Department of Psychology

International Doctor of Philosophy course
in Psychology and Social Neuroscience
XXXII cycle

Towards real world neuroscience: the impact of virtual and augmented reality techniques on the study of human performance and sense of presence.

FINAL DISSERTATION

Candidate

Dott. Matteo Marucci

Supervisor:

Prof.ssa Viviana Betti

Reviewers:

Prof.ssa Sara Bagassi

Prof. Pietro Cipresso

Table of Contents

Preface

Virtual and Augmented reality systems are rapidly evolving. Their technical development is accompanied by a spreading of applications in fields such as architecture, design, digital marketing, education, entertainment, gaming, robotics, fine arts, military and medicine training, engineering and research. Those tools, requiring 8 years ago investment in the order of tens of thousands of US dollars (Bohil, Alicea, & Biocca, 2011) can boast nowadays easier economical accessibility and predicted market size in the next 5 years ranging from 50 to 300 billion (summary of sources). Having regard to the forthcoming large-scale adoption, it is beyond necessary, from a neuroscientific perspective, to investigate the impact of VR and AR technology on human cognition and performance. Therefore, monitoring the interplay between these devices and our brain dynamics in the short and long term and across their development leaps will be crucial to ensure their correct and safe usage in research and beyond. This could pave the way to new research paradigms and commercial applications, granting users the right experience for their needs whether they will need to train for a complex surgical intervention along with an artificial agent or just get the maximum excitement from an immersive sensorial journey. The first introductory chapter will discuss the contribution of virtual and augmented reality as tools for the study of human cognition and performance and the need, arising from an "embodied" vision of cognition, for a transition towards experimental paradigms that allow neuroscientist to study subjects that interact in a realistic environment through spontaneous and natural actions.

1. INTRODUCTION

1.1 Grounded Cognition and Virtual reality: Towards real-world neuroscience

Over the last few decades, in the field of neuroscience, the empirical research related to the theoretical framework of Grounded Cognition (Barsalou, 2008) has been expanded. This approach, diametrically opposed to Cartesian dualism, is in stark contrast to the representational and computational theories that dominated the panorama of cognitive psychology from the 1950s to the 1980s (Fodor, 1983; Pylyshyn, 1984) and rejects their conception of the mind as a computer software based on amodal mental representations. On the contrary, the theoretical proposal of Grounded Cognition emphasizes the causal and constitutive role of the sensorimotor system and the interaction with the physical and social environment in the development of cognitive processes and human behavior. A fundamental impulse for the diffusion of GC theories in the psychological and neuroscientific panorama and in particular in the development of theories on social cognition came from the discovery of the mirror neuron system. Originally identified in the premotor cortex of macaques while observing an experimenter grasping food (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti & Craighero, 2007; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), this system is also present in humans and consists of a neuronal population extended to different brain areas (Rizzolatti & Craighero, 2004). This network is able to activate itself, even if with different intensity (Rochat et al., 2010), both after the production and compared to the mere observation of action. Following its discovery, some studies have shown that the activity of a part of this network is also involved in understanding the intentions that motivate the perceived action (Iacoboni et al., 2005) and constitutes also the neurophysiological mechanism behind the prediction of the purpose of an action (Fogassi et al., 2005). In other words, the sensorimotor system is able, through a simulation of the observed action (mirror neurons) and according to the properties of the objects (canonical neurons) with which it interacts (Murata et al., 1997; Sakata, Taira, Murata, & Mine, 1995) to make inferences about the purpose of a movement (Kohler et al., 2002) and its possible outcome. This mechanism is not only valid for actions, but we are also witnessing an internal simulation of the bodily state and emotions of the agent observed. From this point of view, watching facial expressions that encode a certain emotion produces in the observer the activation of the same brain areas of those involved in the first person (Wicker et al., 2003). The circuits of mirror neurons would, therefore, allow us to put ourselves in the shoes of the other not only on the motor plane but also on the emotional one in a vision that Gallese defines as "embodied simulation" (Gallese, 2005). According to the author, this mechanism is the basis of the intersubjective relationship and makes it possible for two individuals to find in the other the relationship of similarity that Stein defines as empathy (W. Stein, 2012). The results of the scientific resonance that this discovery has constituted, extend even further (e.g. cognitive development and autism), nevertheless, its greatest contribution has certainly been to shift the attention of the scientific community, involved in the study of cognition, towards the body and motor domain. To date, this vision of cognition, as *grounded* to the opportunities offered by the physical, perceptual and motor characteristics of the human body to interact with the environment and other agents, has numerous theoretical declinations (e.g., embodied, extended, enacted and situated cognition) that are characterized by the weight attributed respectively to the body or the environment in shaping our cerebral architecture (Barsalou, 2008; Clark & Chalmers, 1998; Noë & Noë, 2004; Wilson, 2002). These approaches, though relatively distant on some positions, are united by the need that a study of cognition concretely placed in the domain of action, perception and interaction poses: the need to study the brain in a context as natural as possible.

From this perspective, we will understand the importance of virtual and augmented reality technology in the implementation of paradigms of "real-world neuroscience" that investigate the brain in its natural habitat, the realistic interaction.

1.2 Virtual Reality

The term virtual reality (VR) refers to a computer-generated environment (Parsons, Gaggioli, & Riva, 2017), that aims to achieve a realistic sensory substitution. The goal of an ideal VR setup, in fact, is to immerse the user in a virtual scenario that, through the concurrent stimulation of multiple sensory channels, will result in an artificial experience almost indistinguishable from reality. The state of the art in virtual reality is represented by Head-Mounted Displays (HMD). Although also other technology can be used to render virtual 3d environments, like a CAVE (CAVE Automatic Virtual Environment) system (Cruz-Neira, Sandin, & DeFanti, 1993), using projectors and 3d glasses, their price and maintenance costs are usually sufficient reasons to opt for a more portable and less expensive solution, like a helmet.

