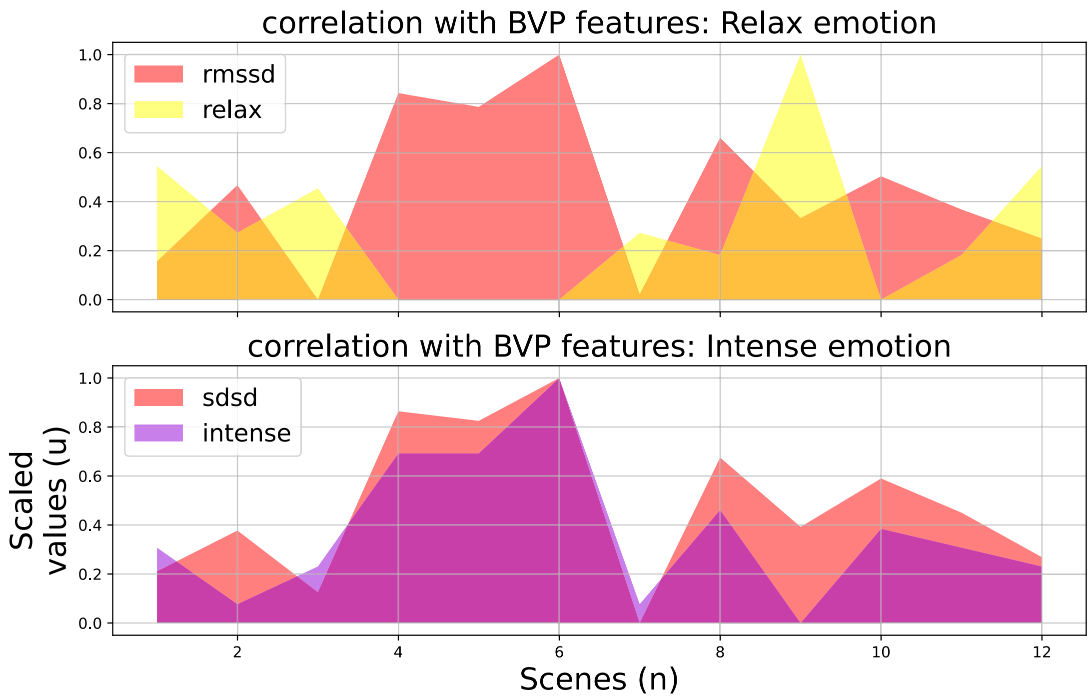

FIGURE 9. **Correlation between two features of the BVP signals with the arousal majority ratings over time. RMMSD is Root Mean Square of successive differences in IBI. SDD is Standard Deviation of IBI differences.**

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

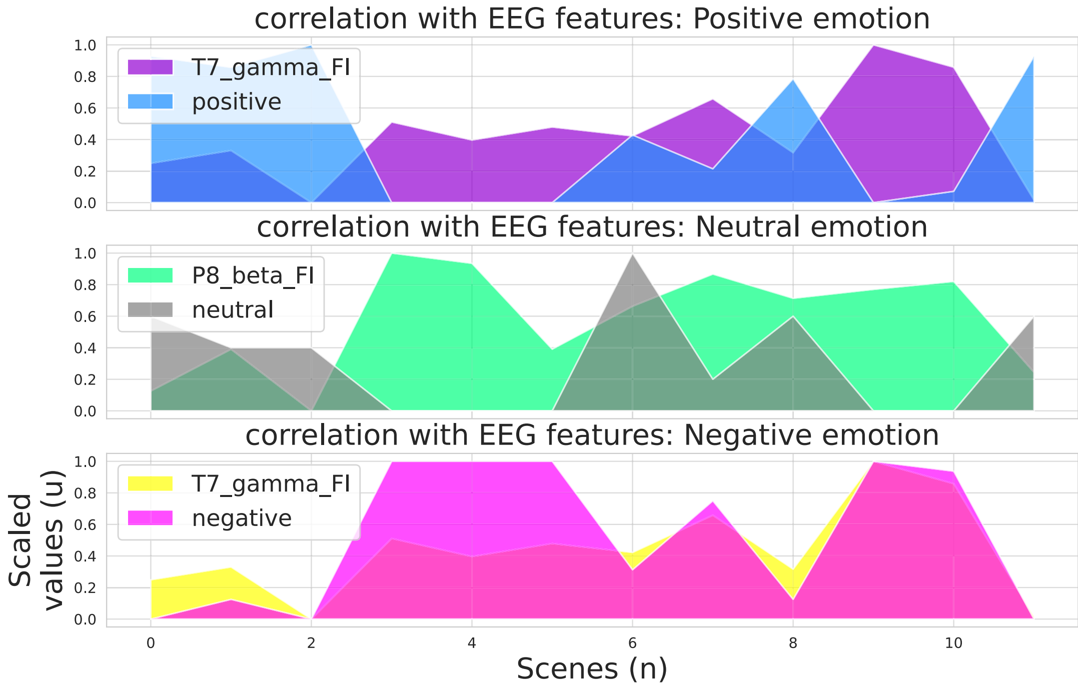

FIGURE 10. EEG most correlated or anti-correlated features across the scenes. Fisher's information of both, T7 temporary electrode for the gamma band (T7_gamma_FI) and P8 parietal electrode for the beta band (P8_beta_FI).

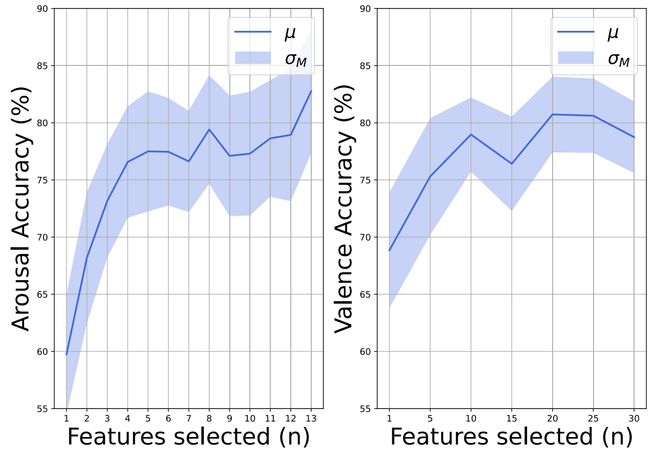

FIGURE 11. Arousal and Valence classification accuracy results. Macro-average F1-score validation metric used. The performance was evaluated with different subsets of features. Solid line is the mean value and shadow area represent de standard error of mean.

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.

