Modelling human emotions using immersive virtual reality, physiological signals and behavioural responses

Author:
Javier Marín Morales

Supervisors:
Dr. Mariano Alcañiz Raya
Dra. María del Carmen Llinares
Dr. Jaime Guixeres Provinciale

April, 2020

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Universitat Politècnica de València

Programa de doctorado en Tecnologías para la Salud y el Bienestar

Ph.D. Thesis

Modelling human emotions using immersive virtual reality, physiological signals and behavioural responses

April 2020

Author:

Javier Marín Morales

Supervisors:

Dr. Mariano Alcañiz Raya

Dra. María del Carmen Llinares Millán

Dr. Jaime Guixeres Provinciale

Abstract

In recent years the scientific community has significantly increased its use of virtual reality (VR) technologies in human behaviour research. In particular, the use of immersive VR has grown due to the introduction of affordable, high performance head mounted displays (HMDs). Among the fields that has strongly emerged in the last decade is affective computing, which combines psychophysiology, computer science, biomedical engineering and artificial intelligence in the development of systems that can automatically recognize emotions. The progress of affective computing is especially important in human behaviour research due to the central role that emotions play in many background processes, such as perception, decision-making, creativity, memory and social interaction.

Several studies have tried to develop a reliable methodology to evoke and automatically identify emotional states using objective physiological measures and machine learning methods. However, the majority of previous studies used images, audio or video to elicit emotional statements; to the best of our knowledge, no previous research has developed an emotion recognition system using immersive VR. Although some previous studies analysed physiological responses in immersive VR, they did not use machine learning techniques for biosignal processing and classification.

Moreover, a crucial concept when using VR for human behaviour research is validity: the capacity to evoke a response from the user in a simulated environment similar to the response that might be evoked in a physical environment. Although some previous studies have used psychological and cognitive dimensions to compare responses in real and virtual environments, few have extended this research to analyse physiological or behavioural responses. Moreover, to our knowledge, this is the first study to compare VR scenarios with their real-world equivalents using physiological measures coupled with machine learning algorithms, and to analyse the ability of VR to transfer and extrapolate insights obtained from VR environments to real environments.

The main objective of this thesis is, using psycho-physiological and behavioural responses in combination with machine learning methods, and by performing a direct comparison between a real and virtual environment, to validate immersive VR as an emotion elicitation tool. To do so we develop an experimental protocol involving emotional 360º environments, an art exhibition in a real museum, and a highly-realistic 3D virtualization of the same art exhibition.

The set of emotional 360º panoramas were four versions of the same virtual room, designed to elicit four possible arousal-valence combinations. In addition, a set of features was extracted from electroencephalography (EEG) and electrocardiography (ECG) signals, which were then input into a support vector machine classifier to recognize subjects' arousal and valence perceptions. The model´s accuracy was 75.00% along the arousal dimension and 71.21% along the valence dimension. The findings validated the use of immersive 360º panoramas to elicit and automatically recognize different emotional states based on neural and cardiac dynamics; in addition, this represents the first emotion recognition system designed to operate in combination with an HMD.

As to the museums, we analysed the psycho-physiological patterns evoked during a free exploration of an actual art exhibition and the same exhibition virtualized through a 3D immersive virtual environment. The majority of the stimuli did not present statistical differences in terms of emotional self-assessment. In addition, an emotion recognition system was developed, using a support vector machine in combination with cardiovascular and linear and nonlinear brain dynamics, in both the real and the virtual museum. The 2-class (high/low) system accuracy was 71.52% and 77.08% along the arousal and valence dimensions, respectively, in the physical museum, and 75.00% and 71.08% in the virtual museum. We also developed a real vs. virtual classifier, achieving an accuracy of 95.27%, using only EEG mean phase coherency features, which demonstrated the high involvement of brain synchronization in emotional virtual reality processes. These insights provide an important contribution at the methodological level and to scientific knowledge, which can guide future emotion elicitation and recognition systems using VR.

Moreover, we compared the navigation patterns of the subjects in the real and virtual museums, as these can radically condition environmental perception and, therefore, alter the various evoked responses. The movement patterns in both museums were, in general, similar, but there were significant differences at the beginning of the exploration, that is, there were time-dependent differences in the patterns during the first 2 minutes of the tours. Subsequently, no significant differences were observed in the navigation patterns between the physical and

the virtual museum. These findings support the use, at navigation level, of immersive virtual environments as empirical tools in human behavioural research.

This thesis provides novel contributions to the use of immersive VR in human behaviour research, particularly in relation to emotions. VR can help in the application of methodologies designed to present more realistic stimuli in the assessment of daily-life environments and situations, thus overcoming the current limitations of affective elicitation, which classically uses images, audio and video. Moreover, it analyses the validity of VR by performing a direct comparison using highly-realistic simulation. We believe that immersive VR will revolutionize laboratory-based emotion elicitation methods. Moreover, its synergy with physiological measurement and machine learning techniques will impact transversely in many other research areas, such as architecture, health, assessment, training, education, driving and marketing, and thus open new opportunities for the scientific community. The present dissertation aims to contribute to this progress.

Resumen

El uso de la realidad virtual (RV) se ha incrementado notablemente en la comunidad científica para la investigación del comportamiento humano. En particular, la RV inmersiva ha crecido debido a la democratización de las gafas de realidad virtual o *head mounted displays* (HMD), que ofrecen un alto rendimiento con una inversión económica. Uno de los campos que ha emergido con fuerza en la última década es el *Affective Computing*, que combina psicofisiología, informática, ingeniería biomédica e inteligencia artificial, desarrollando sistemas que puedan reconocer emociones automáticamente. Su progreso es especialmente importante en el campo de la investigación del comportamiento humano, debido al papel fundamental que las emociones juegan en muchos procesos psicológicos como la percepción, la toma de decisiones, la creatividad, la memoria y la interacción social.

Muchos estudios se han centrado en intentar obtener una metodología fiable para evocar y automáticamente identificar estados emocionales, usando medidas fisiológicas objetivas y métodos de aprendizaje automático. Sin embargo, la mayoría de los estudios previos utilizan imágenes, audios o vídeos para generar los estados emocionales y, hasta donde llega nuestro conocimiento, ninguno de ellos ha desarrollado un sistema de reconocimiento emocional usando RV inmersiva. Aunque algunos trabajos anteriores sí analizan las respuestas fisiológicas en RV inmersivas, estos no presentan modelos de aprendizaje automático para procesamiento y clasificación automática de bioseñales.

Además, un concepto crucial cuando se usa la RV en investigación del comportamiento humano es la validez: la capacidad de evocar respuestas similares en un entorno virtual a las evocadas por el espacio físico. Aunque algunos estudios previos han usado dimensiones psicológicas y cognitivas para comparar respuestas entre entornos reales y virtuales, las investigaciones que analizan respuestas fisiológicas o comportamentales están mucho menos extendidas. Según nuestros conocimientos, este es el primer trabajo que compara entornos físicos con su réplica en RV, empleando respuestas fisiológicas y algoritmos de aprendizaje automático y analizando la capacidad de la RV de transferir y extrapolar las conclusiones obtenidas al entorno real que se está simulando.

El objetivo principal de la tesis es validar el uso de la RV inmersiva como una herramienta de estimulación emocional usando respuestas psicofisiológicas y comportamentales en combinación con algoritmos de aprendizaje automático, así como realizar una comparación directa entre un entorno real y virtual. Para ello, se ha desarrollado un protocolo experimental que incluye entornos emocionales 360º, un museo real y una virtualización 3D altamente realista del mismo museo.

Con respecto al conjunto de entornos emocionales 360º, se diseñaron cuatro alternativas de una habitación virtual para generar las cuatro posibles combinaciones de *arousal-valencia* alto y bajo. Además, se obtuvieron un conjunto de variables de las señales de encefalograma (EEG) y electrocardiografía (ECG), que fueron procesadas junto con un clasificador *Support Vector Machine* para reconocer la percepción del sujeto en términos de *arousal* y *valencia*. Estos resultados validan el uso de los panoramas 360º inmersivos para generar y reconocer automáticamente diferentes estados emocionales utilizando las dinámicas cerebrales y cardiacas, y suponen el primer sistema de reconocimiento emocional utilizando un HMD.

En lo que concierne al análisis del museo, fueron estudiados los patrones psicofisiológicos evocados durante la exploración libre de una exhibición de arte real y durante la virtualización de la misma mediante un escenario de RV 3D inmersivo. La mayoría de los estímulos no presentaron diferencias estadísticas en términos de la autoevaluación emocional de los sujetos. Además, un sistema de reconocimiento emocional fue desarrollado usando un *Support Vector Machine* en combinación con las dinámicas cerebrales y cardiovasculares en el museo real y virtual. La precisión del clasificador de dos clases (alto/bajo) fue de 71.52% y 77.08% en las dimensiones de *arousal* y *valencia* respectivamente en el museo real, y de 75.00% y 71.08% en el museo virtual. Por otro lado, también se ha desarrollado un clasificador para discriminar entre los estímulos reales y los virtuales, que ha alcanzado una precisión del 95.27% utilizando solo variables de *mean phase coherency* del EEG, lo que demuestra la alta implicación de la sincronía cerebral en los procesos emocionales en RV. Estos resultados aportan una importante contribución a nivel metodológico y de conocimiento científico, guiando futuras estimulaciones emocionales y sistemas de reconociendo usando RV.

Asimismo, se han analizado los patrones de navegación tanto en el museo real y como en el virtual, ya que estos pueden condicionar radicalmente la percepción del entorno y, por ello, alterar las respuestas evocadas. Los patrones en ambos museos presentan una alta similitud, mostrando diferencias significativas al principio de la exploración, en términos del área explorada y del tiempo dedicado a visitar la exposición. Los resultados muestran que estas existen durante los dos primeros minutos del recorrido y, a partir de ese momento, no hay diferencias entre el museo real y el virtual en términos de navegación. Estos resultados apoyan el uso de la RV inmersiva como herramienta de investigación del comportamiento humano a nivel de navegación.

La tesis presenta novedosas contribuciones del uso de la RV inmersiva en la investigación del comportamiento humano, en particular en lo relativo al estudio de las emociones. Esta ayudará a aplicar metodologías a estímulos más realistas para evaluar entornos y situaciones de la vida diaria, superando las actuales limitaciones de la estimulación emocional que clásicamente ha incluido imágenes, audios o vídeos. Además, en ella se analiza la validez de la RV realizando una comparación directa usando una simulación altamente realista. Creemos que la RV inmersiva va a revolucionar los métodos de estimulación emocional en entornos de laboratorio. Además, su sinergia junto a las medidas fisiológicas y las técnicas de aprendizaje automático, impactarán transversalmente en muchas áreas de investigación como la arquitectura, la salud, la evaluación psicológica, el entrenamiento, la educación, la conducción o el marketing, abriendo un nuevo horizonte de oportunidades para la comunidad científica. La presente tesis espera contribuir a caminar en esa senda.

Resum

L'ús de la realitat virtual (RV) s'ha incrementat notablement en la comunitat científica per a la recerca del comportament humà. En particular, la RV immersiva ha crescut a causa de la democratització de les ulleres de realitat virtual o *head mounted displays* (HMD), que ofereixen un alt rendiment amb una reduïda inversió econòmica . Un dels camps que ha emergit amb força en l'última dècada és el *Affective Computing*, que combina psicofisiologia, informàtica, enginyeria biomèdica i intel·ligència artificial, desenvolupant sistemes que puguen reconéixer emocions automàticament. El seu progrés és especialment important en el camp de la recerca del comportament humà, a causa del paper fonamental que les emocions juguen en molts processos psicològics com la percepció, la presa de decisions, la creativitat, la memòria i la interacció social.

Molts estudis s'han centrat en intentar obtenir una metodologia fiable per a evocar i automàticament identificar estats emocionals, utilitzant mesures fisiològiques objectives i mètodes d'aprenentatge automàtic. No obstant això, la major part dels estudis previs utilitzen imatges, àudios o vídeos per a generar els estats emocionals i, fins on arriba el nostre coneixement, cap d'ells ha desenvolupat un sistema de reconeixement emocional mitjançant l'ús de la RV immersiva. Encara que alguns treballs anteriors sí que analitzen les respostes fisiològiques en RV immersives, aquests no presenten models d'aprenentatge automàtic per a processament i classificació automàtica de biosenyals.

A més, un concepte crucial quan s'utilitza la RV en la recerca del comportament humà és la validesa: la capacitat d'evocar respostes similars en un entorn virtual a les evocades per l'espai físic. Encara que alguns estudis previs han utilitzat dimensions psicològiques i cognitives per a comparar respostes entre entorns reals i virtuals, les recerques que analitzen respostes fisiològiques o comportamentals estan molt menys esteses. Segons els nostres coneixements, aquest és el primer treball que compara entorns físics amb la seua rèplica en RV, emprant respostes fisiològiques i algorismes d'aprenentatge automàtic i analitzant la capacitat de la RV de transferir i extrapolar les conclusions obtingudes a l'entorn real que s'està simulant.

L'objectiu principal de la tesi és validar l'ús de la RV immersiva com una eina d'estimulació emocional usant respostes psicofisiològiques i comportamentals en combinació amb algorismes d'aprenentatge automàtic, així com realitzar una comparació directa entre un entorn real i virtual. Per a això, s'ha desenvolupat un protocol experimental que inclou entorns emocionals 360º, un museu real i una virtualització 3D altament realista del mateix museu.

Respecte al conjunt d'entorns emocionals 360º, es van dissenyar quatre alternatives d'una habitació virtual per a generar les quatre possibles combinacions d' *arousal-valencia* alt i baix. A més, es van obtenir un conjunt de variables dels senyals d'encefalograma (EEG) i electrocardiografia (ECG), que van ser processades conjuntament amb un classificador *Support Vector Machine* per a reconéixer la percepció del subjecte en termes d'*arousal* i *valència*. Aquests resultats validen l'ús dels panorames 360º immersius per a generar i reconéixer automàticament diferents estats emocionals mitjançant les dinàmiques cerebrals i cardíaques, i suposen el primer sistema de reconeixement emocional utilitzant un HMD.

En el que pertoca a l'anàlisi del museu, van ser estudiats els patrons psicofisiològics evocats durant l'exploració lliure d'una exhibició d'art real i durant la virtualització de la mateixa mitjançant un escenari de RV 3D immersiu. La major part dels estímuls no van presentar diferències estadístiques en termes de l'autoavaluació emocional dels subjectes. A més, un sistema de reconeixement emocional va ser desenvolupat utilitzant un *Support Vector Machine* en combinació amb les dinàmiques cerebrals i cardiovasculars en el museu real i virtual. La precisió del classificador de dues classes (alt/baix) va ser de 71.52% i 77.08% en les dimensions d'*arousal* i *valencia* respectivament en el museu real, i de 75.00% i 71.08% en el museu virtual. D'altra banda, també s'ha desenvolupat un classificador per a discriminar entre els estímuls reals i els virtuals, que ha aconseguit una precisió del 95.27% utilitzant només variables de *mean phase coherency* del EEG, la qual cosa demostra l'alta implicació de la sincronia cerebral en els processos emocionals en RV. Aquests resultats aporten una important contribució en l'àmbit metodològic i de coneixement científic, guiant futures estimulacions emocionals i sistemes de reconeixent usant RV.

Així mateix, s'han analitzat els patrons de navegació tant en el museu real i com en el virtual, ja que aquests poden condicionar radicalment la percepció de l'entorn i, per això, alterar les respostes evocades. Els patrons en tots dos museus presenten una alta similitud, mostrant diferències significatives al principi de l'exploració, en termes de l'àrea explorada i del temps dedicat a visitar l'exposició. Els resultats mostren que aquestes existeixen durant els dos primers minuts del recorregut i, a partir d'aquest moment, no hi ha diferències entre el museu real i el virtual en termes de navegació. Aquests resultats donen suport a l'ús de la RV immersiva com a eina de recerca del comportament humà en l'àmbit de navegació.

La tesi presenta noves contribucions de l'ús de la RV immersiva en la recerca del comportament humà, en particular quant a l'estudi de les emocions. Aquesta ajudarà a aplicar metodologies a estímuls més realistes per a avaluar entorns i situacions de la vida diària, superant les actuals limitacions de l'estimulació emocional que clàssicament ha inclòs imatges, àudios o vídeos. A més, en ella s'analitza la validesa de la RV realitzant una comparació directa usant una simulació altament realista. Creiem que la RV immersiva revolucionarà els mètodes d'estimulació emocional en entorns de laboratori. A més, la seua sinergia al costat de les mesures fisiològiques i les tècniques d'aprenentatge automàtic, impactaran transversalment en moltes àrees de recerca com l'arquitectura, la salut, l'avaluació psicològica, l'entrenament, l'educació, la conducció o el màrqueting, obrint un nou horitzó d'oportunitats per a la comunitat científica. La present tesi espera contribuir a caminar en aquesta senda.

Acknowledgements

A mi padre por enseñarme la importancia del ejemplo y la constancia.

A mi madre por enseñarme lo que significa la valentía y la fuerza.

A mis hermanos y familia por estar a mi lado incondicionalmente.

A Adriana por ser mi apoyo en esta etapa de descubrimiento constante y hacerme feliz cada día.

A mis amigos por todos los buenos momentos que me brindan.

A mis compañeros del i3B por enseñarme tanto y hacerme sentir como en casa.

A mis directores Mª Carmen, Mariano y Jaime, por haberme guiado en este camino.

A Gaetano perché senza il suo contributo questa tesi non sarebbe possibile.

Al IVAM por abrirme sus puertas.

Muchas gracias.

Content

List of Figures

List of Tables

Chapter 1

Introduction

"The good thing about science is that it's true whether or not you believe in it."
Neil deGrasse Tyson

Motivation

Virtual reality (VR) has experienced an increase in its popularity in recent years in scientific and commercial contexts (Cipresso et al., 2018). Its general applications include gaming, training, education, health, marketing, design and many apps that can be downloaded from company platforms, such as HTC Vive, Oculus or Sony PlayStation. This increase is based on the development of a new generation of low-cost headsets which has democratised global purchases of head mounted displays (HMDs) (Castelvecchi, 2016).

Nonetheless, VR has been used in research since the 1990s (Slater & Usoh, 1994). The scientific interest in VR is due to the fact that it provides simulated experiences that create the sensation of being in the real world (Giglioli et al., 2017). In particular, environmental simulations are representations of physical environments that allow researchers to analyse reactions to common concepts (Kwartler, 2005). They are especially important when what they depict cannot be physically represented. VR makes it possible to study these scenarios under controlled laboratory conditions (Vince, 2004). Moreover, VR allows the time and cost-effective isolation and modification of variables, unfeasible in real space (Alcañiz et al., 2003).

Virtual reality set-ups: formats, displays and interfaces

The set-ups that display VR simulations have been progressively integrated into studies as the relevant technologies have evolved; These consist of a combination of three objective features, formats, display devices and user interfaces.

The format describes the structure of the information displayed. The most common are 2D multimedia and 3D environments; the main difference between them is their levels of interactivity (Mengoni et al., 2011). 2D multimedia, including 360° panoramic images and videos, provide non-interactive visual representations. The validity of this format has been extensively explored (Stamps III, 1990). Moreover, the latest advances in computer-generated images simulate light, texture and atmospheric effects to such a degree of photorealism that it is possible to produce a virtual image that is indistinguishable, to the naked eye, from a photograph of a real world scene (Morinaga et al., 2018). This format allows scientists to test static computer-generated environments, with many variations, cheaply and quickly in a laboratory. On the other hand, 3D environments generate interactive representations which allow changes in the user's point of view, navigation and even interaction with objects and persons (Siriaraya & Ang, 2019). Developing realistic 3D environments is more time consuming than developing 360° computer-generated photographs, and their level of realism is limited by the power of the hardware. However, the processing potency of GPUs (graphics processing units) is increasing every year, which will enhance the performance of 3D environments. Moreover, the interaction capacity of 3D environments, which facilitates the simulation of real-world tasks, is a key aspect in the application of virtual reality (Cipresso et al., 2018).

The display devices are the technological equipment used to visualize the formats. They are classified according to the immersion they provide, that is, the sensorimotor contingencies that they support. These are related to the actions that experimental subjects carry out in the perception process, for example, when they bend down and shift the position of their heads, and their gaze direction, to see underneath an object. Therefore, the sensorimotor contingencies supported by a system define a set of valid actions (e.g. turning the head, bending forward) that carry meaning in terms of perception within the virtual environment (Slater, 2009). Since immersion is objective, one system is more immersive than another if it is superior in at least one characteristic while the others are equal.

There are three categories of immersion system, non-immersive, semi-immersive and immersive (Cipresso et al., 2018). Non-immersive systems are simpler devices which use a single-screen, such as a desktop PC, to display environments (Kober et al., 2012). Semi-immersive systems, such as the cave automatic virtual environment (CAVE), or the powerwall screen, use large projections to display environments on walls, enveloping the viewer (Borrego et al., 2016; Clemente et al., 2014). These displays typically provide a stereo image of an environment, using a perspective projection linked to the position of the observer's head. Immersive devices, such as HMDs, provide fully-immersive systems that isolate the user from external world stimuli (Borrego et al., 2018). These provide a complete simulated experience, including a stereoscopic view, which responds to the user's head movements. During the last two decades VR has usually been displayed through desktop PCs or semi-immersive systems, such as CAVEs and powerwalls (Vecchiato et al., 2015). However, improvements in the performance, and availability, of the new generation of HMDs is boosting their use in research (Jensen & Konradsen, 2017).

The user interfaces, which are exclusive to 3D environments which allow this level of interaction, are the functional connections between the user and the VR environment which allow him or her to interact with objects and navigate (Riecke et al., 2018). Regarding interaction with objects, manipulation tasks include selection, that is, acquiring or identifying an object or subset of objects; positioning, that is, changing an object's 3D position; and rotation, that is, changing an object's 3D orientation. In terms of the navigation metaphors in 3D environments, virtual locomotion has been thoroughly analysed (Templeman et al., 1999), and can be classified as physical or artificial. Regarding the physical, there are room-scale based metaphors, such as real-walking, which allow the user to walk freely inside a limited physical space. These are normally used with HMDs; position and orientation are determined by the user's head. They are the most naturalistic of the metaphors, but are highly limited by the physical tracked area (E. Bozgeyikli, L. Bozgeyikli et al., 2016). In addition, there are motion-based metaphors, such as walking-in-place or redirected walking. Walking-in-place is a pseudo-naturalistic metaphor where the user performs a virtual locomotion to navigate, for example, by moving his/her hands as if (s)he was walking, or by performing footstep-like movements, while remaining stationary (Tregillus & Folmer, 2016). Redirected walking is a technique where the user perceives (s)he is walking freely but, in fact, is being unknowingly manipulated by the virtual display: this allows navigation in an environment larger than the actual tracked area (Nescher et al., 2014). Regarding the artificial, controller-based metaphors allow the user to control their movements directly through a joystick or a similar device, such as a keyboard or a trackball (Nabiyouni et al., 2015). In addition, teleportation-based metaphors allow the user to point where (s)he wants to go and teleport him or her there with an instantaneous "jump" (Bozgeyikli, Raij, et al., 2016). Moreover, recent advancements in latest generation HMD devices have increased the performance of navigation metaphors. Point-and-click teleport metaphors have become mainstream technologies implemented in all low-cost devices. However, other techniques have also increased in performance: walking-in-place metaphors have become more user-friendly and robust; room-scale based metaphors now have increased coverage areas, provided by low-cost tracking methods; and controller-based locomotion now addresses virtual sickness through effective, dynamic field-of-view adjustments (Boletsis, 2017).

Navigation in Virtual Reality

The high interactivity offered by 3D environments also causes an issue that must be considered; navigation can radically condition space perception and, therefore, alter evoked responses. Navigation has been divided into wayfinding and travel components (LaViola et al., 2017). Wayfinding is the cognitive process of establishing a route or path from an origin to a destination. It has been analysed in VR with topics where the wayfinding component is critical to task performance, such as firefighting training (Bliss et al., 1997). In addition, researchers have compared the wayfinding performance of users in physical and virtual environments and concluded that their performance in virtual environments is poorer (Richardson et al., 1999; van der Ham et al., 2015). However, other authors have claimed that the technical limitations of VR, which impact on user involvement, might be the cause of this lower performance (Lessels & Ruddle, 2005); thus, further research needs to be carried out using the latest generation HMDs, which are less technically limited than the previous generation.

Second, the travel function is related to the task of moving from one point to another. It is very influenced by the navigation metaphor used to perform the navigation; many have been analysed, such as joysticks, walking-in-place and teleportation (Riecke et al., 2010; Tregillus et al., 2017; Wilson et al., 2014). However, there is no consensus as to which is the standard navigation method (Lee et al., 2018). Moreover, there is a lack of research comparing the influence of different metaphors on the travel component, and comparing how similar is its operation in the virtual and physical scenarios, although this radically affects how subjects perceive environments.

Sense of presence

In addition to the objective features of the set-up, the experience of users in virtual environments can be measured by the concept of presence, understood as the subjective feeling of "being-there" (Slater & Wilbur, 1997). A high degree of presence creates in the user the sensation of physical presence and the illusion of interacting and reacting as if (s)he was in the real world (Heeter, 1992). In the 2000s, the strong illusion of being in a place, in spite of the sure knowledge that one is not actually there, was characterised as "place illusion" (PI), to avoid any confusion that might be caused by the use of multiple meanings of the word "presence". Moreover, just as PI relates to how the world is perceived, and the correlation of movements and concomitant changes in the images that form perceptions, "plausibility illusion" (PsI) relates to what is perceived, in a correlation of external events not directly caused by the participant (Slater, 2009). PsI is determined by the extent to which a system produces events that directly relate to the participant, and the overall credibility of the scenario being depicted in comparison with viewer expectations, for example, when an experimental participant is provoked into giving a quick, natural and automatic reply to a question posed by an avatar.

Although presence plays a critical role in VR experiences, there is limited understanding of what factors affect presence in virtual environments. However, there is consensus that exteroception and interoception factors affect presence. It has been shown that exteroception factors, such as higher levels of interactivity and immersion, which are directly related to the experimental set-up, provoke increased presence, especially in virtual environments not designed to induce particular emotions (Baños et al., 2004; Slater et al., 1994; Usoh et al., 1999). As to the interoception factors, which are defined by the content displayed, participants will have higher presence ratings if they feel emotionally affected; for example, previous studies have found a strong correlation between arousal and presence (Diemer et al., 2015). Recent research has also analysed presence in specific contexts and suggested that, for example, in social environments, it is enhanced when the VR elicits genuine cognitive, emotional, and behavioural responses, and when participants create their own narratives about events. (Riches et al., 2019). On the other hand, presence decreases when users experience physical problems, such as cybersickness (Kiryu & So, 2007).

Virtual Reality in Human Behaviour Research

VR is, thus, proposed as a powerful tool to simulate complex, real situations and environments, offering researchers unprecedented opportunities to investigate human behaviour in closely-controlled designs in controlled laboratory conditions (Diemer et al., 2015). There are now many researchers in the field, who have published many studies, so a strong, interdisciplinary community exists (Cipresso et al., 2018).

Education and training is one field where VR has been much applied. Freina and Ott (2015) showed that VR can offer great educational advantages. It can solve time-travel problems; for example, students can experience different historical periods. It can address physical inaccessibility; for example, students can explore the solar system in the first person. It can circumnavigate ethical problems; students can “perform” serious surgery. Surgical training is now one of the most closely analysed research topics. Interventional surgery lacked satisfactory training methods before the advent of VR, except learning on real patients (Alaraj et al., 2011). Bhagat, Liou and Chang (2016) analysed improvements in military training. These authors suggested that cost-effective 3D VR significantly improved subjects learning motivation and outcomes, and provided a positive impact on their live firing achievement scores. In addition, besides enhancements in cost-effectivity, VR offers a safe training environment, as evidenced by the extensive research into driving and flight simulators (Dols et al., 2016; Yavrucuk et al., 2011). Moreover, de-Juan-Ripoll et al. (2018) proposed that VR is an invaluable tool to assess risk-taking profiles and to train in related skills, due to its transferability to real-world situations.

Several researchers have also demonstrated the effectiveness of VR in therapeutic applications. It offers some distinct advantages over standard therapies, including precise control over the degree of exposure to the therapeutic scenario, the possibility of tailoring scenarios to individual patients’ needs, and even the capacity to provide therapies that might otherwise be impossible (Bohil et al., 2011). Taking some examples, studies, using VR, have analysed the improvement in the training in social skills for persons with mental or behavioural disorders, such as phobias (Peperkorn et al., 2014), schizophrenia (Park et al., 2011) and autism (Didehbani et al., 2016). Lloréns, Noé, Colomer and Alcañiz (2015) showed that VR-based telerehabilitation interventions promoted the reacquisition of locomotor skills associated with balance, in the same way as in-clinic interventions (both complemented with conventional therapy programmes). Moreover, it has been proposed as a key tool for the diagnosis of neurodevelopmental disorders (Alcañiz et al., 2019).

In addition, VR has been applied transversally to many fields, such as architecture and marketing. In architecture, VR has been used as a framework within which to test the overall validity of proposed plans and architectural designs, generate alternatives and conceptualize learning, instruction and the design process itself (Portman et al., 2015). In marketing it has been applied in the analysis of consumer behaviour in laboratory-controlled conditions (Bigné et al., 2015) and as a tool to develop emotionally engaging consumer experiences (Alcañiz et al., 2019).

One of the most important topics in human behaviour research is human emotions, due to the central role that they play in many background processes, such as perception, decision-making, creativity, memory and social interaction (Picard, 2003). Given the presence that VR provokes in users, it has been suggested as a powerful means of evoking emotions in laboratory environments (Alcañiz et al., 2003). In one of the first confirmatory studies into the efficacy of immersive VR as an affective medium, Baños et al. (2004) showed that emotion has an impact on presence. Subsequently, many other similar studies showed that VR can evoke emotions, such as anxiety and relaxation (Riva et al., 2007), positive valence in obese children taking exercise (Guixeres et al., 2013), arousal in natural environments, such as parks (Felnhofer et al., 2015) and different moods in social environments featuring avatars (Lorenzo et al., 2016).

Implicit measures and the neuroscience approach

Traditionally, most theories of human behaviour research have been based on a model of the human mind that assumes that humans can think about and verbalize accurately their attitudes, emotions and behaviours (Brief, 1998). Therefore, classical psychological evaluations used self-assessment questionnaires and interviews to quantify subjects' responses. However, these explicit measures have been demonstrated to be subjective, as stereotype-based expectations can lead to systematically biased behaviour, given that most individuals are motivated to be, or appear, nonbiased (Payne, 2001). The terms used in questionnaires can also be differentially interpreted by respondents; and the outcomes depend on the subjects possessing a wide knowledge of their dispositions, which is not always the case (Schmitt, 1994).

Recent advances in neuroscience show that most of the brain processes that regulate our emotions, attitudes and behaviours are not conscious. In contrast to explicit processes, humans cannot verbalize these implicit processes (Barsade et al., 2009). In recent years, growing interest has developed in "looking" inside the brain to seek solutions to problems that have not traditionally been addressed by neuroscience. In this sense, neuroscience offers techniques that can recognise implicit measurements not controlled by conscious processes (Lieberman, 2007). These developments have provoked the emergence in last decade of a new field called neuroeconomics, which blends psychology, neuroscience, and economics into models of decision-making, rewards, risks, and uncertainties (Camerer et al., 2005). Neuroeconomics addresses human behaviour research, in particular the brain mechanisms involved in economic decision-making, from the point of view of cognitive neuroscience, using implicit measures.

Several implicit measuring techniques have been proposed in recent years. Some examples of their applications in human behaviour research are: heart rate variability (HRV) has been correlated with arousal changes in vehicle drivers to detect critical points in a route (Riener et al., 2009); electrodermal activity (EDA) has been used to measure stress caused by cognitive load in the workplace (Setz et al., 2009); electroencephalogram (EEG) has been used to assess engagement in audio-visual content (Berka et al., 2007); functional magnetic resonance imaging (fMRI) has been used to record the brain activity of participants engaged in social vs. mechanical/analytic tasks (Jack et al., 2013); functional near-infrared spectroscopy (fNIRS) has been used as a direct measure of brain activity related to decision-making processes in approach-avoidance theories (Ernst et al., 2013); eye-tracking (ET) has been used to measure subconscious brain processes that show correlations with information processing in risky decisions (Glöckner & Herbold, 2011); facial expression analysis (FEA) has been applied to detect emotional responses in e-learning environments (Bahreini et al., 2016); and speech emotion recognition (SER) has been used to detect depressive disorders (Huang et al., 2018).

Table 1.1 gives an overview of the implicit measuring techniques that have been used in human behaviour research.

Implicit technique	Biometric signal measured	Sensor	Features	Psychological or behavioural construct inferred
EDA (electro dermal activity)	Changes in skin conductance	Electrodes attached to fingers, palms or soles	Skin conductance response, tonic activity and phasic activity	Attention and arousal (Prokasy, 2012)

HRV (heart rate variability)	Variability in heart contraction intervals	Electrodes attached to chest or limbs or optical sensor attached to finger, toe or earlobe	Time domain, frequency domain, non-linear domain	Stress, anxiety, arousal and valence (Kim et al., 2018; Kreibig, 2010)
EEG (electroencep halogram)	Changes in electrical activity of the brain	Electrodes placed on scalp	Frequency band power, functional connectivity, event-related potentials	Attention, mental workload, drowsiness, fatigue, arousal and valence (Gruzelier, 2014; Lotte et al., 2018)
fMRI (functional magnetic resonance imaging)	Concentrations of oxygenated vs. deoxygenated haemoglobin in the blood vessels of the brain	Magnetic resonance signal	blood-oxygen-level dependent	Motor execution, attention, memory, pain, anxiety, hunger, fear, arousal and valence (Thibault et al., 2018)
fNIRS (functional near-infrared spectroscopy)	Concentrations of oxygenated vs. deoxygenated haemoglobin in the blood	Near-infrared light placed on scalp	blood-oxygen-level dependent	Motor execution, cognitive task (mental arithmetic), decision-making and valence (Naseer & Hong, 2015)
ET (eye tracking)	Corneal reflection & pupil dilation	Infrared cameras point towards eyes	Eye movements (gaze, fixation, saccades), blinks, pupil dilation	Visual attention, engagement, drowsiness and fatigue (Meißner & Oll, 2019)
FEA (facial expression analysis)	Activity of facial muscles	Camera points towards face	Position and orientation of head. Activation of action units	Basic emotions, engagement, arousal and valence (Calvo & Nummenmaa, 2016)
SER (speech emotion recognition)	Voice	Microphone	Prosodic and spectral features	Stress, basic emotions, arousal and valence (Schuller, 2018)

Table 1.1. Overview of the implicit techniques used in human behaviour research and their main applications

In addition, recent researches have highlighted the potential of virtual reality environments for enhancing ecological validity in the clinical, affective, and social neurosciences; these studies have usually involved the use of simple and static stimuli which lack many of the potentially important aspects of real-world activities and interactions (Parsons, 2015). Therefore, VR could play an important role in the future of neuroeconomics by providing a more ecological framework within which to develop experimental studies with implicit measures.

Affective computing and emotion recognition systems

Affective Computing, which analyses human responses using implicit measures, has developed into an important field of study in last decades. Introduced by Rosalind Picard in 1997, it proposed the automatic quantification and recognition of human emotions as an interdisciplinary field based on psychophysiology, computer science, biomedical engineering and artificial intelligence (Picard, 1997). The automatic recognition of human emotion statements using implicit measures can be transversally applied to all human behaviour topics and complement classic explicit measures. In particular, it can be applied to neuroeconomics research as they share the same neuroscientific approach of using implicit measures, and due to the important relationship that has been found between emotions and decision-making (Camerer et al., 2005).

Emotion recognition models can be divided into three: emotional modelling, emotion classification and emotion elicitation.