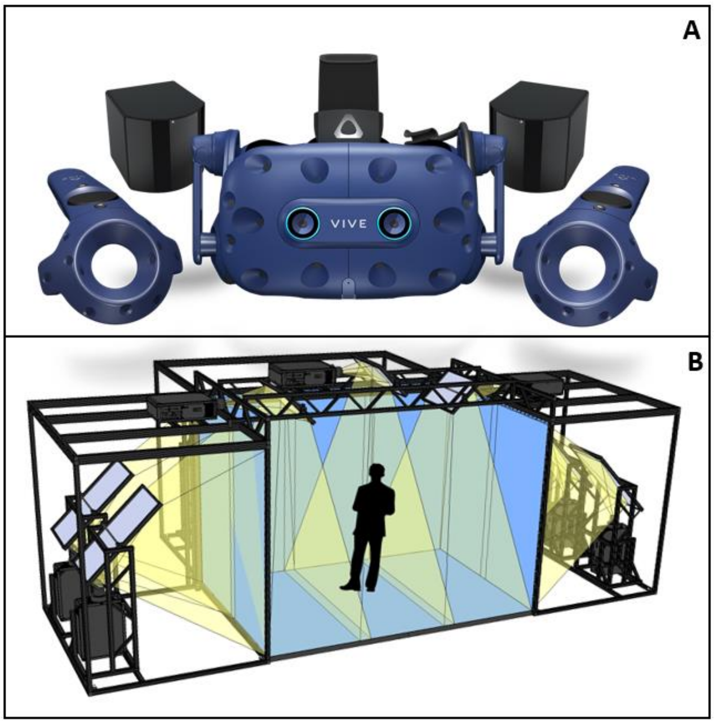

Figure 1 | A) HTC Vive eye pro helmet with depth cameras and eye tracking. B) Cave apparatus employing mirrors, projectors and special 3d googles to give the user the illusion of a 3d space.

Modern head-mounted displays have built-in audio devices, internal tracking sensors and external cameras that track the helmet and the controllers in space. Additionally, In the last decade more and more companies appearing on the market are developing virtual reality control systems such as gloves and suites with haptic and force feedback. Nowadays the creation of virtual experience is becoming increasingly simpler. A virtual world, in fact, can be easily generated with little expertise using free software usually employed in the gaming industry (e.g., Unity3d or Unreal Engine) and free 3d models found on online repositories.

1.2.1 Virtual Reality helmets

Another, more technical, definition (Cruz-Neira et al., 1993), sees VR as a system capable of providing real-time viewer-centered head tracking perspective with a large field of view, interactive control, and binocular display. Inside modern helmets, an Inertial Measurement Unit (IMU) is usually found. The IMU is fundamental for head tracking and relies on an accelerometer and a gyroscope to give the user 6 degrees of freedom tracking in the virtual space. The field of view is defined as the total angular size of the image visible to both the eyes. On average, the human vertical FOV is approximately 130 degrees, while the horizontal binocular FOV is 200, out of which 120 degrees is a binocular overlap(The binocular overlap is especially important for stereopsis). One of the latest VR Headsets of the market (www.Pimax.com) sports a FOVof 200 degrees and a resolution up to 3840x2160x2. Using a binocular display, each eye gets a separate image creating a stereoscopic view. These display types provide the most depth cues and a sense of immersion however, they are the heaviest, most complex and computationally intensive displays. For this reason, only recently we are seeing on the market, standalone Head-Mounted Displays (HMD) that don't need a computer and its powerful GPU to render the stereoscopic images. A VR HMD is defined as fully immersive as it completely blocks the user's view of the outside world.

1.2.2 Presence and immersion in VR

The use of virtual reality is spreading at great speed in research fields such as psychology and cognitive neuroscience. This diffusion can be understood by analyzing the possibility that this technique offers in the design of experimental paradigms that are both realistic (Meehan, Insko, Whitton, & Brooks Jr, 2002) and controlled at the same time. Through a computer

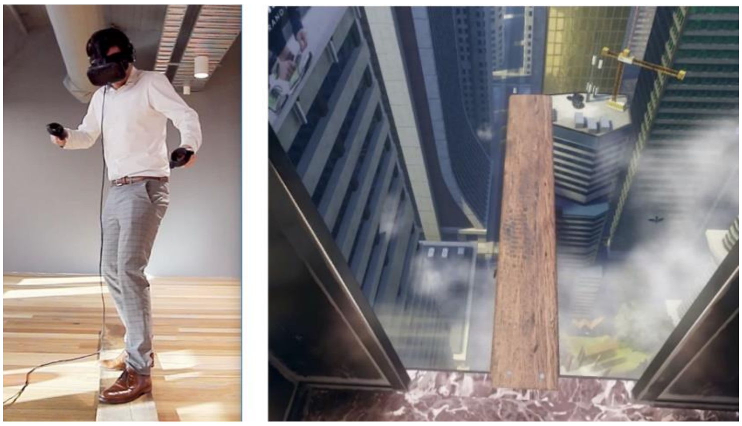

***Figure 2* | A setup similar to that used by Meehan and colleagues (2002) can be taken as an example of ecological, engaging and controlled paradigm.**

programmed simulation, then projected into a helmet for virtual reality, it is possible to immerse an individual in contexts simulated at the detail level. On the contrary, is possible

to create situations where the laws of physics are distorted. For example, Mast and Oman (Mast & Oman, 2004) employed a virtual reality paradigm to explore the illusions of visual reorientation. This phenomenon is experienced by astronauts when the perceived identity of a surface change due to the rotation of the entire field of view. This is hard to replicate in real life, as we are usually immersed in visual cues in our environment that help us to navigate. In both cases VR allows scientists to draw behavioral information or gather physiological data while participants watch move and interact in a much more ecological condition than a task performed in a laboratory setting. Often, in fact, in cognitive, affective or even in social neuroscience subjects are evaluated while pushing keyboard buttons in front of a screen in a dark room.

Among the characteristics that oriented researchers towards VR, one cannot avoid mentioning the two constructs that have most contributed to act as a bridge to neuroscience, namely immersion (Mel Slater, 1999), and the sense of presence. The first term refers to the degree of sensory fidelity achieved by the technological setup (Mel Slater, Brogni, & Steed, 2003) and is fundamental to evoke the illusion of being present in the virtual world (Cummings & Bailenson, 2015). One need only to consider the technological extent reached during the last 5 years by VR devices in domains such as field of view (from 90° to 200°), display resolution (from 1280 x 800 to 3840 x 2160), pixel density (from 215 PPI to 800) and refresh rate (from 60 to 120 HZ) to grasp the pace of technological progress and the consequent immersion provided.

The sense of presence, on the other hand, is the illusion that between the user and his experience there is no mediation (Lombard & Ditton, 1997) of an artificial screen, so he "feels there".

Slater (M. Slater, 2018) defines it as a perceptual but non-cognitive illusion, in which the perceptual system, for example, identifies a threat (the virtual precipice) and the brain-body system reacts automatically and rapidly (this is the safest thing to do), while the cognitive system activates relatively later and concludes "But I know this is not real". Certainly, this illusion can arise in very different contexts due to its intrinsically multidimensional nature (Kalawsky, 2000). However, only a paucity of studies has investigated directly the impact of multisensory stimulation on the sense of presence (Cooper et al., 2018; Dinh, Walker, Hodges, Chang, & Kobayashi, 1999; Hendrix & Barfield, 1996). Nevertheless, interconnected cognitive and psychological factors such as the involvement in the digital experience or the sense of agency and ownership experienced towards a virtual body can play an important role in shaping the sense of perceived presence (Skarbez, Brooks, & Whitton, 2017).