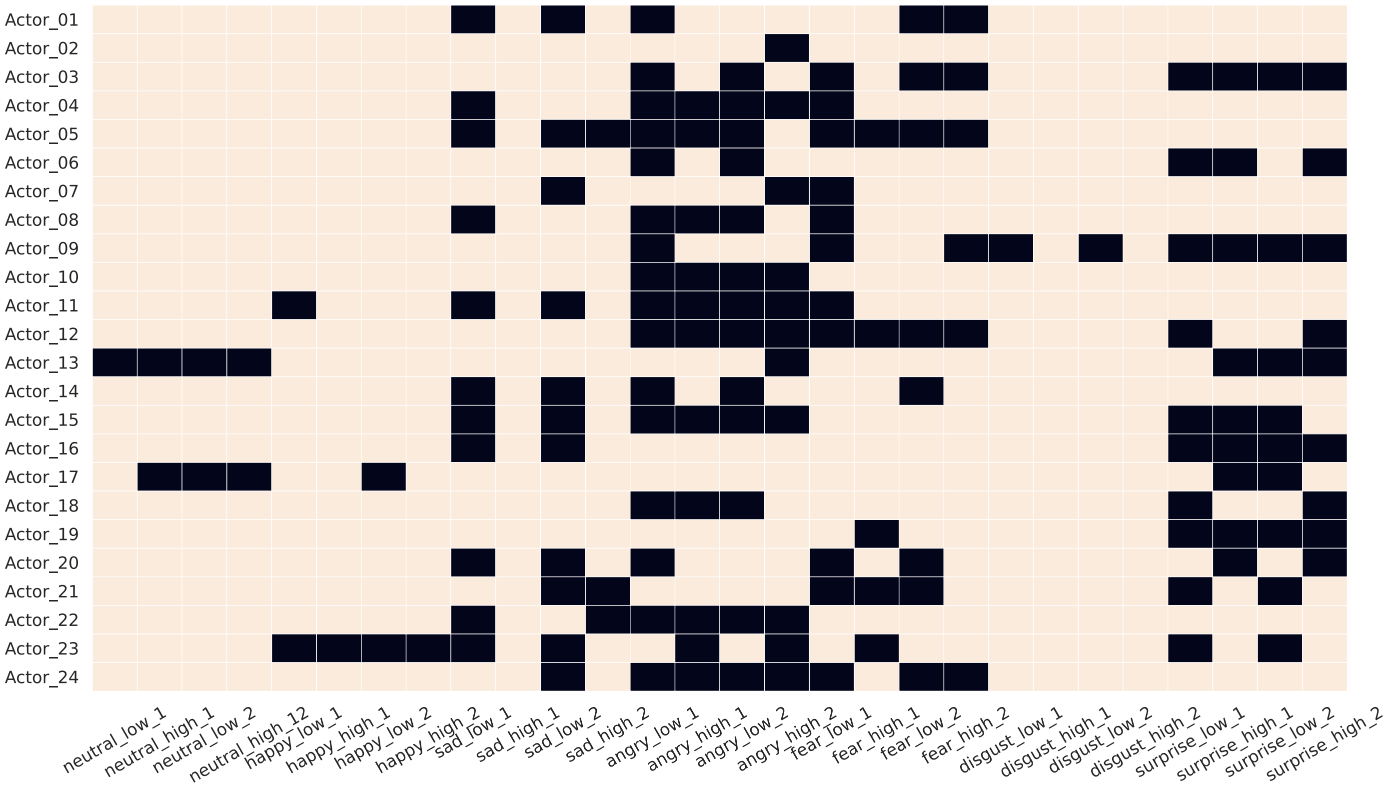

FIGURE 12. Facial expression recognition coherence on the acted out emotions of the RAVDESS database. Black cells show incorrect estimations.

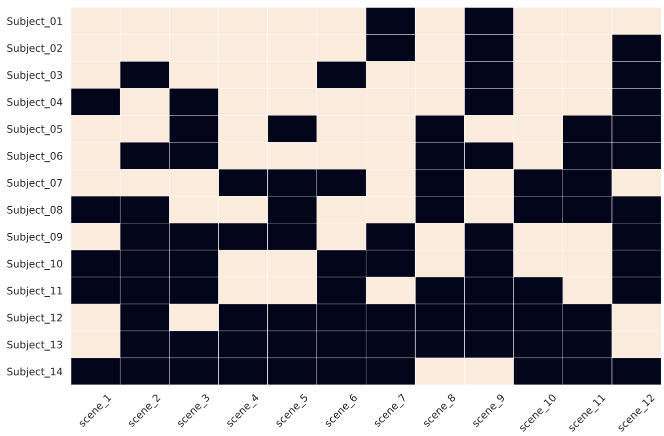

FIGURE 13. Facial expression recognition coherence for the self-expressed emotional reactions during the experiment. Black cells show incorrect estimations.

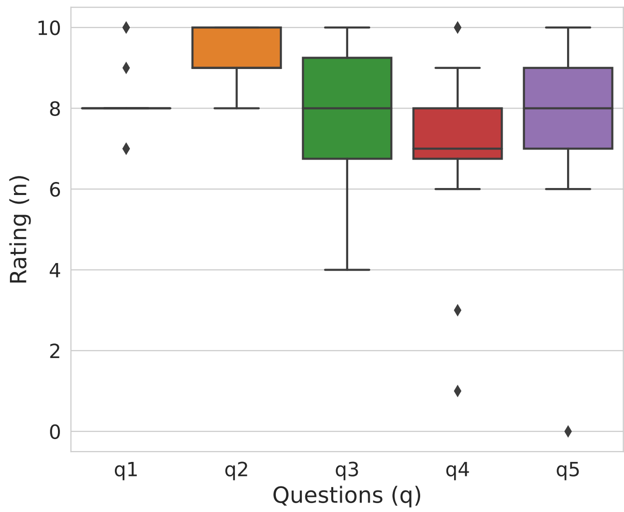

FIGURE 14. Distribution of ratings assigned by all participants to questions regarding the experiment experience (q1, q4, and q5 have some outliers). As q6 is a binary question, it is not represented in these box-plots.

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.

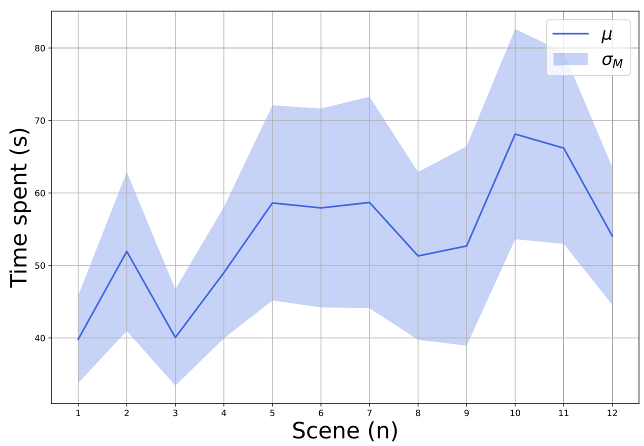

FIGURE 15. Mean and standard deviation of the time spent by all users during the self-expressions of emotions after each scene during the dramatic story. Solid line is the mean value and shadow area represent de standard error of mean.

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.