The emotional modelling approach can be divided into discrete and dimensional. Discrete models characterize the emotion system as a set of basic emotions which includes anger, disgust, fear, joy, sadness and surprise, and the complex emotions that result from combining them (Ekman, 1999). On the other hand, dimensional models suggest that emotional responses can be modelled in a multidimensional space where each dimension represents a fundamental property common to all emotions. The most commonly-used theory is the circumplex model of affect, which proposes a two-dimensional space, consisting of valence, that is, the degree to which an emotion is perceived as positive or negative, and arousal, that is, the intensity of the emotion in terms of activation from low to high (Russell & Mehrabian, 1977).

Affective computing uses biometric signals and machine learning algorithms to automatically classify emotions. Many signals have been used, such as voice, face, neuroimaging and physiological (Calvo & D'Mello, 2010). Notably, one of the main emotion classification topics uses variables associated with central nervous system (CNS) and autonomic nervous system (ANS) dynamics (Calvo & D'Mello, 2010). First, human emotional processing and perception involve cerebral cortex activity, which allows the automatic classification of emotions using the CNS. EEG is one of the techniques most used in this context (Valenza et al., 2016). Second, many emotion recognition studies have used the ANS to analyse changes in cardiovascular dynamics provoked by mood changes; HRV is one of the most used techniques (Valenza et al., 2012). The combination of physiological features with machine learning algorithms, such as in support vector machines, linear discriminant analysis, K-nearest neighbour and neural networks, has achieved high levels of accuracy in inferring subjects' emotional states (Zangeneh et al., 2018).

Finally, emotion elicitation is the ability to reliably and ethically elicit affective states. This elicitation is a critical factor in the development of systems that can detect, interpret and adapt to human affect (Kory & D'Mello, 2015). The many methods that elicit emotions in laboratories can be mainly divided into two groups, active and passive. Active methods involve directly influencing subjects, including behavioural manipulation (Ekman, 2007), social interaction (Harmon-Jones et al., 2007) and dyadic interaction (Roberts et al., 2007). Passive methods usually present external stimuli as images, sound or video. As to the use of images, the International Affective Picture System (IAPS) is among the databases most used as an elicitation tool in emotion recognition methodologies (Valenza et al., 2012). It includes over a thousand depictions of people, objects and events, standardized on the basis of valence and arousal (Kory & D'Mello, 2015). As to audio, the International Affective Digitalised Sound System (IADS) database is the most commonly applied in studies which use sound to elicit emotions (Nardelli et al., 2015). However, some studies directly use music or narrative to elicit emotion (Kim, 2007). With respect to audio-visual stimuli, many studies have used film to induce arousal and valence (Soleymani et al., 2015). As, to the best of our knowledge, the passive elicitation methods so far used in affective computing studies have not included immersive and interactive environments, they have significant limitations due to the importance of achieving high degrees of presence in the simulation of real-world experiences (Baños et al., 2004). Thus, by simulating real-word situations in laboratory environments, VR offers a new emotion elicitation approach for emotion recognition studies.

Affective computing and Virtual Reality

In recent years studies have applied implicit measures to analyse emotions using immersive VR. Table 1.2 provides a summary of studies correlating physiological signals with emotions using HMDs.

Year	Signals	Data analysis	Emotion	Participants	VR Stimuli	Stimuli comparison	Reference
2005	HRV, EDA	T-test	Arousal	67 healthy subjects	3D Training room vs pit room	No	(Meehan et al., 2005)
2010	HRV, EDA	ANOVA	Anxiety	10 healthy subjects and 20 subjects with food disorders	3D photo and real food catering	VR vs photo vs real	(Gorini et al., 2010)
2013	HRV, EDA	ANOVA	Arousal	50 healthy subjects	3D high-mobility wheeled vehicle	No	(Parsons et al., 2013)
2014	EDA	ANOVA	Fear	48 healthy and 48 spider-phobic subjects	3D Virtual lab with time-varying threat (spiders and snakes)	No	(Peperkorn et al., 2014)
2014	HRV	ANOVA	Anxiety	30 high anxiety and 35 low anxiety subjects	3D lecture hall	No	(Felnhofer et al., 2014)
2015	HRV, EDA	Cross-correlations	Arousal	306 healthy subjects	3D Room with time-varying threat (explosions, spiders, gunshots, etc.)	No	(McCall et al., 2015)
2015	EDA	ANOVA	Arousal	120 healthy subjects	3D Park with 5 variations (joy, sadness, boredom, anger and anxiety)	No	(Felnhofer et al., 2015)
2015	HRV, EDA	ANOVA	Anxiety	41 healthy and 42 spider-phobic subjects	3D virtual lab with spiders	No	(Notzon et al., 2015)
2016	HRV, EDA	Regression	Arousal	300 healthy subjects	3D Room with time-varying threats (explosions, spiders, gunshots, etc.)	No	(Hildebrandt et al., 2016)
2016	EDA	Kruskall-Wallis Test and correlations	Stress	12 healthy subjects	3D rooms (neutral, stress and calm)	No	(Higuera et al., 2016)
2016	HRV	Regression	Arousal	36 healthy subjects	3D Flight simulator	No	(Bian et al., 2016)
2016	HRV, EDA	ANOVA	Stress	45 healthy males	3D Trier Social Stress Test	No	(Shiban et al., 2016)
2017	HRV, EDA	ANOVA	Stress	18 healthy subjects	360^{o} Indoor vs natural panoramas	No	(Anderson et al., 2017)
2017	HRV	ANOVA	Arousal	108 healthy subjects	3D cemetery and park	No	(Chittaro et al., 2017)
2017	HRV, EDA	Mann–Whitney U tests and correlations	Pleasantness	100 healthy subjects	3D, 360^{o} and real retail store	3D VR vs 360^{o} VR vs real	(Higuera-Trujillo et al., 2017)
2017	HRV, EDA	ANOVA	Anxiety	100 healthy subjects	Mixed reality (3D VR with real world elements)	No	(Biedermann et al., 2017)
2018	HRV	ANOVA	Anxiety	30 healthy subjects	3D VR claustrophobic environments	AR vs VR	(Tsai et al., 2018)

2019	HRV	t-test, correlations and regressions	Arousal	30 healthy subjects	3D Exposure to a high height	No	(Kisker et al., 2019)
2019	HRV, EDA	ANOVA	Fear	49 height-fearful subjects	3D Forest	No	(Gromer et al., 2019)
2019	HRV	ANOVA	Stress	50 healthy subjects	3D Trier Social Stress Test	Replication of a real study	(Zimmer & Wu, 2019)
2019	EDA	Mann-Whitney U	Stress	60 healthy subjects	3D indoor building on fire	No	(Lin et al., 2019)

Table 1.2. Summary of studies including physiological signals, emotions and head-mounted displays

In terms of implicit measures, these studies used the ANS (HRV and EDA), probably because it is easier to combine this system with HMDs to measure peripheral physiological responses than it is to combine the CNS with HMDs (Gromer et al., 2019). To the best of our knowledge, no previous study has analysed emotions using the CNS with HMDs. Other implicit techniques, such as fMRI and FEA, have not hitherto been applied, probably because of the difficulties of combining them with HMDs.

Physiological signals have been mainly correlated with arousal, anxiety and stress (Anderson et al., 2017; Felnhofer et al., 2014; Kisker et al., 2019), and their main applications have been to phobia (Peperkorn et al., 2014) and risk perception research (McCall et al., 2015). Only one study has analysed the valence dimension itself, in particular the measurement of pleasantness (Higuera-Trujillo et al., 2017). However, while all these analysis included hypothesis testing and correlations, none inferred emotional statements using machine learning methods. Therefore, to the best of our knowledge, no affective computing research has hitherto used immersive virtual reality and applied machine learning algorithms to physiological responses.

The validity of Virtual Reality in human behaviour research

Finally, it is crucial to point out that the usefulness of simulation in human behaviour research has been analysed through the validity concept, that is, the capacity to evoke a response from the user in a simulated environment similar to one that might be evoked by a physical environment (Rohrmann & Bishop, 2002). To this extent, there is a need to perform direct comparisons between virtual and real environments. Some comparisons have studied the validity of virtual environments by assessing psychological responses (Bishop & Rohrmann, 2003) and cognitive performance (de Kort et al., 2003). However, there have been fewer analyses of physiological and behavioural responses (van der Ham et al., 2015; Yeom et al., 2017). Heydarian et al. (2015) analysed user performance in office-related activities, for example, reading texts and identifying objects; Chamilothori, Wienold and Andersen (2018) compared subjective perceptions of daylit spaces; and Kimura et al. (2017) analysed orienteering task performance. Table 1.2 shows the direct comparisons between the virtual and real environments undertaken using HMDs, in emotional response terms. Higuera-Trujillo, López-Tarruella and Llinares (2017) analysed psycho-physiological responses, through EDA, evoked by real-world scenarios and VR scenarios with different immersion levels; Gorini et al. (2010) analysed HRV and EDA responses (to food) of subjects with food disorders using VR, photos and in the real-world;

and Zimmer & Wu (2019) replicated a previous study correlating stress and EDA in VR. Thus, there has not hitherto been a thorough analysis of environmental simulations comparing real and virtual worlds in terms of emotional responses, in particular employing CNS and ANS dynamics and machine learning algorithms.

In conclusion, the increase in VR use by the scientific community in human behaviour research, and the need for further validation of VR, in particular related to emotion elicitation and recognition using physiological signals and machine learning algorithms, is the main motivation to perform the present research.

Objectives

The main objective of this thesis is to analyse the use of immersive VR as an emotion elicitation tool in human behaviour research, in particular its use in emotion recognition systems. To do so we develop: a direct comparison between a real-world environment and its virtualization; and an emotion recognition system for immersive virtual environments, using physiological sensors and machine learning. The specific objectives are:

- SO1. To develop a set of immersive VR environments using computer-generated 360º panoramas displayed through an HMD, and validate, through psychological self-assessment, that they are able to evoke a wide range of emotions.
- SO2. To develop a 3D VR realistic simulation of a real-world emotional environment, displayed through an HMD, and validate its capacity to evoke a high degree of presence and the same emotions as would be evoked in the real world, through a direct comparison using psychological self-assessment.
- SO3. To develop and validate a set of emotion recognition models using physiological signals and machine learning algorithms to automatically infer emotions in 360º panoramas displayed through an HMD, 3D VR simulation displayed through an HMD and a real-world environment.
- SO4. To analyse the capacity of a 3D VR simulation, displayed through an HMD, to evoke the same emotions and navigation pattern as would be evoked in the real world, through a direct comparison of physiological responses and the travel component of navigation.

Thesis structure

The thesis document is structured as follows:

Chapter 1 introduces and describes the motivation behind the thesis. In addition, it includes the objectives and thesis structure.

Chapter 2 presents the paper "Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors", published in Scientific Reports (Q1, 4.01 JCR 2018). It describes the development of a set of four VR environments (360º panoramas) designed to elicit four possible arousal-valence combinations. In addition, it presents an emotion recognition model, using EEG and ECG, developed using sixty HMD-wearing participants in VR environments.

Chapter 3 presents the paper "Real vs. immersive-virtual emotional experience: Exploiting psycho-physiological patterns in a free exploration of an art museum", published in Plos One (Q2, 2.77 JCR 2018). This presents a direct comparison made between a real and virtual museum (3D environment) through psychological measures, such as self-assessment, and physiological measures through an emotion recognition system, using EEG and ECG.

Chapter 4 presents the paper "Navigation comparison between a real and a virtual museum: time-dependent differences using a head mounted display", published in Interacting with Computers (Q4, 0.86 JCR 2018). This analysed the direct comparison between a real a virtual museum (3D environment) in terms of navigation behaviour.

Chapter 5 discusses the results and the major contributions of the thesis.

Chapter 6 provides an overall conclusion and future research directions.

Finally, the manuscript enumerates the publications and research stages derived from this thesis and provides a list of references.

Chapter 2

Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., Alcañiz, M., & Valenza, G. (2018). Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific reports*, 8(1), 13657.

Abstract

Affective Computing has emerged as an important field of study that aims to develop systems that can automatically recognize emotions. Up to the present, elicitation has been carried out with non-immersive stimuli. This study, on the other hand, aims to develop an emotion recognition system for affective states evoked through Immersive Virtual Environments. Four alternative virtual rooms were designed to elicit four possible arousal-valence combinations, as described in each quadrant of the Circumplex Model of Affects. An experiment involving the recording of the electroencephalography (EEG) and electrocardiography (ECG) of sixty participants was carried out. A set of features was extracted from these signals using various state-of-the-art metrics that quantify brain and cardiovascular linear and nonlinear dynamics, which were input into a Support Vector Machine classifier to predict the subject's arousal and valence perception. The model´s accuracy was 75.00% along the arousal dimension and 71.21% along the valence dimension. Our findings validate the use of Immersive Virtual Environments to elicit and automatically recognize different emotional states from neural and cardiac dynamics; this development could have novel applications in fields as diverse as Architecture, Health, Education and Videogames.

Introduction

Affective Computing (AfC) has emerged as an important field of study in the development of systems that can automatically recognize, model and express emotions. Proposed by Rosalind Picard in 1997, it is an interdisciplinary field based on psychology, computer science and biomedical engineering (Picard, 1997). Stimulated by the fact that emotions are involved in many background processes (Picard, 2003) (such as perception, decision-making, creativity, memory, and social interaction), several studies have focused on searching for a reliable methodology to identify the emotional state of a subject by using machine learning algorithms.

Thus, AfC has emerged as an important research topic. It has been applied often in education, healthcare, marketing and entertainment (Gross & Levenson, 1995; Harms et al., 2010; Jerritta et al., 2011; Koolagudi & Rao, 2012), but its potential is still under development. Architecture is a field where AfC has been infrequently applied, despite its obvious potential; the physical-environment has on a great impact, on a daily basis, on human emotional states in general (Lindal & Hartig, 2013), and on well-being in particular (Ulrich, 1984). AfC could contribute to improve building design to better satisfy human emotional demands (Fernández-Caballero et al., 2016).

Irrespective of its application, Affective Computing involves both emotional classification and emotional elicitation. Regarding emotional classification, two approaches have commonly been proposed: discrete and dimensional models. On the one hand, the former posits the existence of a small set of basic emotions, on the basis that complex emotions result from a combination of these basic emotions. For example, Ekman proposed six basic emotions: anger, disgust, fear, joy, sadness and surprise (Ekman, 1999). Dimensional models, on the other hand, consider a multidimensional space where each dimension represents a fundamental property common to all emotions. For example, the "Circumplex Model of Affects" (CMA) (Posner et al., 2005) uses a Cartesian system of axes, with two dimensions, proposed by Russell and Mehrabian (Russell & Mehrabian, 1977): valence, i.e., the degree to which an emotion is perceived as positive or negative; and arousal, i.e., how strongly the emotion is felt.

In order to classify emotions automatically, correlates from, e.g., voice, face, posture, text, neuroimaging, and physiological signals are widely used (Calvo & D'Mello, 2010). In particular, several computational methods are based on variables associated with central nervous system (CNS) and autonomic nervous system (ANS) dynamics (Calvo & D'Mello, 2010). On the one hand, the use of CNS is justified by the fact that human emotions originate in the cerebral cortex, involving several areas in their regulation and feeling. In this sense, the electroencephalogram (EEG) is one of the techniques most used to measure CNS responses (Valenza et al., 2016) , also through the use of wearable devices. On the other hand, a wider class of affective computing studies consider ANS changes elicited by specific emotional states. In this sense, experimental results over the last three decades show that Heart Rate Variability (HRV) analyses can provide unique and non-invasive assessments of autonomic functions on cardiovascular dynamics (Valenza et al., 2012; Valenza, Citi, et al., 2014). To this extent, there has been a great increase over the last decade in research and commercial interest in wearable systems for physiological monitoring. The key benefits of these systems are their small size, lightness, low-power consumption and, of course, their wearability (Valenza, Nardelli, et al., 2014). The state of the art (Kumari et al., 2017; Piwek et al., 2016; Xu et al., 2017) on wearable systems for physiological monitoring highlight that: i) surveys predict that the demand for wearable devices will increase in the near future; ii) there will be a need for more multimodal fusion of physiological signals in the near future; and iii) machine learning algorithms can be merged with traditional approaches. Moreover, recent studies present promising results on the development of emotion recognition systems through using wearable sensors instead of classic lab sensors, through HRV (He et al., 2017) and EEG (Nakisa et al., 2018).

Regarding emotional elicitation, the ability to reliably and ethically elicit affective states in the laboratory is a critical challenge in the process of the development of systems that can detect, interpret and adapt to human affect (Kory & D'Mello, 2014). Many methods of eliciting emotions have been developed to evoke emotional responses. Based on the nature of the stimuli, two types of method are distinguished, the active and the passive. Active methods can involve behavioural manipulation (Ekman, 2007), social psychological methods with social interaction (Harmon-Jones et al., 2007) and dyadic interaction (Roberts et al., 2007). On the other hand, passive methods usually present images, sounds or films. With respect to images, one of the most prominent databases is the International Affective Picture System (IAPS), which includes over a thousand depictions of people, objects and events, standardized on the basis of valence and arousal (Kory & D'Mello, 2014). The IAPS has been used in many studies as an elicitation tool in emotion recognition methodologies (Valenza et al., 2012). With respect to sound, the most used database is the International Affective Digitalised Sound System (IADS) (Nardelli et al., 2015). Some researchers also use music or narrative to elicit emotions (Kim, 2007). Finally, audio-visual stimuli, such as films, are also used to induce different levels of valence and arousal (Soleymani et al., 2015).

Even when, as far we know, elicitation has been carried out with a non-immersive stimulus, it has been shown that these passive methods have significant limitations due to the importance of immersion for eliciting emotions through the simulation of real experiences (Baños et al., 2004). In the present, Virtual Reality (VR) represents a novel and powerful tool for behavioural research in psychological assessment. It provides simulated experiences that create the sensation of being in the real world (Giglioli et al., 2017; Marín-Morales et al., 2017). Thus, VR makes it possible to simulate and evaluate spatial environments under controlled laboratory conditions (Marín-Morales et al., 2017; Vince, 2004), allowing the isolation and modification of variables in a cost and time effective manner, something which is unfeasible in real space (Alcañiz et al., 2003). During the last two decades VR has usually been displayed using desktop PCs or semi-immersive systems such as CAVEs or Powerwalls (Vecchiato et al., 2015). Today, the use of head-mounted displays (HMD) is increasing: these provide fully-immersive systems that isolate the user from external world stimuli. These provide a high degree of immersion, evoking a greater sense of presence, understood as the perceptual illusion of non-mediation and a sense of "being-there" (Slater & Wilbur, 1997). Moreover, the ability of VR to induce emotions has been analysed in studies which demonstrate that virtual environments do evoke emotions in the user (Alcañiz et al., 2003). Other works confirm that Immersive Virtual Environments (IVE) can be used as emotional induction tools to create states of relaxation or anxiety (Riva et al., 2007), basic emotions, (Baños et al., 2006; Baños et al., 2012) and to study the influence of the users cultural and technological background on emotional responses in VR (Gorini et al., 2009). In addition, some works show that emotional content increases sense of presence in an IVE (Gorini et al., 2011) and that, faced with the same content, self-reported intensity of emotion is significantly greater in immersive than in non-immersive environments (Chirico et al., 2017). Thus, IVEs, showing 360º panoramas or 3D scenarios through a HMD (Blascovich et al., 2012), are powerful tools for psychological research, (Blascovich et al., 2012).

Taking advantage of the IVE's potentialities, in recent years some studies have used IVE and physiological responses, such as EEG, HRV and EDA, in different fields. Phobias (Hildebrandt et al., 2016; McCall et al., 2015; Notzon et al., 2015; Peperkorn et al., 2014), disorders (Amaral et al., 2017), driving and orientation (Eudave & Valencia, 2017; Sharma et al., 2017), videogames (Bian et al., 2016), quality of experience (Egan et al., 2016), presence (Meehan et al., 2005) and visualization technologies (Higuera-Trujillo et al., 2017), are some examples of these applications. Particularly in emotion research, arousal and relaxation have been analysed in outdoor (Anderson et al., 2017; Felnhofer et al., 2015) and indoor (Higuera et al., 2016) IVEs using EDA. Therefore, the state of the art presents the following limitations: (1) few studies analyse physiological responses in IVEs and, in particular, using an affective approach; (2) there are few validated emotional IVE sets which include stimuli with different levels of arousal and valence: and, (3) there is no affective computing research that tries to automatically recognize the user's mood in an IVE through physiological signals and machine learning algorithms.

In this study, we propose a new AfC methodology capable of recognizing the emotional state of a subject in an IVE in terms of valence and arousal. Regarding stimuli, IVEs were designed to evoke different emotional states from an architectural point of view, by changing physical features such as illumination, colour and geometry. They were presented through a portable HMD. Regarding emotion recognition, a binary classifier will be presented, which uses effective features extracted from EEG and HRV data gathered from wearable sensors, and combined through nonlinear Support Vector Machine (SVM) (Valenza et al., 2012) algorithms.

Material and methods

Experimental context

This work is part of a larger research project that attempts to characterize the use of VR as an affective elicitation method and, consequently, develop emotion recognition systems that can be applied to 3D or real environments.

An experimental protocol was designed to acquire the physiological responses of subjects in 4 different stimuli presentation cases: 2D desktop pictures, a 360º panorama IVE, a 3D scenario IVE and a physical environment. The experiment was conducted in two distinct phases that presented some differences. Both phases were divided into 3 stages; the results of the experiment are at Figure 2.1. Between each stage, signal acquisition was temporarily halted and the subjects rested for 3 minutes on a chair. Stage 1 consisted of emotion elicitation through a desktop PC displaying 110 IAPS pictures, using a methodology detailed in previous research (Valenza et al., 2012). Stage 2 consisted of emotion elicitation using an HMD based on a new IVE set with four 360º panoramas. Finally, stage 3 consisted of the free exploration of a museum exhibition.

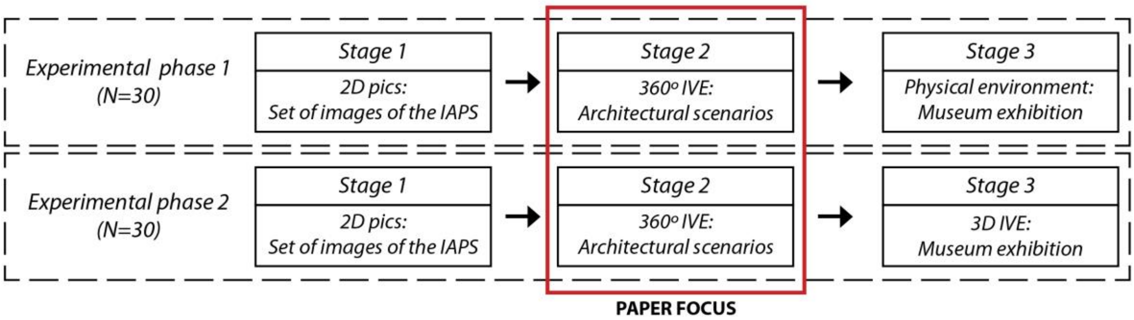

Figure 2.1. Experimental phases of the research

In the present paper we focus on an analysis of stage 2. The experimental protocol was approved by the ethics committee of the Polytechnic University of Valencia and informed consent was obtained from all participants. All methods and experimental protocols were performed in accordance with the guidelines and regulations of the local ethics committee of the Polytechnic University of Valencia.

Participants

A group of 60 healthy volunteers, suffering neither from cardiovascular nor evident mental pathologies, was recruited to participate in the experiment. They were balanced in terms of age (28.9 ± 5.44) and gender (40% male, 60% female). Inclusion criteria were as follows: age between 20 and 40 years; Spanish nationality; having no formal education in art or fine art; having no previous experience of virtual reality; and not having previously visited the particular art exhibition. They were divided into 30 subjects for the first phase and 30 for the second.

To ensure that the subjects constituted a homogeneous group, and that they were in a healthy mental state, they were screened by i) the Patient Health Questionnaire (PHQ-9) (Kroenke et al., 2001) and ii) the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994).

PHQ-9 is a standard psychometric test used to quantify levels of depression (Kroenke et al., 2001). Significant levels of depression would have affected the emotional responses. Only participants with a score lower than 5 were included in the study. The test was presented in the Spanish language as the subjects were native Spanish speakers. SAM tests were used to detect if any subject had an emotional response that could be considered as an outlier, with respect to a standard elicitation, in terms of valence and arousal. A set of 8 IAPS pictures (Lang et al., 1997) (see Table 2.1), representative of different degrees of arousal and valence perception, was scored by each subject after stage 1 of the experiment. The z-score of each subject's arousal and valence score was calculated using the mean and deviation of the IAPS's published scores (Lang et al., 1997). Subjects that had one or more z-scores outside of the range -2.58 and 2.58 (α=0.005) were excluded from further analyses. Therefore, we retained subjects whose emotional responses, caused by positive and negative pictures, in different degrees of arousal, belonged to 99% of the IAPS population. In addition, we rejected subjects if their signals presented issues, e.g., disconnection of the sensors during the elicitation or if artefacts affected the signals. Taking these exclusions into account, the number of valid subjects was 38 (age: 28.42 ± 4.99; gender: 39% male, 61% female).

IAPS picture	AROUSAL	VALENCE
7234	3.41 ± 2.29	4.01 ± 1.32
5201	3.20 ± 2.50	7.76 ± 1.44
9290	4.75 ± 2.20	2.71 ± 1.35
1463	4.61 ± 2.56	8.17 ± 1.48

9181	6.20 ± 2.23	1.84 ± 1.25
8380	5.84 ± 2.34	7.88 ± 1.37
3102	6.92 ± 2.50	1.29 ± 0.79
4652	7.24 ± 2.09	7.68 ± 1.64

Table 2.1. Arousal and valence score of selected IAPS pictures

Set of Physiological Signals and Instrumentation

The physiological signals were acquired using the B-Alert x10 (Advanced Brain Monitoring, Inc., USA) (Figure 2.2). It provides an integrated approach for wireless wearable acquisition and recording of electroencephalographic (EEG) and electrocardiographic (ECG) signals, sampled at 256 Hz. EEG sensors were located in the frontal (Fz, F3 and F4), central (Cz, C3 and C4) and parietal (POz, P3, and P4) regions with electrode placements on the subjects' scalps based on the international 10-20 electrode placement. A pair of electrodes placed below the mastoid was used as reference, and a test was performed to check the conductivity of the electrodes, aiming to keep the electrode impedance below 20kΩ. The left ECG lead was located on the lowest rib and the right lead on the right collarbone.

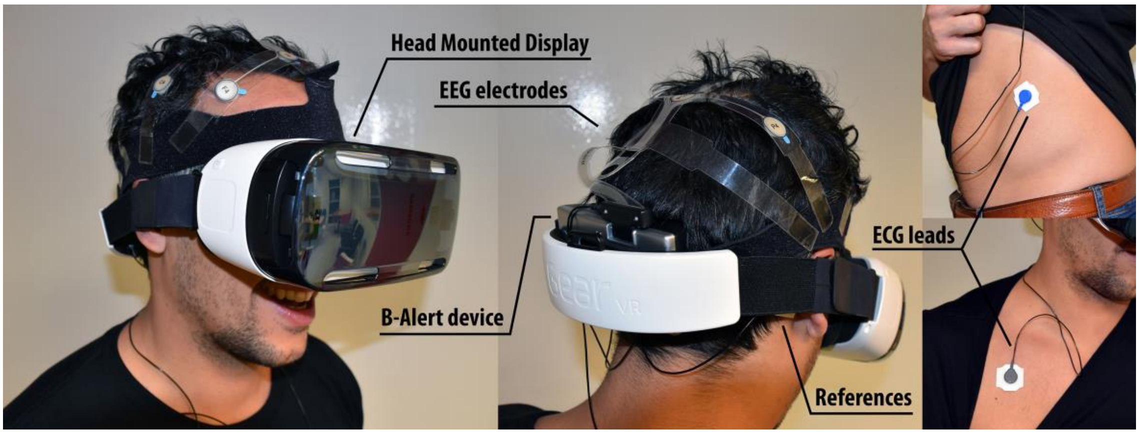

Figure 2.2. Exemplary experimental set-up

Stimulus elicitation

We developed an affective elicitation system by using architectural environments displayed by 360º panoramas implemented in a portable HMD (Samsung Gear VR). This combination of environments and display-format was selected due to its capacity for evoking affective states. The bidirectional influence between the architectural environment and the user's affective-behavioural response is widely accepted: even subtle variations in the space may generate different neurophysiological responses (Nanda et al., 2013). Furthermore, the 360º panorama-format provided by HMD devices is a valid set-up to evoke psychological and physiological responses similar to those evoked by physical environments (Higuera-Trujillo et al., 2017). Thus, following the combination of the arousal and valence dimensions, which gives the four possibilities described in the CMA (Russell, 1980), four architectural environments were proposed as representative of four emotional states.

The four architectural environments were designed based on Kazuyo Sejima´s "Villa in the forest" scenario (Sejima, 1996). This architectural work was considered by the research team as an appropriate base from which to make the modifications designed to generate the different affective states.

The four base-scenario configurations were based on different modifications of the parameters of three design variables: illumination, colour, and geometry. Regarding illumination, the parameters "colour temperature", "intensity", and "position" were modified. The modification of the "colour temperature" was based on the fact that higher temperature may increase arousal, being registrable at the neurophysiological level (Noguchi & Sakaguchi, 1999; Ochiai et al., 2015). "Intensity" was also modified in the same way to try to increase or reduce arousal. The "position" of the light was direct, in order to try to increase arousal, and indirect to reduce it. The modifications of these last two parameters were based on the design experience of the research team. Regarding colour, the parameters "tone", "value", and "saturation" were modified. The modification of these parameters was performed jointly on the basis that warm colours increase arousal and cold ones reduce it, being registrable at the psychological (Küller et al., 2009) and neurophysiological levels (Hogg et al., 1979; Jacobs & Hustmyer, 1974; Jalil et al., 2012; Jin et al., 2009; Yildirim et al., 2011). Regarding geometry, the parameters "curvature", "complexity", and "order" were modified. "Curvature" was modified on the basis that curved spaces generate a more positive valence than angular, being registrable at psychological and neurophysiological levels (Vartanian et al., 2013). The modification of the parameters "complexity" and "order" was performed jointly. This was based on three conditions registrable at the neurophysiological level: (1) high levels of geometric "complexity" may increase arousal and low levels may reduce arousal (Tsunetsugu et al., 2005); (2) high levels of "complexity" may generate a positive valence if they are submitted to "order", and negative valence if presented disorderly (Stamps, 1999); and (3) content levels of arousal generated by geometry may generate a more positive valence (Berlyne, 1970). The four architectural environments were designed on this basis. Table 2.2 shows the configuration guidelines chosen to elicit the four affective states.

		High-Arousal & Negative-Valence (Room 1)	High-Arousal & Positive-Valence (Room 2)	Low-Arousal & Negative-Valence (Room 3)	Low-Arousal & Positive-Valence (Room 4)
Illumination	Colour temperature	7500K	7500K	3500K	3500K
	Intensity	High	High	Low	Low
	Position	Mainly Direct	Mainly Direct	Mainly Indirect	Mainly Indirect
Colour	Tone	Warm colours	Warm colours	Cold colours	Cold colours
	Value				
	Saturation				
Geometry	Curvature	Rectilinear	Curved	Rectilinear	Curved
	Complexity	High	Low-Medium	Medium-High	Low
	Order	Low	High	Low-Medium	High

Table 2.2. Configuration guidelines chosen in each architectural environment configuration

In a technical sense, the four architectural environments were developed in similar ways. The process consisted of modelling and rendering. Modelling was performed by using Rhinoceros v5.0 (www.rhino3d.com). The 3D-models used for the four architectural environments were 3446946, 3490491, 3487660, and 3487687 polygons. On completion of this process, they were exported in .dwg format for later rendering. The rendering was performed using the VRay engine v3.00.08 (www.vray.com), operating with Autodesk 3ds Max v2015 (www.autodesk.es). 15 textures were used for each of the four architectural environments. Configured as 360º panoramas, renders were exported in .jpg format with resolutions of 6000x3000 pixels at 300 dots per inch. These were implemented in the Samsung Gear VR HMD device. This device has a stereoscopic screen of 1280×1440 pixels per eye and a 96° field of view, supported by a Samsung Note 4 mobile telephone with a 2.7GHz quad-core processor and 3GB of RAM. The reproduction of the architectural 360º panoramas was fluid and uninterrupted.

Prior to the execution of the experimental protocol, a pre-test was performed in order to ensure that the architectural 360° panoramas would elicit the affective states for which they had been designed. It was a three-phased test: individual questionnaires, a focus-group session conducted with some respondents to the questionnaire and individual validation-questionnaires. The questionnaires asked the participants to evaluate the architectural 360° panoramas. A SAM questionnaire, embedded in the 360º panorama, was used, with evaluations ranging from -4 (totally disagree) to 4 (totally agree) for all the emotion dimensions. 15 participants (8 men and 7 women) completed the questionnaires. First, the participants freely viewed each architectural environment, then the SAM questionnaires were presented and the answers given orally. Figure 2.3 shows an example of one of these questionnaires. After the questionnaire sessions had been completed, a focus group session, which was a carefully managed group discussion, was conducted (Krueger & Casey, 2000). Five of the participants (3 men and 2 women) with the most unfavourable evaluations in phase 1 were selected as participants and one of the members of the research team, with previous focus-group experience, moderated. The majority of the changes were performed to Room 3, due to the discordances between the self-assessment and their theoretical quadrant. Once the changes were implemented, a similar evaluation to phase 1 was performed. Table 2.3 shows the arousal and valence ratings of the four architectural 360° panoramas of this pre-test phase. After these phases, no new variations were considered necessary. This procedure allowed us to assume some initial reliability in the design of the architectural environments. Figure 2.4 shows these final configurations. High quality images of the stimuli are included in the supplementary material.