1.2.3 Body and Visual Perception in VR

The sense of ownership (Gallagher, 2000), refers to the feeling of belonging that we perceive with respect to our body or its parts, while agency (Haggard & Chambon, 2012) refers to the experience of control that we experience in relation to an action. Ownership and Agency are both concepts that relate to the theoretical domain of Embodiment, or "Incorporation". In this field, numerous evidences, have shown how it is possible to induce in experimental subjects a sense of incorporation towards a virtual avatar (IJsselsteijn, de Kort, & Haans, 2006) following the multisensory approach used by Botvinick & Cohen to obtain the famous "rubber hand illusion" (Botvinick & Cohen, 1998). Virtual reality, for his multimodal nature, is the ideal tool to carry out multisensory induced illusions through the use of body manipulation techniques (V. Petkova & Ehrsson, 2008; Mel Slater, Pérez Marcos, Ehrsson, & Sanchez-Vives, 2009; Mel Slater, Spanlang, Sanchez-Vives, & Blanke, 2010). One the other hand virtual reality gave precious insight also in the field of space perception. For example, using an immersive environment researcher can study the interplay between movement and illusions investigated only in stationary situations, such as change blindness (Bruder, Wieland, Bolte, Lappe, & Steinicke, 2013) automaticity and inattention blindness (Suma, Clark, Finkelstein, & Wartell, 2010). Additionally, VR enables a concurrent and precise control of the photometric relationship of the environment and the stimuli position and orientation. This measure prevents light from bouncing from environmental surfaces from interfering with phenomenon like the Simultaneous Lightness Contrast (SLC)(Soranzo, Lugrin, & Wilson, 2013).

1.2.4 Performance and Workload in VR

Despite a pervasive belief that the sense of presence is causally related to performance (Mel Slater, Linakis, Usoh, & Kooper, 1996; Witmer, Jerome, & Singer, 2005), there is no solid evidence to support this statement. While researchers have recently started to point to a significant correlation between the two factors (i.e. presence and performance(Cooper et al., 2018; Ma & Kaber, 2006; Stevens & Kincaid, 2015), however, as stated by Nash and colleagues (Nash, Edwards, Thompson, & Barfield, 2000), one could argue that the relation between a construct like presence and a measured variable like performance, is easily influenced by external factors and thus is not always easy to draw conclusions. In fact, even when a correlation between the sense of presence and participants' performances is found, it is important to mention that it could be a result worth analysing considering also the cognitive resources depleted in the process. It is easy to imagine, how the same performance achieved in a task could be still obtained at the expenses of a very different mental cost, also depending on the amount of data the brain has to process. Indeed, in operational contexts where the goal of a virtual reality application is not related to mere entertainment, then it is crucial to monitor the workload experienced by the user, even regardless of the performances.

1.3 Augmented Reality

In Augmented Reality (AR) we can witness an enrichment of the information at our disposal thanks to digital data displayed on a virtual layer or spatialized sounds in the environment. Within an AR system, the virtual content is arranged in real-time (Azuma et al., 2001) around users in the same spatial coordinates as the physical objects they have around them and is possible to employ techniques of interaction that simulate those between the physical world and objects (e.g., move, grab and leave). According to Milgram et al. (Milgram, Takemura, Utsumi, & Kishino, 1995), AR places between reality (real environment) and virtuality (virtual environment) on the reality-virtuality continuum. While this definition can refer to the enrichment of reality through the five senses, it is often used to indicate only the visual domain. Most of AR technology can be an Optical see-through or Video see-through (Bimber & Raskar, 2006). In Optical see-through glasses, the user views reality directly through optical elements such as holographic waveguides and other systems that enable graphical overlay on the real world. Microsoft's Hololens, Magic Leap One and the Google Glass are recent examples of optical see-through smart glasses, however, AR applications can be easily programmed using the camera of a modern smartphone, also known as HandHeld Augmented Reality. Using a Video see-through instead, the user views reality that is first captured by one or two cameras mounted on the display. These camera views are then combined with computer-generated imagery for the user to see. The HTC Vive VR headset (Fig. 1), for example, has an inbuilt pair of cameras which is often used for creating AR experiences on the device. The sensory element superimposed on the reality can have an additive role, namely, to provide additional details, or subtractive, and act as a filter or a mask. The virtual elements can partially or completely overlap with the original view of an object. In augmented reality based on overlapping, object recognition plays a fundamental role because the application cannot replace the original view if it is unable to determine coordinates and boundaries. An example of augmented reality overlap could be as follows: by downloading an application and scanning the selected pages in a printed or digital catalog, users can visualize virtual furniture in their apartment. Augmented reality, unlike VR, requires that the viewer has an immersive experience in a concrete context, altering the perception of the environment through superimposed, visual, auditory or tactile elements but without losing the feeling of being in a physically real context. Nevertheless, as seen for VR environments in the previous paragraphs, also in augmented reality a realistic virtual environment can elicit a sense of presence (Juan et al., 2006). Additionally, augmented reality systems can be triggered using a marker as well as by GPS coordinates or by sound detection. For example, Marker-based augmented reality employs image recognition using a camera and a visual marker, such as a QR code, to produce a 3D output only when the marker is detected. As noted by Cipresso and colleagues (Cipresso, Chicchi Giglioli, Alcañiz Raya, & Riva, 2018), despite AR recent appearance in applied context, given its technological immaturity, its technology contribution is already being explored in many fields such as surgery (Ha & Hong, 2016; Thomas, 2016), Architecture (Lin and Hsu, 2017), maintenance (Palmarini, Erkoyuncu, Roy, & Torabmostaedi, 2018), education (Bacca et al., 2014) and air traffic control (P. Aricò et al., 2018; Bagassi, De Crescenzio, Lucchi, & Masotti, 2016).

1.4 Aims

The present work explored experimentally the contribution of multimodal VR and AR systems both in a laboratory environment and in an applied context.