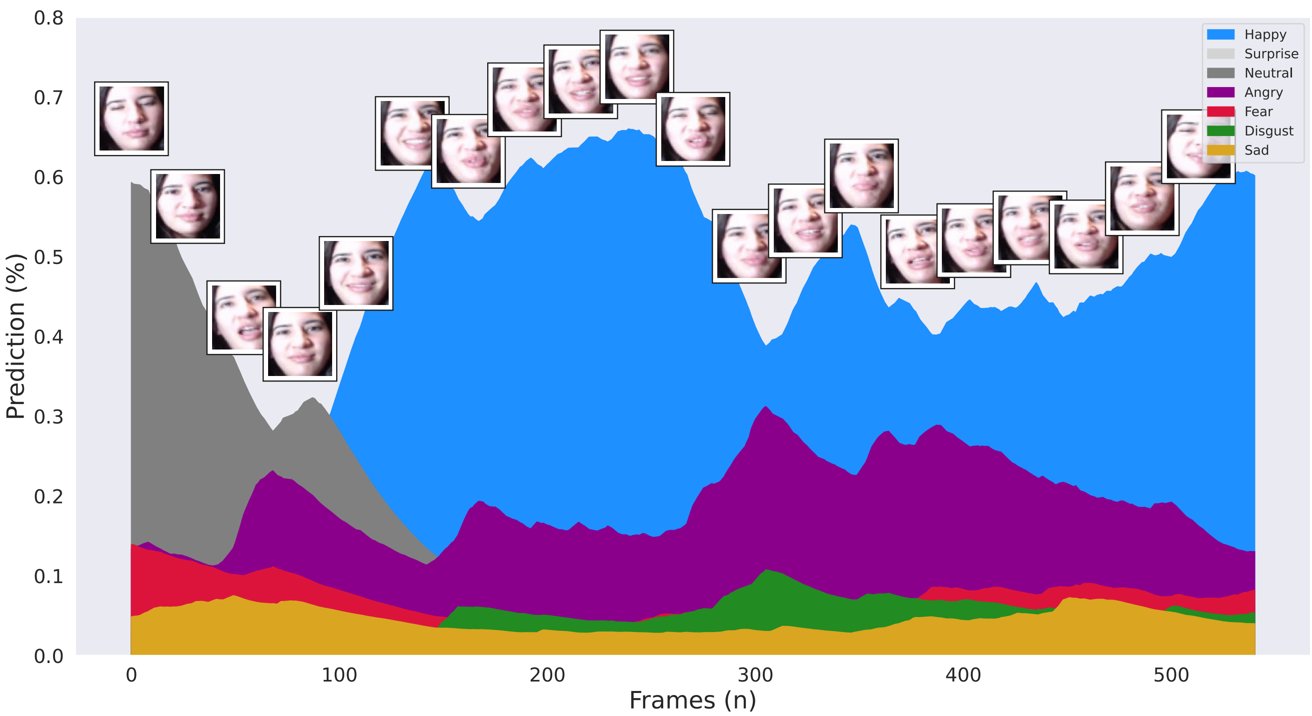

FIGURE 16. Expressive subject self-expressions. POSITIVE facial expression evolution.

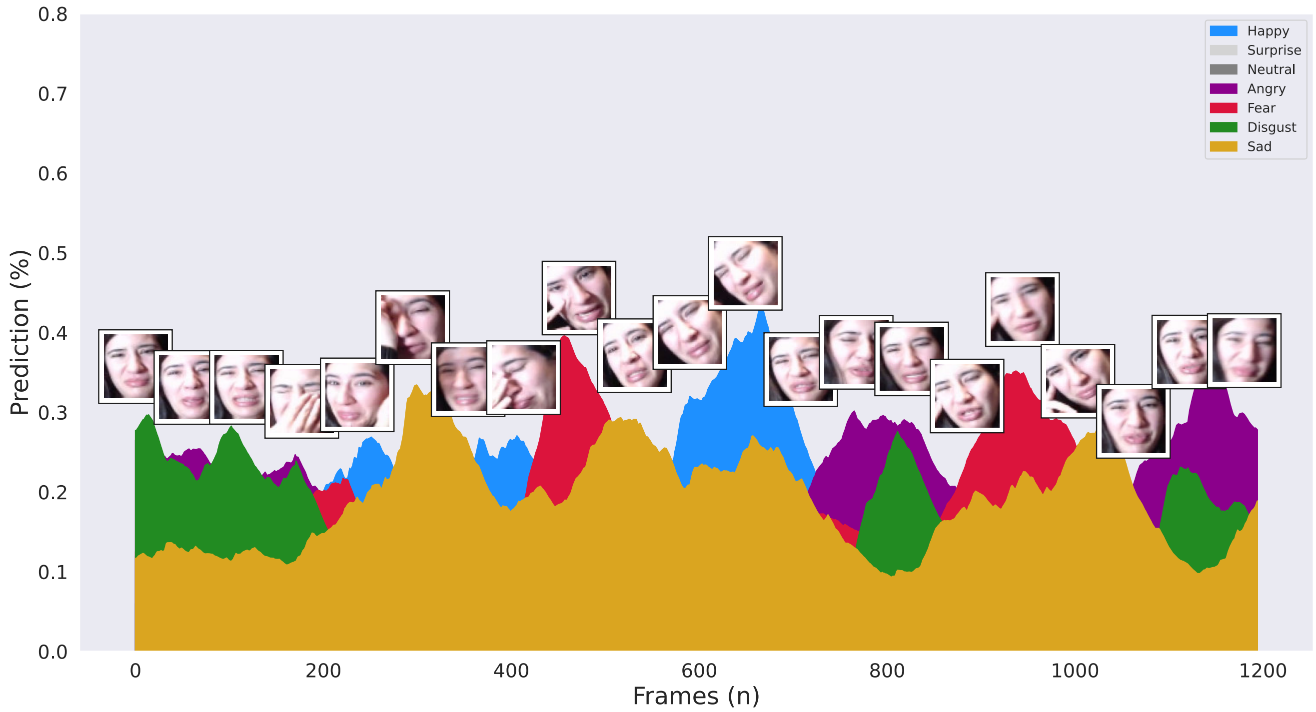

FIGURE 17. **Expressive subject self-expressions. NEGATIVE facial expression evolution.**

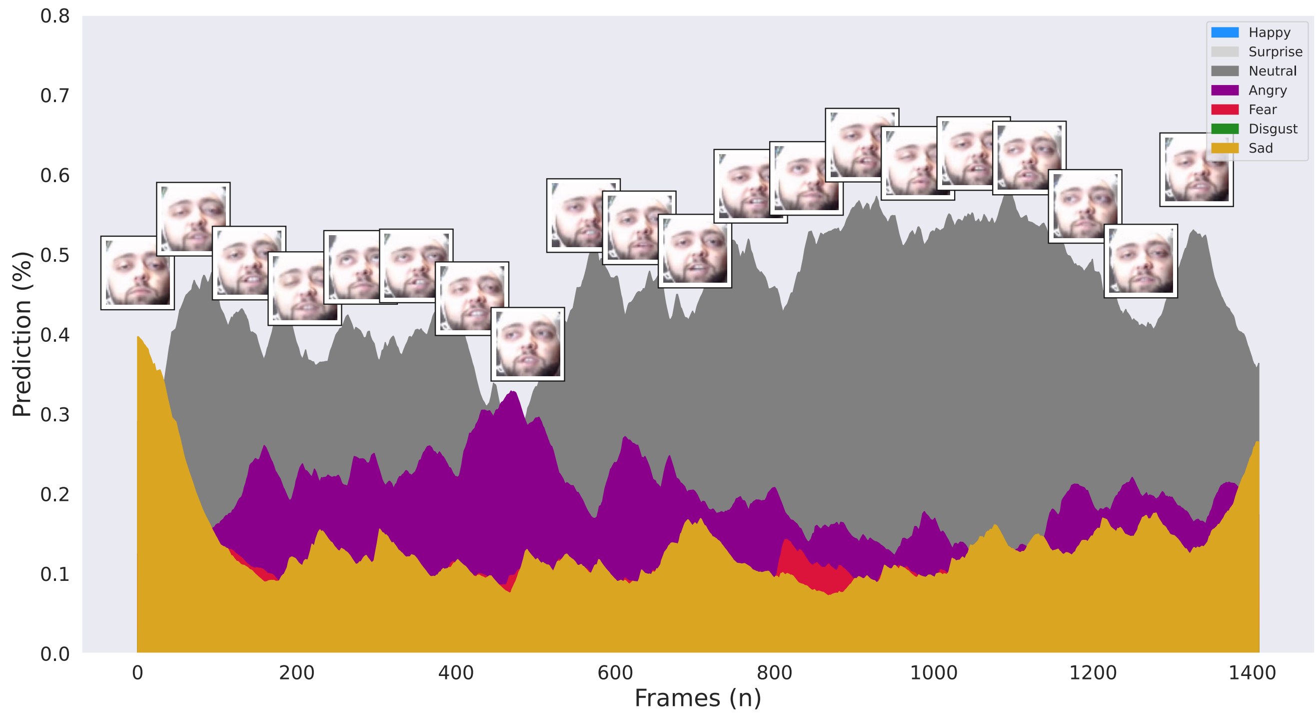

FIGURE 18. **Non-expressive subject self-expressions. POSITIVE facial expression evolution.**

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.