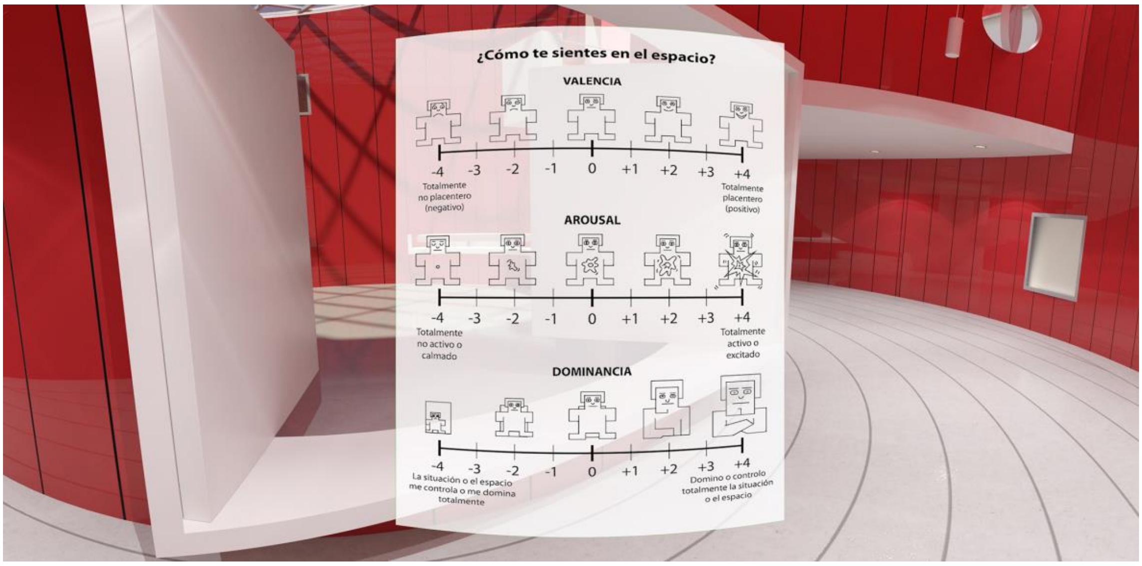

Figure 2.3. Example of SAM questionnaire embedded in the room 1. Simulation developed using Rhinoceros v5.0, VRay engine v3.00.08 and Autodesk 3ds Max v2015

	Arousal	Valence
High-Arousal & Negative-Valence (Room 1)	2.23 ± 1.59	-2.08 ± 1.71
High-Arousal & Positive-Valence (Room 2)	1.25 ± 1.33	1.31 ± 1.38
Low-Arousal & Negative-Valence (Room 3)	-0.69 ± 1.65	-1.46 ± 1.33
Low-Arousal & Positive-Valence (Room 4)	-2.31 ± 1.30	1.92 ± 1.50

Table 2.3. Arousal and Valence resulted in the pre-test with 15 participants. The scores are averaged using mean and standard deviation for a Likert scale between -4 to +4

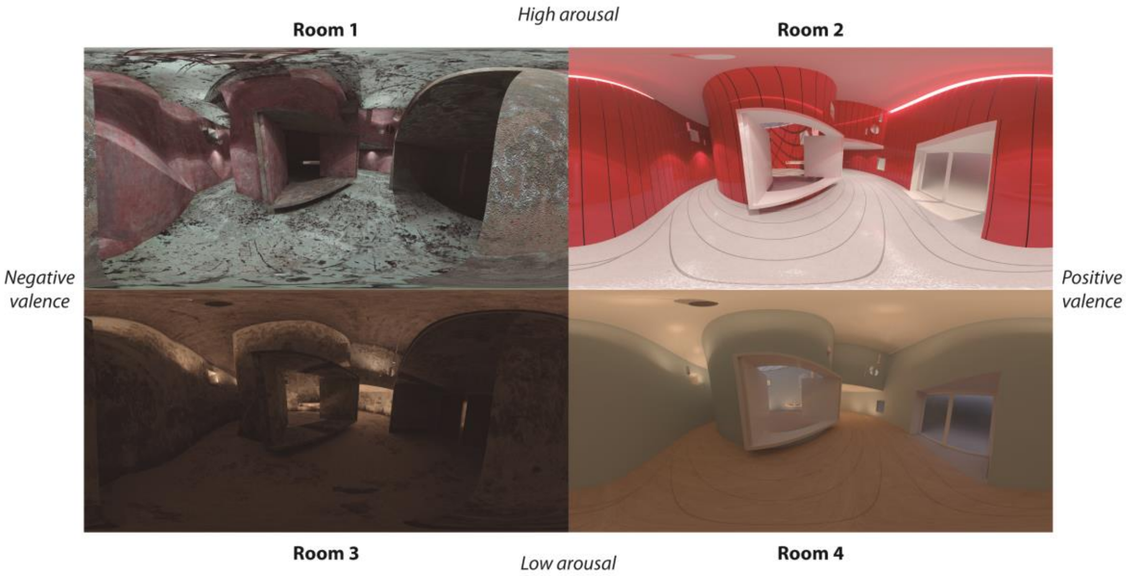

Figure 2.4. 360º panoramas of the four IVEs. Simulations developed using Rhinoceros v5.0, VRay engine v3.00.08 and Autodesk 3ds Max v2015

None of the pre-test participants was included in the main study. Regarding the experimental protocol, each room was presented for 1.5 minutes and the sequence of presentation was counter-balanced using the Latin Square method. After viewing the rooms, the users were asking to orally evaluate the emotional impact of each room using a SAM questionnaire embedded in the 360º photo.

Signal processing

Heart rate variability

The ECG signals were processed to derive HRV series (Acharya et al., 2006). The artefacts were cleaned by the threshold base artefacts correction algorithm included in the Kubios software (Tarvainen et al., 2014). In order to extract the RR series, the well-known algorithm developed by Pan-Tompkins was used to detect the R-peaks (Pan & Tompkins, 1985). The individual trends components were removed using the smoothness prior detrending method (Tarvainen et al., 2002).

We carried out the analysis of the standard HRV parameters, which are defined in the time and frequency domains, as well as HRV measures quantifying heartbeat nonlinear and complex dynamics (Acharya et al., 2006). All features are listed in Table 2.4.

Time domain	Frequency domain	Other
Mean RR	VLF peak	Pointcaré SD1
Std RR	LF peak	Pointcaré SD2

RMSSD	HF peak	Approximate Entropy (ApEn)
pNN50	VLF power	Sample Entropy (SampEn)
RR triangular index	VLF power %	DFA α1
TINN	LF power	DFA α2
	LF power %	Correlation dimension (D2)
	LF power n.u.	
	HF power	
	HF power %	
	HF power n.u.	
	LF/HF power	
	Total power	

Table 2.4. List of used HRV features

Time domain features include average (Mean RR) and standard deviation (Std RR) of the RR intervals, the root mean square of successive differences of intervals (RMSSD), and the ratio between the number of successive RR pairs having a difference of less than 50 ms and the total number of heartbeat analyses (pNN50). The triangular index was calculated as a triangular interpolation of the HRV histogram. Finally, TINN is the baseline width of the RR histogram, evaluated through triangular interpolation.

In order to obtain the frequency domain features, a power spectrum density (PSD) estimate was calculated for the RR interval series by a Fast Fourier Transform based on Welch's periodogram method. The analysis was carried out in three bands: very low frequency (VLF, <0.04 Hz), low frequency (LF, 0.04-0.15 Hz) and high frequency (HF, 0.12-0.4 Hz). For each frequency band, the peak value was calculated, corresponding to the frequency with the maximum magnitude. The power of each frequency band was calculated in absolute and percentage terms. Moreover, for the LF and HF bands, the normalized power (n.u.) was calculated as the percentage of the signals subtracting the VLF to the total power. The LF/HF ratio was calculated in order to quantify sympatho-vagal balance and to reflect sympathetic modulations (Acharya et al., 2006). In addition, the total power was calculated.

Regarding the HRV nonlinear analysis, many measures were extracted, as they are important quantifiers of cardiovascular control dynamics mediated by the ANS in affective computing (Acharya et al., 2006; Valenza et al., 2012, 2016; Valenza, Citi, et al., 2014). Pointcaré plot analysis is a quantitative-visual technique, whereby the shape of a plot is categorized into functional classes. The plot provides summary information as well as detailed beat-to-beat information on heart behaviour. SD1 is related to the fast beat-to-beat variability in the data, whereas SD2 describes the longer-term variability of R–R (Acharya et al., 2006).

Approximate Entropy (ApEn) and Sample Entropy (SampEn) are two entropy measures of HRV. ApEn detects the changes in underlying episodic behaviour not reflected in peak occurrences or amplitudes (Pincus & Viscarello, 1992), whereas SampEn statistics provide an improved evaluation of time-series regularity and provide a useful tool in studies of the dynamics of human cardiovascular physiology (Richman & Moorman, 2000). DFA correlations are divided into short-term and long-term fluctuations through the α1 and α2 features. Whereas α1 represents the fluctuation in the range of 4-16 samples, α2 refers to the range of 16-64 samples (Peng et al., 1995). Finally, the correlation dimension is another method for measuring the complexity or strangeness of the time series; it is explained by the D2 feature. It is expected to give information on the minimum number of dynamic variables needed to model the underlying system (Grassberger & Procaccia, 1983).

Electroencephalographic signals

In order to process the EEG signals, the open source toolbox EEGLAB (Delorme & Makeig, 2004) was used. The complete processing scheme is shown at Figure 2.5.

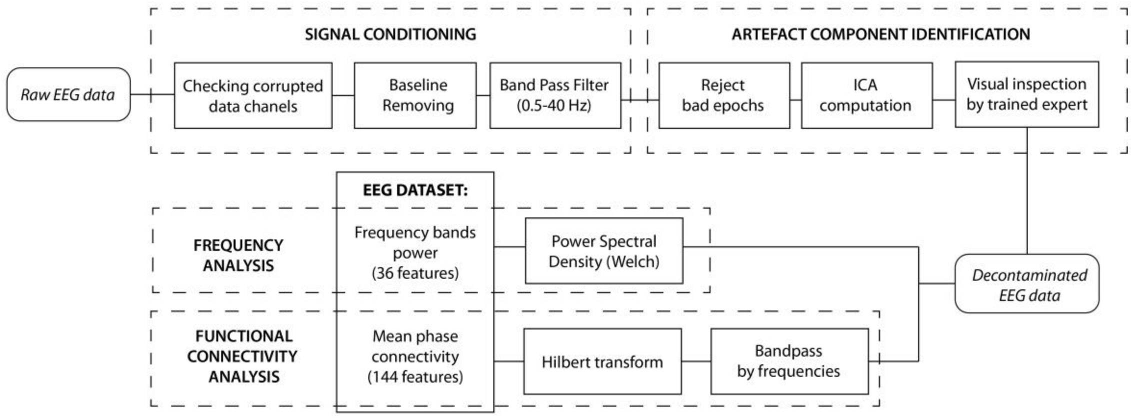

Figure 2.5. Block scheme of the EEG signal processing steps

Firstly, data from each electrode were analysed in order to identify corrupted channels. These were identified by computing the fourth standardized moment (kurtosis) along the signal of each electrode (Colomer et al., 2016). In addition, if the signal was flatter than 10% of the total duration of the experiment, the channel was classified as corrupted. If one of the nine channels was considered as corrupted, it could be interpolated from neighbouring electrodes. If more than one channel was corrupted, the subject would be rejected. Only one channel among all of the subjects was interpolated.

The baseline of EEG traces was removed by mean subtraction and a band pass filter between 0.5 and 40 Hz was applied. The signal was divided into epochs of one second and the intra-channel kurtosis level of each epoch was computed in order to reject the epochs highly damaged by noise (Colomer et al., 2016). In addition, automatic artefact detection was applied, which rejects the epoch when more than 2 channels have samples exceeding an absolute threshold of >100.00 µV and a gradient of 70.00 µV between samples (Kober et al., 2012).

The Independent Component Analysis (ICA) (Hyvärinen & Oja, 2000) was then carried out using infomax algorithm to detect and remove components due to eye movements, blinks and muscular artefacts. Nine source signals were obtained (one per electrode). A trained expert manually analysed all the components, rejecting those related to artefacts. The subjects who had more than 33% of their signals affected by artefacts were rejected.

After the pre-processing, spectral and functional connectivity analyses were performed.

EEG spectral analysis, using Welch's method (Welch, 1967), was performed to estimate the power spectra in each epoch, with 50% overlapping, within the classical frequency bandwidth θ (4-8 Hz), α (8-12 Hz), β (13-25 Hz), γ (25-40 Hz). Frequency band δ (less than 4Hz) was not taken into account in this study because it relates to deeper stages of sleep. In total, 36 features were obtained from the nine channels and 4 bands.

A functional connectivity analysis was performed using Mean Phase Coherence (Mormann et al., 2000), for each pair of channels:

$$R^2 = E[\cos(\Delta\phi)]^2 + E[sin(\Delta\phi)]^2 \qquad (1)$$

Where R is the MPC, $\Delta\phi$ represents the relative phase diference between two channels derived from the instantaneous difference of the analytics signals from the Hilbert transform, and E is the expectation operator. By definition, MPC values ranged between 0 and 1. In the case of strong phase synchronization between two channels, the MPC is close to 1. If the two channels are not synchronized, the MPC remains low. 36 features were derived from each possible combination of a pair of 9 channels in one specific band. In total, 144 features were created using the 4 bands analysed.

Feature reduction and machine learning

Each room was presented for 1.5 minutes and was considered as an independent stimulus. In order to characterize each room, all HRV features were calculated using this time window. In the case of EEG, in both the frequency band power and mean phase connectivity analyses, we considered the mean of all the epochs of each stimulus as the representative value of the stimulus time window. Altogether, 209 features described each stimulus for each subject. Due to the high-dimensional feature space obtained, a feature reduction strategy was adopted for decreasing this dimension. We implemented the well-known

Principal Component Analysis method (PCA) (Jolliffe, 2002). This mathematical method is based on the linear transformation of the different variables in the principal components, which can be assembled in clusters. We select the features that explain 95% of the variability of the dataset. The PCA was applied three times: (1) in the HRV set, reducing the features from 29 to 3; (2) in the frequency band power analysis of the EEG, reducing the features from 36 to 4; and (3) in the mean phase coherency analysis of the EEG, reducing the features from 144 to 12. Hence, the feature reduction strategy reduces our features to a total of 19.

The machine learning strategy could be summarized as follows:

(1) To divide the dataset into training and test sets

(2) The development of the model (parameter tuning and feature selection) using cross-validation in the training set

(3) To validate the model using the test set

Firstly, the dataset was sliced randomly into 15% for the test set (5 subjects) and 85% for the training set (33 subjects). In order to calibrate the model, the Leave-One-Subject-Out (LOSO) cross-validation procedure was applied to the training set using Support Vector Machine (SVM)-based pattern recognition (Schöllkopf et al., 2000). Within the LOSO scheme, the training set was normalized by subtracting the median value and dividing this by the median absolute deviation over each dimension. In each of the 36 iterations, the validation set consisted of one specific subject and he/she was normalized using the median and deviation of the training set.

Regarding the algorithm, we used a C-SVM optimized using a sigmoid kernel function, changing the parameters of cost and gamma using a vector with 15 parameters logarithmically spaced between 0.1 and 1000. Additionally, in order to explore the relative importance of all the features in the classification problem we used a support vector machine recursive feature elimination (SVM-RFE) procedure in a wrapper approach (RFE was performed on the training set of each fold and we computed the median rank for each feature over all folds). We specifically chose a recently developed, nonlinear SVM-RFE, which includes a correlation bias reduction strategy in the feature elimination procedure (Yan & Zhang, 2015). After the cross-validation, using the parameters and feature set obtained, the model was applied to the test set that had not previously been used. The self-assessment of each subject was used as the output of the arousal and valence model. The evaluation was bipolarized in positive/high (>0) and negative/low (<=0). All the algorithms were implemented by using Matlab© R2016a, endowed with an additional toolbox for pattern recognition, i.e., LIBSVM (Chang & Lin, 2011). A general overview of the analysis is shown in Figure 2.6.

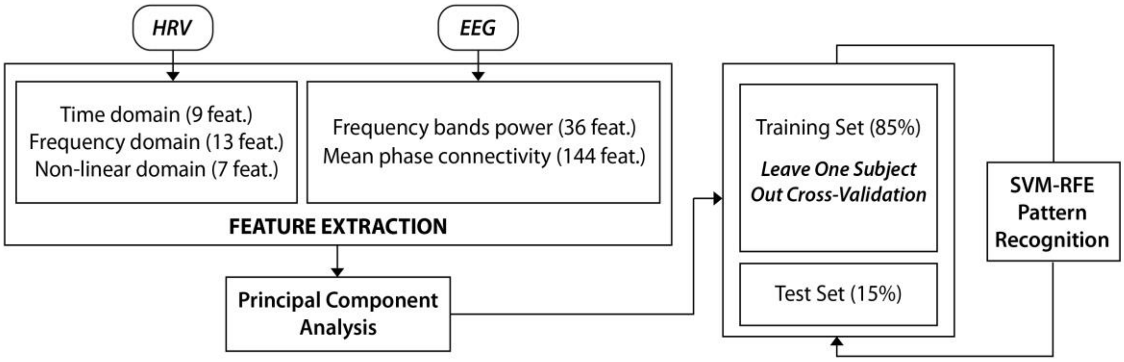

Figure 2.6. Overview of the feature reduction and classification chain

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding authors on reasonable request.

Results

Subjects' self-assessment

Figure 2.7 shows the self-assessment of the subjects for each IVE averaged using mean and standard deviation in terms of arousal (Room 1: 1.17 ± 1.81, Room 2: 2.10 ± 1.59, Room 3: 0.05 ± 2.01, Room 4: -0.60 ± 2.11) and valence (Room 1: -1.12 ± 1.95, Room 2: 1.45 ± 1.93, Room 3: -0.40 ± 2.14, Room 4: 2.57 ± 1.42). The representation follows the CMA space. All rooms are located in the theoretical emotion quadrant for which they were designed, except for Room 3 that evokes more arousal than hypothesized. Due to the non-Gaussianity of data ($p < 0.05$ from the Shapiro-Wilk test with null hypothesis of having a Gaussian sample), Wilcoxon signed-rank tests were applied.
Table 2.5 presents the result of multiple comparisons using Tukey's Honestly Significant Difference Procedure. Significant differences were found in the valence dimension between the negative-valence rooms (1 and 3) and the positive-valence rooms (2 and 4). Significant differences were found in the arousal dimension between the high-arousal rooms (1 and 2) and the low-arousal rooms (3 and 4), but not for pairs 1 and 3. Therefore, the IVEs statistically achieve all the desired self-assessments except for arousal perception in Room 3, which is higher than we hypothesized. After the bipolarization of scores (positive/high >0), they are balanced (61.36% high arousal and 56.06% positive valence).

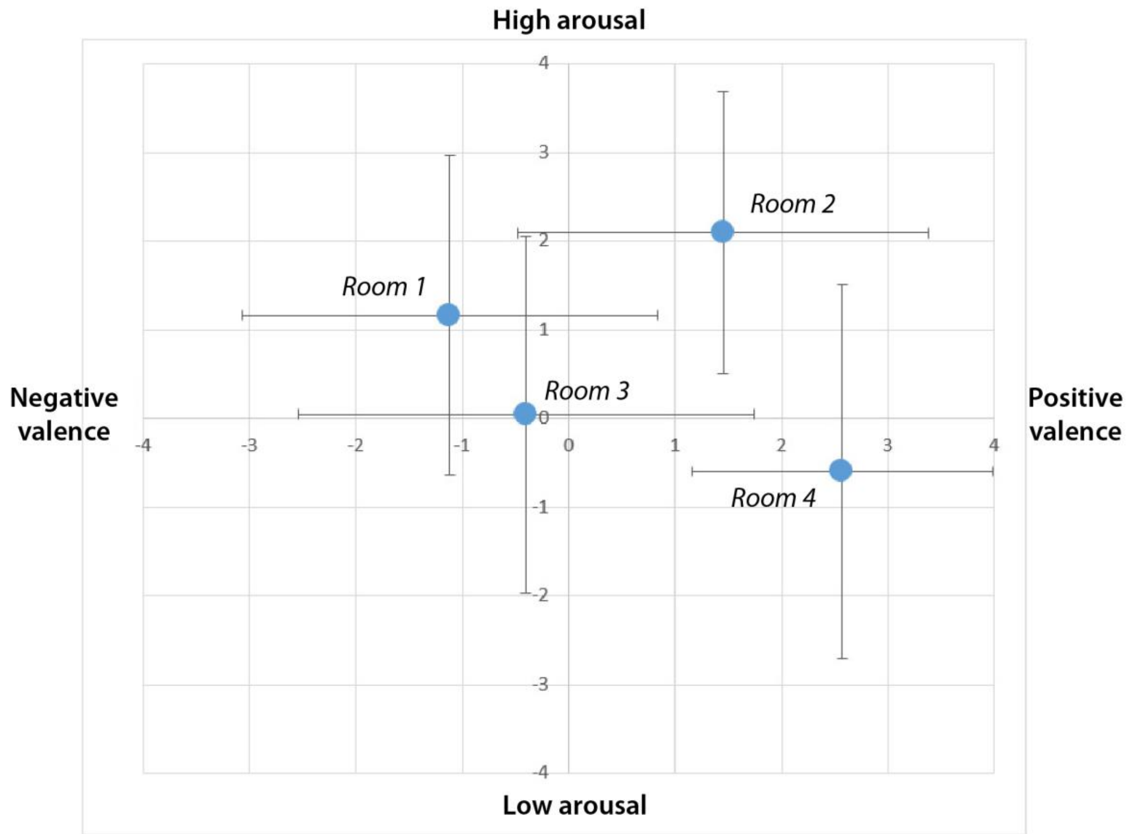

Figure 2.7. Self-assessment score in the IVEs using SAM and a Likert scale between -4 and +4. Blue dots represent the mean whereas horizontal and vertical lines represent standard deviation

		p-value	
IVE		Arousal	Valence
1	2	0.052	10-6 (***)
1	3	0.195	0.152
1	4	0.007 (**)	10-9 (***)
2	3	10-5 (***)	0.015 (*)
2	4	10-8 (***)	0.068
3	4	0.606	10-7 (***)

Table 2.5. Signification test of the self-assessment of the emotional rooms

Arousal classification

Table 2.6 shows the confusion matrix of cross validation and the total average accuracy (75.00%), distinguishing two levels of arousal using the first 15 features selected by the nonlinear SVM-RFE algorithm. The F-Score of arousal classification is 0.75. The changes in accuracy depending on number of features are shown in Figure 2.8, and

Table 2.7 presents the list of features used. Table 2.8 shows the confusion matrix of the test set and the total average accuracy (70.00%) using the parameters and the feature set defined in the cross-validation phase. The F-score of arousal classification is 0.72 in the test set.

Arousal	High	Low
High	82.72	17.28
Low	37.25	62.75

Table 2.6. Confusion matrix of cross-validation using SVM classifier for arousal level. Values are expressed as percentages. Total Accuracy: 75.00%

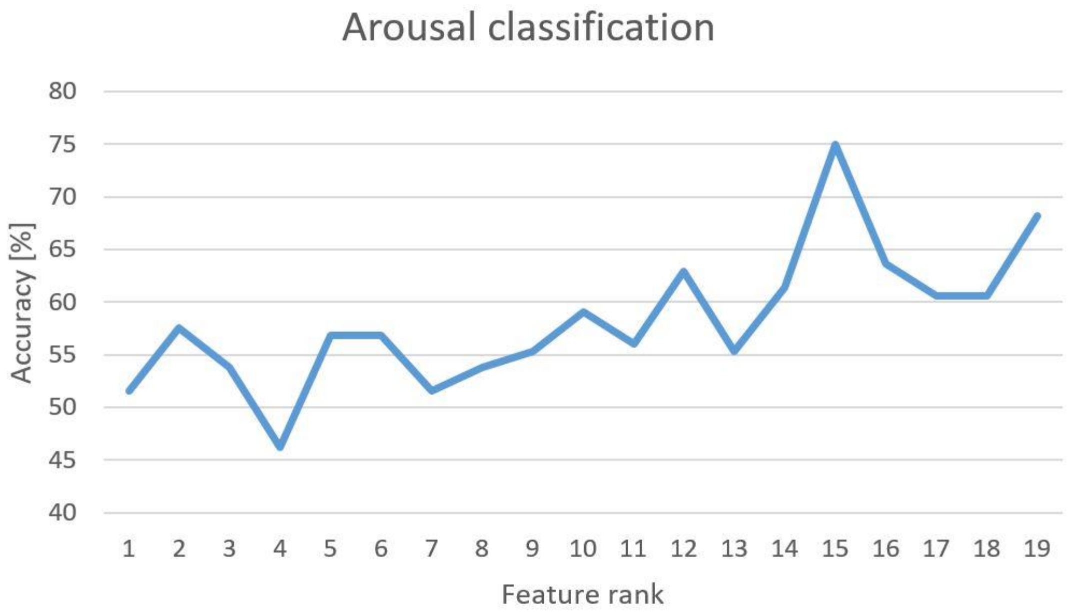

Figure 2.8. Recognition accuracy of arousal in cross-validation as a function of the feature rank estimated through the SVM-RFE procedure

Rank	Feature
1	EEG MPC PCA 8
2	EEG MPC PCA 9
3	EEG MPC PCA 11
4	EEG MPC PCA 10
5	EEG MPC PCA 7
6	EEG MPC PCA 12
7	EEG Band Power PCA 3
8	EEG Band Power PCA 1
9	HRV PCA 1

10	EEG Band Power PCA 4
11	EEG Band Power PCA 2
12	HRV PCA 3
13	EEG MPC PCA 4
14	HRV PCA 2
15	EEG MPC PCA 5

Table 2.7. Selected features ordered by their median rank over every fold computed during the LOSO procedure for arousal classification

Arousal	**High**	**Low**
High	75.00	25.00
Low	33.33	66.67

Table 2.8. Confusion matrix of test set using SVM classifier for arousal level. Values are expressed as percentages. Total Accuracy: 70.00%

Valence classification

Table 2.9 shows the confusion matrix of the cross validation and total average accuracy (71.21%), distinguishing two levels of valence using the first 10 features selected by the nonlinear SVM-RFE algorithm. The F-Score of the valence classification is 0.71. The changes in accuracy depending on the number of features are shown in Figure 2.9, and Table 2.10 presents the list of features used. Table 2.11 shows the confusion matrix of the test set and total average accuracy (70.00%), using the parameters and the feature set defined in the cross-validation phase. The F-score of the valence classification was 0.70 in the test set.

Valence	**Positive**	**Negative**
Positive	71.62	28.38
Negative	29.31	70.69

Table 2.9. Confusion matrix of cross-validation using SVM classifier for valence level. Values are expressed as percentages. Total Accuracy: 71.21%

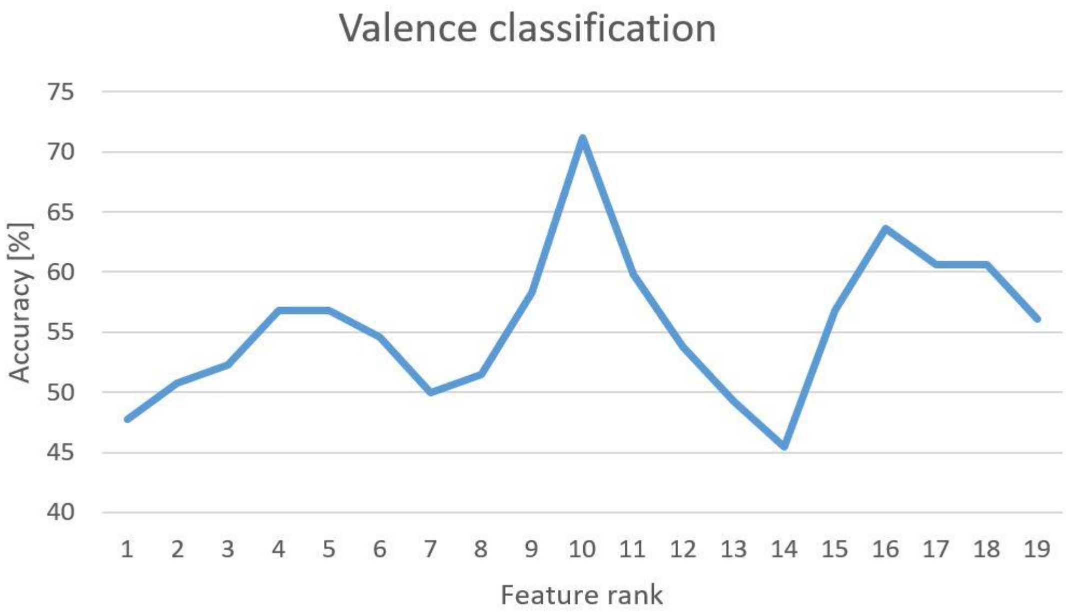

Figure 2.9. Recognition accuracy of valence in cross-validation as a function of the feature rank estimated through the SVM-RFE procedure

Rank	Feature
1	EEG MPC PCA 8
2	EEG MPC PCA 6
3	EEG MPC PCA 11
4	EEG MPC PCA 7
5	EEG MPC PCA 10
6	EEG MPC PCA 12
7	EEG MPC PCA 9
8	EEG Band Power PCA 3
9	EEG Band Power PCA 4
10	EEG MPC PCA 2

Table 2.10. Selected features ordered by their median rank over every fold computed during the LOSO procedure for valence classification

Valence	Positive	Negative
Positive	75.00	25.00
Negative	37.50	62.50

Table 2.11. Confusion matrix of test set using SVM classifier for valence level. Values are expressed as percentages. Total Accuracy: 70.00%

Discussion

The purpose of this study is to develop an emotion recognition system able to automatically discern affective states evoked through an IVE. This is part of a larger research project that seeks to analyse the use of VR as an affective elicitation method, in order to develop emotion recognition systems that can be applied to 3D or real environments. The results can be discussed on four levels: (1) the ability of IVEs to evoke emotions; (2) the ability of IVEs to evoke the same emotions as real environments; (3) the developed emotion recognition model; and (4), the findings and applications of the methodology.

Regarding the ability of the IVEs to evoke emotions, four versions of the same basic room design were used to elicit the four main arousal-valence combinations related to the CMA. This was achieved by changing different architectural parameters, such as illumination, colour and geometry. As shown in Figure 2.7 and

Table 2.5, proper elicitation was achieved for Room 1 (high arousal and negative valence), Room 2 (high arousal and positive valence) and Room 4 (low arousal and positive valence), but it overlapped somewhat with the arousal-valence representation in Room 3: despite the satisfactory pre-test, in the event it evoked higher arousal and valence than expected. This is due to the difficulties we experienced in designing a room to evoke negative emotion with low arousal. It should be noted that IAPS developers may also have experienced this problem because only 18.75% of the pics are situated in this quadrant (Lang et al., 1997). Other works based on processing valence and arousal using words show that a U-model exists in which arousal increases in agreement with valence intensity regardless of whether it is positive or negative (Lewis et al., 2007). Hence, for future works, Room 3 will be redesigned to decrease its arousal and valence and a self-assessment with a larger sample will be performed, by questionnaire, to robustly assess the IVE. Nonetheless, after thresholding the individual self-assessment scores to discern 2 classes (high/low), the IVE set was balanced in arousal and valence. Therefore, we could conclude that the proposed room set can satisfactorily evoke the four emotions represented by each quadrant of the CMA.

To this extent, although previous studies have presented IVEs capable of evoking emotional states in a controlled way (McCall et al., 2016), to the best of our knowledge we have presented the first IVE suite capable of evoking a variety of levels of arousal and valence based on CMA. Moreover, the suite was tested through a low-cost portable HMD, the Samsung Gear, therefore increasing the possible applications of the methodology. High quality images of the stimuli are included in the supplementary material. This represents a new tool that can contribute in the field of psychology, in general, and in the affective computing field, in particular, fostering the development of novel immersive affective elicitation using IVEs.

There are still some topics that need to be researched, relating to the capacity of the IVE display formats, to ensure that they evoke the same emotions as real environments. Studies comparing display formats show that the 360º IVEs offer results closer to reality, according to the participants' psychological responses, and 3D IVEs do so according to their physiological responses (Higuera-Trujillo et al., 2017). Moreover, it is quite possible that IVEs will offer the best solutions at both psychological and physiological levels as they become even more realistic, providing a real improvement not only at the visual and auditory levels but also at the haptic (Blake & Gurocak, 2009). In addition, 3D IVEs allow users to navigate and interact with the environment. Hence, there are reasons to think that they could be powerful tools for developing applications for affective computing, but studies comparing human responses in real and simulated IVE are scarce (Heydarian et al., 2015; Kuliga et al., 2015; Yeom et al., 2017), especially regarding emotional responses; these studies are required. Moreover, every year the resolution of Head Mounted Displays is upgraded, which brings them closer to eye resolution. Thus, it is possible that in some years the advances in Virtual Reality hardware will make the present methodology more powerful. In addition, works comparing VR devices with different levels of immersion are needed in order to give researchers the best set-ups to achieve their aims. In future works, we need to consider all these topics to improve the methodology.

Regarding the emotion recognition system, we present the first study that develops an emotion recognition system using a set of IVEs as a stimulus elicitation and proper analyses of physiological dynamics. The accuracy of the model was 75.00% along the arousal dimension and 71.21% along the valence dimension in the phase of cross-validation, with average of 70.00% along both dimensions in the test set. They all present a balanced confusion matrix. The accuracies are considerably higher than the chance level, which is 58% in brain signal classification and statistical assessment (n=152, 2-classes, p=0.05) (Combrisson & Jerbi, 2015). Although the accuracy is lower than other studies of emotion recognition in images (Valenza et al., 2012) and sounds (Nardelli et al., 2015), our results present a first proof of concept that suggests that it is possible to recognize the emotion of a subject elicited through an IVE. The research was developed with a sample of 60 subjects, who were carefully screened to demonstrate agreement with a "standard" population reported in the literature[46]. It should be noted that the possible overfitting of the model was controlled using: (1) a feature reduction strategy with a PCA; (2) a feature selection strategy using a SVM-RFE; (3) a first validation of the model using LOSO cross-validation; and (4) a test validation using 5 randomly chosen subjects (15%), who had not been used before to train or perform the cross-validation of the model. In the arousal model, features derived from three-signal analyses were selected: 3/3 of HRV, 4/4 of EEG BandPower and 8/12 of EEG MPC. However, in the valence model only the EEG analysis was used: 0/3 of HRV, 2/4 of EEG BandPower and 8/12 EEG MPC. Moreover, in both models, the first six features selected by RFE-SVM were derived from an EEG MPC analysis. This suggests that cortical functional connectivity provides effective correlates of emotions in an IVE. Furthermore, according to recent evidence (He et al., 2017; Nakisa et al., 2018), the reliability of emotion recognition outside of the laboratory environment is improved by wearables. In future experiments, these results could be optimized using further, maybe multivariate signal analyses and alternative machine learning algorithms (Colomer et al., 2016). In addition, the design of new, controlled IVEs that can increase the number of stimuli per subject, using more combinations of architectural parameters (colour, illumination and geometry), should also improve the accuracy and robustness of the model. In future studies, we will improve the set of stimuli presented including new IVEs in order to develop a large set of validate IVE stimuli to be used in emotion research.

The findings presented here mark a new step in the field of affective computing and its applications. Firstly, the methodology involved in itself a novel trial to overcome the limitations of passive methods of affective elicitation, in order to recreate more realistic stimuli using 360º IVEs. Nevertheless, the long-term objective is to develop a robust pre-calibrate model that could be applied in two ways: (1) in 3D environments that would allow the study of emotional responses to "real" situations in a laboratory environment through VR simulation using HMD devices and (2) in physical spaces. We hypothesize in both cases that the emotion recognition models developed through controlled 360º IVEs will work better than the models calibrated by non-immersive stimuli, such as IAPS. This approach will be discussed in future studies using stage 3 of the experimental protocol.

Regarding the implications for architecture, the methodology could be applied in two main contexts, research and commercial. On the one hand, researchers could analyse and measure the impact of different design parameters on the emotional responses of potential users. This is especially important due to the impossibility of developing researches in real or laboratory environments (e.g. analysing arousal changes caused by the pavement width on a street). The synergy of affective computing and virtual reality allows us to isolate a parameter design and measure the emotional changes provoked by making changes to it, while keeping the rest of the environment identical. This could improve the knowledge of the emotional impact that might be made by different design parameters and, consequently, facilitate the development of better practices and relevant regulations. On the other hand, this methodology could help architects and engineers in their decision-making processes for the design of built environments before construction, aiding their evaluations and the selection of the options that might maximize the mood that they want to evoke:

for example, positive valence in a hotel room or a park, low arousal in a schoolroom or in a hospital waiting room and high arousal in a shop or shopping centre. Nevertheless, these findings could be applied to any other field that needs to quantify the emotional effects of spatial stimuli displayed by Immersive Virtual Environments. Health, psychology, driving, videogames and education might all benefit from this methodology.

Chapter 3

Real vs. immersive-virtual emotional experience: Exploiting psycho-physiological patterns in a free exploration of an art museum

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., Alcañiz, M. & Valenza, G. (2019). Real vs. immersive-virtual emotional experience: Analysis of psycho-physiological patterns in a free exploration of an art museum. *PloS one*, 14(10).

Abstract

Virtual reality is a powerful tool in human behaviour research. However, few studies compare its capacity to evoke the same emotional responses as in real scenarios. This study investigates psycho-physiological patterns evoked during the free exploration of an art museum and the museum virtualized through a 3D immersive virtual environment (IVE). An experiment involving 60 participants was performed, recording electroencephalographic and electrocardiographic signals using wearable devices. The real vs. virtual psychological comparison was performed using self-assessment emotional response tests, whereas the physiological comparison was performed through Support Vector Machine algorithms, endowed with an effective feature selection procedure for a set of state-of-the-art metrics quantifying cardiovascular and brain linear and nonlinear dynamics. We included an initial calibration phase, using standardized 2D and 360° emotional stimuli, to increase the accuracy of the model. The self-assessments of the physical and virtual museum support the use of IVEs in emotion research. The 2-class (high/low) system accuracy was 71.52% and 77.08% along the arousal and valence dimension, respectively, in the physical

museum, and 75.00% and 71.08% in the virtual museum. The previously presented 360º stimuli contributed to increasing the accuracy in the virtual museum. Also, the real vs. virtual classifier accuracy was 95.27%, using only EEG mean phase coherency features, which demonstrates the high involvement of brain synchronization in emotional virtual reality processes. These findings provide an important contribution at a methodological level and to scientific knowledge, which will effectively guide future emotion elicitation and recognition systems using virtual reality.