The first study will explore the usage of virtual reality in combination with EEG to investigate the dynamics of multisensory stimuli combinations and perceptual load on human workload and performance. Multisensory integration effects on attention, in fact, have traditionally been studied in a laboratory environment, far from our everyday multimodal experience. This work instead, researched whether a virtual immersive system can reproduce previous findings under more ecological circumstances. Therefore, we asked subjects to drive on a virtual racetrack using pedals and a real steering wheel with the goal of collecting 3D targets appearing on the road. They were cued by auditory, vibrotactile or both signals, in two different conditions of multimodal perceptual load: Low and High. In addition to assessing their performance, we collected subjective ratings on workload and the sense of presence derived by the virtual experience. Furthermore, we recorded skin conductance and electroencephalography to select the spectral features related to the workload.

In the second and third studies (chapters 3 and 4) through an augmented reality setup, we investigated the contribution of an extra layer of information applied to a real-world context such as the air traffic control domain, in combination with electrophysiological measures. The project had the specific aim to evaluate, through neurophysiological measures, the impact of a novel solution such as a remote air traffic control tower. In order to improve cost-efficiency in air traffic provision, in fact, many countries are currently considering the remote tower air traffic control. The general idea is that in the future, the air traffic controller (ATCO) will no more need to be present in the actual control tower, but he will be able to operate from another location. With respect to standard control towers, this new solution will allow monitoring the traffic in small airports thanks to high-resolution cameras, advanced sensors and radio transmissions.

2. STUDY 1 - Multisensory Integration and Perceptual Load: The Impact on Performances, Workload, and Presence in Virtual Reality.

2.1 Abstract

Evidence indicates that multisensory integration, although subject to perceptual load modulation, enhances attentional capture. This effect, however, has always been investigated in front of a screen, far from our everyday multimodal experience. Since virtual reality can help scientists to bridge the gap between experimental control and ecological validity, we devised an immersive driving task in combination with electroencephalography to investigate *the dynamics of multimodal cueing and multisensory load modulation on attention, mental workload, and presence.* Subjects drove on the virtual track collecting 3D targets appearing on the road, cued by auditory, vibrotactile or both signals, in two different conditions of multimodal perceptual load: Low and High. We assessed their performance and recorded skin conductance and electroencephalography to select the spectral features related to their workload. Furthermore, we collected subjective ratings on the workload and the sense of presence derived from the virtual experience. In the High load task, we found higher performances associated with the bimodal and trimodal stimulations and an increase in the presence and in the GSR. In the trimodal condition (Visual-Audio-Vibrotactile) participants felt an overall increase in the sense of presence and a decrease in workload, both subjective and EEG based, regardless of the load.

2.2 Introduction

Although the visual channel is the major source of information on which we rely to navigate in the external environment, natural stimuli are typically multimodal.

The process through which the brain combines information from independent, but temporally aligned, signals derived from multiple sensory sources (e.g., vision, audition)(L. E. Miller, Longo, & Saygin, 2017; Barry E. Stein, Stanford, & Rowland, 2014) into a coherent representation is termed multisensory integration. The spatiotemporal concordance of visual and auditory signals derived by this integration is known to enhance neuronal responses (Meredith & Stein, 1983; R. L. Miller, Pluta, Stein, & Rowland, 2015; Barry E Stein & Meredith, 1993), *BOLD* signal (Stevenson, Geoghegan, & James, 2007) and ERPs activity (Talsma, Doty, & Woldorff, 2007). This enhanced activation is also reflected at the behavioral level. For instance, multimodal stimuli induce faster and more accurate responses than the summed probability of two unisensory stimuli (Lunn, Sjoblom, Ward, Soto-Faraco, & Forster, 2019; Pannunzi et al., 2015)Also, multisensory cues reduce visual search latencies, not only when spatially informative and colocalized with visual targets (Ngo & Spence, 2010; Charles Spence, Pavani, & Driver, 2000), but also when task-irrelevant or uninformative (Matusz & Eimer, 2011; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). Benefits from multisensory stimuli are also found at a perceptive level. A broad-band auditory stimulus significantly enhances the perceived intensity of a visual stimulus (Barry E. Stein, London, Wilkinson, & Price, 1996) also, task-irrelevant tactile stimulations increase auditory intensity ratings (Gillmeister & Eimer, 2007). However, evidence shows that the facilitation effects of the multisensory integration seem to partially rely on the stimuli to be attended (Talsma et al., 2007).

When evaluating whether multisensory integration can modulate exogenous orienting mechanisms, an important role is played by the perceptual/attentional load (N. Lavie, Hirst, de Fockert, & Viding, 2004; Santangelo & Spence, 2007). In fact, during a peripheral spatial cueing task the subjects' performance is not particularly benefitting from multisensory integration, compared to the unisensory stimulation. By contrast, under a condition of concurrent visual perceptual load (e.g. Rapid Sequential Visual Presentation task - RSVP) the effectiveness of multimodal stimuli improved in enhancing subjects' performance, compared to unisensory signals (Santangelo & Spence, 2007). However, a limit of previous studies consists of using only the visual source to modulate the perceptual load. This is in line with Spence and Santangelo's remark (C. Spence & Santangelo, 2009) on the great number of studies that failed to shed light on the modulatory effects of load on multisensory stimuli using a silent and dark laboratory environment.

Real-life sensory processing and behavior are complex and dynamic, involving interactions of many context-dependent stimuli through different sensory modalities. A growing number of studies highlight that complex and multisensory stimuli produce highly reliable, selective and time-locked activity in many brain regions compared to conventional experimental protocols (for a review, see Hasson et al., 2009 (Hasson, Nusbaum, & Small, 2009)). However, naturalistic stimulation is often questioned because of its uncontrolled nature (Felsen & Dan, 2005). Virtual Reality (VR) experiments can bridge the gap between the control granted by laboratory experiments and the realism needed for a real-world neuroscience approach (Matusz, Dikker, Huth, & Perrodin, 2019). An interesting example is the usage of a VR flanker task with realistic stimuli (Olk, Dinu, Zielinski, & Kopper, 2018). From a methodological perspective, VR allows researchers to maintain a high degree of control on the experiment, while at the same time immersing participants in highly realistic multisensory environments (Blascovich et al., 2002). Furthermore, subjects' reactions correspond very closely to those aroused in a real situation, even being aware they are not (Gonzalez-Franco, Perez-Marcos, Spanlang, & Slater, 2010; Pavone et al., 2016; V. Petkova & Ehrsson, 2008). This phenomenon is explained mainly by the sense of presence (SoP), that is, the illusion that between the user and his experience there is no mediation of a device [Lombard, 1997 #387] (Lombard & Ditton, 1997), he then "feels there"(Sanchez-Vives & Slater, 2005).