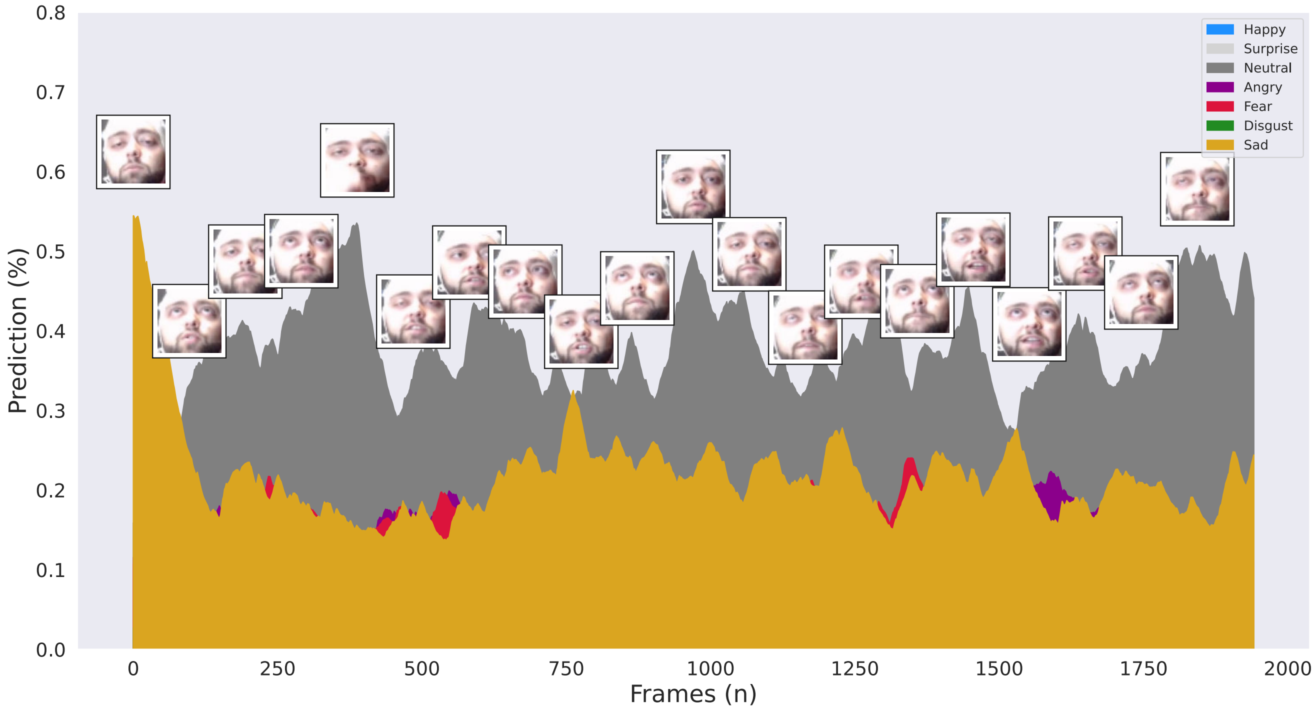

FIGURE 19. Non-expressive subject self-expressions. NEGATIVE facial expression evolution.

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.

4.4 Related publications

Partial results of this thesis were also published by the author in collaborations with other research studies:

- In collaboration with Bonomini et al. [2019, 2020], where a study was made on the effect of deep breathing and cognitive task performance, the Biosignals application for the recording of physiological signals [Val-Calvo, 2020a] was developed, and also some of the experiments presented in those articles were performed as part of the thesis to acquire skills as an experimenter with data from humans.

- The development of a methodology for real-time facial expression estimation, which has been later used for the present thesis, was done in collaboration with Benamara et al. [2019, 2020]. Through this collaboration, item 4 of the objectives has been covered. The main collaboration related to this thesis was done by the author in the search for a technical solution that could be run in real-time and the definition of the deep learning model.

- An alternative methodology for real-time emotion estimation using EEG signals was presented at the IWINAC 2019 conference by Val-Calvo et al. [2019b]. The development of such a methodology helped the author to understand that a proper validation methodology must be performed because simple cross-validation generates over-optimistic results yielding incorrect conclusions. Therefore, this paper helped to cover item 6 of the objectives.

4.5 Results discussion

The development of an emotion estimation methodology has several concerns that must be carefully faced. On one hand, emotion estimation is not a straight forward task, it requires the implementation of several disciplines which have to be gracefully coordinated to achieve a successful outcome. In the affective-HRI domain, it is crucial to ensure that real-time emotion estimation is a quick and versatile process and to obtain accurate results the acquisition methodologies must perform reliably. In this research, the main purpose is to evaluate the feasibility of the emotion estimation methodology under dynamic conditions but without taking into account optimization processes regarding the classification algorithms used, as this would require further research into the models and hyper-parameter optimization. In that context, the developed strategies showed to be robust even taking into account the dynamic properties of the used stimuli, therefore, emotion estimation in such a paradigm is not only possible but has a wide margin for optimization.

On the other hand, the importance of an emotional model must be highlighted, which must include a range of emotions. As explained previously, this work has used both, discrete and continuous models. Concerning the physiological signals, EEG signals have been used for valence prediction, and GSR with BVP, for the arousal prediction in the continuous emotional model, while facial expressions have been used for the estimation of the discrete emotional model. Physiological signals reflect involuntary changes in the CNS and autonomous nervous system and therefore can not be subjected to inhibition, as is the case for facial expressions. It has been shown by Val-Calvo et al. [2020b] that depending on the expressiveness of the user, a property which mainly depends on cultural and environmental learning, the estimation becomes feasible or not.

Besides, bias is one of the main questions for any experimental design. Such a bias affects the results obtained through databases. For example, when dealing with facial expressions, the RAVDESS database consists of a set of actors expressing emotions. 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 [Tsai et al., 2009, Chen et al., 2018b]. This causes volunteers to overreact their facial expressions, which is also the case for the FER-2013 database, with more than twenty thousand facial expressions that are overreacted, causing bias in any result obtained using this type of database. 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. In contrast, the self-developed recordings[1] deal with a completely realistic scenario where no acting is performed. To create a more realistic scenario, both, a dramatic film and a dramatic story have been chosen as emotional drives. First, a film has a varied distribution of all types of arousal and valence. For the paradigm of an affective-HRI, a similar approach has been conducted, since, to allow the robot to drive emotions, a dramatic story has been created that talks about philosophical questions to make volunteers think about them and express their deeper insights both verbally and emotionally, that is to evoke emotions on them.

A proper validation scheme is important for every novel methodology to be comparable to those in literature. Predefined cross-validation schemes for supervised learning algorithms are not suitable for model performance evaluations when using temporal series [Tashman, 2000], as the temporal correlation must be taken into account and, therefore, reasonable independence must be maintained between training and test sets. Different validation schemes were found suitable for the evaluation of temporal series as they are based on Leave-One-Out strategies, leave-one-trial-out, leave-one-session-out or leave-one-subject-out methodologies have been used. Furthermore, the F1 score is chosen since, when the performance of the test set is biased towards a specific result, it shows the highest penalization, and it is desirable to build models that properly generalize over the objective task.

The set of selected features chosen for this methodology is easy to compute, allowing the quick development of portable systems with high significant results, which, also, are far from black-box techniques such as deep-learning approaches or very complex features with difficult interpretation in biological terms. The feature processing also must consider how the feature space is considered. To allow the management of the intrinsic complexities of the feature space, smoothing techniques can often improve the results, as it is the case for label smoothing or feature space smoothing before the classification step is produced. The first allows for algorithms to minimize the error in the case of mislabeled samples while the second stabilized the variability of the feature space making it more suitable for the classification process.