Introduction

The automatic quantification and recognition of human emotions is a research area known as "Affective Computing", which combines knowledge in the fields of psychophysiology, computer science, biomedical engineering and artificial intelligence (Picard, 1997). Due to the central role that emotions play in many background processes, such as perception, decision-making, creativity, memory and social interaction, several studies have focused on trying to obtain a reliable methodology to evoke and automatically identify emotional states from objective psychometric measures (Picard, 2003). Major exploitations of computational machines with affective intelligence focus on healthcare, education, marketing and entertainment (Harms et al., 2010; Jerritta et al., 2011), as well as on environmental psychology, i.e. the study of the effect of the environment on humans (Lindal & Hartig, 2013).

Irrespective of the application, two approaches have commonly been proposed to model emotions: discrete and dimensional models. The former proposes that there is a small set of basic emotions, assuming that complex emotions result from a combination of these basics, including anger, disgust, fear, joy, sadness and surprise (Ekman, 1999). Although discrete models are more easily understood by the non-expert, they are strongly criticized for lacking consistency and objective correlates (e.g. psychobiological and psychophysiological specific correlates) (Barrett, 2017). Dimensional models propose a multidimensional space where each dimension represents a fundamental property common to all emotions. The “Circumplex Model of Affect” (CMA) is one of the most used model, and refers to a Cartesian system of axes with two dimensions (Russell & Mehrabian, 1977): valence, i.e. how much an emotion is perceived as positive or negative; arousal, i.e. the intensity of the emotion in terms of activation from low to high.

To automatically classify emotions, correlates from, e.g. voice, face, posture, text, neuroimaging and physiological signals are widely employed (Calvo & D’Mello, 2010). In particular, several computational methods are based on variables associated with Central Nervous System (CNS) and Autonomic Nervous System

(ANS) dynamics (Calvo & D'Mello, 2010). On the one hand, the use of the CNS to automatically classify emotion is justified by the fact that human emotional processing and perception involve activity of the cerebral cortex. In this regard, the electroencephalogram (EEG) is one of the techniques most used to measure CNS responses (Valenza et al., 2016). On the other hand, a wider class of affective computing studies exploits ANS changes on cardiovascular dynamics as elicited by specific emotional states, especially through Heart Rate Variability (HRV) analyses (Valenza et al., 2012). To this extent, recently proposed emotion recognition systems exploit wearable systems (Valenza, Nardelli, et al., 2014), allowing physiological monitoring in physical real-world environments through both HRV (He et al., 2017) and EEG (Nakisa et al., 2018).

Concerning the experimental emotional manipulation, the ability to reliably and ethically elicit affective states has proven to be a challenging task (Kory & D'Mello, 2015). Based on the nature of the stimuli used to evoke emotional responses, two types are distinguished: active and passive. Active methods may involve behavioural manipulation (Ekman, 2007), social psychological methods with social interaction (Harmon-Jones et al., 2007), or dyadic interaction (Roberts et al., 2007). On the other hand, passive methods can fundamentally present images, sounds or films. Of note, regarding the images, the International Affective Picture System (IAPS) is one of the most prominent databases. It includes over a thousand depictions of people, objects and events standardized on the basis of valence and arousal (Kory & D'Mello, 2015). IAPS has been used in many researches as an elicitation tool in emotion recognition methodologies (Valenza et al., 2012).

Although many computational models have been successfully developed in lab environments using controlled stimuli, the influence of the level of immersion of the set-up (i.e. the objective description referring to the physical extent of the sensory information) has often been underestimated, thus eliciting emotional experiences not similar to real-world scenarios (Baños et al., 2004). To overcome these limitations, researchers propose environmental-simulation technologies to replicate the experience of physical environments (Lange, 2001).

At present, Virtual Reality (VR) is one of the most powerful technologies that simulate experiences and provide the sensation of being in real situations (Baños et al., 2006). In fact, 3D immersive virtual environments (IVE) have successfully been applied to phobias (Peperkorn et al., 2014), presence (Meehan et al., 2005), visualization technologies (Higuera-Trujillo et al., 2017), quality of experience (Egan et al., 2016) and videogames (Bian et al., 2016). Specifically, the main advantages of this technology are that: i) it allows us to isolate and modify variables in an efficient and low-cost way, something which is very difficult, or even impossible, in real environments (Alcañiz et al., 2003); and ii) it allows us to analyse an environment before its construction or environments far distant from the lab. Of note, VR can profitably be used to evoke emotions (Alcañiz et al., 2003; Baños et al., 2012) and states of relaxation or anxiety (Riva et al., 2007). Many VR researches have been performed using desktop or semi-immersive systems such as Powerwalls or caves (Vecchiato et al., 2015). Nowadays, the use of head-mounted displays (HMD) is growing due to their improved performance and decreased price. They are fully immersive devices that isolate the user from the external world. These devices, in fact, provoke a high sense of presence, understood as the illusion of "being-there" (Slater & Wilbur, 1997). Note that HMDs have two main formats for displaying IVEs: 360º panoramas and 3D VR environments. 360º panoramas offer results closer to reality in terms of the participants' psychological responses, while 3D VR in terms of their physiological responses (Higuera-Trujillo et al., 2017). In addition, 3D VR allows the user to freely interact with the environment.

The comparison between the responses evoked by physical environments and their virtual simulations has been studied to some degree through the assessment of psychological responses (Bishop & Rohrmann, 2003), cognitive performance (de Kort et al., 2003) and - to a much lesser extent – physiological and behavioural responses (Yeom et al., 2017; van der Ham et al., 2015). Although differences have been found, environmental simulations achieve a considerable level of general validity (Villa & Labayrade, 2012). However, in the variety of studies undertaken, simulations have not yet been comprehensively compared with the real world in the analysis of emotional experiences, especially by employing a thorough analysis of CNS and ANS dynamics.

To this end, the main aim of the present study is to perform an exploratory research to comparatively and quantitatively investigate the psychological and physiological patterns evoked during, first, free exploration of a real art museum and, second, where they visualize a virtualization of the museum through a 3D IVE.

Three specific hypotheses are proposed in the present study:

H1. Psychological self-assessments do not show significant differences between the real and the virtual museums.

H2. Physiological signals allow prediction of the self-assessment in both cases.

H3. Undertaking an initial calibration phase, using standardized 2D and 360° emotional stimuli, increases the accuracy of the emotion recognition models in real-world environments and their simulations using VR.

To this extent, firstly, we undertake a psychological comparison, using self-assessment tests, for both real and virtual environments. Secondly, we perform a comprehensive physiological comparison using brain and cardiovascular linear and nonlinear dynamics to build arousal and valence-specific classifiers. Thirdly, we analyse the inclusion of 2D (i.e. IAPS images) and 360° standardized emotional stimuli as a part of the calibration phase of the classifier. Moreover, at an exploratory level, we also investigate differences and similarities in psychophysiological responses elicited by real and virtual environments. To this end, we develop emotion recognition models for real vs. immersive-virtual scenario comparison to determine if the subject is experiencing a virtual or real emotional experience. Classification accuracies are gathered from nonlinear Support Vector Machine algorithms and a set of EEG and HRV features extracted using various state-of-the-art metrics. Methodological details, the experimental results, and the discussion and conclusion follow below.

Materials and Methods

Experimental design

An experiment was conducted in two different phases, including two prior stages using controlled stimuli. Each stage was presented consequently (Figure 3.1), with signal acquisition independently recorded. Between each stage, the subjects rested for 3 minutes, sitting on a chair. Stage 1 consisted of showing the subjects 2D pictures based on IAPS. Stage 2 consisted of a 360º panorama emotion IVE presented in an HMD. Finally, the last stage in both phases consisted of the free exploration of a museum exhibition. However, in Stage 3.1, the subjects explored a real museum exhibition and in Stage 3.2 the subjects explored the 3D virtual reality simulation of the same exhibition. Each subject was randomly assigned to undergo either Stage 3.1 or Stage 3.2.

The ethics committee of the Polytechnic University of Valencia approved the experimental protocol. All methods and experimental protocols were performed in accordance with the guidelines and regulations of the local ethics committee of the Polytechnic University of Valencia. Written informed consent was obtained from all participants involved in the experiment. In particular, the individual in this manuscript has given written informed consent (as outlined in the PLOS consent form) to publish these case details.

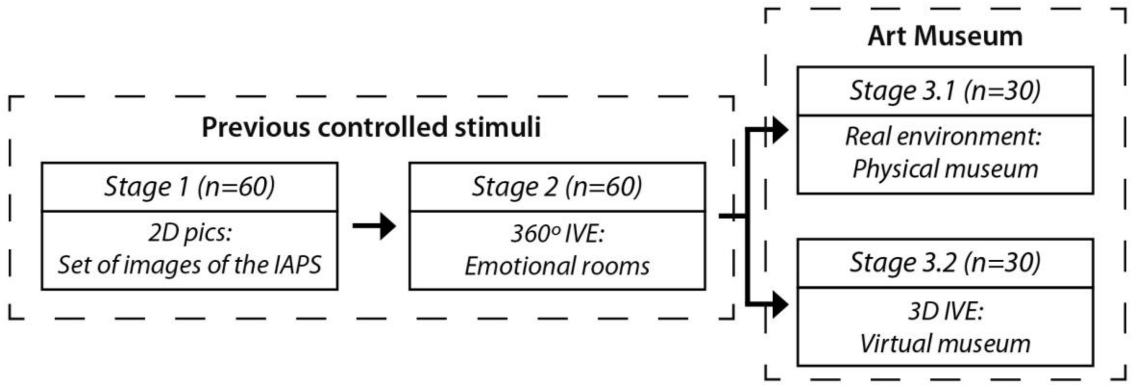

Figure 3.1. Experimental phases of the research. n represents the number of subjects involved in each stage

Stimulus Elicitation

Previous Controlled Stimuli

In *Stage 1* we developed an affective elicitation using standardized 2D pictures. This was achieved by projecting a set of images onto a monitor (Dell E198FPb, LCD, 19-inch, 1280x1024 @ 75Hz). At first, the users were asked to rest for 4 minutes while looking at a blank image (B), in order to start the experiment from a relaxed status. This period was divided into: one open eye minute; one closed eye minute; one open eye minute; and one closed eye minute. Thereafter, the affective elicitation began. We took inspiration from the elicitation methodologies reported in previous works (Valenza et al., 2016; Valenza et al., 2012), with minor changes. Briefly, the slideshow comprised of 9 image sessions, alternating neutral sessions (from N1 to N5) and arousal sessions (from A1 to A4). The order of presentation of the images was random. One-minute resting-state sessions (from R1 to R8) were placed between each neutral/arousal session. Each arousal session was divided into 3 blocks of valence (from V1 to V3). Thus, 1 block of neutral pictures and 12 blocks of non-neutral pictures were displayed. Further details are reported in the Supplementary Material. The overall protocol used 110 images. Each image was presented for 10 seconds for the whole duration of the experiment, 18 minutes and 20 seconds.

In *Stage 2*, we developed an affective elicitation using architectural environments displayed by 360º panoramas implemented in a portable HMD. The stimuli had been analysed and validated in previous research (Marín-Morales et al., 2018). This type of environment was chosen as the influence of architectural environments on affective-behavioural responses is widely accepted (Eberhard, 2009). Previous research shows that subtle variations in the space may generate different neurophysiological responses (Nanda et al., 2013). In addition, previous works show that the 360º panorama-format using HMD devices is a valid set-up for evoking psychological and physiological responses similar to those that physical environments evoke (Higuera-Trujillo et al., 2017). Hence, four architectural environments were proposed as representative of four emotional states (Figure 3.2), following the CMA (Russell, 1980). The emotional rooms were designed based on different variations of the same base-scenario, “Villa in the forest”, by Kazuyo Sejima (Sejima, 1996). The research team, which included architects, considered this an appropriate base from which to make modifications to generate different moods. The architectural parameters used to modify the base-scenario were illumination, colour and geometry.

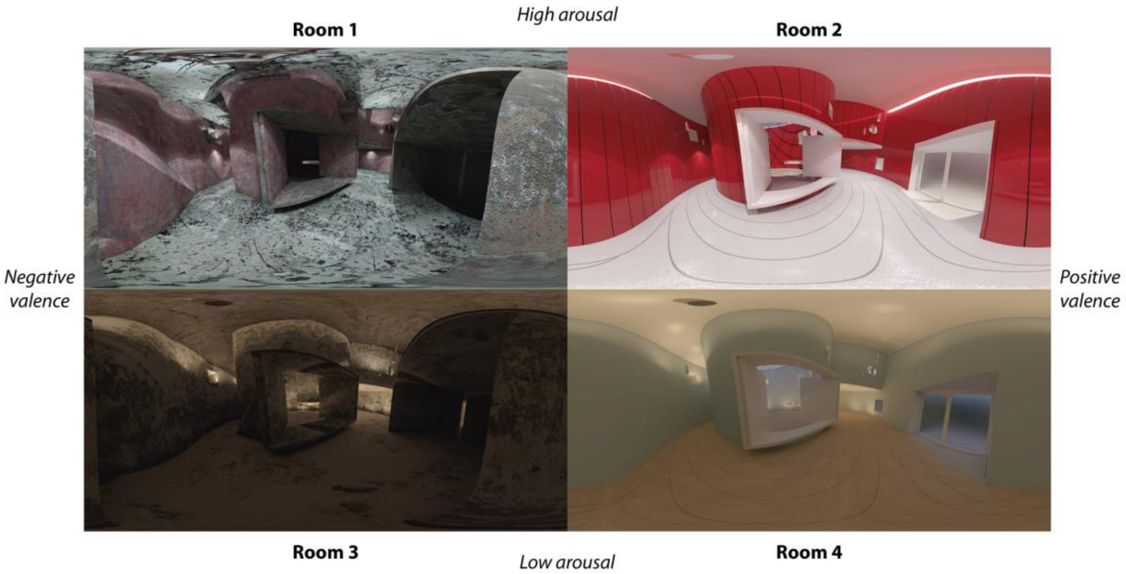

Figure 3.2. 360º panoramas used in stage 2

Technically, the process of developing the four architectural environments consisted of modelling and rendering. Modelling was performed by using Rhinoceros v5.0 (www.rhino3d.com) and rendering was performed using the VRay engine v3.00.08 (www.vray.com), operating from Autodesk 3ds Max v2015 (www.autodesk.es). Renders were exported in .jpg format with resolutions of 6000x3000 pixels at 300 dots per inch. The 360º panoramas were implemented in Samsung Gear VR HMDs and the reproduction was fluid and uninterrupted. The Samsung HMD has a stereoscopic screen of 1280×1440 pixels per eye and a 96° field of view, supported by a Samsung Note 4 mobile telephone with a 2.7GHz quad-core processor and 3GB of RAM. Figure 3.3 shows an example of experimental set-up of Stages 1 and 2.

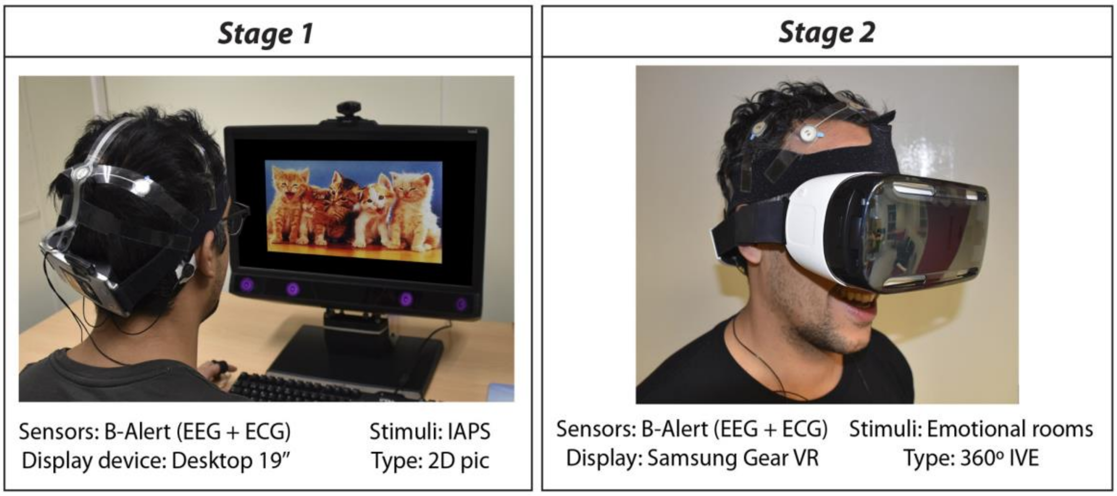

Figure 3.3. Example of experimental set-up of stage 1 and 2

Regarding the protocol, each room was presented for 1.5 minutes and the sequence was counter-balanced using the Latin Square method. After viewing each room, the users were asked to orally self-assess the emotional state evoked by each room using a SAM questionnaire embedded in the 360º photo, ranging from -4 to 4, for arousal and valence dimensions.

Physical Museum Exhibition

In *Stage 3.1*, we performed an affective elicitation using a physical environment. An art exhibition was chosen in order to evoke an intense emotional experience. The Institut Valencià d'Art Modern (IVAM) offered us their facilities to undertake our study. We selected the art-exhibition "Départ-Arrivée", by Christian Boltanski, because it had a very emotional topic, the Nazi holocaust. The exhibition had 5 rooms and an area of approximately 750 m^2 (Figure 3.4).

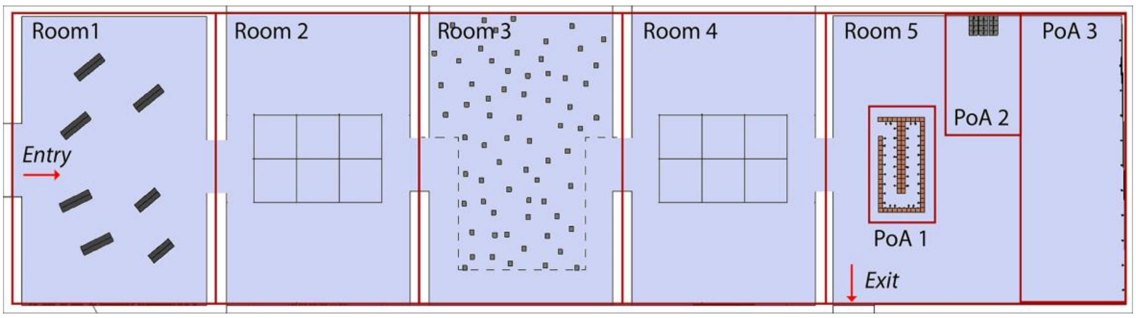

Figure 3.4. Plan of the art-exhibition with the 5 rooms and 3 pieces of art

The subjects were asked to explore freely the first four rooms. When they entered the fifth room, they had also to explore it freely, but they were, in addition, required to stop and study the three pieces of art in detail. The researcher waited for the subject at the exit door, allowing the subject to freely explore the exhibition.

In order to track the position of the subjects, therefore being sure that she/he visited all rooms, we used a GoPro camera, that subjects carried attached to their chests by means of a suitable harness. The physiological signals were recording on a laptop that the subject carried in a backpack. Figure 3.5 shows an exemplary experimental set-up of Stage 3.1.

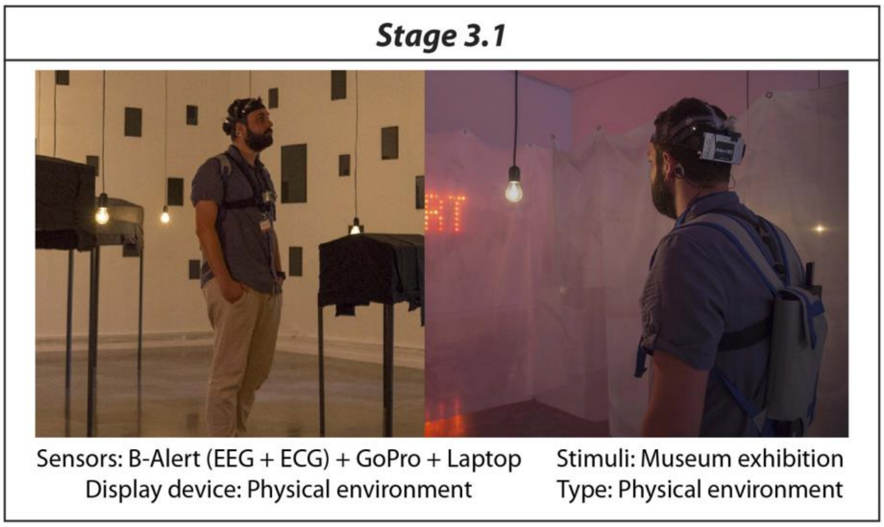

Figure 3.5. Example of experimental set-up of Stage 3.1

After the museum exploration, the subjects were asked to complete two questionnaires. In the first, they evaluated the emotional impact of each of the five rooms and the three pieces of art, using a SAM questionnaire and a photo of each room. In the second, we presented two questions to evaluate the subjective impact of the sensors: (1) "During the test, did you feel annoyed by the sensors?"; (2) "During the test, was there ever a time when you forgot that you were sensorized?". The subjects who reported feeling "moderately" or "a lot" annoyed were excluded from further analyses.

Virtual Museum Exhibition

In *Stage 3.2*, an affective elicitation was performed through the 3D VR representation of the museum exhibition visited in phase 1. The Unity 5.1 game engine (www.unity3d.com) was used. A three-dimensional representation of the museum exhibition was provided by Rhinoceros v5.0. Textures partially extracted from the physical environment were imported to achieve maximum realism. This involved exhaustively and methodologically drawing and photographing the entire exhibition. Exemplary photographs of the real environment and screenshots of the virtual environment are shown in Figure 3.6. Further examples are in the Supplementary Materials. Regarding the 3D VR simulation, the developed scenario was compiled for HTC Vive (www.vive.com). This system allows visual and displacement simulations. On the one hand, visualization is performed using an HMD with 2160x1200 pixels (1080×1200 per eye) and a field of view of 110 degrees working at 90Hz refresh rate. On the other hand, displacements are performed using tracking technology, two controllers and two base stations that, together, allow the subject to interact with their environment and physically move within an area of a 2x2 metres. Specifically, the teleport navigation metaphor included in the HTC Vive developed tools was used, with a 2.5 metres from the subject maximum teleportation radio. It was chosen in order to achieve pseudo-naturalistic navigation, allowing the subjects to take large steps. The entire system was connected to the research PC (Predator G6, www.acer.com) via DisplayPort 1.2 and USB 3.0, running smoothly and without interruptions. Figure 3.7 shows an exemplary experimental set-up of Stage 3.2.

Figure 3.6. Comparison between the physical museum (left) and the virtual museum (right). The photos represent Room 1 and Room 5

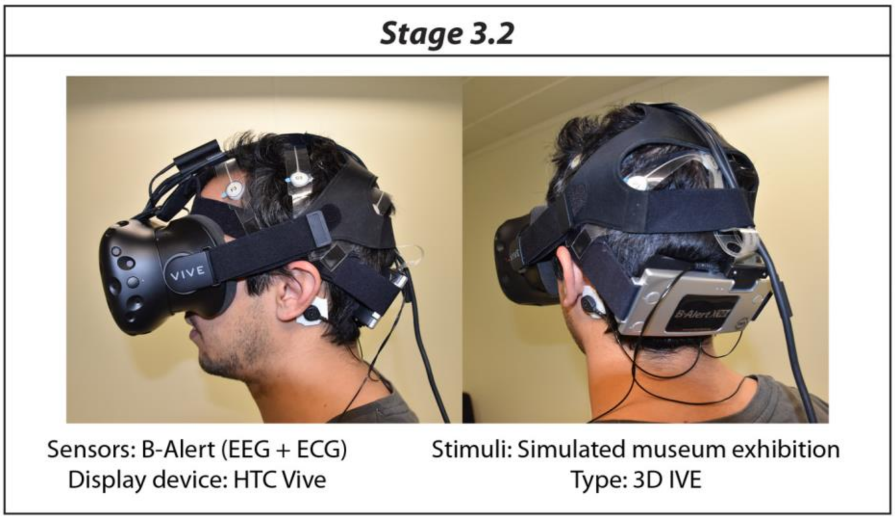

Figure 3.7. Exemplary experimental setup of Stage 3.2

Before starting this stage of the experiment, the subjects were placed in a neutral scenario, which displayed only a floor, without any texture. A screenshot of this scenario is included in the Supplementary Material. They were asked to undertake a period of training in this place. They could take all the time that they needed inside this scenario, until they considered their adaptation to VR and the navigation metaphor complete. After this, the instructions for the virtual museum exhibition were exactly the same as for the physical exhibition. The subject´s navigation was also displayed in real time on a desktop and the researcher used this to note when the subject arrived at the exit, in order to stop the recording and remove the HMD.

Following the exploration of the virtual museum, the subjects were asked to answer the same two questionnaires as for Stage 3.1: (1) affective self-assessment evaluation of the rooms and pieces of art; (2) impact of the sensors in the behaviour of the subjects. In addition, in this phase the subjects had to answer a questionnaire about presence in the virtual museum. We used the well-known "SUS questionnaire" (Slater et al., 1994). Its current version consists of six items, rated on 1-to-7 Likert scale, measuring three aspects of the subject's senses: the experience of being inside the simulation; the consideration of the simulation as the dominant reality; and the memory of the simulation as a place.

Participants' Eligibility and Group Homogeneity

A homogeneous population of 60 healthy subjects (age 28.9 ± 5.44, 40% male, 60% female), suffering neither from cardiovascular nor obvious mental pathologies, was recruited to participate in the experiment. They were divided into 30 subjects for the first phase and 30 for the second. The following were the criteria to participate in the study: age between 20 and 40 years; Spanish nationality; not having formal education in art or a fine-art background; not having any previous virtual reality experience; and not having previously visited this particular art exhibition.

Two questionnaires were included to ensure that the subjects were in a healthy mental state and constituted a homogeneous group. In the first, all participants were screened by a Patient Health Questionnaire (PHQ) (Kroenke et al., 2001). Only participants with a score lower than 5 were included in the study to avoid the presence of either middling or severe personality disorders. In the second, a self-assessment, based on a selection of IAPS pictures (Lang et al., 1997) using the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994), was administered. The presented set consisted of different degrees of arousal and valence perception (arousal from 3.41 to 7.24; valence from 1.29 to 8.17; pictures selected: 7234, 5201, 9290, 1463, 9181, 8380, 3102, 4652).

The self-assessment values were used to analyse if any subject had an emotional response that could be considered as an outlier with respect to standard elicitations. To this end, the arousal and valence of each subject were standardized through a z-score using the mean and deviation of the IAPS published scores (Lang et al., 1997). Standardized evaluations outside of the range -2.58 to 2.58 (i.e., α=0.01) were designated as outliers (Cousineau & Chartier, 2010). Subjects with outliers were excluded from further analyses, while we retained the emotional responses that belong statistically to 99% of the population as published in the IAPS. In addition, the subjects whose signal recording experienced errors were rejected, e.g. because of disconnection of the sensors during the elicitation. The participants had successfully to complete all the stages.

Physiological Signals and Instrumentation Set

The electroencephalographic (EEG) and electro-cardiographic (ECG) signals were acquired using B-Alert x10 (Advanced Brain Monitoring, Inc. USA). This provides an integrated approach for wireless wearable acquisition, sampled at 256 Hz. Regarding the EEG, the location of the sensors was in the frontal (Fz, F3 and F4), central (Cz, C3 and C4) and parietal (POz, P3, and P4) regions based on international 10-20 electrode placement. A pair of electrodes placed below the mastoid was used as a reference. A test was performed to check that the impedances of the electrodes were below 20kΩ. In order to check the proper conductivity of the electrodes, a test was performed. Concerning the ECG, the left lead was located on the lowest rib and the right lead on the right collarbone. Data from 15 subjects out of 60 were rejected due to poor quality.

Signal Processing

Firstly, the signals were synchronized and segmented for each stage. The methodology used is detailed in the Supplementary Materials. Then, HRV and EEG signal processing methods were applied to extract features to characterize the physiological responses to the stimuli.

Heart Rate Variability

To obtain the RR series from the ECG, we implemented the Pan-Tompkins's algorithm for QRS complex detection. The individual trends components were removed using the smoothness prior detrending method (Tarvainen et al., 2002). Artefacts and ectopic beats were corrected through the use of Kubios HRV software (Tarvainen et al., 2014). From the RR series, we performed the analysis of the standard HRV parameters in the time and frequency domains. In addition, we included other HRV measures quantifying heartbeat nonlinear and complex dynamics (Acharya et al., 2006). Table 3.1 presents a list of features included.

The time domain analysis includes the following features: average and standard deviation of the RR intervals, the root mean square of successive differences of intervals (RMSSD), the number of successive differences of intervals which differ by more than 50 ms (pNN50), the triangular interpolation of the HRV histogram and the baseline width of the RR histogram evaluated through triangular interpolation (TINN). The features of the frequency domain were calculated using a power spectrum density (PSD), applying Fast Fourier Transform. The analysis was performed in three bands: VLF (very low frequency, <0.04 Hz), LF (low frequency, 0.04-0.15 Hz) and HF (high frequency, 0.12-0.4 Hz).

Time domain	Frequency domain	Other
Mean RR	VLF peak	Poincaré SD1
Std RR	LF peak	Poincaré SD2
RMSSD	HF peak	Approximate Entropy (ApEn)
pNN50	VLF power	Sample Entropy (SampEn)
RR triangular index	VLF power %	DFA α1
TINN	LF power	DFA α2
	LF power %	Correlation dimension (D2)
	LF power n.u.	
	HF power	
	HF power %	
	HF power n.u.	
	LF/HF power	
	Total power	

Table 3.1. List of HRV features used

For each of the three frequency bands we calculated the peak value (corresponding to the frequency having maximum magnitude) and the power of each frequency band in absolute and percentage terms. Normalized power (n.u.) was calculated for the LF and HF bands as the percentage of total power, subtracting previously the power of VLF to the total power. The LF/HF ratio was calculated to quantify sympatho-vagal balance and to reflect sympathetic modulations (Acharya et al., 2006). Moreover, the total power was calculated.

Finally, many features were extracted using nonlinear analysis, as they were shown to be important quantifiers of cardiovascular control dynamics mediated by the ANS in affective computing (Valenza et al., 2012). Firstly, Poincaré plot analysis was applied. It is a quantitative-visual technique, whereby the shape of a plot is categorized into functional classes, providing summary information of the behaviour of the heart. SD1 is associated with fast beat-to-beat variability and SD2 analyses the longer-term variability of R–R (Acharya et al., 2006). An entropy analysis was included, using Sample Entropy (SampEn) and Approximate Entropy (ApEn). SampEn provides an evaluation of time-series regularity (Richman & Moorman, 2000) and ApEn detects changes in underlying episodic behaviour not reflected in peak occurrences or amplitudes (Pincus & Viscarello, 1992). DFA correlations analyse short-term and long-term fluctuations through the $\alpha1$ and $\alpha2$ features, where $\alpha1$ represents the fluctuation in range of 4-16 samples and $\alpha2$ refers to the range of 16-64 samples (Peng et al., 1995). Finally, the D2 feature measures the complexity or strangeness of the time series. This is expected to provide information on the minimum number of dynamic variables needed to model the underlying system (Grassberger & Procaccia, 1983).

Electroencephalographic Signals

Figure 3.8 shows the complete EEG processing scheme, which is performed using the open source toolbox EEGLAB (Delorme & Makeig, 2004).

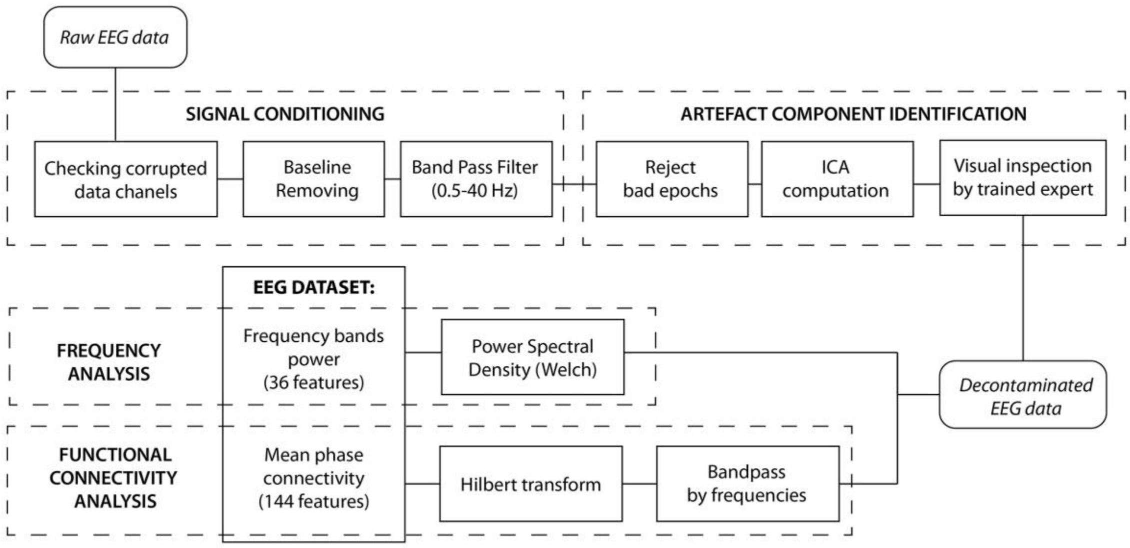

Figure 3.8. Block scheme of the EEG signal processing steps

Firstly, the data from each channel was analysed to identify corrupted channels using the fourth standardized moment (kurtosis) along the signal of each electrode (Colomer et al., 2016). Moreover, the channel was also classified as corrupted if the signal was flatter than 10% of the total stage duration. If a channel was considered as corrupted, it could be interpolated from its neighbouring electrodes. The subject would be rejected if more than one channel was corrupted. Among all of the subjects, only one channel was interpolated.

The EEG baseline was removed by mean subtraction and a band pass filter between 0.5 and 40 Hz. The signal was segmented in epochs of one-second duration. Moreover, an automatic artefact detection was applied, rejecting epochs when more than 2 channels contained samples which exceeded an absolute threshold of 100.00 µV and a gradient of 70.00 µV between samples (Kober et al., 2012). The Independent Component Analysis (ICA) (Hyvärinen & Oja, 2000) with an infomax algorithm was performed to identify and remove components due to blinks, eye movements and muscular artefacts. The components were analyzed by a trained expert to identify and reject those related to artefacts. The effectiveness of the algorithms used to detect and remove artefacts was carefully checked by visual inspection. The subjects who had more than one third of their signals affected by artefacts were rejected. Spectral and functional connectivity analyses were performed after the pre-processing.

An EEG spectral analysis was performed to estimate the power spectra in each epoch, within the frequency bandwidth: θ (4-8 Hz), α (8-12 Hz), β (13-25 Hz), γ (25-40 Hz). Frequency band δ (< 4Hz) was not taken into account in this study. It was performed using Welch's method with 50% overlapping. 36 features were obtained from the 9 channels and 4 bands. The functional connectivity analysis was performed using Mean Phase Coherence (Mormann et al., 2000). It was performed for each pair of channels in each band:

$$R^2 = E[\cos(\Delta\phi)]^2 + E[sin(\Delta\phi)]^2 \qquad (1)$$

Where R is the MPC, $\Delta\phi$ is the relative phase difference between two channels derived by the instantaneous difference of the analytics signals from the Hilbert transform, and E is the expectation operator. MPC values can oscillate between 0 and 1. The MPC is close to 1 when a strong phase synchronization exists between two channels. Alternatively, MPC is close to 0 if the two channels are not synchronized. From each combination of a pair of 9 channels in one specific band, 36 features were extracted. Consequently, 144 features were developed from the 4 bands analysed.