Given the role of the perceptual load in modulating multisensory stimuli processing and consistently with the multisensory nature of our daily experience, the aim of the present study was to investigate the effects of multisensory cueing on a target detection task under two different conditions of multisensory perceptual load. Here, different multimodal cues (e.g., auditory and vibrotactile) were concurrently presented, alone or in combination, with a visual target and with the perceptually corresponding environmental noises. To do so, we devised a novel naturalistic visual search task in VR, in which participants drove a car on a virtual racetrack with the aim to hit the highest number of spheres (i.e., the targets). During the experiment, we acquired electroencephalography (EEG) signal and galvanic skin response (GSR) along with subjective measures to assess the perceived workload and the sense of presence. The electrodermal response permits to assess the level of arousal; previous studies demonstrated its utility as an objective measure of anxiety under threating situations or to provide physiological evidence of body illusions (Gonzalez-Franco et al., 2010; V. Petkova & Ehrsson, 2008; Mel Slater et al., 2010). Thought EEG measures we extracted the spectral features related to the workload using an already validated algorithm (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; G. Di Flumeri et al., 2018). For the first time, here we investigate the facilitation effect of multisensory cueing by exploring whether and how auditory and tactile stimuli (alone or combined) presented concurrently with the visual targets in two different conditions of perceptual load (e.g. Low and High) could i) improve the participants' detection performance, ii) enhance their sense of presence in the virtual environment and iii) modulate their mental workload.

2.3 Materials and methods

2.3.1 Participants

Eighteen healthy male volunteers (mean ± SD, 26.5 ± 3.2 years old) without significant psychiatric or neurologic diseases and with normal or corrected-to-normal visual acuity participated in the study. We chose to involve only male subjects in order to reduce the inter-subjects' variability due to putative gender effects associated with the 3D driving task, and consequently, to make the statistical results more reliable. In particular, since video games are played more by males than by females (e.g., Griffiths, 1991b; Kaplan, 1983; Phillips, Rolls, Rouse, & Griffiths, 1995; Wright et al., 2001), and our experimental task resembled very closed the setup of a videogame, we didn't want to risk introducing such a bias in our study. It is worth noting, however, that this phenomenon could be partially explained in terms of accessibility to videogames consoles (Woodard & Gridina, 2000). The experimental protocol was approved by the ethics committee of Sapienza University of Rome and was carried out in accordance with the ethical standards of the Declaration of Helsinki of 1964. All participants gave their written informed consent before the experiment.

2.3.2 Experimental setup

Inspection of Figure 3A depicts the experimental setup. A virtual scenario consisting of a car on a racetrack was designed in 3ds Max 2015 (Autodesk Inc.) and implemented in Unity3d 5.3.1 (https://unity3d.com). Subjects explored the Virtual Reality (VR) environment from the driver's first-person perspective by means of an Oculus Rift Dk2 head mounted display (HMD) (www.oculus.com). The HMD has a 100° field of view, a resolution of 960 x 1080 per eye and internal sensors to track head movements. To control the car, a Logitech g27 racing wheel with no shifter was used. Participants were instructed to use only the throttle pedal and to maintain it pressed despite the speed of the virtual car was fixed to reduce the performance variability across subjects. The vibrotactile feedback was delivered using two DC vibrating motors applied to a wearable belt, developed by Braintrends (www.braintrends.it) for the experiment purposes. The audio feedback was delivered using headphones. To record skin sweating activity, Galvanic skin response (GSR) was monitored by means of the NeXus-10 MKII system (MindMedia BV, Netherlands, *www.mindmedia.com*) and its dedicated software BioTrace+. GSR sensors were applied to the index and middle fingers of the participant's non-dominant hand (Boucsein, 2012). The EEG recordings were carried out using 38 channels. The amplifier adopted in the experiment was the Galileo BEPlus (EBNeuro Spa, Italy, *www.EBNeuro.biz*).

2.3.3 Experimental procedure

Before the experiment, participants were asked to become familiar with the steering wheel and the pedals. The familiarization procedure was performed in the virtual reality setting during a driving task similar to that employed in the experimental protocol, without any other task to perform. The familiarization procedure was considered as completed when participants were confident with the driving task and the Head Mounted Display (Tieri et al, 2015). Subjects began the experiment with two 1 min runs in which they were instructed to rest, respectively with eyes open and closed. Then, two 2 min baseline conditions were performed while subjects drove in two different conditions of cognitive load (i.e., Low and High). These two conditions were coherent to those ones employed during the experimental tasks. For the Low condition, the virtual scenario consisted in sunny weather and maximum visibility, for the High condition rainy weather and mist were employed, limiting the visibility and providing a high environmental noise due to the presence of rain and thunders (Fig.3B-C). After the baseline tasks, subjects were asked to perform 8 runs, 3-minute long each. All these baseline conditions were necessary to calibrate the EEG-based algorithms (see below for further details). The goal of the virtual task was to drive a car around a racetrack (the same employed during the baselines) and to hit slightly transparent spheres-like objects, spawned randomly on the left or right side of the track. The goal of this task was to achieve the highest possible score by gathering as many targets possible while completing a lap as fastest as possible. The collectible items could arise from the ground or fall from the sky and were fixed in number (i.e., 20 in Low, 40 in High) but not in the position on the horizontal plane (i.e. right or left), which was randomized, to avoid expectations or habituation effects.