Also related to the dimensionality of the feature space and the set of selected features,

[1] The recordings, used to build the specific databases of articles 2 and 3, were obtained during the experiments carried out at the Instituto de Bioingeniería of the Univ. Miguel Hernández, authorized by its Ethics Committee and complying with the legal requirements of the European General Data Protection Regulation, since no other personal information was collected and all data were anonymized. All participants were volunteers and provided their informed written consent. The other databases used in the articles were publicly available.

EEG signals often involve the use of high dimensional spaces, as they are produced by a large set of electrodes. As our objective is to build easy to wear and lightweight computational methods, a small set of electrodes have been chosen that are spatially highly correlated with emotional processing in the CNS. The chosen set of electrodes showed to be a good choice since it offered a proper balance between the more informative electrodes and redundancy for the classifying step and, as a result, the robustness of the methodology has proved to be higher.

Artifact removal is the first step of the whole process and therefore plays a very important role in the outcome. A modified version of EAWICA, constrained to the delta band, was used to reduce artifacts (EOG), and therefore reduce computation time. EEG artifact removal and successive feature processing steps were previously probed to perform in real-time constraints. Besides, the processes involved for both, BVP peak detection and GSR preprocessing and feature processing, does not require heavy computations. Therefore, the proposed methodology can be applied in real-time scenarios with meaningful results taking into account a near realistic scenario.

The obtained recognition results are in good agreement with other studies, positive and negative valence are highly anti-correlated, for BVP features, neutral and intense arousal states are correlated and anti-correlated respectively with HF and for GSR features, each arousal state has a different correlation sequence. For the case of EEG features, highly correlated features are mostly accumulated in the beta and gamma frequency ranges for positive-valence and negative-valence, while neutral-valence achieves its highest correlation values for the beta frequency range, hinting at a possibly different behavior of the neural flow. All those findings are mostly coherent with previous results, regardless of the experimental paradigm and culture differences.

Chapter 5

Conclusions

- Online artifact removal to denoise EOG artifacts from EEG recordings can be performed with significant efficiency using a strategy based on wavelet decomposition and independent component analysis (ICA), as stated in article 1.

- Valence emotional dimension is feasible using an EEG-based methodology for three discrete categories ("negative", "neutral", "positive") and with a low density of electrodes. A set of relevant and partially invariant features have been found which are based on complexity measurements of the signal, and feature space smoothing has been demonstrated in article 1 to improve the results in terms of time consumption and accuracy, with the use of the Savitzky-Golay filtering method.

- Arousal estimation can be achieved by the design of a GSR and BVP based methodology using three discrete categories ("relax", "neutral", "intense"), as probed in article 2.

- Multimodal emotion recognition based on both, physiological signals and facial expressions, fulfills in article 3 the objective of minimizing the uncertainty of the robot regarding the users' emotional state.

- The use of more realistic scenarios, first in article 2, using dramatic films as a dynamic stimulus, and second in article 3, by an affective-HRI paradigm under a realistic scenario that merges a dramatic story in combination with constant emotional stimuli and users' emotional feedback, has proved the robustness of emotion estimation methodologies.

- Real-time emotion estimation using lightweight methodologies can be achieved with the use of proper optimization strategies, as stated in the three articles.

- The developed applications: BIOSIGNALS and GEERT, used for article 2, and GePHYCAM, used for article 3, will help to further the research in the direction of affective brain-computer interfaces.

Future work

Future work will include multimodal integration of different sources using innovative techniques based on deep learning strategies that take into account temporal and spatial information. For example, using long and short-term memory networks combined with convolutional neural networks embedded in the time scale, with a strategy of merging multimodal features and with the combination of attention mechanisms. In this sense, the BIOSIGNALS, GEERT and GePHYCAM applications continue to be a work in progress with the objective of being a platform that allows the integration of diverse multimodal sources and that optimizes the visualization, signal processing, and pattern detection processes in real-time.

As for the experimental design, the next step will consider the use of the emotion estimated by the robot to dynamically modify the affective-HRI. This approach would allow the robot to dynamically adapt to the user's emotional changes, or even modulate it according to a predefined objective.

Finally, and in relation to the capacity of the robot to interact with the environment, in this context, with the emotional reactions of the users, it could be approached a cognitive architecture based on the continuous learning and using the methods of hierarchical reinforcement learning developed more recently.

Bibliography

Hadi Ratham Al Ghayab, Yan Li, Siuly Siuly, and Shahab Abdulla. Epileptic seizures detection in eegs blending frequency domain with information gain technique. *Soft Computing*, 23(1):227–239, 2019. (↩ 7)

Nadir Kamel Benamara, Mikel Val-Calvo, José Ramón Álvarez-Sánchez, Alejandro Díaz-Morcillo, José Manuel Ferrández Vicente, Eduardo Fernández-Jover, and Tarik Boudghene Stambouli. Real-time emotional recognition for sociable robotics based on deep neural networks ensemble. In *Understanding the Brain Function and Emotions - IWINAC 2019, Proceedings, Part I*, volume 11486 of *Lecture Notes in Computer Science*, pages 171–180. Springer, 2019. URL https://doi.org/10.1007/978-3-030-19591-5_18. (↩ 11, 14, and 76)

Nadir Kamel Benamara, Mikel Val-Calvo, José Ramón Álvarez-Sánchez, Alejandro Díaz-Morcillo, José Manuel Ferrández Vicente, Eduardo Fernández, and Tarik Boudghene Stambouli. Real-time facial expression recognition using smoothed deep neural network ensemble. *Integrated Computer-Aided Engineering*, Pre-press:1–15, 2020. URL https://doi.org/10.3233/ICA-200643. (↩ 11, 14, and 76)

Hans Berger. Über das elektrenkephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1):527–570, December 1929. doi: 10.1007/bf01797193. URL https://doi.org/10.1007/bf01797193. (↩ 7)

Maria Paula Bonomini, Mikel Val-Calvo, Alejandro Díaz-Morcillo, José Manuel Ferrández Vicente, and Eduardo Fernández-Jover. Autonomic modulation during a cognitive task using a wearable device. In *Understanding the Brain Function and Emotions - IWINAC 2019, Proceedings, Part I*, volume 11486 of *Lecture Notes in Computer Science*, pages 69–77. Springer, 2019. URL https://doi.org/10.1007/978-3-030-19591-5_8. (↩ 76)

Maria Paula Bonomini, Mikel Val Calvo, Alejandro Diaz Morcillo, Maria Florencia Segovia, Jose Manuel Ferrandez Vicente, and Eduardo Fernández Jover. The effect of breath pacing on task switching and working memory. *International Journal of Neural Systems*, 30(6):2050028, June 2020. URL https://doi.org/10.1142/S0129065720500288. (↩ 76)

Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012. (↩ 8)

Jerome S Bruner and Renato Tagiuri. The perception of people. Technical report, Ft. Belvoir : Defense Technical Information Center, 1954. URL https://doi.org/10.21236/ad0024982. (↩ 9)

Darius Burschka. Task representation in robots for robust coupling of perception to action in dynamic scenes. In *Springer Proceedings in Advanced Robotics*, pages 25–31. Springer International Publishing, November 2019. doi: 10.1007/978-3-030-28619-4_4. URL https://doi.org/10.1007/978-3-030-28619-4_4. (←˒ 1)

György Buzsáki, Costas A Anastassiou, and Christof Koch. The origin of extracellular fields and currents - eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, 13(6): 407–420, 2012. (←˒ 5)

Delphine Caruelle, Anders Gustafsson, Poja Shams, and Line Lervik-Olsen. The use of electrodermal activity (eda) measurement to understand consumer emotions–a literature review and a call for action. *Journal of Business Research*, 104:146–160, 2019. (←˒ 8)