Data Fusion and Pattern Recognition

An overview of the emotion recognition classification scheme is shown at Figure 3.9. For each stimulus, HRV features were calculated using the time windows defined in the segmentation methods. Concerning EEG, we considered the mean of the time-windows of the stimuli as the representative value in both analyses. Therefore, each stimulus (pictures, emotional rooms and museum rooms/pieces of art) was described by 209 features.

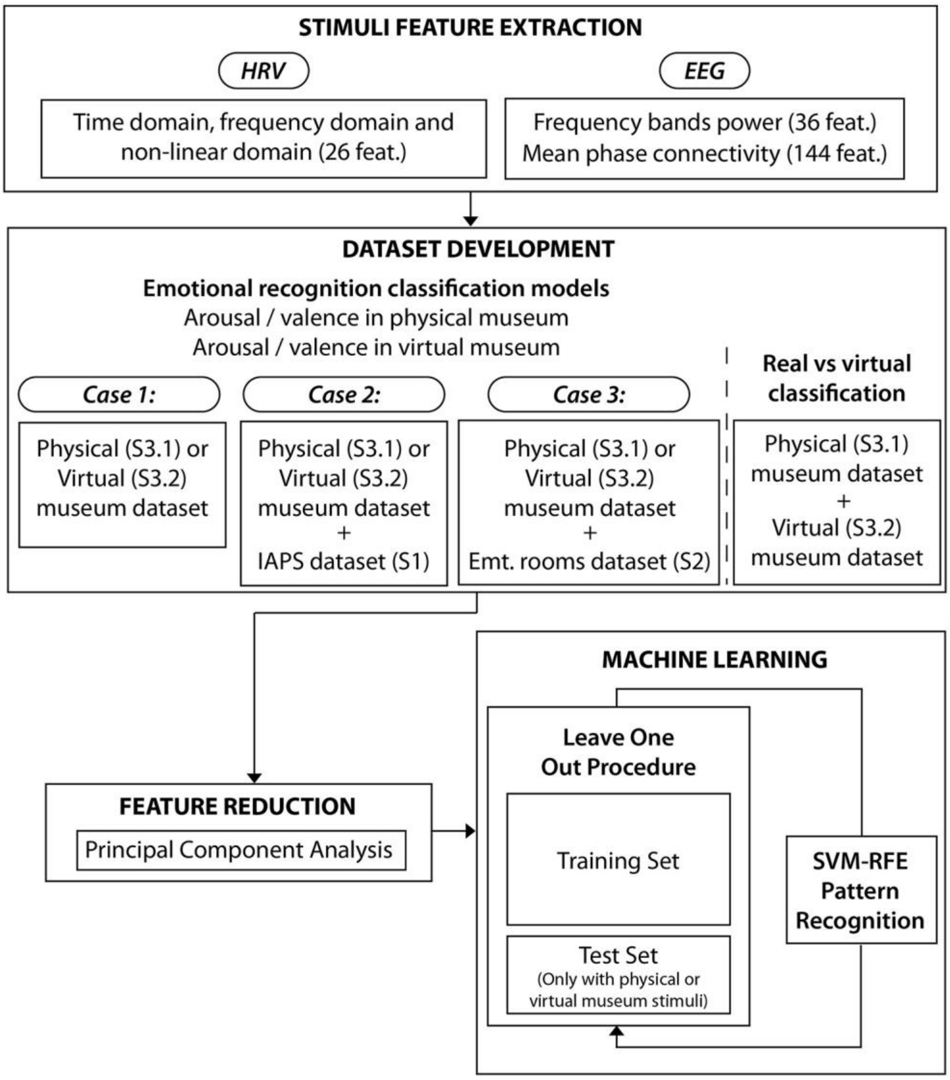

Figure 3.9. Overview of the data analysis and emotional pattern recognition

Four classification models were independently developed: arousal level in the physical museum, valence level in the physical museum, arousal level in the virtual museum, and valence level in the virtual museum.

For each model, the datasets of stimuli were created using three analytical cases: (1) using only the museum data, (2) including IAPS data and (3) also including the 360º data. The different cases were performed to test the following analyses.

Case 1: Physical or Virtual Museum. The features of stimuli from Stage 3.1 were used in order to analyse emotion recognition in the physical museum. In addition, the features of the stimuli from Stage 3.2 were independently used to analyse emotion recognition in the virtual museum.

Case 2: Physical or Virtual Museum + IAPS. In order to analyse the influence of including standardized 2D image responses in the emotional models, the features of Stage 1 stimuli were concatenated with the physical museum feature stimuli (Stage 3.1) or virtual museum feature stimuli (Stage 3.2).

Case 3: Physical or Virtual Museum + Emotional Rooms. In order to analyse the influence of including 360º IVE responses in the models, the features of Stage 2 stimuli were concatenated with the physical museum feature stimuli (Stage 3.1) or virtual museum feature stimuli (Stage 3.2).

In each emotional model, the class label was bipolarized into high/positive (>0) and low/negative (<=0) for both arousal/valence.

Finally, a 2-class pattern recognition algorithm discerning between real vs. virtual museum exploration was developed, i.e. a classifier that aims to recognize if the emotional experience is elicited from a virtual or real scenario.

In all the classification models, including the emotional and real vs. virtual classifier, a feature reduction strategy was adopted to decrease the dimension of the dataset due to the high-dimensional feature space obtained. We implemented the Principal Component Analysis method (PCA) (Jolliffe, 2002), which is based on the linear transformation of the different variables in the principal components. We included the features that explained 95% of the variability of the dataset. The PCA was independently applied in the three analyses. In order to validate the machine learning models, the Leave-One-Subject-Out (LOSO) cross-validation procedure was applied, using Support Vector Machine (SVM)-based pattern recognition (Schöllkopf et al., 2000). For the LOSO scheme, the training set was normalized by subtracting the median value and dividing by the median absolute deviation over each dimension.

In each iteration, the validation set consisted of the stimuli of the physical or virtual museums of one specific subject; it was normalized using the median and deviation of the training set. Regarding the learning model, a C-SVM with sigmoid kernel function was used. The parameters of cost and gamma were optimized using a vector with 15 parameters logarithmically spaced between 0.1 and 1000. Moreover, we performed a feature selection strategy to explore the relative importance of each feature. A support vector machine recursive feature elimination (SVM-RFE) procedure, in a wrapper approach, was included. It was performed on the training set of each fold and we computed the median rank for each feature over all folds.

We specifically chose a recently developed nonlinear SVM-RFE which includes a correlation bias reduction strategy in the feature elimination procedure (Yan & Zhang, 2015). The model was optimized to achieve best accuracy whenever it has a balanced confusion matrix. We consider a model balanced when its confusion matrix has a true positive and a true negative over 60%. The algorithms were implemented using Matlab© R2016a and LIBSVM toolbox (Chang & Lin, 2011).

Results

Subjects' Self-assessment

No subjects showed depressive symptoms according to their PHQ-9 scores. The mean and standard deviations of the PHQ-9 questionnaires were 3.31 ± 2.57. Considering the IAPS self-assessment, a total of 8 subjects were considered outliers with respect to standard emotion elicitations.

Regarding the Stage 2, the evaluation of the subjects for each IVE averaged using mean and standard deviation in terms of arousal were (Room 1: 1.17 ± 1.81, Room 2: 2.10 ± 1.59, Room 3: 0.05 ± 2.01, Room 4: -0.60 ± 2.11) and valence (Room 1: -1.12 ± 1.95, Room 2: 1.45 ± 1.93, Room 3: -0.40 ± 2.14, Room 4: 2.57 ± 1.42), achieving the emotion statement for which they were designed, except in the case of arousal in Room 3.

Concerning Stages 3.1 and 3.2, Figure 3.10 shows the self-assessment of the subjects for the museum stimuli (rooms and pieces of art), using mean and standard deviations in terms of arousal and valence. Due to the non-Gaussianity of data ($p < 0.05$ from the Shapiro-Wilk test with null hypothesis of having a Gaussian sample), the Mann-Whitney U test was applied ($\alpha<0.05$). Along the arousal dimension, no significant differences were found. Regarding valence, only room 1 showed a significant difference (p-value=0.006). In addition, we analyse the stimuli considering a second alpha threshold ($\alpha<0.1$) in order to decrease the probability of perform a type II error. In this case, room 1 (p-value=0.084) and room 4 (p-value=0.053) show higher arousal in virtual condition, and room 1 (p-value=0.006) and room 3 (p-value=0.051) present higher valence in virtual condition. After the bipolarization of scores (positive/high >0), the physical museum presents 59.72% of high arousal and 40.97% of positive valence values; and the virtual museum presents 71.71% of high arousal and 61.84% of positive valence values.

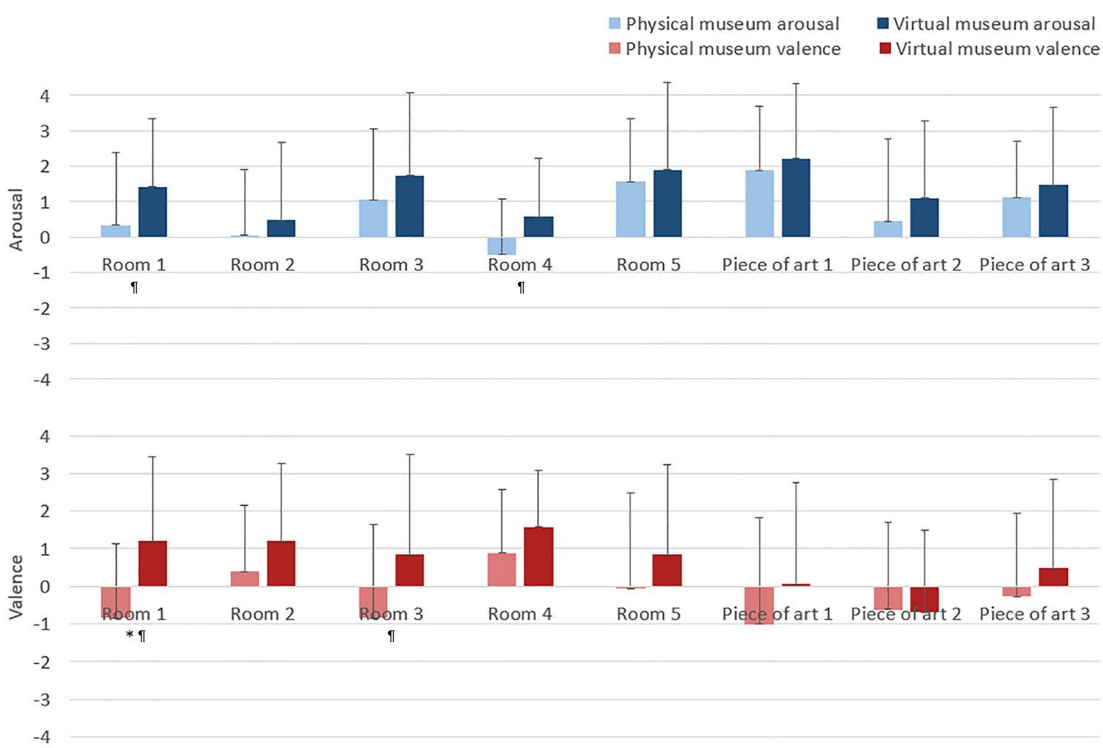

Figure 3.10. Self-assessment scores in physical and virtual museums using SAM and a Likert scale between -4 and +4. Bars represent the means, vertical lines represent the standard deviation of the means, blue represents arousal, and red valence. (indicates significant differences with p < 0.05, ¶ indicates significant differences with p < 0.1)*

Emotion Recognition Classification

Table 3.2 shows an overview of the results of the four emotion recognition models in three analysis cases. Regarding arousal recognition in the physical museum, not including IAPS or emotional rooms data, the accuracy is 68.05%. In the case where IAPS and emotional rooms data were included, the accuracy increases by 3.47%, reaching 71.52% in both cases. In all cases the model used features of the three analyses and the confusion matrices were balanced.

					Confusion matrix				Featured used			
Analysis cases	**Feature**	**Accuracy**	**F-Score**	**ΔAccuracy**	**True high/pos**	**False high/pos**	**False low/neg**	**True low/neg**	**Total**	**HRV**	**EEG Band**	**EEG MPC**
(1) Physical museum	Arousal	68.05 %	0.68	-	70.93	29.06	36.2	63.79	9/14	1/3	1/1	7/10
(2) Physical museum + IAPS	**Arousal**	**71.52 %**	**0.72**	**+3.47 %**	75.58	24.41	34.48	65.5	10/18	1/2	1/4	8/12
(3) Physical museum + Em. Rooms	**Arousal**	**71.52 %**	**0.72**	**+3.47 %**	79.06	20.93	39.65	60.34	16/18	3/3	3/3	10/12
(1) Physical museum	Valence	74.30 %	0.74	-	74.57	25.42	25.88	74.11	9/14	1/3	1/1	7/10
(2) Physical museum + IAPS	**Valence**	**77.08 %**	**0.76**	**+2.78 %**	72.88	27.11	20	80	10/18	2/2	1/4	7/12
(3) Physical museum + Em. Rooms	Valence	76.38 %	0.74	+2.08 %	69.49	30.5	18.82	81.17	3/18	0/3	0/3	3/12

(1) Virtual museum	Arousal	(TN< 60)	-	-	88.07	11.92	44.18	55.81	-	-	-	-
(2) Virtual museum + IAPS	Arousal	71.05 %	0.70	-	75.22	24.77	39.53	60.46	4/22	0/2	0/3	4/17
(3) Virtual museum + Em. Rooms	**Arousal**	**75.00 %**	**0.75**	-	76.14	23.85	27.9	72.09	1/25	0/4	0/3	1/18
(1) Virtual museum	Valence	67.10 %	0.68	-	71.27	28.72	39.65	60.34	4/26	0/3	0/5	4/18
(2) Virtual museum + IAPS	Valence	67.10 %	0.68	+0.00 %	71.27	28.72	39.65	60.34	3/22	0/2	0/3	3/17
(3) Virtual museum + Em. Rooms	**Valence**	**71.05 %**	**0.71**	**+3.95 %**	74.46	25.53	34.48	65.51	3/25	0/4	0/3	3/18

Table 3.2. Level of emotion recognition

Level of recognition of arousal and valence in physical/virtual museum exhibition using (1) only physical/virtual museum dataset (2) including IAPS dataset and (3) also including emotional rooms dataset. Average of accuracy in percentage, F-score, increment of accuracy when IAPS and Rooms datasets were included in each case, confusion matrix and features used in each analysis. Bold indicates cases with the highest accuracy.

Concerning the valence recognition in the physical museum, not including IAPS or the emotional rooms data, the accuracy is 74.30%. The best accuracy is obtained by including IAPS data, achieving 77.08%. The confusion matrix is balanced in all cases and features of all analyses were included.

Regarding arousal in the virtual museum, it was not possible to develop a balanced model without including IAPS or emotional room data, because the true negative was below 60%. Including the IAPS data, the accuracy was 71.05%. However, the best accuracy is obtained by including the emotional rooms, achieving 75.00%. Moreover, this model presents a more balance confusion matrix. Both cases only use EEG MPC features.

Concerning valence in the virtual museum, not including IAPS or the emotional rooms, the accuracy is 67.10%. The model including IAPS data presents the same results. The best accuracy includes emotional room data, achieving 71.05% of accuracy. All cases used only EEG MPC features.

Real vs. Virtual Classification

Table 3.3 shows the level of recognition of the nature of the stimuli in the museum, classifying if the stimuli are real or virtual. The accuracy is 95.27% and the confusion matrix is balanced. The model uses only one feature of EEG MPC, the first component of the PCA, to achieve this level of recognition.

Analysis cases	Feature	Accuracy	F-Score	Confusion matrix True high	False high	False low	True low	Featured used Total	HRV	EEG Band	EEG MPC
Real vs. Virtual	Nature	95.27 %	0.95	94.07	5.92	3.47	96.52	1/17	0/3	0/2	1/12

Table 3.3. Level of recognition of nature of stimuli

Level of recognition of nature of stimuli (real or virtual), including average of accuracy in percentage, F-score, confusion matrix and features used from each analysis

Discussion and Conclusion

The purpose of this novel and exploratory research was to quantitatively compare psychological and physiological patterns during an emotional experience in a physical environment and their virtualization through a 3D IVE, guiding future emotion elicitation and recognition systems using VR. With this aim in mind, we developed a realistic 3D IVE simulation of an art museum and performed a comparative study involving 60 subjects in a real art museum and its simulation, while they were performing a free exploration of an exhibition. In addition, we included two prior phases including controlled stimuli using 2D pictures and 360º IVEs, in order to study the influence of this data on the accuracy and robustness of the emotional models. The results can be discussed on four levels: i) a comparison of the psychometric scores, ii) a comparison of the physiological patterns, iii) a comparison of the level of emotion recognition and the influence of previously (standardized) controlled stimuli, iv) a comparison of emotional subjective and psychophysiological correlates in VR and real scenarios and its meaning in the framework of the different theories and models of emotion and iv) a methodological assessment.

Self-assessment results were used to compare the psychometric patterns. The virtual museum presents slightly more arousal and valence levels than the physical museum. This slight bias could be due to the subjects having no previous VR experience, and the novelty could increase arousal and valence. This should be taken account of in future experiments with these types of subjects. However, only Room 1 presents significant differences in valence considering the usual alpha threshold (α=0.05). Moreover, considering that this conservative threshold is focused to avoid type I error, we analyse a second threshold (α=0.1) to decrease the probabilities to perform a type II error and claim incorrectly the null hypothesis of equal means. The vast majority of the stimuli (93.75%) do not present statistically significant differences in self-assessment considering the first alpha threshold (α=0.05). However, two rooms (1 and 4) present higher arousal, and two rooms (1 and 3) show higher valence in virtual condition considering the second alpha threshold (α=0.1) (Howell, 2009). Room 1 presents the biggest differences in the evoked emotion and it could be provoked by a 'wow' effect derived also by the novelty and the lack of previous experience in VR. This effect will need to be consider in future research. The results suggest that 3D IVEs are powerful tools for emotional elicitation, since the majority of stimuli do not present significant differences in affective statements reported by the subjects in comparison to those evoked by physical environments, and are appropriate for emotion research, thus supporting H1. The results also support the use of VR to elicit emotion and are in accordance with previous research (Baños et al., 2004; Riva et al., 2007; Gorini et al., 2011), but more confirmatory research is needed in the future, especially considering the new VR devices.

Regarding the physiological pattern comparison, the automatic feature selection of the SVM-RFE algorithm was used. The influence of the features of each analysis on the models could be analysed as the PCA was applied independently for HRV, EEG Band Power and EEG MPC. In emotion recognition of the physical museum, all the models (except for the valence with Emotional Room data) used features of all the analyses to predict mood. Thus, all analyses contributed with information about emotional status. However, the emotion recognition models developed for the virtual museum used only few EEG MPC features. Moreover, the real vs virtual classification model used only the first component of the EEG MPC PCA to discriminate between the real and virtual museum stimuli. These results reveal the important role that brain synchronization plays in the neuro-physiological processes involved in VR, as they can discriminate between virtual and real environments with a level of recognition of over 95% accuracy. In addition, the use of EEG MPC features to recognize emotions in VR suggests that brain synchronization is deeply involved in emotional processes in VR environments. The measures of nonlinear interdependency in EEG have become in the last years an emerging field and they have been applied to analyse perceptual processes, cognitive tasks and disorders (Glass, 2001; Stam, 2005). Even when these have been applied in virtual reality studies (Baumgartner et al., 2006; Kober et al., 2012; Zhao et al., 2009), to our knowledge we present the first evidence of their influence in immersive virtual emotional experiences. In future studies, the correlations between emotions in VR and the synchronization of each brain region should be analysed in depth, since in this exploratory study we use PCA as a feature reduction method in order to perform classification models to test the hypotheses.

The study has some limitations that might affect the physiological responses. We use wearable sensors which allow us to undertake research in real-world, but they have a limited number of sources. In addition, the recording could be affected by artefacts, especially those caused by movement. Although many researches use EEG data gathered in combination with HMDs (Zhang et al., 2017), or wearables in real-world (Debener et al., 2012), these set-ups need to be further examined and improved for use in future research. In addition, the real and the virtual environments have intrinsic differences in unavoidable physical features such as light, colour and complexity, and these may affect physiological responses. Furthermore, the time of the exploration of each room/piece of art need to be considered since it can radically condition space perception and therefore alter the emotion evoked. The navigation behaviour will be analysed in future research.

Concerning the level of emotion recognition, we present the first study that develops an emotion recognition system in a 3D IVE, comparing results with the physical environment model. Firstly, we presented the results of the models in the physical and virtual museum without the IAPS or Emotional Rooms dataset, using features extracted from EEG and HRV series gathered from wearable sensors, and properly combined through nonlinear SVM algorithms. The models were validated using LOSO cross-validation, which has been extensively performed in emotion recognition research to validate models. The accuracies of the model in the physical museum without IAPS or Emotional Rooms datasets achieve 68.05% in arousal and 74.30% in valence, both balanced in confusion matrix. These results are considerably higher than the level of chance, which is 58% in statistical assessment classification with brain signals (p=0.05, n>100, 2-classes) (Kim & André, 2008; Koelstra et al., 2012; Lin et al., 2010). The accuracy of the model in the virtual museum, without including IAPS or Emotional Rooms datasets, is 67.10% in valence and is balanced. However, the model of arousal in the virtual museum does not exceed the balance threshold (>60% of true high and true low), invalidating its accuracy. Therefore, the 3D IVEs show an initial limitation for use in evoking stimuli in emotion recognition systems, especially in arousal recognition.

The emotion stimuli habitually applied in the methodologies of affective computing studies, such as IAPS, include a large number of stimuli to elicit a wide range of emotions with different levels of intensity. This wide range of moods allows the emotion recognition systems to improve their accuracy. However, real-world environment (physical or simulated) stimuli are not created to evoke different ranges of valence and arousal and cannot cover different mood intensity. Thus, the responses to a set of controlled emotional stimuli are included in the emotion models to test if they improve the accuracy of the models. Thus, we analyse the addition of datasets of pre-performed controlled, standardized stimuli which are designed to evoke a range of arousal and valence, including 2D pictures (IAPS) and 360º IVEs (Emotional Rooms).

As can be seen in Table 3.2, accuracy improves in all models when using IAPS or Emotional Rooms information, supporting H3. Regarding the physical museum, the IAPS and Emotional Rooms datasets provide better accuracy in terms of arousal (71.52%), increasing the accuracy by 3.47% in both cases. The inclusion of IAPS datasets maximizes recognition in terms of valence, achieving 77.08%.

Therefore, the inclusion of IAPS works slightly better than the Emotional Rooms phase in physical environments. Regarding the virtual museum, the Emotional Rooms dataset provides better accuracy in terms of arousal (75.00%). In this case, the Emotional Rooms provide 4 points of accuracy more than IAPS and the museum dataset doesn´t achieve a balanced result. The Emotional Rooms dataset also provides better accuracy in terms of valence in the virtual museum (71.05%). The good performance related to the inclusion of the Emotional Rooms dataset in the virtual museum could be because the 360º IVEs provide important information for the recognition of arousal in 3D environments, because both use an HMD. Moreover, the initial accuracy limitation of the model with no previous data is exceeded with the inclusion of the Emotional Room dataset. Thus, a prior phase with 360º IVE controlled stimuli is shown as a powerful methodology to develop emotion recognition models in 3D IVEs. Therefore, the results support H2, since the physiological signals allow us to predict the self-assessment in both cases. In future experiments, these results could be optimized using alternative machine learning algorithms and multivariate signal analyses (Colomer et al., 2016) and a confirmatory analysis need to be performed to corroborate the hypotheses stated.

Although the arousal and valence self-evaluations were to some extent similar in the virtual and real museums, the two conditions appear to be different in terms of psychophysiological parameters. Moreover, the psychophysiological-based emotional classifiers, virtual and real environments, although they had similar performances with high accuracy, used different features and were affected differently by the introduction of features acquired in stage 1 and 2. Interestingly, the classifier for the virtual environment needs less features than the classifier for the real museum, which suggests that the psychophysiological reaction in the latter was more complex than the former. Our data, therefore, highlights a possible limitation of the application of the circumplex model of emotions to psychophysiological data, since similar subjective experiences (in terms of arousal and valence) did not show unique psychophysiological patterns. For instance, the model does not take in account where the emotion takes place: a VR environment is necessarily unfamiliar and the degree of familiarity does not follow a linear relationship with the similarity to reality (for instance, see (de Borst & de Gelder, 2015) for a detailed review of the uncanny valley phenomenon and related issues). Understanding reality in its context is analysed by the Theory of Mind (ToM) and several models suggest that the ToM may modulate emotional perception (Mitchell & Phillips, 2015): even for phobias and their treatment, patients tend to prefer VR because they are cognitively aware that the phobic stimulation is similar but not identical to the real scenario (Powers & Emmelkamp, 2008). Recently, an uncanny-valley of the mind reaction was theorized to describe a scenario where VR agents performed in a very similar (but not identical) emphatic way (Stein & Ohler, 2017). Similarly, it is possible that being in an environment which is very similar (but not identical) to a real environment will elicit a sense of eeriness. Such a sense of eeriness may interact with psychophysiological responses, but to a lesser extent with the arousal-valence subjective evaluation. It is possible to imagine that, by introducing further dimensions, such as emotional embodiment (Critchley, 2009; Niedenthal, 2007) or emotional presence, to the circumplex model of emotion may overcome the current limitations of the model. Regarding emotional presence there are several pieces of evidence that suggest how vividness of emotional experience can affect arousal and valence. For instance, patients with Post Traumatic Stress Disorder (PTSD) report very vivid traumatic emotional memories with high arousal and negative valence. On the contrary, techniques designed to reduce the vividness of such memories also reduce arousal and valence (Leer et al., 2014). Finally, our results may also be explained by reference to constructed emotion theory (Barrett, 2017). According to this theory, emotions are predictive and not reactive systems, therefore they depend on what the brain/mind considers the most probable outcome in terms of previous knowledge and sensorial input. Emotional labelling, as we know it, is just an approximation to something similar we have experienced in the past and therefore is not particularly reliable. It is not, therefore, unexpected that VR and real museum experiences are subjectively similar, but different in terms of psychophysiological correlates. Future studies might test the fit of constructed emotion theory to VR data. In this sense, a switch of paradigm may be needed. For instance, as proposed for the psychophysiological correlates of mental disorders (Gentili, 2017), we might adopt a data driven approach, based on unsupervised learning algorithms, to identify hidden similarities in psychophysiological reactivity to emotional states.

At a methodological level, the proposed signal processing and machine learning techniques using data from healthcare wearables provide satisfactory levels of recognition, achieving accuracies over 70%. They are presented as a powerful software and hardware equipment to extend the applications of emotion recognition systems including physical real-world environments, and they are in accordance with recent studies using HRV (Jo et al., 2017) and EEG (Hassib et al., 2017). However, concerning signal recordings, even when the physical environments allow the analysis of the real impact of one specific environment, these also present the following limitations: i) it is difficult to keep ambient features constant; ii) it is difficult, or even impossible in some cases, to change some environmental features in order to analyse their impact; iii) the extra-cost of developing studies of environments situated far distant and; iv) it is impossible to analyze the impact of an environment before it is constructed. On the other hand, the capacity of virtual simulation to evoke the same emotions as physical environments could be essential in the near future, taking into account the rise of virtuality and the central role that emotion plays in many background processes.

Moreover, the capacity of IVEs to be used as stimuli could significantly improve the application of emotion recognition in simulated real-world tasks.

Some possible caveats should be mentioned. This exploratory study aimed at investigating human psycho-physiological patterns of emotions during a free exploration of virtual and real art museums. We used wearable sensors allowing to translate our research to real scenarios, although such sensors a limited number of physiological sensors. In addition, the biosignals could be affected by artefacts especially caused by head movement in the case of virtual museum, and by walking in the case of real museum where we recorded biosignals "in the wild", i.e. outside of the highly constrained and tightly controlled laboratory paradigms. This is especially true for the EEG series, although many researches have successfully employed such data in combination with HMDs or other wearable devices in naturalistic conditions (Debener et al., 2012; Marín-Morales et al., 2018; Zhang et al., 2017). Nonetheless, our results point to the significance of brain synchronization for the emotion recognition in both real and virtual museum scenarios. The psychological self-assessment was performed using retrospective reports, leading to possible bias such as recency, primacy and memory, although our experimental paradigm replicates a real scenario. Note also that the user's emotional perception could be biased by stopping the real or virtual museum exploration. The real and the virtual environments have intrinsic differences in unavoidable physical features such as light, colour and complexity, and these may affect physiological responses. Furthermore, the time of the exploration for each room/piece of art would need to be considered as a confounding/critical factor in future studies because of its possible role in evoking emotions. In particular, it could affect the real vs virtual museum discrimination in case of differences in the time of exploration. In this regards, we recently found that Room 1 and Room 2 of the virtual museum are associated with lower time of visit than the real exhibition (Marín-Morales et al., 2019). On the other hand, the other 6 stimuli do not show differences in terms of exploration time between real and virtual museums.

This study marks new steps in the discipline of affective computing and its application to environmental physiology and other fields, providing evidence through psychological and physiological comparisons during an emotional experience in real and virtual environments. This exploratory study tries to contribute to overcome passive methods' limitations of affective elicitation classically used in emotion recognition models, such as pictures, sounds or videos, supporting the use of VR in emotion elicitation. The methodology has implications at commercial and research levels in many disciplines as health, architectural design, urban planning and aesthetics. It could be applied to study the emotional responses of subjects in many specific environments, such as hospitals, schools and factories, where the emotional responses of users play a critical role in daily wellbeing. More specifically, new emotion recognition models will strongly contribute to the development of ambient assisted living, smart environments that change depending on human responses. On the other hand, the new VR set-up allows the analysis of the influence of one parameter, changing it while maintaining the remainder of the environment in a steady state. This will help to develop many studies, impossible to undertake in real environments for physical reasons (e.g. architectural modification of spaces) or security reasons (e.g. phobias therapy). Moreover, it will allow the analysis of environments before their construction, helping in the decision-making process of creating new environments oriented to wellbeing.

Supporting information

IAPS experimental protocol

Table 3.4 shows the rating of the images used in each session of the experimental protocol of Stage 1.

Arousal level	Valence level	N pics.	Valence rating	Valence range	Arousal rating	Arousal range
N	N	6	4.99 ± 0.14	4.82 ÷ 5.21	2.80 ± 0.26	2.35 ÷ 3.03
A1	V1	7	4.26 ± 0.33	3.71 ÷ 4.75	3.59 ± 0.22	3.31 ÷ 3.88
A1	V2	6	5.41 ± 1.07	4.38 ÷ 6.50	3.67 ± 0.17	3.53 ÷ 3.95
A1	V3	7	7.56 ± 0.51	6.81 ÷ 8.11	3.58 ± 0.28	3.20 ÷ 3.92
A2	V1	7	3.06 ± 0.71	2.09 ÷ 4.02	4.71 ± 0.22	4.28 ÷ 4.97
A2	V2	6	5.21 ± 1.20	4.09 ÷ 6.40	4.62 ± 0.31	4.20 ÷ 4.94
A2	V3	7	7.70 ± 0.70	6.45 ÷ 8.30	4.45 ± 0.22	4.07 ÷ 4.62
A3	V1	7	2.34 ± 0.58	1.84 ÷ 3.17	5.94 ± 0.22	5.65 ÷ 6.20
A3	V2	6	4.80 ± 1.52	3.27 ÷ 6.31	5.54 ± 0.19	5.35 ÷ 5.90
A3	V3	7	7.57 ± 0.60	6.53 ÷ 8.06	5.69 ± 0.26	5.27 ÷ 5.96
A4	V1	7	2.01 ± 0.55	1.29 ÷ 2.70	6.61 ± 0.26	6.29 ÷ 6.94
A4	V2	6	4.90 ± 1.71	2.96 ÷ 6.62	6.78 ± 0.16	6.55 ÷ 6.99
A4	V3	7	7.44 ± 0.33	6.87 ÷ 7.88	6.81 ± 0.53	6.23 ÷ 7.39

Table 3.4. Rating of IAPS images used in Stage 1

Physiological signal segmentation and synchronization

Previous controlled stimuli

In Stage 1, the IAPS images and the physiological signal recorded were synchronized using the software iMotions (iMotions A/S, Denmark). The segmentation of IAPS is explained in the protocol. In order to use the same number of blocks as in the emotional rooms, we used, for the classification model, only the four outermost blocks, combining low arousal (A1), high arousal (A4), negative valence (V1) and positive valence (V3). Each block was considered as an independent stimulus and had a duration of 70 seconds.

Regarding Stage 2, the software allowed the insertion of live markers in the physiological signal recording. The researcher inserted a live marker when each room began to display in the Samsung Gear HMD, so each room (stimuli) was associated with a specific live marker. Then, each time window of each stimulus was segmented using these markers. In conclusion, in the IAPS and Emotional room stages, a subject had four stimuli and each was theoretically situated in one quadrant of the CMA, providing a controlled stimulus set that included all arousal and valence combinations.

Physical museum exhibition

The physiological signals were recorded using iMotions software (iMotions A/S, Denmark), running in the laptop carried by the subjects. In order to record the positions of the subjects, we used a GoPro camera Figure 3.11. To synchronize the video with the physiological signals, we needed a synchronization point. When the signal and video recording started, the camera was focused onto the laptop and the researcher inserted a live marker software. By using this marker in the signals, and the frame of the video where the researcher inserted it, we could synchronize the video and the physiological signals.

Figure 3.11. Additional environment comparison between the physical museum (left) and the virtual museum (right). The photos represent Room 2 and Room 3

To obtain the data about the navigation of the subjects from the video, we designed a tool using Microsoft Virtual Studio in C++ language. The software simultaneously showed two items: a video of the recorded exploration of the exhibition and a plan of the exhibition. The device includes two buttons to advance and rewind the video with 1-second jumps. In addition, it allows us to enter the position of the subjects in each frame of the plan through "clicks", using the video as a reference. The researcher watched the videos and positioned the subjects in the plan at one second intervals. Finally, the navigation was saved in a file with the route sampled every second. The timeline of the video was synchronized with the physiological signal timeline using the previously inserted live marker.

8 stimuli were defined, 5 when the subjects visited each room and 3 when they viewed the pieces of art. A visit to an area starts when a subject enters the area and finishes when the subject leaves the area. The area of each room was defined by its walls and the area of each piece of art was defined by an area of influence, shown at Figure 3.12. If an area has only one visit, the stimuli are defined by the time taken to visit the area. If the subject makes more than one visit to an area and the time between the visits is less than 15 seconds, the visits were merged and considered as a single visit. After this pre-process, if there were more than one visit to the same room/stopping point, the stimuli were defined by the longest visit. In addition, a visit needed to have a duration of at least 40 seconds to be considered as a valid stimulus.

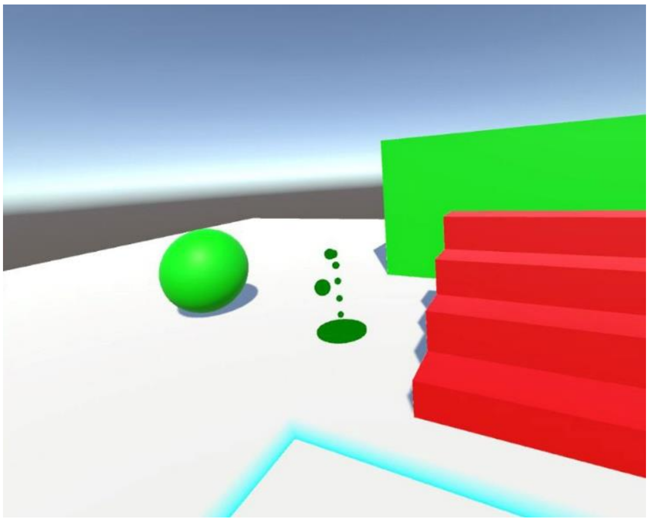

Figure 3.12. Screenshot of the training environment

Virtual museum exhibition

The physiological signals were recorded using iMotions software. A script in Unity was developed and inserted into the scenario with 2 tasks: (1) recording the position of the subject in each frame of the virtual environment and exporting it in a csv; (2) sending a live marker automatically to the iMotions software when the scenario and the recording of the position started, allowing synchronization between the navigation and the physical signal.