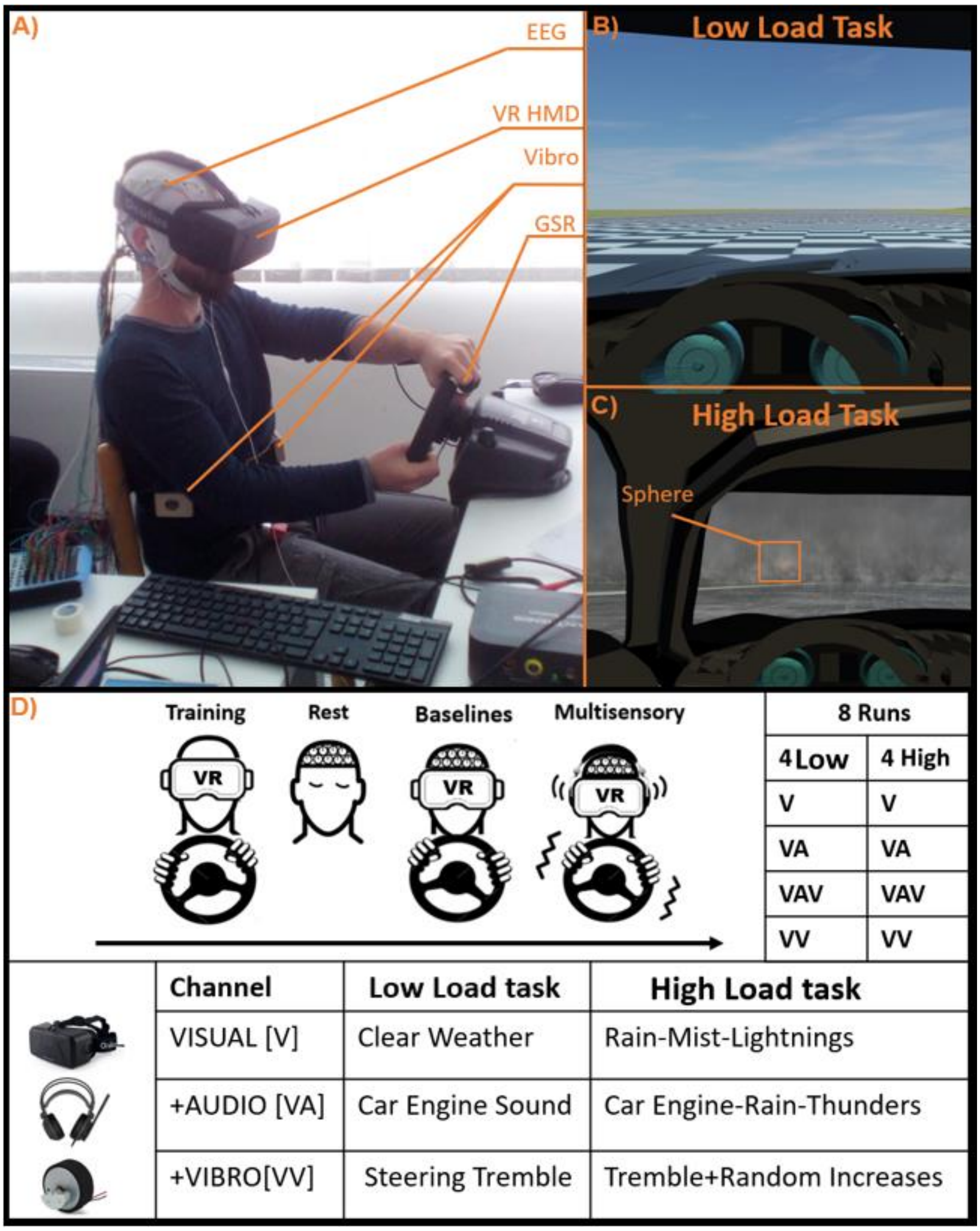

Figure 3 A) Experimental Setup; B) the First-person perspective of Low and C) High Load Conditions. The target object (the sphere) is represented in the insert; D) Summary of the experimental procedure with the detail for each Task.

The perceptual load of the task was increased (i.e. High condition) thanks to a faster-spawning rate of the spheres and to the visual and audio presence of rain, mist, thunders, and lightning. For each Load Condition, each subject performed 4 runs, which differed in terms of sensory channels enabled: V for the only Visual channel, VA for Visual and Acoustical channels, VV for Visual and Vibro-tactile ones, VAV with all of them. In the V task, participants experienced just visual stimuli, and they could not feel any sound or vibration from the surrounding environment. In the VA task, the Acoustic channel provided a cue of the spheres' appearance through a *"beep"* in a stereo mode (i.e. providing also information about the side, right or left). Participants could also hear the engine of the car running and, for which concerns the High load task, the sound of the rain and thunders. In the VAV task, subjects guided towards the stimuli also from the stereo vibrotactile feedback provided through a vibrating belt, tied to their abdomen. In this task, also the steering wheel vibrated following the car engine modulation and thunders. The VV task was the same as for the VAV, except for the sound, muted in this case. Figure 3B and C illustrate examples taken from the Low and High Condition, while Figure 3D provides a summary of the employed multimodal perceptual task. At the end of each task, the HMD was removed from subjects' heads to avoid any possible sense of nausea and to allow them to fill the workload and the 3 items questionnaire on the sense of presence, both described in the next paragraph. Subjects started with the Low or High Load Condition in a counterbalanced order. Both tasks were comprehensive of the 4 sensory input conditions, presented in a counterbalanced way across participants.

2.3.4 Subjective evaluation

After each experimental run, subjects gave their answers to the presence and workload questionnaires (Table 1). The sense of presence questionnaire was adapted from previous work (Sanchez-Vives & Slater, 2005) and consisted of three questions scored on a scale from 1 to 7. To achieve a final presence score, we averaged the three answers. The Nasa TLX Questionnaire (Hart & Staveland, 1988) assesses workload based on 6 subscales(Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration). They are rated for each task within a 100-

points range with 5-point steps. The scoring and the analysis related to the NASA-TLX were carried out using a Digital TLX software that automatically calculated the weighted average index of workload index using the weights and the parameters established by Hart and Staveland's questionnaire (Hart & Staveland, 1988).

Presence	
Question 1.	How much did you get the feeling of being inside the racetrack?
Question 2.	How much did you get the feeling that the virtual experience was the reality for you and you almost forgot the real world of the lab?
Question 3.	When you think back to this experience, do you think of the virtual track as an image you saw or like a place you visited?
Workload	
Mental Demand	How mentally demanding was the task?
Physical Demand	How physically demanding was the task?
Temporal Demand	How hurried or rushed was the pace of the task?
Performance	How successful were you in accomplishing what you were asked to do?
Effort	How hard did you have to work to accomplish your level of performance?
Frustration	How insecure , discouraged, irritated, stressed and annoyed were you?

Table 1. Questions assessing the perceived feeling of presence and workload.

2.3.5 Neurophysiological measurements

2.3.5.1 EEG recordings

For each subject, scalp EEG signals have been recorded by the digital monitoring *BEplus* system (EBNeuro system, Italy) with a sampling frequency of 256 (Hz) by 38 Ag/AgCl passive wet electrodes covering all the scalp sites (Fp1, Fpz, Fp2, F3, F4, F7, Fz, F8, AF3, AF4, AF7, AF8, FC3, FC4, FCz, C3, Cz, C4, T3, T4, CPz, C5, C6, CP3, CP4, P3, Pz, P4, P7, P8, PO7, PO8, PO3, PO4, O1, Oz, O2, POz) referenced to both the earlobes and grounded to the AFz electrode, according to the 10-20 standard (Jurcak, Tsuzuki, and Dan 2007). In addition, both vertical and horizontal EOG signals have been recorded concurrently with the EEG, and with the same sampling frequency (256 Hz), by two bipolar channels surrounding the right eye, in order to collect the eye-related activity (i.e. eyes blinks and saccades) of the subjects during the execution of the task. Two different analyses have been performed to evaluate the experienced workload of the subject across the different Sensory Tasks and Load Conditions.