Ginevra Castellano, Iolanda Leite, André Pereira, Carlos Martinho, Ana Paiva, and Peter W Mcowan. Multimodal affect modeling and recognition for empathic robot companions. *International Journal of Humanoid Robotics*, 10(01):1350010, 2013. (←˒ 11)

Xin Chai, Qisong Wang, Yongping Zhao, Xin Liu, Ou Bai, and Yongqiang Li. Unsupervised domain adaptation techniques based on auto-encoder for non-stationary eeg-based emotion recognition. *Computers in biology and medicine*, 79:205–214, 2016. (←˒ 11)

Hu Chen, Ye Gu, Fei Wang, and Weihua Sheng. Facial expression recognition and positive emotion incentive system for human-robot interaction. In *2018 13th World Congress on Intelligent Control and Automation (WCICA)*, pages 407–412. IEEE, 2018a. (←˒ 11)

JX Chen, PW Zhang, ZJ Mao, YF Huang, DM Jiang, and YN Zhang. Accurate eeg-based emotion recognition on combined features using deep convolutional neural networks. *IEEE Access*, 7:44317–44328, 2019. (←˒ 10)

Luefeng Chen, Min Wu, Mengtian Zhou, Zhentao Liu, Jinhua She, and Kaoru Hirota. Dynamic emotion understanding in human-robot interaction based on two-layer fuzzy svr-ts model. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017. (←˒ 11)

Luefeng Chen, Mengtian Zhou, Wanjuan Su, Min Wu, Jinhua She, and Kaoru Hirota. Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction. *Information Sciences*, 428:49–61, 2018b. (←˒ 11 and 77)

Shizhe Chen and Qin Jin. Multi-modal dimensional emotion recognition using recurrent neural networks. In *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*, pages 49–56. ACM, 2015. (←˒ 10)

Vincent T. Cialdella, Emilio J. C. Lobato, and J. Scott Jordan. Wild architecture. In *Natural Language Processing*, pages 791–807. IGI Global, 2020. doi: 10.4018/978-1-7998-0951-7.ch038. URL https://doi.org/10.4018/978-1-7998-0951-7.ch038. (←˒ 2)

Felipe Cid, Luis J Manso, and Pedro Núnez. A novel multimodal emotion recognition approach for affective human robot interaction. In *Proceedings of the Workshop on Multimdoal Semantics For Robotics Systems (MuSRobS)*, CEUR Workshop Proceedings, pages 1–9, 2015. URL http://ceur-ws.org/Vol-1540/#paper_01.pdf. (←˒ 11)

Antonio R Damasio. Emotion in the perspective of an integrated nervous system. *Brain research reviews*, 26(2-3):83–86, 1998. (↩ 5)

Charles Darwin. *The expression of the emotions in man and animals*. John Murray, 1872. (↩ 9)

Liyanage C De Silva and Pei Chi Ng. Bimodal emotion recognition. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 332–335. IEEE, 2000. (↩ 11)

William Dement and Nathaniel Kleitman. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, 9(4):673–690, November 1957. doi: 10.1016/0013-4694(57)90088-3. URL https://doi.org/10.1016/0013-4694(57)90088-3. (↩ 7)

Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, November 2011. doi: 10.1109/iccvw.2011.6130508. URL https://doi.org/10.1109/iccvw.2011.6130508. (↩ 14)

Vincenzo Donadio, Alex Incensi, Veria Vacchiano, Rossella Infante, Martina Magnani, and Rocco Liguori. The autonomic innervation of hairy skin in humans: an in vivo confocal study. *Scientific reports*, 9(1):1–7, 2019. (↩ 8)

Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for EEG-based emotion classification. In *6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 81–84. IEEE, 2013. (↩ 14 and 27)

Robert Edelberg. Electrical activity of the skin: Its measurement and uses in psychophysiology. *Handbook of psychophysiology*, pages 367–418, 1972. (↩ 8)

Paul Ekman. Pictures of facial affect. *Consulting Psychologists Press*, 1976. (↩ 9)

Paul Ekman. Facial expressions of emotion: an old controversy and new findings. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 335 (1273):63–69, 1992. (↩ 9)

Paul Ekman and Wallace V Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971. (↩ 9)

Paul Ekman, Wallace V Friesen, and Maureen O'sullivan. Smiles when lying. *Journal of personality and social psychology*, 54(3):414, 1988. (↩ 9)

Rosenberg Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997. (↩ 9)

Olov Engwall, José Lopes, and Anna Åhlund. Robot interaction styles for conversation practice in second language learning. *International Journal of Social Robotics*, March 2020. doi: 10.1007/s12369-020-00635-y. URL https://doi.org/10.1007/s12369-020-00635-y. (↩ 1)

Diego R Faria, Mario Vieira, Fernanda CC Faria, and Cristiano Premebida. Affective facial expressions recognition for human-robot interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 805–810. IEEE, 2017a. (↩ 11)

Diego Resende Faria, Mario Vieira, and Fernanda CC Faria. Towards the development of affective facial expression recognition for human-robot interaction. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 300–304, 2017b. (↩ 11)

Joshua J. Foster, David W. Sutterer, John T. Serences, Edward K. Vogel, and Edward Awh. Alpha-band oscillations enable spatially and temporally resolved tracking of covert spatial attention. *Psychological Science*, 28(7):929–941, May 2017. doi: 10.1177/0956797617699167. URL https://doi.org/10.1177/0956797617699167. (↩ 7)

J Frois-Wittman. The judgment of facial expression. *Journal of Experimental Psychology*, 13(2):113, 1930. (↩ 9)

Martin Gjoreski, Hristijan Gjoreski, Mitja Luštrek, and Matjaz Gams. Deep ensembles for inter-domain arousal recognition. *Proceedings of Machine Learning Research*, 86: 52–65, 2018. (↩ 10)

Ian J. Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio. Challenges in representation learning: A report on three machine learning contests. In *Neural Information Processing*, pages 117–124. Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-42051-1_16. URL https://doi.org/10.1007/978-3-642-42051-1_16. (↩ 14)

Katalin M. Gothard. The amygdalo-motor pathways and the control of facial expressions. *Frontiers in Neuroscience*, 8, March 2014. doi: 10.3389/fnins.2014.00043. URL https://doi.org/10.3389/fnins.2014.00043. (↩ 9)

Dirk Hagemann, Shari R Waldstein, and Julian F Thayer. Central and autonomic nervous system integration in emotion. *Brain and cognition*, 52(1):79–87, 2003. (↩ 5)

Naoya Hasegawa and Yoshihiko Takahashi. Incorporation of human facial expression into robot control. In *Handbook of Research on Advanced Mechatronic Systems and Intelligent Robotics*, pages 310–322. IGI Global, 2020. doi: 10.4018/978-1-7998-0137-5.ch012. URL https://doi.org/10.4018/978-1-7998-0137-5.ch012. (↩ 2)

Rachael E Jack, Oliver GB Garrod, Hui Yu, Roberto Caldara, and Philippe G Schyns. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19):7241–7244, 2012. (↩ 9)

Jesin James, Li Tian, and Catherine Inez Watson. An open source emotional speech corpus for human robot interaction applications. In *Interspeech*, pages 2768–2772, 2018. (↩ 11)

Peter Jönsson. Respiratory sinus arrhythmia as a function of state anxiety in healthy individuals. *International journal of psychophysiology*, 63(1):48–54, 2007. (← 8)

Hye-Won Jung, Yong-Ho Seo, Michael S Ryoo, and Hyun Seung Yang. Affective communication system with multimodality for a humanoid robot, ami. In *4th IEEE/RAS International Conference on Humanoid Robots, 2004.*, volume 2, pages 690–706. IEEE, 2004. (← 11)