The navigation recording was resampled to the same frequency as the physical environment navigation (1 Hz). Following this, applying the same methodology, we defined the stimuli in exactly the same way as for the real museum.

Real vs. Virtual Classification

Regarding the real-virtual classification, a dataset was created concatenating the features of the stimuli in the physical museum (Stage 3.1) and the virtual museum (Stage 3.2), mixing virtual and physical stimuli in the same dataset. The output of this dataset is the nature of the stimulus (real or virtual). Thus, the pattern recognition classifier algorithm tries to recognize if the stimuli are virtual or real, by seeking to analyse which features enable the classifier to determine the nature of the stimuli.

Chapter 4

Navigation comparison between a real and a virtual museum: time-dependent differences using a head mounted display

Marín-Morales, J., Higuera-Trujillo, J. L., de-Juan, C., Llinares, C., Guixeres, J., Iñarra, S. & Alcañiz, M. (2019). Navigation comparison between a real and a virtual museum: time-dependent differences using a head mounted display. *Interacting with Computers*, 31(2), 208-220.

Abstract

The validity of environmental simulations depends on their capacity to replicate responses produced in physical environments. However, very few studies validate navigation differences in immersive virtual environments, even though these can radically condition space perception and therefore alter the various evoked responses. The objective of this paper is to validate environmental simulations using 3D environments and head-mounted display devices, at behavioural level through navigation. A comparison is undertaken between the free exploration of an art exhibition in a physical museum and a simulation of the same experience. As a first perception validation, the virtual museum shows a high degree of presence. Movement patterns in both “museums” show close similarities, and present significant differences at the beginning of the exploration in terms of the percentage of area explored and the time taken to undertake the tours. Therefore, the results show there are significant time-dependent differences in navigation patterns during the first 2 minutes of the tours. Subsequently, there are no significant differences in navigation in physical and virtual museums. These findings support the use of immersive virtual

environments as empirical tools in human behavioural research at navigation level.

Introduction

Environmental simulations are representations of physical environments that allow researchers to compare reactions to common concepts (Kwartler, 2005). They are particularly important when what they depict cannot be physically represented. Therefore, they are widely employed in different areas related to human behaviour. Similarly, the emergence of virtual reality has generated a wide range of possibilities, both at the scientific and the commercial level.

Virtual reality allows the development of environmental simulations in which users can perform as if they are in the real world (Alcañiz et al., 2004). These simulations have a great variety of set-ups, involving a combination of formats and supports (Mengoni et al., 2011). They have been progressively integrated into studies as the relevant technologies have evolved. On the one hand, among the formats - understood as the codification standard - photography and 3D environments are highlighted. Photographs, including panoramic images, provide us with non-interactive visual representations, whereas 3D environments can generate interactive representations. On the other hand, display devices - the technological devices used to visualize the formats - can be classified according to their capacity to isolate the user from physical reality (Rangaraju & Terk, 2001), also known as immersion. Immersion is defined as the objective level of fidelity that a virtual reality system provides and, while it is related to human perception, it is inherent in each technology (Slater, 2003). Thus, virtual reality can be displayed by: non-immersive systems, usually single-screen, such as desktop PCs; semi-immersive, surround-screen systems, such as the cave automatic virtual environment (CAVE); and fully-immersive systems, such as the head-mounted display (HMD).

The environments displayed through immersive devices are called immersive virtual environments (IVE) (Blascovich et al., 2002). Today, the tendency is to use virtual reality environments through immersive displays. Their synergy offers a higher sense of presence (Sanchez-Vives & Slater, 2005), understood as the illusion of 'being there' (Steuer, 1992). While these set-ups were, in the past, difficult to implement, they are now much more accessible (Parsons, 2015) and have improved performance. Furthermore, the progress in virtual reality set-ups has led researchers to involve the human body in simulation experiences: introducing full embodiment in a virtual environment, understood as the sense of using our body coherently as we do in the real world (Dourish, 1999), enhances sense of presence by the incorporation of natural interactions.

Virtual reality is being increasingly used in the area of natural phenomena and social interaction simulation, due to its ability to activate brain mechanisms that are similar to those in real life (Alcañiz et al., 2009). Since VR allows the measurement of performance in real-time, it has become an important investigative tool in the field of human behaviour research. Specifically, VR is widely employed in psychological assessment (Freeman et al., 2017), medical treatment (Dascal et al., 2017), education (Jensen & Konradsen, 2017; Babu et al., 2018), emotion recognition (Marín-Morales et al., 2018) and architecture (Portman et al., 2015), among other areas.

The usefulness of simulation for human behaviour research has been analysed through the concept of validity: the capacity to evoke a response from the user in a simulated environment similar to one that might be evoked by a physical environment (Rohrmann & Bishop, 2002). Comparisons between physical spaces and their simulations through 3D IVEs have been made at different levels. At a physiological level it is found that 3D IVEs evoke responses more similar to those elicited by physical environments than formats with lower interactivity, although at the psychological level the validity decreases when compared to other formats, due to its lower realism (Higuera-Trujillo et al., 2017). In addition, measured by psychological response, a relation is found between sense of presence and the immersive capacity of HMDs (Baños et al., 2004) and the navigation metaphor (Usoh et al., 1999). Other studies have carried out comparisons between real and virtual spaces, analysing user performance in sets of everyday office-related activities (e.g. reading texts and identifying objects in an office environment) (Heydarian et al., 2015), physiological responses in different thermal conditions (Yeom et al., 2017), subjective perception of daylit spaces (Chamilothori et al., 2018) and orienteering tasks (Kimura et al., 2017).

Navigation encompasses travel and wayfinding components (LaViola et al., 2017). On the one hand, the travel function is related to the task of moving from one point to another and, therefore, to the metaphors employed for executing displacements. On the other hand, wayfinding is the cognitive process of establishing a route or path from an origin to a destination.

Regarding the travel component, interaction and navigation techniques are especially important aspects. They may, even, influence sense of presence (Slater & Usoh, 1994; Usoh et al., 1999). Different metaphors are discussed in terms of their efficiency in navigation. The benefits of full free-physical motion are frequently emphasized, although there is no consensus as to which is the standard method of navigation (Lee et al., 2018). However, it is suggested that performing only body rotations might be useful: users adopt other strategies to increase the surface that they cover (Riecke et al., 2010). Other metaphors, such as those based on head motions to indicate forward movements, have also been studied with relative success (Tregillus et al., 2017). In addition, there seems to be consensus that the exclusive use of joysticks is not the most efficient navigation metaphor (Wilson et al., 2014), despite its familiarity. However, in terms of their ability to generate a user navigation experience similar to that of a physical space, no research intensively compares different metaphors and navigation devices in a controlled environment.

The wayfinding component is fundamental to the user's performance. This applies equally to virtual and physical environments, as the knowledge acquired in both has a similar structure (Ruddle et al., 1997). Hence, for example, wayfinding component analysis has been used successfully for firefighting training (Bliss et al., 1997). Several studies compare virtual and physical environments, generally concluding that the virtual offers worse performance than the physical (Richardson et al., 1999; van der Ham et al., 2015). Some authors claim that these differences can be attributed to a lack of user involvement caused by technical limitations (Lessels & Ruddle, 2005), fundamentally of the display (as field of view, or photorealism) and navigation (as metaphors, or degrees of freedom) systems. These studies focus on non-immersive or semi-immersive systems. However, very few studies compare physical space and its virtualization displayed through immersive systems. Taking into account the growing use of IVEs, and the fact that HMDs offer better performance - in terms of speed of navigation - than desktop screens (Ruddle et al., 1999), there is a clear need to make this comparison using the most modern HMDs.

The present work addresses these limitations. Specifically, the objective is to validate environmental simulations by means of 3D IVEs at a navigation level. The research question to be answered is: are there differences between navigation in a physical space and its virtualization using a 3D IVE and a latest generation HMD? A comparative study was conducted of a free exploration of an art exhibition in an actual museum and a virtual museum simulated by means of an HTC Vive. The results may be of interest to researchers and content developers and are applicable to different fields.

Material and methods

Participants

A homogeneous group of 60 healthy subjects (age 28.9 ± 5.44, 40% men, 60% women) were recruited. They all received financial compensation. The criteria were as follows:

(i) Age between 20 and 40
(ii) Spanish nationality
(iii) Not having formal, or informal, training in fine arts
(iv) Not having previous experience of HMDs
(v) Not having previously visited this particular exhibition

(vi) Having normal vision, or corrected to normal with contact lenses

Two questionnaires were included to ensure that the subjects had healthy mental conditions and homogeneous emotional responses. First, the subjects were analysed by the Patient Health Questionnaire (PHQ) (Kroenke et al., 2001). Only subjects with a value inferior to 5 were included to avoid including individuals in states of depression. Second, a selection of IAPS images (Lang et al., 1997) were evaluated by the participants using the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994). The images had a range of 3.41 to 7.24 in arousal and of 1.29 to 8.17 in valence (selected images: 7234, 5201, 9290, 1463, 9181, 8380, 3102, 4652). The participants' self-assessments were normalized by means of a z-score using the means and deviation published in the IAPS. Participants with evaluations outside of the range -2.58 and 2.58 (α=0.005) were excluded since they were considered outliers.

Physical museum

During the first phase of the study, 30 subjects visited an actual museum to perform the test. The Institut Valencià d'Art Modern (IVAM) offered its facilities for the study.

The exhibition "Départ-Arrivée" by Christian Boltanski was selected due to its high emotional content, since its setting is the Nazi holocaust and because it was spacious enough to allow users to freely navigate. It consisted of five rooms with an approximate total floor surface area of 750 m² (Figure 4.1). Each room is considered to be a single piece of art. In addition, the last room contained three art pieces that could be analysed independently. Furthermore, the rooms presented information boards with the artist's notes on the works. Finally, in Room 3 there was a path laid out from which subjects could not deviate.

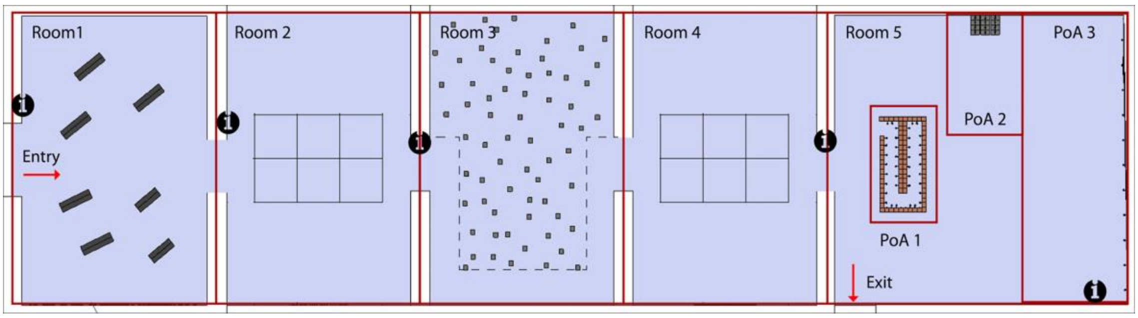

Figure 4.1. Plan of the art exhibition. Circles with an '"i" represent the artwork information boards. In room 3, the dashed line represents a limit that could not be crossed by the subjects

The subjects were told, before starting the experiment, that they could freely explore the first four rooms. In the last room, while they also could explore it freely, they had also to view, in detail, the three pieces of art in the room. The researcher waited for the subject at the exhibition exit, allowing the subjects to explore the space without any external influence.

A GoPro camera was used to record the subjects' navigation. The subjects carried this attached to their chests by means of a harness (Figure 4.2).

Figure 4.2. Example of a subject in the physical museum

Virtual museum

During the second phase of the study the museum was virtualized using a 3D scenario. For this we used the Unity 5.1 game engine (www.unity3d.com). In order to achieve a scenario with maximum realism, we imported a three-dimensional copy of the exhibition created by Rhinoceros v5.0 and textures partially derived from the physical environment. This process required the exhaustive and methodical drawing and photographing of the whole exhibition. A team of architects visited the physical exhibition and carried out a validation of virtualization at a general level, and of the level of lighting and texturing. Virtualization was considered complete after the appropriate changes. Figure 4.3 shows photographs and screenshots of the virtual environment.

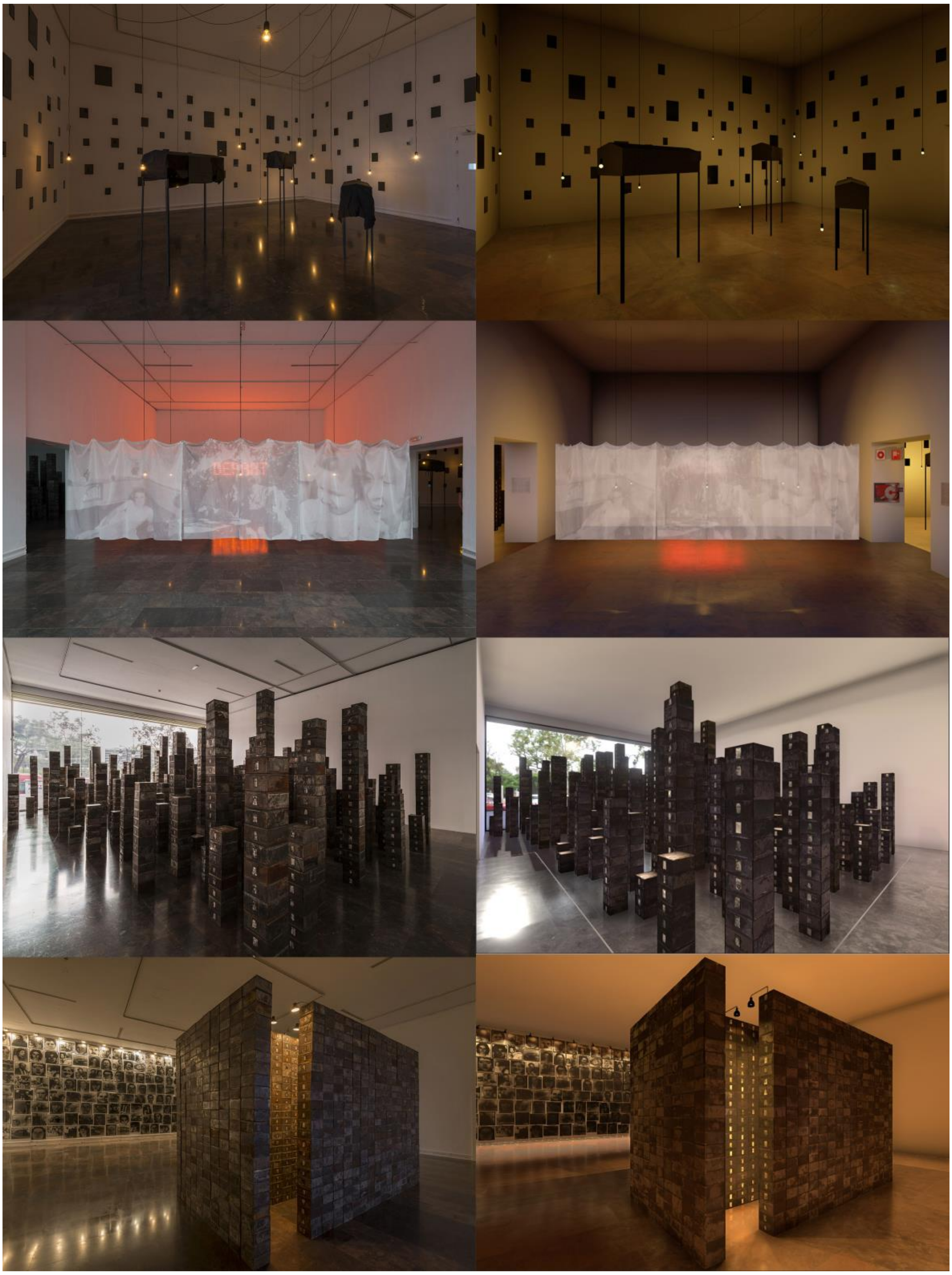

Figure 4.3. Comparison between a photograph of the physical museum (left) and a screenshot of the virtual museum (right). From top to bottom, Room 1, Room 2, Room 3 and Room 5

For the simulation of the 3D VR, we compiled the scenario for HTC Vive (www.vive.com), which enabled us to carry out visual and displacement simulations. Visualization was conducted by means of an HMD with 2160x1200 pixels (1080×1200 per eye), and a field of view of 110 degrees working at 90Hz refresh rate. On the other hand, we conducted the displacements by means of a tracking technology made up of two controllers and two base stations that, together, enabled the subject to interact with the environment and physically move within an area of a 2x2 metres. Specifically, the metaphor used was the teleport navigation metaphor incorporated into the HTC Vive with a maximum teleportation radio of 2.5 metres from the subject. This was selected because we hypothesize that it will allow us to achieve pseudo-naturalistic navigation. The equipment was connected to the research computer (Predator G6, www.acer.com) by means of a DisplayPort 1.2 and USB 3.0 and ran smoothly and without interruption.

After the environmental simulation of the art museum had been created, the study was replicated using the 3D IVE with the second group of 30 subjects in a lab environment. Figure 4.4 shows a subject exploring the virtual museum.

Figure 4.4. Example of a subject viewing the virtual museum

Before starting the experiment, the subjects carried out several tasks to adapt themselves to the HMD device and to the navigation metaphors, in a neutral scenario, without textures. The researcher ensured that the subject adapted correctly and navigated fluently. The subjects could stay in this scenario for as long as they wanted, until they considered that they could use the device without difficulty. During this period, the researcher addressed any doubts raised by the subjects about the HMD device, given that they had no previous HMD experience. Figure 4.5 shows a screenshot of a training environment.

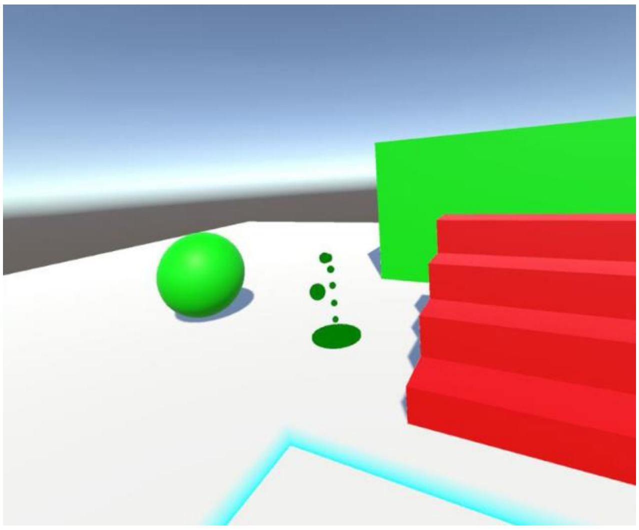

Figure 4.5. Screenshot of the training environment

After the training, the researcher gave the subjects the same instructions as were given to the subjects who had visited the physical museum. During the virtual experiment, the researcher remained in an adjoining room. Finally, the researcher was able to note through the monitor when the subject arrived at the exit of the "museum" and then removed the HMD device. Following the test, the subject had to answer a presence questionnaire, the "SUS questionnaire" (Usoh et al., 2000). This consists of six items assessed from 1 to 7 using a Likert scale to measure three aspects of presence:

(i) The experience of being inside the simulation.
(ii) The consideration of the simulation as the dominant reality.
(iii) The memory of the simulation as a place.

Signal synchronization

Regarding navigation in the physical museum, software was developed using the "Microsoft Virtual Studio" in C++ language to synchronize the data. The software simultaneously shows two items: a video recording of the subjects' exploration of the exhibition and a plan of the exhibition. It includes two buttons that can advance and rewind the video with 1-second jumps. In addition, it allows the position of the subjects to be manually entered into the plan, using the video as a reference. The researcher reviewed all the videos, positioning the subjects in the plans at 1-second intervals. Finally, the navigation path was saved to a file with the route sampled every second.

Regarding navigation in the virtual museum, a script in Unity was developed which recorded the subjects' positions at a frequency of 7Hz while they were exploring the scenario and exported them to a csv file at the end of the test.

Finally, the recorded navigation path was resampled to the same frequency as the path generated in the physical space (1 Hz).

Spatial segmentation and analysis

The analysis of the subjects' navigation was based on the framework developed by Marín-Morales et al. (Marín-Morales et al., 2017). This proposes the segmentation of space into Areas of Interest (AOIs) on which several indicators are calculated to characterize navigation. The exhibition is comprised of twenty-three AOIs. The first five AOIs are defined by the area of the five rooms. Moreover, each room was divided into several internal AOIs. Rooms 1, 2 and 4 are divided into four symmetrical AOIs. Room 3 is divided into three AOIs covering the walking area. Room 5 is divided into three AOIs, each of them including a specific piece of art. The analysis was carried out based on three items: the heatmaps, the percentage of area explored and the length of time of the visits.

Heatmaps were created using every point of the subjects' trajectories at 1 Hz. Subsequently, a radius of 0.75 m was applied to each position, defining that each subject's presence spans a circle of 1.5m in diameter. Considering that heatmaps are usually relative to themselves, i.e. they adapt the colours to the maximum and minimum values that they represent in each case, both heatmaps were constructed according to the same linear representation scale, allowing them to be comparable between themselves. The highest valued 5% of the pixels were dismissed and were saturated in red to increase the sensitivity of the heatmap. On the other hand, we calculated the percentage of area explored in each AOI by each user, considering that the area explored is calculated with the centroids of the subject's navigation points with a radius of 0.75 metres.

Regarding the length of the visits, a visit is defined as the period of time from when a subject enters an AOI to the moment he or she leaves it. In particular, the variable being analysed is the length of time of the main visit to each of the AOIs, defined as that visit with the longest duration in the case that an AOI was visited more than once by the same subject. Processing assured that the variable included the main visit to the room or piece of art: if there were less than 15 seconds between two visits to the same AOI, the visits were put together and considered as just one visit.

Results

Self-assessment: presence and cybersickness

Table 4.1 shows the results of presence provided by the SUS questionnaire. Two items are between 6 and 7: "I had a sense of 'being there' in the museum space" and “During the experience I often thought that I was really in the museum space”. Another two items are between 5.50 and 6: “There were times during the experience when the museum space was the reality for me" and "During the experience you felt you were in the museum space”. Finally, the two remaining items are below 5. The total average of the set of items is 5.47 out of 7 so the level of presence of the simulation is high. The subjects did not report any level of cybersickness.

Question	Score
1. I had a sense of ‘‘being there’’ in the museum space	*6.17 (0.95)*
2. There were times during the experience when the museum space was the reality for me	*5.86 (0.95)*
3. The museum space seems to me to be like somewhere that I visited before	*4.87 (1.91)*
4. During the experience you felt you were in the museum space	*5.87 (1.17)*
5. I think of the museum space as a place similar to other places that I’ve been today	*3.93 (2.26)*
6. During the experience I often thought that I was really in the museum space	*6.13 (0.94)*

Table 4.1. Results of presence (SUS Questionnaire). The results are presented using the mean and the standard deviation

Heatmaps

Figure 4.6 shows the heatmaps of the trajectories in the physical and virtual museums. Carrying out a descriptive analysis room by room, it is observed that:

(i) In Room 1, the exploration is more dispersed in the physical museum. In addition, the participants are very focused on the information board in the physical museum.

(ii) In Room 2, the trajectories are similar. However, taking into account that the natural path is that presented in the physical museum, the subjects´ trajectories deviated slightly from these natural paths in the virtual museum, being a bit more dispersed. Similarly, the information board was scarcely visited in the virtual museum.

(iii) In Room 3, the trajectories are very similar and there are no differences, except that some subjects ignored to a small extent the limitations set by the exhibition organizers.

(iv) In Room 4, there are no differences among the trajectories, except for the same effect mentioned for Room 2.

(v) In Room 5, there are no differences, except for the trajectory relevant to piece of art 3, where the subjects notably stopped at the information board in the physical museum, whereas the trajectories were much more dispersed in the virtual museum. It is noteworthy that, both in the physical and the virtual museum, a light spot can be seen in the middle of the room, caused by the subjects who visualised the room from its central point.

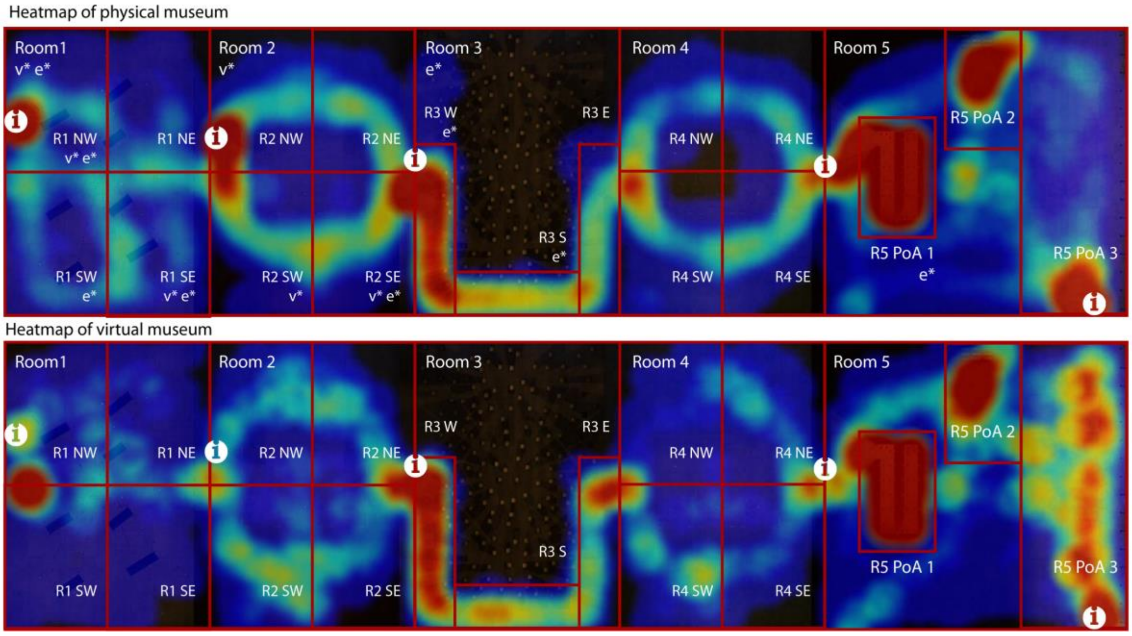

Figure 4.6. Heatmaps of the physical museum (top) and the virtual museum (bottom). They include the names and positions of the AOIs, and the positions of the information boards, indicated by an "i". Below the name of the AOI, v indicates significant differences in visit time and e* significant differences in percentage of area explored with p < 0.05 in each AOI. The indicators are included in the heatmap condition with major value in each case*

Length of visit time and percentage of area explored

Figure 4.7 shows the length of time of the main visits and the percentage of area explored in each room in the physical and virtual museums, using mean and standard deviations. Due to the Gaussianity of data ($p > 0.05$ from Shapiro-Wilk test with null hypothesis of having a Gaussian sample), a T-Test was applied. Regarding the visit times, only in Room 1 (p-value=0.00) and Room 2 (p-value=0.03) were there significant differences, subjects staying less time in the virtual museum. The percentage of area explored is higher in the physical museum for all rooms, significant differences being observed in Room 1 (p-value=0.00) and Room 3 (p-value=0.00).

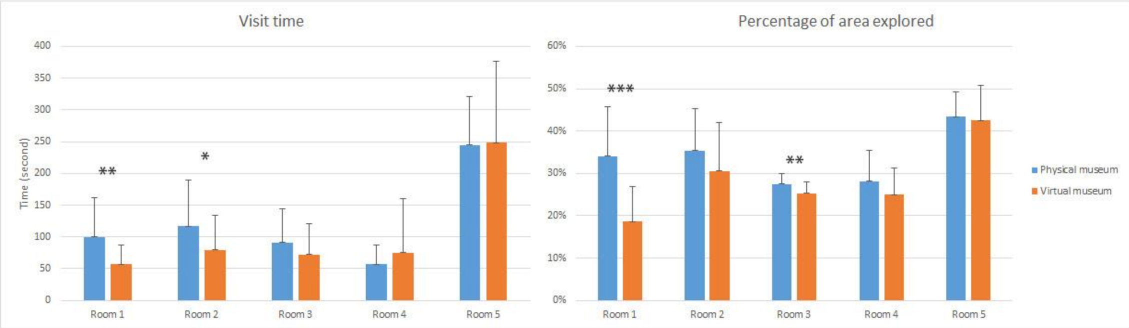

Figure 4.7. Representation of timings for the main visit and the percentage of area explored for each AOI. The bars represent the average and the lines represent the standard deviation. (indicates significant differences with p<0.05, ** indicates significant differences with p < 0.01 and *** indicates significant differences with p < 0.001)*

Table 4.2 shows the visit time and the percentage of the area explored of each internal AOI and the difference between both conditions. Due to the Gaussianity of data ($p > 0.05$ from Shapiro-Wilk test with null hypothesis of having a Gaussian sample), a T-Test was applied. Considering the visit time, the main differences were found in Room 1 (R1 NW and R1 SE) and Room 2 (R2 SW and R2 SE). The percentages of area explored were different in Room 1 (R1 NW, R1 SW and R1 SE), Room 2 (R2 SE), Room 3 (R3 W and R3 S) and Room 5 (R5 PoA 1).

		Visit time				Percentage of area explored			
Room	**AOI**	**Physical museum**	**Virtual museum**	**Difference**	**pvalue**	**Physical museum**	**Virtual museum**	**Difference**	**pvalue**
Room 1	R1 NW	**51.93 (56.71)**	**20.51 (18.89)**	**+31.43**	**0.007 (**)**	**31.83% (4.51)**	**27.24% (4.27)**	**+4.59%**	**0.000 (***)**
	R1 SW	25.10 (37.29)	22.42 (29.15)	+2.68	0.562	**32.12% (19.03)**	**13.64% (11.81)**	**+18.48%**	**0.000 (***)**
	R1 NE	32.17 (54.53)	17.84 (18.12)	+14.33	0.906	31.28% (18.78)	24.62% (14.89)	+6.66%	0.147
	R1 SE	**45.63 (56.76)**	**8.74 (20.17)**	**+36.90**	**0.000 (***)**	**39.23% (21.95)**	**14.09% (13.46)**	**+25.14%**	**0.000 (***)**
Room 2	R2 NW	54.23 (49.73)	37.80 (55.63)	+16.44	0.077	32.81% (15.22)	26.37% (20.32)	+6.44%	0.185
	R2 SW	**69.17 (67.84)**	**30.20 (39.86)**	**+38.97**	**0.011 (*)**	35.98% (14.59)	33.80% (20.52)	+2.17%	0.650
	R2 NE	25.57 (29.10)	33.07 (50.15)	-7.51	0.691	33.63% (14.25)	30.35% (13.90)	+3.27%	0.107
	R2 SE	**71.07 (55.00)**	**27.65 (38.33)**	**+43.42**	**0.001 (***)**	**39.34% (12.19)**	**31.94% (13.37)**	**+7.40%**	**0.035 (*)**
Room 3	R3 W	62.13 (51.61)	43.07 (48.41)	+19.06	0.118	**81.70% (5.53)**	**74.15% (10.08)**	**+7.55%**	**0.001 (***)**
	R3 S	13.40 (7.86)	14.60 (11.54)	-1.20	0.644	**78.63% (4.70)**	**74.55% (8.65)**	**+4.08%**	**0.032 (*)**
	R3 E	13.57 (6.44)	17.19 (7.19)	-3.62	0.052	71.41% (9.07)	69.27% (8.42)	+2.14%	0.365
Room 4	R4 NW	15.07 (18.88)	12.25 (20.00)	+2.82	0.229	24.55% (17.07)	19.08% (15.17)	+5.47%	0.210
	R4 SW	27.00 (30.51)	17.68 (14.70)	+9.32	0.674	33.95% (16.55)	29.31% (15.29)	+4.64%	0.281

	R4 NE	14.60 (18.04)	16.34 (14.10)	-1.74	0.686	24.19% (16.91)	25.80% (15.46)	-1.61%	0.952
	R4 SE	26.60 (26.41)	37.39 (87.28)	-10.79	0.416	30.36% (12.99)	25.44% (9.62)	+4.91%	0.114
Room 5	R5 PoA 1	76.13 (32.89)	78.44 (51.58)	-2.31	0.839	**99.98% (0.06)**	**94.87% (18.55)**	**+5.10%**	**0.007 (**)**
	R5 PoA 2	48.57 (44.62)	71.93 (96.36)	-23.36	0.840	63.03% (12.10)	57.40% (11.20)	+5.63%	0.076
	R5 PoA 3	112.90 (62.77)	115.73 (95.35)	-2.83	0.894	43.53% (11.47)	40.64% (13.23)	+2.90%	0.386

Table 4.2. Results of the visit time and percentage of area explored in each internal AOI. The results are presented using the mean and the standard deviation, the difference between both conditions and the p-value of the T-Test

Discussion

The aim of this study is to validate environmental simulations made by means of 3D IVEs and a latest generation HMD, at presence and navigation levels. The results can be discussed on four levels: i) level of presence and cybersickness, ii) differences in navigation, iii) methodological analysis and iv) comparison with previous works.

Regarding sense of presence, an overall average of 5.47 (out of 7) was shown by the SUS questionnaire. It reached more than 6 for the questions "I had a sense of ''being there'' in the museum space" and "During the experience I often thought that I was really in the museum space". The results are considered to be high, taking into account the previous results reported by studies using similar technologies (Sas & O'Hare, 2003). In addition, since no subject reported any cybersickness, the self-assessment supports the use of the present set-up at a perception level.

With regard to navigation, we analysed trajectories and the lengths of visits to AOIs. The heatmaps show similar navigation patterns in the physical and the virtual museums. Room 1 shows the biggest differences between the exploration and the heatmaps, with 15.40% more area being explored in the physical environment**.** The heatmaps do not show the same pattern. Moreover, AOIs R1 NW, R1 SW and R1 SE present significant differences. These differences may be because, in the first room, the subject is still adapting to the IVE. The second room shows no significant differences at either the exploration level or with the path patterns in the heatmap, except for AOI R2SE. This room has a central object that tends to be avoided (Figures 1 and 3). In the physical museum, the subjects walk around this object, forming a clear rectangle in their trajectories. In the virtual museum, they tend to do the same, but not so markedly. There are subtle differences in patterns when the subjects try to avoid obstacles. Room 3 shows a similar pattern. Nevertheless, there are significant differences in the percentages of areas explored, in both the general room and in the internal AOIs (R3 W and R3 S). These differences could be due to the navigation metaphor used. The teleport metaphor, which allows the user to perform "jumps" of a maximum of 2.5 metres, decreases the percentage of area explored when the subjects walk along a narrow aisle. This suggests that the navigation is less naturalistic in narrow environments. There are no significant differences among the patterns in Room 4, not even in the area explored percentage. No significant differences were observed in the last room in the heatmap patterns at the exploration or trajectory levels. Moreover, it was noted that the same pattern is seen in Figure 5 for the pieces of art, in the central point of the room and at the exit from the exhibition. For piece of art 1 (R5 PoA 1), we find significant differences, although, in both conditions, the percentage of area explored is more than 90%, so this does not represent a substantial difference. Moreover, this effect is similar to that observed in Room 3, where the percentage of area explored decreased, being in a narrow aisle, as the piece of art is a labyrinth. For piece of art 2, there are no differences. For piece of art 3, there are no significant differences in the exploration, but a more dispersed pattern is detected in the heatmap. Therefore, in terms of exploration, the major difference is observed in Room 1. In addition, differences were found for all the information boards. Current HMD technologies are limited for reading medium or small texts. This problem might be resolved by an increase in the resolution of the HMD systems.