2.3.5.2 Time-domain analysis

The first analysis has been performed in the time domain, by using Event-Related Potentials (ERPs) elicited by the hit targets. Since such kind of analysis requires a strong synchronization between the EEG recording system and the target appearance, all the experimental task-related events (i.e. starting and ending time for each condition, the appearance of each target to hit) have been synchronized with the recording systems by using dedicated trigger station (BrainTrends, www.braintrends.it). ERPs represent the EEG voltage fluctuations that are associated in time with some physical or mental occurrences (Picton et al., 2000). Specifically, the event-related potential (ERP) P300 is a positive deflection of the EEG signal elicited at around 300 ms after the occurrence of a stimulus that the user is paying attention to (Fabiani, Gratton, Karis, & Donchin, 1987). The P300 component is normally evaluated in terms of amplitude and latency. Amplitude is defined as the difference between the largest positive-going peak of the ERP waveform within a time window (e.g., 250–500 ms) and the average of the time-epoch voltage chosen for the analysis (e.g. 0-600ms). Latency (in ms) is defined as the time from stimulus onset to the point of maximum positive amplitude within the same time windows. The evoked response is extracted from the EEG as the average of a series of single responses (synchronized averaging) in order to remove the random fluctuations of the EEG. Although P300 amplitude and latency could be affected by several factors (e.g. attention, fatigue, age, gender (Polich & Kok, 1995) it has received much attention as a potential indicator of mental workload. In this regard, recent studies demonstrated a reduced P300 amplitude under higher workload conditions in ecologically valid tasks. Unlike P300 amplitude, the latency of the P300 is often increased if the categorization of the eliciting stimulus becomes more difficult, thus, representing the timing of mental processing (Arico et al., 2014; Kathner, Wriessnegger, Muller-Putz, Kubler, & Halder, 2014). To summarize, for each subject we calculated P300 amplitude and latency values, averaged for all the hit targets, for each condition (2 load conditions and 4 sensory tasks). In the following, the algorithm steps to calculate P300 amplitude and latency starting from recorded EEG signals were reported.

First of all, the EEG and EOG signals have been firstly band-pass filtered with a fifth-order Butterworth filter (low-pass filter cut-off frequency: 30 (Hz), high-pass filter cut-off frequency: 1 (Hz)). Independent Components Analysis (ICA(Lee, Girolami, Bell, & Sejnowski, 2000)) has been performed to remove eye blinks and eye saccades artifact contributions that could affect the morphology of the evoked P300 potentials. After that, the EEG signals have been segmented in epochs of 600 ms starting from the onset of the target appearance. For other sources of artifacts (e.g. bio amplifier saturation, muscular activity) specific procedures of the EEGLAB toolbox have been used (Delorme & Makeig, 2004). In particular, three criteria have been applied. *Threshold criterion*: if the EEG signal amplitude exceeds ±100 (μV), the corresponding epoch would be marked as an artifact. *Trend criterion*: each EEG epoch has been interpolated in order to check the slope of the trend within the considered epoch. If such slope was higher than 10 (μV/s) the considered epoch would be marked as an artifact. *Sample-to-sample difference criterion*: if the amplitude difference between consecutive EEG samples was higher than 25 (μV), it meant that an abrupt variation (no-physiological) happened and the EEG epoch would be marked as artifact. At the end, all the EEG epochs marked as artifact have been rejected from the EE withG dataset with the aim to have an artifact-free EEG signal from which estimate the brain variations along the different conditions (Gianluca Di Flumeri, Borghini, et al., 2019). All the previously mentioned numeric values have been chosen following the guidelines reported in Delorme and Makeig (Delorme & Makeig, 2004).

At this point, the remaining time epochs have been filtered by using a method described in Aricò et al., (Arico et al., 2014) based on the use of wavelet transformation, to increase the signal to noise ratio (SNR) of the P300 potentials recorded during the experimental tasks.

Such an algorithm was necessary to compute as accurately as possible amplitude and latency of P300 potentials, starting from a not too high number of stimuli (i.e. 20 targets for the Low tasks and 40 for the High ones). Amplitude and latency studies on P300 potential are normally performed on the Cz electrode, where such component assumes the higher amplitude (Arico et al., 2014), so, just P300 epochs on this electrode have been taken into account for the following analyses. In particular, we decomposed each single target epoch into its time-frequency representation by evaluating the continuous wavelet transform (CWT). We used a complex Morlet mother wavelet, with frequency content ranging from 1 to 20 Hz with a frequency resolution of 0.5 Hz and a time window of 600 ms. We computed the power spectrum (PWT) for each transformed single epoch, defined as the squared magnitude of the CWT. Finally, we computed the average PWT over all epochs, to identify the wavelet coefficients with the highest power. Coefficients below a specified power threshold were filtered out, according to the following procedure: the empirical cumulative distribution function (CDF) of the power spectrum was calculated through the Kaplan–Meier estimation (Lawless, 2003) the filtering model consisted of a matrix (PMask) whose time-frequency elements were set to 1 when the CDF of the corresponding wavelet coefficient was greater than the threshold and set to 0 otherwise. We computed the best threshold value referring to the original method used in a previous study (Arico et al., 2014), aiming to eliminate as much noise as possible while preserving the shape of the P300 potential. A filtered version of the target single epochs was finally obtained by evaluating the inverse CWT (ICWT) of the coefficient of every single epoch, multiplied for the PMask. At this point, filtered epochs have been averaged for each subject, load condition, and sensory modality condition. Finally, as stated before, the amplitude has been estimated as the difference between the largest positive-going peak of the P300 waveform within a time window (e.g., 250–500 ms) and the average of the time-epoch voltage chosen for the analysis (i.e. 0-600ms). Latency (ms) has been calculated as the time from stimulus onset to the point of maximum positive amplitude.