David Kelley and Mathew Twyman. Biasing in an independent core observer model artificial general intelligence cognitive architecture. *Procedia Computer Science*, 169: 535–541, 2020. doi: 10.1016/j.procs.2020.02.213. URL https://doi.org/10.1016/j.procs.2020.02.213. (← 2)

Jonghwa Kim and Elisabeth André. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30 (12):2067–2083, 2008. (← 13)

Kyung Hwan Kim, Seok Won Bang, and Sang Ryong Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing*, 42(3):419–427, 2004. (← 13)

Dana Kulic and Elizabeth A. Croft. Estimating intent for human-robot interaction. *Advanced Robotics*, 2003. URL https://api.semanticscholar.org/CorpusID:1120056. (← 11)

Dana Kulic and Elizabeth A Croft. Affective state estimation for human–robot interaction. *IEEE Transactions on Robotics*, 23(5):991–1000, 2007. (← 11)

Yatindra Kumar, ML Dewal, and RS Anand. Epileptic seizures detection in eeg using dwt-based apen and artificial neural network. *Signal, Image and Video Processing*, 8 (7):1323–1334, 2014. (← 7)

Egor Lakomkin, Mohammad Ali Zamani, Cornelius Weber, Sven Magg, and Stefan Wermter. On the robustness of speech emotion recognition for human-robot interaction with deep neural networks. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 854–860. IEEE, 2018. (← 11)

Zirui Lan, Olga Sourina, Lipo Wang, Reinhold Scherer, and Gernot R Müller-Putz. Domain adaptation techniques for eeg-based emotion recognition: a comparative study on two public datasets. *IEEE Transactions on Cognitive and Developmental Systems*, 11 (1):85–94, 2018. (← 11)

Carney Landis. Studies of emotional reactions. ii. general behavior and facial expression. *Journal of Comparative Psychology*, 4(5):447, 1924. (← 9)

Kent M. Lee, Kristen A. Lindquist, Nathan L. Arbuckle, Samantha M. Mowrer, and B. Keith Payne. An indirect measure of discrete emotions. *Emotion*, 20(4):659–676, June 2020. doi: 10.1037/emo0000577. URL https://doi.org/10.1037/emo0000577. (← 2)

Jinpeng Li, Shuang Qiu, Changde Du, Yixin Wang, and Huiguang He. Domain adaptation for eeg emotion recognition based on latent representation similarity. *IEEE Transactions on Cognitive and Developmental Systems*, 2019. (← 11)

Changchun Liu, Pramila Rani, and Nilanjan Sarkar. Human-robot interaction using affective cues. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 285–290. IEEE, 2006. (←˒ 11)

Guan-Horng Liu and Evangelos A Theodorou. Deep learning theory review: An optimal control and dynamical systems perspective. *arXiv preprint arXiv:1908.10920*, 2019. (←˒ 1)

Yong-Jin Liu, Minjing Yu, Guozhen Zhao, Jinjing Song, Yan Ge, and Yuanchun Shi. Real-time movie-induced discrete emotion recognition from eeg signals. *IEEE Transactions on Affective Computing*, 9(4):550–562, 2017a. (←˒ 11)

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 Journal of Automatica Sinica*, 4(4):668–676, 2017b. (←˒ 11)

U. Livneh, J. Resnik, Y. Shohat, and R. Paz. Self-monitoring of social facial expressions in the primate amygdala and cingulate cortex. *Proceedings of the National Academy of Sciences*, 109(46):18956–18961, October 2012. doi: 10.1073/pnas.1207662109. URL https://doi.org/10.1073/pnas.1207662109. (←˒ 9)

Choubeila Maaoui and Alain Pruski. Emotion recognition through physiological signals for human-machine communication. *Cutting Edge Robotics*, 2010:317–332, 2010. (←˒ 11)

Ruhi Mahajan and Bashir I. Morshed. Unsupervised eye blink artifact denoising of EEG data with modified multiscale sample entropy, kurtosis, and wavelet-ICA. *IEEE Journal of Biomedical and Health Informatics*, 19(1):158–165, January 2015. doi: 10.1109/jbhi.2014.2333010. URL https://doi.org/10.1109/jbhi.2014.2333010. (←˒ 27)

Nadia Mammone and Francesco Morabito. Enhanced automatic wavelet independent component analysis for electroencephalographic artifact removal. *Entropy*, 16(12):6553–6572, 2014. (←˒ 27)

William Ecuador Martínez Albán, María José Briceño Ruperti, Dubal Edison Salvatierra Tumbaco, and María Elena Moya Martínez. Brain and emotions on learning process. *International journal of health & medical sciences*, 3(1):17–20, Oct. 2019. doi: 10.31295/ijhms.v3n1.108. URL https://sloap.org/journal/index.php/ijhms/article/view/108. (←˒ 2)

Jinpeng Mi, Song Tang, Zhen Deng, Michael Goerner, and Jianwei Zhang. Object affordance based multimodal fusion for natural human-robot interaction. *Cognitive Systems Research*, 54:128–137, 2019. (←˒ 9)

Ahmad Mirzaei, Ahmad Ayatollahi, Parisa Gifani, and Leili Salehi. Eeg analysis based on wavelet-spectral entropy for epileptic seizures detection. In *2010 3rd International Conference on Biomedical Engineering and Informatics*, volume 2, pages 878–882. IEEE, 2010. (←˒ 5)

Dung Nguyen, Kien Nguyen, Sridha Sridharan, David Dean, and Clinton Fookes. Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition. *Computer Vision and Image Understanding*, 174:33–42, 2018. (←˒ 10)

Peter Nickel and Friedhelm Nachreiner. Sensitivity and diagnosticity of the 0.1-hz component of heart rate variability as an indicator of mental workload. *Human factors*, 45 (4):575–590, 2003. (↩ 8)

Hasan Ocak. Automatic detection of epileptic seizures in eeg using discrete wavelet transform and approximate entropy. *Expert Systems with Applications*, 36(2):2027–2036, 2009. (↩ 5)

Matthew Keith Xi-Jie Pan. *Towards enabling human-robot handovers: exploring nonverbal cues for fluent human-robot handovers.* PhD thesis, University of British Columbia, 2018. (↩ 11)

Erik Peper, Rick Harvey, I-Mei Lin, Hana Tylova, and Donald Moss. Is there more to blood volume pulse than heart rate variability, respiratory sinus arrhythmia, and cardiorespiratory synchrony? *Biofeedback*, 35(2), 2007. (↩ 8)

Luis-Alberto Perez-Gaspar, Santiago-Omar Caballero-Morales, and Felipe Trujillo-Romero. Multimodal emotion recognition with evolutionary computation for human-robot interaction. *Expert Systems with Applications*, 66:42–61, 2016. (↩ 11)

Nicola Petrocchi, Cristina Ottaviani, and Alessandro Couyoumdjian. Compassion at the mirror: Exposure to a mirror increases the efficacy of a self-compassion manipulation in enhancing soothing positive affect and heart rate variability. *The Journal of Positive Psychology*, 12(6):525–536, 2017. (↩ 8)

Stephen W Porges. Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. a polyvagal theory. *Psychophysiology*, 32(4):301–318, 1995. (↩ 4)

Benjamin Recht. A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:253–279, 2019. (↩ 1)

Samantha Reig, Michal Luria, Janet Z. Wang, Danielle Oltman, Elizabeth Jeanne Carter, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. Not some random agent. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, March 2020. doi: 10.1145/3319502.3374795. URL https://doi.org/10.1145/3319502.3374795. (↩ 1)

Ira J Roseman, Martin S Spindel, and Paul E Jose. Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of personality and social psychology*, 59 (5):899, 1990. (↩ 9)