With regard to the lengths of the visits, there were significant differences only in Room 1 (p-value=0.00) and Room 2 (p-value=0.03), including in both cases two internal AOIs (R1 NW, R1 SE, R2 SW and R2 SE). Visits to both rooms in the virtual museum were considerably shorter than for the physical museum. The rest of the rooms do not present significant differences. Thus, taking the visual analysis patterns of the heatmaps, the percentage of area explored and the length of the visits, if we exclude the effect of the narrow aisle in Room 3 and R5 PoA1, the environment shows time-dependent differences. Consequently, and despite the implementation of a training phase which used a different scenario, it is necessary to allow a period of adaptation in the actual museum scenario until navigation behaviours do not show significant differences to those in the space that it simulates. In this study, the adaptation period was approximately 2 minutes, which is the sum of the average length of visits to Room 1 and Room 2 as shown in Figure 7. This might be explained by the fact that the training room is a non-realistic environment, since it presents abstract objects without textures (Figure 5). When the subject enters the virtual museum (Figure 3), he/she is affected by the "wow" effect caused by the realism of the VR, and thus needs some time to start to behave without displaying significant differences to the real museum. In future research, we suggest that more realistic training rooms, with textures and lighting similar to the simulated environment, should be used, to try to avoid the "wow" effect. However, it must be borne in mind that these results come from subjects with no prior experience of HMDs. In the future, when more of the population have wider exposure to these devices, this adaptation period will probably be reduced or even be unnecessary.

At a methodological level, we make a comparison using a HTC Vive, which gives good performance in terms of working area, accuracy and jitter in a room-sized environment for serious games, rehabilitation and health-related applications (Borrego et al., 2018). The role that recently developed low-cost VR devices can play in scientific research is thoroughly analysed by Cipresso et al. (2018), who argue that they may be the next significant stepping stones in technological innovation. Therefore, as it is likely that these devices will be widely used in next years, there is an urgent need to validate them. Regarding the metaphor, some recent reviews analyse the role of the classic navigation metaphors in the new era of VR (Boletsis, 2017; Nilsson et al., 2018), dividing them into three main categories: repositioning systems, locomotion based on proxy gestures and redirected walking. There is consensus that redirected walking is the most natural way to simulate walking. However, there is also consensus that travel techniques must be developed to mimic better the actual experience of walking without requiring a physical space the same size as the virtual environment; this is the weakest point of redirect walking, as it requires large physical spaces. In the present work we suggest the use of repositioning systems, in particular teleport metaphors with a limit of 2.5 metres radius and a movement area of 2x2 metres, to provide pseudo-naturalistic walking. This set-up provides a low cost framework that requires only a small room with two base stations, which are included in the basic HTC Vive pack. The results support the use of this set-up, but future researches should compare this with other navigation metaphors, and also analyse which value of radius teleport limitation better simulates walking. In particular, the reduction of the teleport radius might provide more realistic walking by reducing the large "jumps", but it might also reduce travel performance. Some research uses a teleport without restriction to enhance task performance and usability (LaViola et al., 2017), but this probably reduces the similarities with natural walking. The environment used in the research is chosen to analyse the travel component of the navigation, thus excluding the wayfinding component. Thus, we use a large environment (750 m^2), which allows the performance of very different types of travel, but at the same time imposes an obligatory sequence for the room visits, from 1 to 5, thereby excluding the possible influence of subjects taking different routes on the travel component results. The analysis of the wayfinding process using the present set-up should in the future be undertaken in other environments which allow different routes to be taken from origin to destination.

Much previous research compares navigation between real-world and virtual reality using non-immersive and semi-immersive devices. Richardson et al. (1999) perform a navigation-performance task on a large screen; van der Ham et al. (2015) analyse a route memory task on a computer screen; and Claessen et al. (2016) analyse the wayfinding of chronic stroke patients using videos on a screen. Moreover, the vast majority of the comparisons are focused on the wayfinding component of the navigation, not the travel component. The few researches that use HMDs analyse navigation tasks in small spaces from an orientation and task-goal perspective. Kimura et al. (2017) perform a comparison of orientation-tasks using a HMD, and demonstrate that participants in a VR room show less facility with spatial geometry. Lessels & Ruddle (2005) analyse a searching-task performance using an HMD and suggest that visually photorealistic environments allow navigation to take place almost as efficiently as in a real-world setting. Therefore, previous comparisons between real and virtual environments have the following limitations: i) they do not analyse the new generation of HMDs, ii) they focus on the wayfinding component of navigation, and iii) they use the goal-task approach. This present work aims to contribute to the knowledge in the field by addressing these limitations, using the new generation of HMDs, which can change the paradigm of the use of VR in research, and by analysing the travel component in a free exploration of a real-world environment and task, visiting a museum.

Conclusions

The virtual museum shows a high degree of sense of presence. This outcome supports the use of 3D IVEs with devices, such as HTC Vive, at the perception level and, particularly, in environments with a high emotional content, such as museums. In terms of navigation, the physical museum was explored more, although there are significant differences only in Room 1. The trajectory patterns shown by the heatmaps are very similar, although there were differences with the information boards since their medium and small sized lettering is still not easily read in HMDs. Regarding the length of the visits, the first 2 rooms show significant differences. There are significant time-dependent differences in the navigation during the first 2 minutes of the experiment, even though there was a training room. We advise that future studies using the current set-up with subjects with no experience of HMDs should include an initial adaptation period and use realistic training environments. These conclusions support the use of environmental simulations by means of 3D IVEs and HMDs as empirical tools to study human behaviour at navigation level and raise interesting questions for future commercial and research studies.

Chapter 5

Discussion

"Nothing in life is to be feared. It is only to be understood."
Marie Curie

In this chapter we discuss the major implications of the work. We consider the use of immersive virtual reality in human behaviour research, focussing on the role that VR might play in the next years as an emotion elicitation tool. We have analysed VR's synergy with emotion recognition systems using psychological signals and machine learning; and its validity by performing a direct comparison between a real museum and its virtualization in a 3D environment.

Immersive VR as an emotion elicitation methodology

The objective of the thesis is to validate the use of immersive VR as an emotion elicitation tool in human behaviour research. Given the performance improvements in the latest generation of HMDs, we fixed on the use of HMDs as display devices. They mark a new step in terms of resolution, field of view, level of immersion and price, democratising the purchase of HMDs around the world (Castelvecchi, 2016) and boosting their research applications (Jensen & Konradsen, 2017). To explore different types of immersive stimuli, we analysed them in combination with the two most used immersive formats, 360° panoramas and 3D environments (Mengoni et al., 2011), as they offer pros and cons depending on the research case. In Chapter 2 we used a portable HMD (Samsung Gear VR) combined with 360° panoramas. This offers, with a high degree of realism, a portable solution for analysing static environments, using computer-generated images (Higuera-Trujillo et al., 2017). This set-up is very effective for updating classic affective computing methodologies; it presents users with a series of non-interactive stimuli, such as IAPS (Valenza et al., 2012) and IADS (Greco et al., 2016), and increasing degrees of presence, due to immersion (Baños et al., 2004). In chapters 3 and 4 we used a high-performance HMD (HTC Vive) combined with a 3D environment. This set-up has higher hardware requirements and it is not portable, but it offers a highly-interactive environment which can simulate real-world activities, such as visiting a museum. Previous studies used 3D scenarios (McCall et al., 2016) as their interactivity levels can be very useful in more applied research, they display more naturalistic and interactive environments, and facilitate research into decision-making. Thus, in this thesis we cover the two main approaches, realistic 360º and interactive 3D, to analyse immersive VR as an emotion elicitation method in human behaviour research.

New affective computing methods using 360º immersive VR

In Chapter 2 we developed a new set of emotional immersive VR environments using computed-generated 360º panoramas. Previous studies analysed correlations between HRV and EDA and stress using indoor and outdoor 360º real-world photographs (Anderson et al., 2017), and pleasantness in a retail store (Higuera-Trujillo et al., 2017). In contrast, we developed a new methodology based on a classic affective computing approach, using images, audio and video as emotion elicitation methods (Zangeneh et al., 2018), and added an immersive perspective. Normally, this research uses validated stimuli with the aim of covering a wide range of emotions, that is, stimuli balanced in terms of arousal and valence, IAPS being the set most used (Valenza et al., 2012). However, there is a need to develop new validated sets that can be used in immersive devices, such as HMDs, to enhance users' sense of presence in laboratory environments (Slater & Wilbur, 1997). To the best of our knowledge, this is the first study to validate a set of emotional 360º panoramas in a controlled way, that is, by changing specific parameters of the scenario to elicit emotions using aspects of environmental psychology, such as colours, geometry and illumination. The set presented a wide range of valence-arousal measures through psychological self-assessment. We hope that this set will be a first step in the development of large sets of validated 360º emotional images that can be used in the future by the scientific community. Moreover, as far as we know, we have developed the first emotion recognition model using physiological signals with immersive VR. In particular, EEG and ECG signals, combined with SVM algorithms, have been demonstrated to be good indicators of users' emotional statements, achieving

75.00% accuracy along the arousal dimension and 71.21% along the valence dimension. It is important to note that some studies have combined affective computing with virtual reality (Wu et al., 2010), but all used non-immersive scenarios. Our approach represents represents a new step in the affective computing state-of-the-art, both in terms of its methodological contribution and in terms of the scientific insights it provides to the physiological dynamics of VR.

The power of 3D real-world simulations for evoking emotions

In chapters 3 and 4 we developed a realistic simulation of an art exhibition, chosen for the analysis as it is a very emotional environment. The VR scenario was created, following strict procedures, to replicate the real environment. Moreover, we performed a direct comparison between the real and the virtual environments, which is a novel contribution to VR research as there is still a scarcity of research comparing real-world scenarios with their simulations in laboratory settings (Yeom et al., 2017). Previous studies have analysed emotional responses to food (Gorini et al., 2010) and perception in daylit spaces (Chamilothori et al., 2018). However, more experimental research is needed into immersive VR, so that its capacity to evoke the same moods as real environments can be validated. We validated a 3D VR simulation using psychological self-assessment, and showed that the vast majority of rooms in the exhibition do not present significant differences in terms of arousal and valence. In addition, the simulation achieved a high degree of presence. These results support the capacity of VR to recreate real-world environments. We have been able to draw a very strong and effective comparison due to the complexity of the real-world/virtual environments used, the naturalistic and non-guided task analysed, the ecological method, and the level of realism achieved. However, since the 3D scenario used was an art exhibition, more research is needed to analyse activities in more commonplace environments, such as houses, offices, hospitals, schools, stores, etc. Moreover, the present thesis does not consider the influence of social stimuli, that is, we did not analyse the emotional responses to avatars in VR. As in real life we are very influenced by our social environment, and VR has been shown to be a powerful tool for recreating plausible illusions of social stimuli, the affective computing field in future could take new steps by applying emotion recognition models to virtual social environments.

Using psychological signals and machine learning in 3D VR

In addition to psychological assessments, we developed an emotion recognition system able to automatically infer emotions in real-world emotional environments and their simulations. Previous studies have explored the correlations between physiological responses and fear (Gromer et al., 2019), anxiety (Tsai et al., 2018), arousal (Kisker et al., 2019), stress (Zimmer & Wu, 2019) and pleasantness (Higuera-Trujillo et al., 2017). We used EEG and ECG signals gathered from subjects using wearable sensors, combining them with SVM algorithms. In addition, we analysed the inclusion of previous responses to emotional stimuli in two formats (2D images and 360º panoramas), and showed that these data improved the emotion recognition models in both cases. The emotion recognition systems achieved an accuracy of 71.52% for arousal and 77.08% for valence in the physical museum, and 75.00% for arousal and 71.05% for valence in the virtual museum. The results also showed some differences in the physiological responses in both environments. The emotion recognition models in the real museum used HRV and EEG features, but in the virtual museum they used only EEG MPC features. Moreover, these results revealed the important role that brain synchronization features play in the neuro-physiological processes involved in VR, as they allow us to identify whether subjects are in virtual or real environments to over 95% accuracy. This is in accordance with previous research where measures of nonlinear interdependency in EEG have been applied to analyse perceptual processes, cognitive tasks and disorders (Glass, 2001; Stam, 2005). We used wearable sensors, which allowed us to undertake research in the real-world, but the limited number of sources (only 9 electrodes were used in the EEG) needs to be taken into account**.** This research presents a new methodological framework for assessing the application of emotion recognition systems in 3D environments and in the real world. It is of note that this thesis presents, to the best of our knowledge, the first emotion recognition system which uses immersive 3D VR in elicitation, and one of the most complex applications of affective computing to the real-world, using EEG and ECG. It represents a new step on the long research road to develop emotion recognition models using physiological signals that can be applied to real-world tasks, especially in their laboratory simulations, using VR.

The influence of navigation in 3D environments

Navigation can condition the perception of a space. To analyse the validity of the VR we compared the navigation patterns in the real and virtual museums. Navigation encompasses travel and wayfinding components. The travel component is related to the task of moving from one point to another, and the wayfinding component is related to the cognitive process of establishing a path from an origin to a destination (LaViola et al., 2017). Several previous studies have compared the wayfinding component in virtual and physical environments, generally concluding that the virtual offers poorer performance than the physical (Richardson et al., 1999; van der Ham et al., 2015). However, no research has compared the navigation experience between a real and virtual environment in terms of the travel component. Thus, in Chapter 4 we analyse the travel component of the navigation experience by performing a direct comparison between the virtual and the real museum. The museum chosen allowed us to analyse the travel component as the exhibition is composed of 5 consecutive rooms, so there is only one possible route to follow and, thus, the wayfinding component is excluded. The results showed that there are time-dependent differences in the first 2 minutes, probably due to the 'wow' effect of the VR. Subsequently, the navigation patterns were very similar. As previous studies used screens and most focused on the wayfinding component, taking goal-task approaches (Claessen et al., 2016; van der Ham et al., 2015), we present, to the best of our knowledge, the first direct comparison of free navigation in a real-world environment and its simulation using immersive VR. The results support the use of 3D VR as an emotion elicitation method in human behaviour research, and present some guidelines for consideration for use in future research. However, the present research analysed only the teleport metaphor, and the metaphor used can have a strong influence on navigation patterns (Lee et al., 2018). Thus, further research is needed to analyse results in other environments, adding new behavioural responses, and taking into account the influence of wayfinding in the travel component.

Chapter 6

Conclusion

The thesis presents a novel approach to the use of VR in human behaviour research, in particular in relation to emotion elicitation. The framework developed includes 360º panoramas and 3D environments as display formats, and has implications for affective computing research itself, in that it can improve the methodologies applied in laboratory environments. It can help classic methodologies develop more realistic stimuli to assess daily-life environments and situations, thereby overcoming the current limitations of passive methods of affective elicitation, which traditionally include images, audio and/or video. HRV and EEG responses to affective VR stimuli, in combination with supervised machine learning methods, successfully recognized the mood of users in a suite of 360º affective stimuli, a real museum and its 3D virtualization. Therefore, brain and heart dynamics have been proven to be powerful for analysing emotions in VR, and this result increases our knowledge of the physiological responses related to emotion processes, in particular of measures of EEG nonlinear synchronization. The VR stimuli were displayed using HMDs, a new step in the combination of affective computing and immersive VR. Moreover, the museum was validated in a direct comparison, in terms of psychological patterns using self-assessment and navigation responses, which showed that, in both cases, the majority of stimuli did not show statistical differences. The results will help researchers to analyse and measure the impact of different parameters of the emotional responses of potential users, as they can facilitate the use of more ecological stimuli in VR. This is especially important where it is physically very difficult or impossible to carry out research in actual environments; for example, by analysing the arousal responses provoked by changes in ceiling height; in dangerous situations, for example firefighting training and phobia therapies; and, for example, in the decision-making process for new buildings, where designs can be tested prior to construction. In addition, the capacity of VR to isolate individual environmental parameters, while keeping the rest of the environment identical, provides high synergy in their application with affective computing.

The insights presented can help explain the psycho-physiological responses of human beings to many different stimuli, and facilitate the development of better practices in many fields. For example, in architecture they could help in the design of public and private buildings to optimize, before construction, emotions that designers wish to evoke (e.g. relaxation in a library or positive valence in a hospital). They could help also to develop ambient-assisting living, that is, environments that change depending on the demands of their users. They also offer many health and psychology applications, from assistance in therapies (e.g. phobia recovery by modulating exposure to stimuli based on the emotional responses of the patients), to diagnosis (e.g. evaluating children with autistic spectrum disorders by analysing their emotional responses to social stimuli). In assessments the methodology can help classify different types of personality through subjects' emotional responses (e.g. leadership assessment during a team-work task or public speaking). In the training field, it can offer safe and adaptive environments (e.g. modulating the difficulty of a military training exercise based on the trainees' stress levels). In education, it could help explain which environments and techniques optimize the emotional engagement of students (e.g. by evaluating students' arousal based on student numbers in classrooms). In driving, it could help to evaluate stress in drivers (e.g. detecting the stressful points in the geometry of a road before its construction). In marketing, it could help to optimize the customer experience while shopping (e.g. optimizing store layouts to minimize frustration). However, more research is needed to achieve models of emotion recognition in VR that might be extrapolated to other environments. This opens a new sub-field in affective computing; further research is needed to address the current limitations. Future studies should analyse new emotion recognition models in other environments, using larger datasets of emotional immersive stimuli and higher numbers of participants. Moreover, the use of EEG devices with large numbers of electrodes, in combination with HMDs, remains an under-analysed topic which needs further investigation. The inclusion of other physiological signals, such as EDA and fNIRs, needs to be addressed, as do other implicit techniques, such as eye-tracking, body posture and voice analysis. In addition, the physiological response differences between VR and real environments need further analysis to improve the understanding of the validity of immersive VR. Finally, the inclusion of social stimuli, such as avatars, is a strong point of immersive VR due to the presence levels they provide in comparison to classic methods; future studies also need to address the development of emotion recognition models in social contexts.

In conclusion, emotions play a critical role in our daily lives, so an understanding and recognition of emotional responses is crucial for human research. We believe that VR will revolutionize emotion elicitation methods in laboratory environments. Moreover, its synergy with physiological measurements and machine learning techniques will impact transversely in many areas of research, opening new opportunities for the scientist community. We hope that the present work marks a new step in that direction.

Research activities

Associated projects

Research for new metrics in Neuroarquitectura using Immersive Virtual Environments (TIN2013-45736-R). Ministerio de Economía y Competitividad (Madrid, Spain).

Contrato predoctoral FPI (BES-2014-069449). Ministerio de Economía y Competitividad (Madrid, Spain).

Research interships

Bioengineering and Robotics Research Centre E. Piaggio & Department of Information Engineering, University of Pisa, Pisa, Italy. (April 1, 2017 – July 31, 2017). Supervisor: Gaetano Valenza.

Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom (March 30, 2019 – June 1, 2019). Supervisor: Luca Citi.

Journal papers

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., Alcañiz, M., & Valenza, G. (2019). Real vs. immersive-virtual emotional experience: Analysis of psycho-physiological patterns in a free exploration of an art museum. *PloS one,* 14(10).

Marín-Morales, J., Higuera-Trujillo, J. L., De-Juan-Ripoll, C., Llinares, C., Guixeres, J., Iñarra, S., & Alcañiz, M. (2019). Navigation Comparison between a Real and a Virtual Museum: Time-dependent Differences using a Head Mounted Display. *Interacting with Computers*, 31(2), 208–220.

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., Alcañiz, M., & Valenza, G. (2018). Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports*, *8*(1), 13657.

Marín-Morales, J., Torrecilla, C., Guixeres, J., & Llinares, C. (2017). Methodological bases for a new platform for the measurement of human behaviour in virtual environments, *DYNA Ingeniería e Industria,* 92(1), 34-38.

Conference papers

Marín-Morales, J., Higuera-Trujillo, J. L., Llinares, C., Guixeres J., Alcañiz M., & Valenza G. (2020). Real vs. Immersive Virtual Emotional Museum Experience: a Heart Rate Variability Analysis during a Free Exploration Task, *11th conference of the European Study Group on Cardiovascular Oscillations.*

Marín-Morales, J., Higuera-Trujillo, J.L., Juan-Ripoll, C., Llinares, C., Guixeres, J., Iñarra, S., & Alcañiz, M. (2018). Presence and navigation: a comparison between the free exploration of a real and a virtual museum, *32nd Human Computer Interaction Conference (HCI 2018).*

Higuera-Trujillo, J.L., **Marín-Morales, J.**, Rojas, J.C., López-Tarruella, J., Llinares, C., Guixeres, J., & Alcañiz, M. (2016). Emotional cartography in design: A novel technique to represent emotional states altered by spaces, *10th International Conference on Design & Emotion 2016* (pp 667-685, ISSN 978-94-6186-725-4).

Higuera-Trujillo, J.L., **Marín-Morales, J.**, Rojas, J.C., & López-Tarruella Maldonado, J. (2016). Emotional maps: neuro architecture and design applications, *6th International Forum of Design as a Process. Systems & Design Beyond Processes and Thinking,* Valencia, España. DOI: http://dx.doi.org/10.4995/IFDP.2016.

Conference posters

Marín-Morales, J., Higuera-Trujillo, J.L., Juan-Ripoll, C., Iñarra, S., Guixeres, J., & Llinares, C. (2018). Emotion recognition in virtual environment: introducing an immersive virtual environment set. *International itinerant exhibition research in building engineering EXCO'18.*

Llinares, C., Montaña, A., Llonres, R., **Marín-Morales, J**., Higuera-Trujillo, J. L., & Iñarra, S. (2018). Methodological proposal to analyse pedestrian’s safety perception in urban areas. *International itinerant exhibition research in building engineering EXCO'18* .

Castilla-Cabanes, N., Higuera-Trujillo, J. L., **Marín-Morales, J.,** Llinares, C., & López-Tarruella, J. (2017). Validation of Lighting Design trough the Emotional and Cognitive Effect of the Architectual Space. *International exhibition research in building engineering EXCO 2017.*

Marín-Morales, J., Higuera-Trujillo, J. L., Juan-Ripoll, C., Iñarra, S., & Llinares, C. (2017). Metodología para evaluar el impacto emocional de un espacio arquitectónico en escenarios virtuales. *II Exposición de Jóvenes Arquitectos Colegio Oficial de Arquitectos de Sevilla.*

Marín-Morales, J., Higuera-Trujillo, J. L., Juan-Ripoll, C., Iñarra, S., López-Maldonado, J., & Llinares, C. (2017). Development of new metrics to evaluate the impact of architecture on an emotional level in virtual environments. *International exhibition research in building engineering EXCO 2017.*

Marín-Morales, J., Higuera-Trujillo, J. L., López-Maldonado, J., & Llinares, C. (2017). EEG-Index of stress generated by the environment: towards the neuroscience-based Architectural design. *International exhibition research in building engineering EXCO 2017*.

López-Maldonado, J., Higuera-Trujillo, J. L., Llinares, C., & **Marín-Morales, J**. (2017). Neuroarchitecture: Prediction of emotional well-being provoked by spaces by indirect measurement of brain activity. *International exhibition research in building engineering EXCO 2017*.

Marín-Morales, J., Guixeres, J., Llinares, C., Ausin, J. M., Torecilla-Moreno, C., & Garcia-Carbonell, N. (2016). Nueva plataforma tecnológica de medida del comportamiento humano en entornos virtuales en el estudio de la arquitectura. *XXX Salón tecnológico de la construcción EXCO 2016*.

References

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: A review. *Medical and Biological Engineering and Computing*, *44*(12), 1031–1051. https://doi.org/10.1007/s11517-006-0119-0

Alaraj, A., Lemole, M. G., Finkle, J. H., Yudkowsky, R., Wallace, A., Luciano, C., Banerjee, P. P., Rizzi, S. H., & Charbel, F. T. (2011). Virtual reality training in neurosurgery: review of current status and future applications. *Surgical Neurology International*, *2*.

Alcañiz, M., Baños, R., Botella, C., & Rey, B. (2003). The EMMA Project : Emotions as a Determinant of Presence. *PsychNology Journal*, *1*(2), 141–150.

Alcañiz, M., Bigné, E., & Guixeres, J. (2019). Virtual Reality in Marketing: A Framework, Review, and Research Agenda. *Frontiers in Psychology*, *10*(July), 1–15. https://doi.org/10.3389/fpsyg.2019.01530

Alcañiz, M. L., Olmos-Raya, E., & Abad, L. (2019). Use of virtual reality for neurodevelopmental disorders. A review of the state of the art and future agenda. *Medicina*, *79*, 77–81.

Alcañiz, M., Lozano, J. A., & Rey, B. (2004). Technological background of VR. *Studies in Health Technology and Informatics*, *99*, 199–214.

Alcañiz, M., Rey, B., Tembl, J. J., & Parkhutik, V. (2009). A Neuroscience Approach to Virtual Reality Experience Using Transcranial Doppler Monitoring. *Presence: Teleoperators and Virtual Environments*, *18*(2), 97–111. https://doi.org/doi:10.1162/pres.18.2.97

Amaral, C. P., Simões, M. A., Mouga, S., Andrade, J., & Castelo-Branco, M. (2017). A novel Brain Computer Interface for classification of social joint attention in autism and comparison of 3 experimental setups: A feasibility study. *Journal of Neuroscience Methods*, *290*, 105–115. https://doi.org/10.1016/j.jneumeth.2017.07.029

Anderson, A. P., Mayer, M. D., Fellows, A. M., Cowan, D. R., Hegel, M. T., & Buckey, J. C. (2017). Relaxation with Immersive Natural Scenes Presented Using Virtual Reality. *Aerospace Medicine and Human Performance*, *88*(6), 520–526. https://doi.org/10.3357/AMHP.4747.2017

Babu, S. K., Krishna, S., R, U., & Bhavani, R. R. (2018). Virtual Reality Learning Environments for Vocational Education: A Comparison Study with Conventional Instructional Media on Knowledge Retention. *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, 385–389. https://doi.org/10.1109/ICALT.2018.00094

Bahreini, K., Nadolski, R., & Westera, W. (2016). Towards multimodal emotion recognition in e-learning environments. *Interactive Learning Environments*, *24*(3), 590–605.

Baños, R. M., Liaño, V., Botella, C., Alcañiz, M., Guerrero, B., & Rey, B. (2006). Changing induced moods via virtual reality. In H. Springer, Berlin (Ed.), *International Conference on Persuasive Technology* (pp. 7–15). https://doi.org/10.1007/11755494_3

Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., & Rey, B. (2004). Immersion and Emotion: Their Impact on the Sense of Presence. *CyberPsychology & Behavior*, *7*(6), 734–741. https://doi.org/10.1089/cpb.2004.7.734

Baños, R. M., Etchemendy, E., Castilla, D., García-palacios, A., Quero, S., & Botella, C. (2012). Positive mood induction procedures for virtual environments designed for elderly people. *Interacting with Computers*, *24*(3), 131–138. https://doi.org/10.1016/j.intcom.2012.04.002

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barsade, S. G., Ramarajan, L., & Westen, D. (2009). Implicit affect in organizations. *Research in Organizational Behavior*, *29*, 135–162.

Baumgartner, T., Valko, L., Esslen, M., & Jäncke, L. (2006). Neural Correlate of Spatial Presence in an Arousing and Noninteractive Virtual Reality: An EEG and Psychophysiology Study. *CyberPsychology & Behavior*, *9*(1), 30–45. https://doi.org/10.1089/cpb.2006.9.30

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, *78*(5), B231--B244.

Berlyne, D. E. (1970). Novelty, Complexity, and Hedonic Value. *Perception & Psychophysics*, *8*(5-A), 279–286.

Bhagat, K. K., Liou, W. K., & Chang, C. Y. (2016). A cost-effective interactive 3D virtual reality system applied to military live firing training. *Virtual Reality*, *20*(2), 127–140.

Bian, Y., Yang, C., Gao, F., Li, H., Zhou, S., Li, H., Sun, X., & Meng, X. (2016). A framework for physiological indicators of flow in VR games: construction and preliminary evaluation. *Personal and Ubiquitous Computing*, *20*(5), 821–832. https://doi.org/10.1007/s00779-016-0953-5

Biedermann, S. V, Biedermann, D. G., Wenzlaff, F., Kurjak, T., Nouri, S., Auer, M. K., Wiedemann, K., Briken, P., Haaker, J., Lonsdorf, T. B., & others. (2017). An elevated plus-maze in mixed reality for studying human anxiety-related behavior. *BMC Biology*, *15*(1), 125.

Bigné, E., Llinares, C., & Torrecilla, C. (2015). Elapsed time on first buying triggers brand choices within a category: A virtual reality-based study. *Journal of Business Research*. https://doi.org/10.1016/j.jbusres.2015.10.119

Bishop, I. D., & Rohrmann, B. (2003). Subjective responses to simulated and real environments: a comparison. *Landscape and Urban Planning*, *65*(4), 261–277.

Blake, J., & Gurocak, H. B. (2009). Haptic glove with MR brakes for virtual reality. *IEEE/ASME Transactions on Mechatronics*, *14*(5), 606–615. https://doi.org/10.1109/TMECH.2008.2010934

Blascovich, J. J., Loomis, J. M., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., Bailenson, N., & Bailenson, J. N. (2012). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, *7965*(February), 103–124. https://doi.org/10.1207/S15327965PLI1302

Bliss, J. P., Tidwell, P. D., & Guest, M. A. (1997). The Effectiveness of Virtual Reality for Administering Spatial Navigation Training to Firefighters. *Presence: Teleoperators and Virtual Environments*, *6*(1), 73–83.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, *12*(12). https://doi.org/10.1038/nrn3122

Boletsis, C. (2017). The New Era of Virtual Reality Locomotion: A Systematic Literature Review of Techniques and a Proposed Typology. *Multimodal Technologies and Interaction*, *1*(4), 24. https://doi.org/10.3390/mti1040024

Borrego, A., Latorre, J., Alcañiz, M., & Llorens, R. (2018). Comparison of Oculus Rift and HTC Vive: Feasibility for Virtual Reality-Based Exploration, Navigation, Exergaming, and Rehabilitation. *Games for Health Journal*, *7*(2), g4h.2017.0114. https://doi.org/10.1089/g4h.2017.0114

Borrego, A., Latorre, J., Llorens, R., Alcañiz, M., & Noé, E. (2016). Feasibility of a walking virtual reality system for rehabilitation: Objective and subjective parameters. *Journal of NeuroEngineering and Rehabilitation*, *13*(1). https://doi.org/10.1186/s12984-016-0174-1

Bozgeyikli, E., Bozgeyikli, L., Raij, A., Katkoori, S., Alqasemi, R., & Dubey, R. (2016). Virtual reality interaction techniques for individuals with autism spectrum disorder: design considerations and preliminary results. *International Conference on Human-Computer Interaction*, 127–137.

Bozgeyikli, E., Raij, A., Katkoori, S., & Dubey, R. (2016). Locomotion in virtual reality for individuals with autism spectrum disorder. *Proceedings of the 2016 Symposium on Spatial User Interaction*, 33–42.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brief, A. P. (1998). *Attitudes in and around organizations* (Vol. 9). Sage.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, *30*(6), 1081–1106. https://doi.org/10.1080/02699931.2015.1049124

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37. https://doi.org/10.1109/T-AFFC.2010.1

Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, *43*(1), 9–64.

Castelvecchi, D. (2016). Low-cost headsets boost virtual reality's lab appeal. *Nature*, *533*, 153–154. https://doi.org/10.1038/533153a

Chamilothori, K., Wienold, J., & Andersen, M. (2018). Adequacy of Immersive Virtual Reality for the Perception of Daylit Spaces: Comparison of Real and Virtual Environments. *LEUKOS - Journal of Illuminating Engineering Society of North America*, *00*(00), 1–24. https://doi.org/10.1080/15502724.2017.1404918

Chang, C. C., & Lin, C. J. (2011). Libsvm: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–27. https://doi.org/10.1145/1961189.1961199

Chirico, A., Cipresso, P., Yaden, D. B., Biassoni, F., Riva, G., & Gaggioli, A. (2017). Effectiveness of Immersive Videos in Inducing Awe: An Experimental Study. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-01242-0

Chittaro, L., Sioni, R., Crescentini, C., & Fabbro, F. (2017). Mortality salience in virtual reality experiences and its effects on users' attitudes towards risk. *International Journal of Human-Computer Studies*, *101*, 10–22.

Cipresso, P., Chicchi, I. A., Alcañiz, M., & Riva, G. (2018). The Past, Present, and Future of Virtual and Augmented Reality Research: A network and cluster analysis of the literature. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2018.02086

Claessen, M. H. G., Visser-Meily, J. M. A., De Rooij, N. K., Postma, A., & Van Der Ham, I. J. M. (2016). A Direct Comparison of Real-World and Virtual Navigation Performance in Chronic Stroke Patients. *Journal of the International Neuropsychological Society*, *22*(4), 467–477. https://doi.org/10.1017/S1355617715001228

Clemente, M., Rodríguez, A., Rey, B., & Alcañiz, M. (2014). Assessment of the influence of navigation control and screen size on the sense of presence in virtual reality using EEG. *Expert Systems with Applications*, *41*(4 PART 2), 1584–1592. https://doi.org/10.1016/j.eswa.2013.08.055

Colomer Granero, A., Fuentes-Hurtado, F., Naranjo Ornedo, V., Guixeres Provinciale, J., Ausín, J. M., & Alcañiz Raya, M. (2016). A Comparison of Physiological Signal Analysis Techniques and Classifiers for Automatic Emotional Evaluation of Audiovisual Contents. *Frontiers in Computational Neuroscience*, *10*(July), 1–14. https://doi.org/10.3389/fncom.2016.00074

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, *250*, 126–136. https://doi.org/10.1016/j.jneumeth.2015.01.010

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psy Chological Research*, *3*(31), 58–67. https://doi.org/10.21500/20112084.844

Critchley, H. D. (2009). Psychophysiology of neural, cognitive and affective integration: fMRI and autonomic indicants. *International Journal of Psychophysiology*, *73*(2), 88–94. https://doi.org/10.1016/j.ijpsycho.2009.01.012

Dascal, J., Reid, M., Ishak, W. W., Spiegel, B., Recacho, J., Rosen, B., & Danovitch, I. (2017). Virtual reality and medical inpatients: A systematic review of randomized, controlled trials. *Innovations in Clinical Neuroscience*, *14*(1–2), 14–21.

de-Juan-Ripoll, C., Soler-Domínguez, J. L., Guixeres, J., Contero, M., Gutiérrez, N. Á., & Alcañiz, M. (2018). Virtual reality as a new approach for risk taking assessment. *Frontiers in Psychology*, *9*(DEC), 1–8. https://doi.org/10.3389/fpsyg.2018.02532

de Borst, A. W., & de Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: An affective cognitive neuroscience perspective. *Frontiers in Psychology*, *6*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2015.00576

de Kort, Y. A. W., Ijsselsteijn, W. A., Kooijman, J., & Schuurmans, Y. (2003). Virtual laboratories: comparability of real and virtual environments for environmental psychology. *Presence: Teleoperators and Virtual Environments*, *12*(4), 360–373.

Debener, S., Minow, F., Emkes, R., Gandras, K., & de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, *49*(11), 1617–1621. https://doi.org/10.1111/j.1469-8986.2012.01471.x

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Didehbani, N., Allen, T., Kandalaft, M., Krawczyk, D., & Chapman, S. (2016). Virtual reality social cognition training for children with high functioning autism. *Computers in Human Behavior*, *62*, 703–711.

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, *6*(JAN), 1–9. https://doi.org/10.3389/fpsyg.2015.00026

Dols, J. F., Molina, J., Camacho, F. J., Marín-Morales, J., Pérez-Zuriaga, A. M., & Garcia, A. (2016). Design and development of driving simulator scenarios for road validation studies. *Transportation Research Procedia*, *18*, 289–296.

Dourish, P. (1999). Embodied Interaction: Exploring the Foundations of a New Approach to HCI. *Work*, 1–16.