James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980. (↩ 9)

ME Saab and Jean Gotman. A system to detect the onset of epileptic seizures in scalp eeg. *Clinical Neurophysiology*, 116(2):427–442, 2005. (↩ 5)

Alexei V. Samsonovich. Socially emotional brain-inspired cognitive architecture framework for artificial intelligence. *Cognitive Systems Research*, 60:57–76, May 2020. doi: 10.1016/j.cogsys.2019.12.002. URL https://doi.org/10.1016/j.cogsys.2019.12.002. (↩ 2)

Luz Santamaria-Granados, Mario Munoz-Organero, Gustavo Ramirez-Gonzalez, Enas Abdulhay, and NJIA Arunkumar. Using deep convolutional neural network for emotion detection on a physiological signals dataset (amigos). *IEEE Access*, 7:57–67, 2018. (↩ 10)

Claude E Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948. (↩ 5)

David Silver, Aja Huang, Christopher J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–503, 2016. URL http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html. (↩ 1)

Wolf Singer. Synchronization of cortical activity and its putative role in information processing and learning. *Annual review of physiology*, 55(1):349–374, 1993. (↩ 5)

Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 2018. URL https://doi.org/10.1109/TAFFC.2018.2817622. (↩ 10)

Steven J Stroessner. On the social perception of robots: measurement, moderation, and implications. In *Living with Robots*, pages 21–47. Elsevier, 2020. (↩ 1)

Norman L Strominger, Robert J Demarest, and Lois B Laemle. *Noback's human nervous system: structure and function*. Springer Science & Business Media, 2012. (↩ 5)

Michael Suguitan, Randy Gomez, and Guy Hoffman. MoveAE. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, March 2020. doi: 10.1145/3319502.3374807. URL https://doi.org/10.1145/3319502.3374807. (↩ 2)

Leonard J Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, 16(4):437–450, 2000. (↩ 77)

Jose F Téllez-Zenteno and Lizbeth Hernández-Ronquillo. A review of the epidemiology of temporal lobe epilepsy. *Epilepsy research and treatment*, 2012, 2012. (↩ 5)

Rebecca M. Todd, Vladimir Miskovic, Junichi Chikazoe, and Adam K. Anderson. Emotional objectivity: Neural representations of emotions and their interaction with cognition. *Annual Review of Psychology*, 71(1):25–48, January 2020. doi: 10.1146/annurev-psych-010419-051044. URL https://doi.org/10.1146/annurev-psych-010419-051044. (↩ 2)

Samarth Tripathi, Shrinivas Acharya, Ranti Dev Sharma, Sudhanshu Mittal, and Samit Bhattacharya. Using deep and convolutional neural networks for accurate emotion classification on deap dataset. In *Proceedings of the Twenty-Ninth AAAI Conference on Innovative Applications*, pages 4746–4752, 2017. URL https://aaai.org/ocs/index.php/IAAI/IAAI17/paper/view/15007. (↩ 10)

Ching-Chih Tsai, You-Zhu Chen, and Ching-Wen Liao. Interactive emotion recognition using support vector machine for human-robot interaction. In *2009 IEEE International Conference on Systems, Man and Cybernetics*, pages 407–412. IEEE, 2009. (↩ 77)

Christiana Tsiourti, Astrid Weiss, Katarzyna Wac, and Markus Vincze. Multimodal integration of emotional signals from voice, body, and context: Effects of (in) congruence on emotion recognition and attitudes towards robots. *International Journal of Social Robotics*, 11(4):555–573, 2019. (↩ 11)

Mikel Val-Calvo. Biosignals: General BVP, GSR, TMP and ACC experimentation in real-time, an open source software for physiological real-time experimental designs, 2020a. URL https://zenodo.org/record/3759262. (↩ 14 and 76)

Mikel Val-Calvo. Geert: General EEG experimentation in real-time, an open source software for BCI real-time experimental designs, 2020b. URL https://zenodo.org/record/3759306. (↩ 14)

Mikel Val-Calvo. GePHYCAM: General physiological and camera for real-time recordings and processing, 2020c. URL https://zenodo.org/record/3727503. (↩ 14)

Mikel Val-Calvo, José R Álvarez-Sánchez, Jose M Ferrández-Vicente, and Eduardo Fernández. Optimization of real-time eeg artifact removal and emotion estimation for human-robot interaction applications. *Frontiers in Computational Neuroscience*, 13:80, 2019a. URL https://doi.org/10.3389/fncom.2019.00080. (↩ 14)

Mikel Val-Calvo, José Ramón Álvarez-Sánchez, Alejandro Díaz-Morcillo, José Manuel Ferrández Vicente, and Eduardo Fernández-Jover. On the use of lateralization for lightweight and accurate methodology for eeg real time emotion estimation using gaussian-process classifier. In *Understanding the Brain Function and Emotions - IWINAC 2019, Proceedings, Part I*, volume 11486 of *Lecture Notes in Computer Science*, pages 191–201. Springer, 2019b. URL https://doi.org/10.1007/978-3-030-19591-5_20. (↩ 76)

Mikel Val-Calvo, José Ramón Álvarez-Sánchez, Jose Manuel Ferrández-Vicente, Alejandro Díaz-Morcillo, and Eduardo Fernández-Jover. Real-time multi-modal estimation of dynamically evoked emotions using electroencephalography, heart rate and galvanic skin response. *International Journal of Neural Systems*, 30(4):2050013, april 2020a. URL http://dx.doi.org/10.1142/S0129065720500136. (↩ 14)

Mikel Val-Calvo, José Ramón Álvarez-Sánchez, Jose Manuel Ferrández-Vicente, and Eduardo Fernández-Jover. Affective robot story-telling human-robot interaction: Exploratory real-time emotion estimation analysis using facial expressions and physiological signals. *IEEE-access*, 8:134051–134066, 2020b. URL https://doi.org/10.1109/ACCESS.2020.3007109. (↩ 14 and 76)

Biel Piero E. Alvarado Vásquez and Fernando Matía. A tour-guide robot: Moving towards interaction with humans. *Engineering Applications of Artificial Intelligence*, 88:103356, February 2020. doi: 10.1016/j.engappai.2019.103356. URL https://doi.org/10.1016/j.engappai.2019.103356. (↩ 2)

Guang-Zhong Yang, Jim Bellingham, Pierre E. Dupont, Peer Fischer, Luciano Floridi, Robert Full, Neil Jacobstein, Vijay Kumar, Marcia McNutt, Robert Merrifield, Bradley J. Nelson, Brian Scassellati, Mariarosaria Taddeo, Russell Taylor, Manuela Veloso, Zhong Lin Wang, and Robert Wood. The grand challenges ofScience robotics. *Science Robotics*, 3(14):eaar7650, January 2018. doi: 10.1126/scirobotics.aar7650. URL https://doi.org/10.1126/scirobotics.aar7650. (↩ 1)

Zhong Yin, Mengyuan Zhao, Yongxiong Wang, Jingdong Yang, and Jianhua Zhang. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer methods and programs in biomedicine*, 140:93–110, 2017. (↩ 10)

Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. From facial expression recognition to interpersonal relation prediction. *International Journal of Computer Vision*, 126(5):550–569, November 2017. doi: 10.1007/s11263-017-1055-1. URL https://doi.org/10.1007/s11263-017-1055-1. (↩ 14)

Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, 2015. (↩ 10, 14, and 27)

Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*, (99):1–13, 2018. (↩ 10)

Wei-Long Zheng, Jia-Yi Zhu, and Bao-Liang Lu. Identifying stable patterns over time for emotion recognition from eeg. *IEEE Transactions on Affective Computing*, 10(3): 417–429, 2019. (↩ 10)