Eberhard, J. P. (2009). Applying Neuroscience to Architecture. *Neuron*, *62*(6), 753–756. https://doi.org/10.1016/j.neuron.2009.06.001

Egan, D., Brennan, S., Barrett, J., Qiao, Y., Timmerer, C., & Murray, N. (2016). *An evaluation of Heart Rate and Electrodermal Activity as an Objective QoE Evaluation method for Immersive Virtual Reality Environments*. 3–8. https://doi.org/10.1109/QoMEX.2016.7498964

Ekman, P. (2007). The directed facial action task. In *Handbook of emotion elicitation and assessment* (pp. 47–53).

Ekman, P. (1999). Basic Emotions. In *Handbook of cognition and emotion* (pp. 45–60). https://doi.org/10.1017/S0140525X0800349X

Ernst, L. H., Plichta, M. M., Lutz, E., Zesewitz, A. K., Tupak, S. V, Dresler, T., Ehlis, A.-C., & Fallgatter, A. J. (2013). Prefrontal activation patterns of automatic and regulated approach--avoidance reactions--a functional near-infrared spectroscopy (fNIRS) study. *Cortex*, *49*(1), 131–142.

Eudave, L., & Valencia, M. (2017). Physiological response while driving in an immersive virtual environment. *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, *July*, 145–148. https://doi.org/10.1109/BSN.2017.7936028

Felnhofer, A., Kothgassner, O. D., Hetterle, T., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2014). Afraid to be there? Evaluating the relation between presence, self-reported anxiety, and heart rate in a virtual public speaking task. *Cyberpsychology, Behavior, and Social Networking*, *17*(5), 310–316.

Felnhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A. K., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2015). Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International Journal of Human Computer Studies*, *82*, 48–56. https://doi.org/10.1016/j.ijhcs.2015.05.004

Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J. M., Castillo, J. C., Lozano-Monasor, E., Lopez, M. T., Zangróniz, R., Latorre, J. M., & Fernández-Sotos, A. (2016). Smart environment architecture for emotion detection and regulation. *Journal of Biomedical Informatics*, *64*, 55–73. https://doi.org/10.1016/j.jbi.2016.09.015

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, *47*(14), 2393–2400. https://doi.org/10.1017/S003329171700040X

Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: state of the art and perspectives. *The International Scientific Conference ELearning and Software for Education*, *1*, 133.

Gentili, C. (2017). Why do we keep failing in identifying reliable biological markers in depression? *Journal of Evidence-Based Psychotherapies*, *17*(2), 69–84. https://doi.org/10.24193/jebp.2017.2.4

Giglioli, I. A. C., Pravettoni, G., Martín, D. L. S., Parra, E., & Alcañiz, M. (2017). A novel integrating virtual reality approach for the assessment of the attachment behavioral system. *Frontiers in Psychology*, *8*(JUN), 1–7. https://doi.org/10.3389/fpsyg.2017.00959

Glass, L. (2001). Synchronization and rhythmic processes in physiology. *Nature*, *410*(6825), 277–284. https://doi.org/10.1038/35065745

Glöckner, A., & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98.

Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., & Riva, G. (2011). The Role of Immersion and Narrative in Mediated Presence: The Virtual Hospital Experience. *Cyberpsychology, Behavior, and Social Networking*, *14*(3), 99–105. https://doi.org/10.1089/cyber.2010.0100

Gorini, A., Griez, E., Petrova, A., & Riva, G. (2010). Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders. *Annals of General Psychiatry*, *9*(1), 30.

Gorini, A., Mosso, J. L., Mosso, D., Pineda, E., Ruíz, N. L., Ramíez, M., Morales, J. L., & Riva, G. (2009). Emotional Response to Virtual Reality Exposure across Different Cultures: The Role of the Attribution Process. *CyberPsychology & Behavior*, *12*(6), 699–705. https://doi.org/10.1089/cpb.2009.0192

Grassberger, P., & Procaccia, I. (1983). Characterization of strange attractors. *Physical Review Letters*, *50*(5), 346–349. https://doi.org/10.1103/PhysRevLett.50.346

Greco, A., Valenza, G., Citi, L., & Scilingo, E. P. (2016). Arousal and valence recognition of affective sounds based on electrodermal activity. *IEEE Sensors Journal*, *17*(3), 716–725.

Gromer, D., Reinke, M., Christner, I., & Pauli, P. (2019). Causal Interactive Links Between Presence and Fear in Virtual Reality Height Exposure. *Frontiers in Psychology*, *10*, 141.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, *9*(1), 87–108. https://doi.org/10.1080/02699939508408966

Gruzelier, J. H. (2014). EEG-neurofeedback for optimising performance. I: a review of cognitive and affective outcome in healthy participants. *Neuroscience & Biobehavioral Reviews*, *44*, 124–141.

Guixeres, J., Saiz, J., Alcañiz, M., Cebolla, A., Escobar, P., Baños, R., Botella, C., Lison, J. F., Alvarez, J., Cantero, L., & Lurbe, E. (2013). Effects of virtual reality during exercise in children. *Journal of Universal Computer Science*, *19*(9), 1199–1218.

Harmon-Jones, E., Amodio, D. M., & Zinner, L. R. (2007). Social psychological methods of emotion elicitation. *Handbook of Emotion Elicitation and Assessment*, *January*, 91–105. https://doi.org/10.2224/sbp.2007.35.7.863

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322. https://doi.org/10.1007/s11065-010-9138-6

Hassib, M., Schneegass, S., Eiglsperger, P., Henze, N., Schmidt, A., & Alt, F. (2017). EngageMeter: A System for Implicit Audience Engagement Sensing Using Electroencephalography. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5114–5119. https://doi.org/10.1145/3025453.3025669

He, C., Yao, Y., & Ye, X. (2017). An Emotion Recognition System Based on Physiological Signals Obtained by Wearable Sensors. In C. Yang, G. S. Virk, & H. Yang (Eds.), *Wearable Sensors and Robots: Proceedings of International Conference on Wearable Sensors and Robots 2015* (pp. 15–25). Springer Singapore. https://doi.org/10.1007/978-981-10-2404-7_2

Heeter, C. (1992). Being There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments*, *1*(2), 262–271. https://doi.org/10.1162/pres.1992.1.2.262

Heydarian, A., Carneiro, J. P., Gerber, D., Becerik-Gerber, B., Hayes, T., & Wood, W. (2015). Immersive virtual environments versus physical built environments: A benchmarking study for building design and user-built environment explorations. *Automation in Construction*, *54*, 116–126. https://doi.org/10.1016/j.autcon.2015.03.020

Higuera-Trujillo, J. L., López-Tarruella, J., & Llinares, M. C. (2017). Psychological and physiological human responses to simulated and real environments: A comparison between Photographs, 360° Panoramas, and Virtual Reality. *Applied Ergonomics*, *In Press*. https://doi.org/10.1016/j.apergo.2017.05.006

Higuera, J. L., Marín-Morales, J., Rojas, J. C., Tarruella, J., Llinares, C., Guixeres, J., & Alcañiz, M. (2016). Emotional cartography in design: A novel technique to represent emotional states altered by spaces. *D and E 2016: 10th International Conference on Design and Emotion*, 561–566.

Hildebrandt, L. K., Mccall, C., Engen, H. G., & Singer, T. (2016). Cognitive flexibility, heart rate variability, and resilience predict fine-grained regulation of arousal during prolonged threat. *Psychophysiology*, *53*(6), 880–890. https://doi.org/10.1111/psyp.12632

Hogg, J., Goodman, S., Porter, T., Mikellides, B., & Preddy, D. E. (1979). Dimensions and determinants of judgements of colour samples and a simulated interior space by architects and non- architects. *British Journal of Psychology*, *70*(2), 231–242.

Howell, D. C. (2009). *Statistical methods for psychology*. Cengage Learning.

Huang, K. Y., Wu, C. H., Su, M. H., & Kuo, Y. T. (2018). Detecting unipolar and bipolar depressive disorders from elicited speech responses using latent affective structure model. *IEEE Transactions on Affective Computing*.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4–5), 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

Jack, A. I., Dawson, A. J., Begany, K. L., Leckie, R. L., Barry, K. P., Ciccia, A. H., & Snyder, A. Z. (2013). fMRI reveals reciprocal inhibition between social and physical cognitive domains. *NeuroImage*, *66*, 385–401.

Jacobs, K. W., & Hustmyer, F. E. (1974). Effects of four psychological primary colors on GSR, heart rate and respiration rate. *Perceptual and Motor Skills*, *38*(3), 763–766.

Jalil, N. A., Yunus, R. M., & Said, N. S. (2012). Environmental Colour Impact upon Human Behaviour: A Review. *Procedia - Social and Behavioral Sciences*, *35*, 54–62.

Jensen, L., & Konradsen, F. (2017). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 1–15. https://doi.org/10.1007/s10639-017-9676-0

Jerritta, S., Murugappan, M., Nagarajan, R., & Wan, K. (2011). Physiological signals based human emotion Recognition: a review. *Signal Processing and Its Applications (CSPA), 2011 IEEE 7th International Colloquium On*, 410–415. https://doi.org/10.1109/CSPA.2011.5759912

Jin, H. R., Yu, M., Kim, D. W., Kim, N. G., & Chung, A. S. W. (2009). Study on Physiological Responses to Color Stimulation. In S. Poggenpohl (Ed.), *International Association of Societies of Design Research* (pp. 1969–1979). Korean Society of Design Science.

Jo, Y., Lee, H., Cho, A., & Whang, M. (2017). Emotion Recognition Through Cardiovascular Response in Daily Life Using KNN Classifier. *In Advances in Computer Science and Ubiquitous Computing*, 1451–1456.

Jolliffe, I. T. (2002). Principal Component Analysis, Second Edition. *Encyclopedia of Statistics in Behavioral Science*, *30*(3), 487. https://doi.org/10.2307/1270093

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation*, *15*(3), 235.

Kim, J. (2007). Emotion Recognition Using Speech and Physiological Changes. *Robust Speech Recognition and Understanding*, *June*, 265–280.

Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(12), 2067–2083. https://doi.org/10.1109/TPAMI.2008.26

Kimura, K., Reichert, J. F., Olson, A., Pouya, O. R., Wang, X., Moussavi, Z., & Kelly, D. M. (2017). Orientation in Virtual Reality Does Not Fully Measure Up to the Real-World. *Scientific Reports*, *7*(1), 6–13. https://doi.org/10.1038/s41598-017-18289-8

Kiryu, T., & So, R. H. Y. (2007). *Sensation of presence and cybersickness in applications of virtual reality for advanced rehabilitation*. BioMed Central.

Kisker, J., Gruber, T., & Schöne, B. (2019). Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. *Psychological Research*, 1–14.

Kober, S. E., Kurzmann, J., & Neuper, C. (2012). Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: An EEG study. *International Journal of Psychophysiology*, *83*(3), 365–374. https://doi.org/10.1016/j.ijpsycho.2011.12.003

Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A dataset for emotion analysis using physiological and audiovisual signals. *IEEE Transactions on Affective Computing*, *3*, 1–15. http://www.eecs.qmul.ac.uk/mmv/datasets/deap/

Koolagudi, S. G., & Rao, K. S. (2012). Emotion recognition from speech: A review. *International Journal of Speech Technology*, *15*(2), 99–117. https://doi.org/10.1007/s10772-011-9125-1

Kory, J. M., & D'Mello, S. K. (2015). Affect Elicitation for Affective Computing. In *The Oxford Handbook of Affective Computing* (pp. 371–383).

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Krueger, R. A., & Casey, M. (2000). *Focus groups: a practical guide for applied research.* Sage Publications.

Kuliga, S. F., Thrash, T., Dalton, R. C., & Hölscher, C. (2015). Virtual reality as an empirical research tool - Exploring user experience in a real building and a corresponding virtual model. *Computers, Environment and Urban Systems*, *54*, 363–375. https://doi.org/10.1016/j.compenvurbsys.2015.09.006

Küller, R., Mikellides, B., & Janssens, J. (2009). Color, arousal, and performance—A comparison of three experiments. *Color Research & Application*, *34*(2), 141–152.

Kumari, P., Mathew, L., & Syal, P. (2017). Increasing trend of wearables and multimodal interface for human activity monitoring: A review. *Biosensors and Bioelectronics*, *90*(September 2016), 298–307. https://doi.org/10.1016/j.bios.2016.12.001

Kwartler, M. (2005). Visualization in support of public participation. In I. Bishop & E. Lange (Eds.), *Visualization in landscape and environmental planning: technology and applications* (pp. 251–260). Taylor & Francis.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International Affective Picture System (IAPS): Technical Manual and Affective Ratings. *NIMH Center for the Study of Emotion and Attention*, 39–58. https://doi.org/10.1027/0269-8803/a000147

Lange, E. (2001). The limits of realism: Perceptions of virtual landscapes. *Landscape and Urban Planning*, *54*(1–4), 163–182. https://doi.org/10.1016/S0169-2046(01)00134-7

LaViola, J. J., Kruijff, E., McMahan, R., Bowman, D. A., & Poupyrev, I. (2017). *3D user interfaces: theory and practice*. Addison-Wesley Professional.

Lee, J., Ahn, S., & Hwang, J. I. (2018). A Walking-in-Place Method for Virtual Reality Using Position and Orientation Tracking. *Sensors*, *18*(9), 2832.

Leer, A., Engelhard, I. M., & Van Den Hout, M. A. (2014). How eye movements in EMDR work: Changes in memory vividness and emotionality. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(3), 396–401. https://doi.org/10.1016/j.jbtep.2014.04.004

Lessels, S., & Ruddle, R. A. (2005). Movement Around Real and Virtual Cluttered Environments. *Presence: Teleoperators and Virtual Environments*, *14*(5), 580–596.

Lewis, P. A., Critchley, H. D., Rotshtein, P., & Dolan, R. J. (2007). Neural correlates of processing valence and arousal in affective words. *Cerebral Cortex*, *17*(3), 742–748. https://doi.org/10.1093/cercor/bhk024

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.*, *58*, 259–289.

Lin, J., Cao, L., & Li, N. (2019). Assessing the influence of repeated exposures and mental stress on human wayfinding performance in indoor environments using virtual reality technology. *Advanced Engineering Informatics*, *39*, 53–61.

Lin, Y. P., Wang, C. H., Jung, T. P., Wu, T. L., Jeng, S. K., Duann, J. R., & Chen, J. H. (2010). EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, *57*(7), 1798–1806. https://doi.org/10.1109/TBME.2010.2048568

Lindal, P. J., & Hartig, T. (2013). Architectural variation, building height, and the restorative quality of urban residential streetscapes. *Journal of Environmental Psychology*, *33*, 26–36. https://doi.org/10.1016/j.jenvp.2012.09.003

Lloréns, R., Noé, E., Colomer, C., & Alcañiz, M. (2015). Effectiveness, usability, and cost-benefit of a virtual reality--based telerehabilitation program for balance recovery after stroke: A randomized controlled trial. *Archives of Physical Medicine and Rehabilitation*, *96*(3), 418–425.

Lorenzo, G., Lledó, A., Pomares, J., & Roig, R. (2016). Design and application of an immersive virtual reality system to enhance emotional skills for children with autism spectrum disorders. *Computers & Education*, *98*, 192–205.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain--computer interfaces: a 10 year update. *Journal of Neural Engineering*, *15*(3), 31005.

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., Alcañiz, M., & Valenza, G. (2018). Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports*, *8*(1), 13657. https://doi.org/10.1038/s41598-018-32063-4

Marín-Morales, J., Higuera-Trujillo, J. L., Juan, C. de, Llinares, C., Guixeres, J., Iñarra, S., & Alcañiz, M. (2019). Navigation comparison between a real and a virtual museum: time-dependent differences using a head mounted display. *Interacting with Computers*. https://doi.org/10.1093/iwc/iwz018

Marín-Morales, J., Torrecilla, C., Guixeres, J., & Llinares, C. (2017). Methodological bases for a new platform for the measurement of human behaviour in virtual environments. *DYNA*, *92*(1), 34–38. https://doi.org/http://dx.doi.org/10.6036/7963

McCall, C., Hildebrandt, L. K., Bornemann, B., & Singer, T. (2015). Physiophenomenology in retrospect: Memory reliably reflects physiological arousal during a prior threatening experience. *Consciousness and Cognition*, *38*, 60–70. https://doi.org/10.1016/j.concog.2015.09.011

McCall, C., Hildebrandt, L. K., Hartmann, R., Baczkowski, B. M., & Singer, T. (2016). Introducing the Wunderkammer as a tool for emotion research: Unconstrained gaze and movement patterns in three emotionally evocative virtual worlds. *Computers in Human Behavior*, *59*, 93–107. https://doi.org/10.1016/j.chb.2016.01.028

Meehan, M., Razzaque, S., Insko, B., Whitton, M., & Brooks, F. P. (2005). Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. *Applied Psychophysiology Biofeedback*, *30*(3), 239–258. https://doi.org/10.1007/s10484-005-6381-3

Meißner, M., & Oll, J. (2019). The promise of eye-tracking methodology in organizational research: A taxonomy, review, and future avenues. *Organizational Research Methods*, *22*(2), 590–617.

Mengoni, M., Germani, M., & Peruzzini, M. (2011). Benchmarking of virtual reality performance in mechanics education. *International Journal on Interactive Design and Manufacturing*, *5*(2), 103–117.

Mitchell, R. L. C., & Phillips, L. H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, *70*, 1–10. https://doi.org/10.1016/j.neuropsychologia.2015.02.018

Morinaga, A., Hara, K., Inoue, K., & Urahama, K. (2018). Classification between natural and graphics images based on generalized Gaussian distributions. *Information Processing Letters*, *138*, 31–34. https://doi.org/10.1016/j.ipl.2018.05.010

Mormann, F., Lehnertz, K., David, P., & E. Elger, C. (2000). Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D: Nonlinear Phenomena*, *144*(3), 358–369. https://doi.org/10.1016/S0167-2789(00)00087-7

Nabiyouni, M., Saktheeswaran, A., Bowman, D. A., & Karanth, A. (2015). Comparing the performance of natural, semi-natural, and non-natural locomotion techniques in virtual reality. *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, 3–10.

Nakisa, B., Rastgoo, M. N., Tjondronegoro, D., & Chandran, V. (2018). Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors. *Expert Systems with Applications*, *93*, 143–155. https://doi.org/10.1016/j.eswa.2017.09.062

Nanda, U., Pati, D., Ghamari, H., & Bajema, R. (2013). Lessons from neuroscience: form follows function, emotions follow form. *Intelligent Buildings International*, *5*(sup1), 61–78.

Nardelli, M., Valenza, G., Greco, A., Lanata, A., & Scilingo, E. P. (2015). Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, *6*(4), 385–394. https://doi.org/10.1109/TAFFC.2015.2432810

Naseer, N., & Hong, K. S. (2015). fNIRS-based brain-computer interfaces: a review. *Frontiers in Human Neuroscience*, *9*, 3.

Nescher, T., Huang, Y. Y., & Kunz, A. (2014). Planning redirection techniques for optimal free walking experience using model predictive control. *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, 111–118.

Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*(5827), 1002–1005. https://doi.org/10.1126/science.1136930

Nilsson, N. C., Serafin, S., Steinicke, F., & Nordahl, R. (2018). Natural Walking in Virtual Reality. *Computers in Entertainment*, *16*(2), 1–22. https://doi.org/10.1145/3180658

Noguchi, H., & Sakaguchi, T. (1999). Effect of illuminance and color temperature on lowering of physiological activity. *Applied Human Science*, *18*(4), 117–123.

Notzon, S., Deppermann, S., Fallgatter, A., Diemer, J., Kroczek, A., Domschke, K., Zwanzger, P., & Ehlis, A. C. (2015). Psychophysiological effects of an iTBS modulated virtual reality challenge including participants with spider phobia. *Biological Psychology*, *112*, 66–76. https://doi.org/10.1016/j.biopsycho.2015.10.003

Ochiai, H., Ikei, H., Song, C., Kobayashi, M., Takamatsu, A., Miura, T., Kagawa, T., Li, Q., Kumeda, S., Imai, M., & Miyazaki, Y. (2015). Physiological and Psychological Effects of Forest Therapy on Middle-Aged Males with High-Normal Blood Pressure. *International Journal of Environmental Research and Public Health*, *12*(3), 2532–2542.

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *Biomedical Engineering, IEEE Transactions On*, *1*(3), 230–236.

Park, K. M., Ku, J., Choi, S. H., Jang, H. J., Park, J. Y., Kim, S. I., & Kim, J. J. (2011). A virtual reality application in role-plays of social skills training for schizophrenia: a randomized, controlled trial. *Psychiatry Research*, *189*(2), 166–172.

Parsons, T. D. (2015). Virtual Reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*, 660.

Parsons, T. D., Courtney, C. G., & Dawson, M. E. (2013). Virtual reality Stroop task for assessment of supervisory attentional processing. *Journal of Clinical and Experimental Neuropsychology*, *35*(8), 812–826.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*(2), 181.

Peng, C. K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, *5*(1), 82–87. https://doi.org/10.1017/CBO9781107415324.004

Peperkorn, H. M., Alpers, G. W., & Mühlberger, A. (2014). Triggers of fear: Perceptual cues versus conceptual information in spider phobia. *Journal of Clinical Psychology*, *70*(7), 704–714. https://doi.org/10.1002/jclp.22057

Picard, R. W. (1997). *Affective computing*. MIT press.

Picard, R. W. (2003). Affective Computing: Challenges. *International Journal of Human-Computer Studies*, *59*(1), 55–64. https://doi.org/10.1007/BF01238028

Pincus, S., & Viscarello, R. (1992). Approximate Entropy A regularity measure for fetal heart rate analysis. *Obstetrics & Gynecology*, *79*(2), 249–255.

Piwek, L., Ellis, D. A., Andrews, S., & Joinson, A. (2016). The Rise of Consumer Health Wearables: Promises and Barriers. *PLoS Medicine*, *13*(2), 1–9. https://doi.org/10.1371/journal.pmed.1001953

Portman, M. E., Natapov, A., & Fisher-Gewirtzman, D. (2015). To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. *Computers, Environment and Urban Systems*, *54*, 376–384. https://doi.org/10.1016/j.compenvurbsys.2015.05.001

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715–734. https://doi.org/10.1017/S0954579405050340

Powers, M. B., & Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, *22*(3), 561–569. https://doi.org/10.1016/j.janxdis.2007.04.006

Prokasy, W. (2012). *Electrodermal activity in psychological research*. Elsevier.

Rangaraju, N., & Terk, M. (2001). Framework for immersive visualization of building analysis data. In E. Banissi, F. Khosrowshahi, M. Sarfraz, & A. Ursyn (Eds.), *Fifth International Conference on Information Visualisation* (pp. 37–42). IEEE.

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition*, *27*, 741–750.

Riches, S., Elghany, S., Garety, P., Rus-Calafell, M., & Valmaggia, L. (2019). Factors Affecting Sense of Presence in a Virtual Reality Social Environment: A Qualitative Study. *Cyberpsychology, Behavior, and Social Networking*, *22*(4). https://doi.org/10.1089/cyber.2018.0128

Richman, J., & Moorman, J. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol*, *278*, H2039–H2049.

Riecke, B. E., Bodenheimer, B., McNamara, T. P., Williams, B., Peng, P., & Feuereissen, D. (2010). Do We Need to Walk for Effective Virtual Reality Navigation? Physical Rotations Alone May Suffice. In C. Hölscher, T. F. Shipley, M. Olivetti Belardinelli, J. A. Bateman, & N. S. Newcombe (Eds.), *Spatial Cognition 2010. Lecture Notes in Computer Science* (pp. 234–247). Springer.

Riecke, B. E., LaViola Jr, J. J., & Kruijff, E. (2018). 3D user interfaces for virtual reality and games: 3D selection, manipulation, and spatial navigation. *ACM SIGGRAPH 2018 Courses*, 13.

Riener, A., Ferscha, A., & Aly, M. (2009). Heart on the road: HRV analysis for monitoring a driver's affective state. *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 99–106.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., & Alcañiz, M. (2007). Affective Interactions Using Virtual Reality: The Link between Presence and Emotions. *CyberPsychology & Behavior*, *10*(1), 45–56. https://doi.org/10.1089/cpb.2006.9993

Roberts, N. A., Tsai, J. L., & Coan, J. A. (2007). Emotion elicitation using dyadic interaction task. In *Handbook of Emotion Elicitation and Assessment* (pp. 106–123).

Rohrmann, B., & Bishop, I. D. (2002). Subjective responses to computer simulations of urban environments. *Journal of Environmental Psychology*, *22*, 319–331. https://doi.org/10.1006/jevp.2001.0206

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1997). Navigating buildings on "desk-top" virtual environments: experimental investigations using extended navigational experience. *Journal of Experimental Psychology: Applied*, *3*(2), 143–159.

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1999). Navigating Large-Scale Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *8*(2), 157–168.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(3), 273–294.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.

Sanchez-Vives, M. V. V., & Slater, M. (2005). From presence to consciousness through Virtual Reality. *Nature Reviews Neuroscience*, *6*(4), 332–339.

Sas, C., & O'Hare, G. M. (2003). Presence Equation: An Investigation into Cognitive Factors Underlying Presence. *Presence: Teleoperators and Virtual Environments*, *12*(5), 523–537.

Schmitt, N. (1994). Method bias: The importance of theory and measurement. *Journal of Organizational Behavior*, *15*(5), 393–398.

Schöllkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Comput*, *12*(5), 1207–1245. https://doi.org/doi:10.1162/089976600300015565

Schuller, B. W. (2018). Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. *Communications of the ACM*, *61*(5), 90–99.

Sejima, K. (1996). Kazuyo Sejima 1988-1996. *El Croquis*, *15*(77).

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2009). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 410–417.

Sharma, G., Kaushal, Y., Chandra, S., Singh, V., Mittal, A. P., & Dutt, V. (2017). Influence of landmarks on wayfinding and brain connectivity in immersive virtual reality environment. *Frontiers in Psychology*, *8*(JUL), 1–12. https://doi.org/10.3389/fpsyg.2017.01220

Shiban, Y., Diemer, J., Brandl, S., Zack, R., Mühlberger, A., & Wüst, S. (2016). Trier Social Stress Test in vivo and in virtual reality: Dissociation of response domains. *International Journal of Psychophysiology*, *110*, 47–55.

Siriaraya, P., & Ang, C. S. (2019). The Social Interaction Experiences of Older People in a 3D Virtual Environment. In S. Sayago (Ed.), *Perspectives on Human-Computer Interaction Research with Older People* (pp. 101–117). Springer International Publishing. https://doi.org/10.1007/978-3-030-06076-3_7

Slater, M., Usoh, M., & Steed, A. (1994). Depth of Presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *3*(2), 130–144.

Slater, M., & Usoh, M. (1994). Body centred interaction in immersive virtual environments. *Artificial Life and Virtual Reality*, 125–148. http://web.cs.wpi.edu/~gogo/hive/papers/Slater_Presence_1994_Depth.pdf

Slater, M. (2003). A Note on Presence Terminology. *Emotion*, *3*, 1–5.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Usoh, M., & Steed, A. (1994). Depth of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *3*(2), 130–144. https://doi.org/10.1162/pres.1994.3.2.130

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1007/s10750-008-9541-7

Soleymani, M., Pantic, M., & Pun, T. (2015). Multimodal emotion recognition in response to videos (Extended abstract). *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, *3*(2), 491–497. https://doi.org/10.1109/ACII.2015.7344615

Stam, C. J. (2005). Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clinical Neurophysiology*, *116*(10), 2266–2301. https://doi.org/10.1016/j.clinph.2005.06.011

Stamps III, A. E. (1999). Physical Determinants of Preferences for Residential Facades. *Environment and Behavior*, *31*(6), 723–751. https://doi.org/10.1177/00139169921972326

Stamps III, A. E. (1990). Use of photographs to simulate environments: a meta-analysis. *Percept. Mot. Ski.*, *71*, 907–913.

Stein, J. P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, *160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Steuer, J. (1992). Defining Virtual Reality: dimensions determining telepresence. *Journal of Communication*, *42*(4), 73–93.

Tarvainen, M. P., Niskanen, J. P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV - Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Tarvainen, M. P., Ranta-aho, P. O., & Karjalainen, P. A. (2002). An advanced detrending method with application to HRV analysis. *IEEE Transactions on Biomedical Engineering*, *49*(2), 172–175. https://doi.org/10.1109/10.979357

Templeman, J. N., Denbrook, P. S., & Sibert, L. E. (1999). Virtual locomotion: Walking in place through virtual environments. *Presence*, *8*(6), 598–617.

Thibault, R. T., MacPherson, A., Lifshitz, M., Roth, R. R., & Raz, A. (2018). Neurofeedback with fMRI: A critical systematic review. *NeuroImage*, *172*(September 2017), 786–807. https://doi.org/10.1016/j.neuroimage.2017.12.071

Tregillus, S., Al Zayer, M., & Folmer, E. (2017). Handsfree omnidirectional VR navigation using head tilt. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 4063–4068.

Tregillus, S., & Folmer, E. (2016). Vr-step: Walking-in-place using inertial sensing for hands free navigation in mobile vr environments. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1250–1255.

Tsai, C. F., Yeh, S. C., Huang, Y., Wu, Z., Cui, J., & Zheng, L. (2018). The Effect of Augmented Reality and Virtual Reality on Inducing Anxiety for Exposure Therapy: A Comparison Using Heart Rate Variability. *Journal of Healthcare Engineering*, *2018*.

Tsunetsugu, Y., Miyazaki, Y., & Sato, H. (2005). Visual effects of interior design in actual-size living rooms on physiological responses. *Building and Environment*, *40*(10), 1341–1346. https://doi.org/10.1016/j.buildenv.2004.11.026

Ulrich, R. (1984). View through a window may influence recovery from surgery. *Science*, *224*(4647), 420–421. https://doi.org/10.1126/science.6143402

Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., & Brooks, F. P. (1999). Walking > walking-in-place > flying, in virtual environments. In W. Waggenspack (Ed.), *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99* (pp. 359–364). ACM Press/Addison-Wesley Publishing. https://doi.org/10.1145/311535.311589

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments*, *9*(5), 497–503. https://doi.org/10.1162/105474600566989

Valenza, G., Greco, A., Gentili, C., Lanata, A., Sebastiani, L., Menicucci, D., Gemignani, A., & Scilingo, E. P. (2016). Combining electroencephalographic activity and instantaneous heart rate for assessing brain–heart dynamics during visual emotional elicitation in healthy subjects. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2067), 20150176. https://doi.org/10.1098/rsta.2015.0176

Valenza, G., Citi, L., Lanatá, A., Scilingo, E. P., & Barbieri, R. (2014). Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. *Scientific Reports*, *4*, 4998. https://doi.org/10.1038/srep04998

Valenza, G., Lanata, A., & Scilingo, E. P. (2012). The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE Transactions on Affective Computing*, *3*(2), 237–249. https://doi.org/10.1109/T-AFFC.2011.30

Valenza, G., Nardelli, M., Lanata, A., Gentili, C., Bertschy, G., Kosel, M., & Scilingo, E. P. (2016). Predicting Mood Changes in Bipolar Disorder Through Heartbeat Nonlinear Dynamics. *IEEE Journal of Biomedical and Health Informatics*, *20*(4), 1034–1043. https://doi.org/10.1109/JBHI.2016.2554546

Valenza, G., Nardelli, M., Lanatà, A., Gentili, C., Bertschy, G., Paradiso, R., & Scilingo, E. P. (2014). Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis. *IEEE Journal of Biomedical and Health Informatics*, *18*(5), 1625–1635. https://doi.org/10.1109/JBHI.2013.2290382

van der Ham, I. J., Faber, A. M., Venselaar, M., van Kreveld, M. J., & Löffler, M. (2015). Ecological validity of virtual environments to assess human navigation ability. *Frontiers in Psychology*, *6*, 637.

Vartanian, O., Navarrete, G., Chatterjee, A., Brorson, L., Leder, H., & Modroño, C. (2013). Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture. *Proceedings of the National Academy of Sciences*, *110*, 1–8. https://doi.org/10.1073/pnas.1301227110

Vecchiato, G., Jelic, A., Tieri, G., Maglione, A. G., De Matteis, F., & Babiloni, F. (2015). Neurophysiological correlates of embodiment and motivational factors during the perception of virtual architectural environments. *Cognitive Processing*, *16*(1), 425–429. https://doi.org/10.1007/s10339-015-0725-6

Villa, C., & Labayrade, R. (2012). Validation of an online protocol for assessing the luminous environment. *Lighting Research and Technology*, *0*, 1–20.

Vince, J. (2004). *Introduction to virtual reality*. Media, Springer Science & Business.

Welch, P. D. (1967). The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Aver. aging Over Short, Modified Periodograms. *IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS*, *15*(2), 70–73.

Wilson, P. T., Nguyen, K., Harris, A., & Williams, B. (2014). Walking in place using the Microsoft Kinect to explore a large VE. *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, 27–33.

Wu, D., Courtney, C. G., Lance, B. J., Narayanan, S. S., Dawson, M. E., Oie, K. S., & Parsons, T. D. (2010). Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task. *IEEE Transactions on Affective Computing*, *1*(2), 109–118. https://doi.org/10.1109/T-AFFC.2010.12

Xu, J., Mitra, S., Van Hoof, C., Yazicioglu, R., & Makinwa, K. A. A. (2017). Active Electrodes for Wearable EEG Acquisition: Review and Electronics Design Methodology. *IEEE Reviews in Biomedical Engineering*, *3333*(c), 1–1. https://doi.org/10.1109/RBME.2017.2656388

Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators, B: Chemical*, *212*(September), 353–363. https://doi.org/10.1016/j.snb.2015.02.025

Yavrucuk, I., Kubali, E., & Tarimci, O. (2011). A low cost flight simulator using virtual reality tools. *IEEE Aerospace and Electronic Systems Magazine*, *26*(4), 10–14.

Yeom, D., Choi, J. H., & Zhu, Y. (2017). Investigation of the Physiological Differences between Immersive Virtual Environment and Indoor Enviorment in a Building. *Indoor Adn Built Enviornment*, *0*(0), Accept. https://doi.org/10.1177/1420326X17731945

Yildirim, K., Hidayetoglu, M. L., & Capanoglu, A. (2011). Effects of interior colors on mood and preference: comparisons of two living rooms. *Perceptual and Motor Skills*, *112*(2), 509–524.

Zangeneh Soroush, M., Maghooli, K., Setarehdan, S. K., & Motie Nasrabadi, A. (2018). A Review on EEG Signals Based Emotion Recognition. *International Clinical Neuroscience Journal*, *4*(4), 118–129. https://doi.org/10.15171/icnj.2017.01

Zhang, S., Zhang, Y., Sun, Y., Member, I., & Thakor, N. (2017). *Graph theoretical analysis of EEG functional network during multi-workload flight simulation experiment in virtual reality environment*. 3957–3960.

Zhao, Q. Bin, Zhang, L. Q., & Cichocki, A. (2009). EEG-based asynchronous BCI control of a car in 3D virtual reality environments. *Chinese Science Bulletin*, *54*(1), 78–87. https://doi.org/10.1007/s11434-008-0547-3

Zimmer, P., & Wu, C. (2019). Same same but different? Replicating the real surroundings in a virtual Trier Social Stress Test (TSST-VR) does not enhance presence or the psychophysiological stress response. *Physiology & Behavior*, 112690.

In recent years the scientific community has significantly increased its use of virtual reality (VR) technologies in human behaviour research. Among the fields that has strongly emerged in the last decade is affective computing, which combines psychophysiology, computer science, biomedical engineering and artificial intelligence in the development of systems that can automatically recognize emotions. The progress of affective computing is especially important in human behaviour research due to the central role that emotions play in many background processes, such as perception, decision-making, creativity, memory and social interaction. Several studies have tried to develop a reliable methodology to evoke and automatically identify emotional states using objective physiological measures and machine learning methods. However, the majority of previous studies used images, audio or video to elicit emotional statements. The main objective of this thesis is, using psycho-physiological and behavioural responses in combination with machine learning methods, and by performing a direct comparison between a real and virtual environment, to validate immersive VR as an emotion elicitation tool.