2.3.5.3 Frequency-domain analysis

The second kind of analysis has been performed in the EEG frequency domain. With respect to the former analysis (ERPs-based), that was strictly locked to the processing of targets appearance, such analysis in frequency domain was employed to highlight the overall workload experienced by the subjects along the whole running lap, for each experimental condition. In this regard, most of the studies showed that the brain electrical activities mainly involved in the mental workload analysis are the theta and alpha brain rhythms typically gathered from the *Pre-Frontal Cortex* (PFC) and the *Posterior Parietal Cortex* (PPC) regions. Previous studies demonstrated as the EEG theta rhythm over the PFC presents a positive correlation with the mental workload (Smit et al. 2005). Moreover, published literature stressed the inverse correlation between the EEG power in the alpha frequency band over the PPC and the mental workload (Jausovec & Jausovec, 2012). Only a few studies have reported significant results about the modulation of the EEG power in other frequency bands, i.e. the delta, beta and gamma (Gevins & Smith, 2005). More specifically, most of the studies are focalized on the EEG power modulation occurring in theta (4 – 8 Hz) and alpha (8 – 12 Hz) frequency bands, usually associated with cognitive processes such as working memory and attention, typically involved in mental workload. Mental workload is also known to suppress EEG alpha rhythm and to increase theta rhythm during the activity of information encoding and retrieval (Vecchiato et al., 2014). Depending on such evidence, theta EEG rhythms over frontal sites, and alpha EEG rhythms over parietal sites have been used for this kind of analysis. As for the time domain analysis, all the processing and artifact removing algorithms have been applied in order to i) avoid eyes blinks and saccades contribution, that could even in this case frequency bands related to workload and introduce a bias, and ii) remove all the other sources of artifacts. The EEG signal has been segmented into epochs of 2 seconds, shifted of 0.125 seconds (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016). The *Power Spectral Density* (PSD) was calculated for each EEG epoch using a Hanning window of the same length of the considered epoch (2 seconds length (that means

0.5 (Hz) of frequency resolution). Then, the EEG frequency bands of interest have been defined for each subject by the estimation of the *Individual Alpha Frequency* (IAF) value (Klimesch, 1999).

In order to have a precise estimation of the alpha peak and, hence of the IAF, as stated before the subjects have been asked to keep the eyes closed for a minute before starting with the experiments. Finally, a spectral features matrix (EEG channels x Frequency bins) has been obtained in the frequency bands directly correlated to the mental workload. In particular, only the theta rhythm (IAF-6 ÷ IAF-2), over the EEG frontal channels (F3, F4, F7, Fz and F8), and the alpha rhythm (IAF-2 ÷ IAF+2), over the EEG parietal channels (P3, Pz, P4, P7, and P8) have been considered as variables for the mental workload evaluation.

2.3.5.4 Workload assessment: The EEG Workload Index

In order to select the subjective discriminant EEG spectral features related to the workload, a linear classification algorithm (*automatic stop Stepwise Linear Discriminant Analysis* - asSWLDA, (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016) has been used. This algorithm has been already validated and successfully employed for EEG-based workload assessment in operational environments, such as air traffic management (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016) and car driving (G. Di Flumeri et al., 2018). Once trained with specific “calibration data”, the algorithm can be used to compute a workload index (i.e. W_{EEG} index) on other data by combining the selected EEG features with specific weights in output from the model itself. In particular, for each Load Condition, the classifier has been calibrated by using the baseline data (i.e. BasLOW or BasHIGH and the respective 1st half of the “V” Task (i.e. V Low or V High). We used just such a task (i.e., V) to train the algorithm because the baseline data (i.e. BasLOW orBasHIGH) have also been performed without any sensory input except the visual one, like for the V task. In other words, in this way, we were sure to use two conditions which differed just for the workload (in fact in the Baseline condition there were no targets to collect, in the respective V task the “environmental conditions” were the same but the driver had to hit as many targets as possible), and not because of the presence of other sensory modalities. At this point, we calculated the W_{EEG} index for the 2nd half of each condition, for each Load Condition (i.e. V Low, VA Low, VAV Low, VV Low or V High, VA High, VAV High, VV High), and *vice versa* (i.e. calibration on 2nd half and W_{EEG} evaluation on the 1st half). Of course, the WEEG index over Low and High related Conditions was not comparable among load conditions, because of the different normalization. In conclusion, z-score transformation (Zhang, Chung, & Oldenburg, 1999), has been used to compute a normalization of W_{EEG} index distribution.

2.3.5.5 Arousal assessment

The signal related to the Skin Conductance named hereafter Galvanic Skin Response (GSR), has been recorded with a sampling frequency of 128 Hz (fs = 64 Hz) through the NeXus-10 MKII device (MindMedia BV, Netherlands), a high-quality device consisting in a wireless amplifier with specific GSR sensors applied at the no-dominant hand: by means of two electrodes on the first phalanx of the index and middle fingers, a constant potential was applied in order to induce a skin electrical current. The variations of such current are functions of the skin conductance variations.

The recorded signal was then entirely processed by using the MATLAB software. First, the signal was downsampled to 32 Hz, in order to reduce the data amount. Secondly, the signal was filtered through a 5th order Butterworth low-pass filter, with the cut-off frequency at 4 Hz, in order to remove all the higher frequency components that are not related to the electrodermal activity. Then, the signal was processed by using the Ledalab suite, a specific open-source toolbox implemented within MATLAB for the GSR processing (please visit the web site for further information): the Continuous Decomposition Analysis (Benedek & Kaernbach, 2010) has been applied in order to separate the Tonic (Skin Conductance Level - SCL) and the Phasic (Skin Conductance Response - SCR) components of the GSR. In this regard, Boucsein (Boucsein, 2012) provided the most exhaustive review about GSR physiological interpretation and GSR analysis techniques. Briefly, there are two major components for GSR analysis: the SCL (tonic component) is a slowly changing part of the GSR signal, mostly related with the global arousal of a subject during a situation, whilst the SCR (phasic component) is the fast-changing part of the GSR signal, which occurs in relation to single stimuli reactions.

In the following analysis, the mean value of the SCL and the mean amplitude of the SCR peaks during the experimental conditions have been investigated.

2.3.6 Performance Assessment

At the end of each run, our software produced a log-file with information related to the number of hit targets (#HitT) and to the total time duration (TimeD). Such information has been collected for each subject and experimental condition. A task performance index has been calculated by combining #HitT and TimeD, normalized respectively to the total number of targets (#TotT) and the minimum time required to run a complete lap (TimeMin, Equation 1). The latter has been calculated by dividing the track length by the maximum car speed.

$$\frac{HitT}{TotT} * \frac{TimeD}{TimeMin}$$

Equation 1. Performance Index

2.4 Results

2.4.1 Statistical analysis

Data analysis was performed with R-studio (Version 1.1.463), a free software programming language and software environment for statistical computing (RStudio Team, 2016). We performed a multilevel mixed linear regression analysis (LMM or “mixed-effects models”; Pinheiro and Bates, 2000) through the package Afex (Singmann, Bolker, Westfall, & Aust, 2015). Unlike traditional statistical methods, LMM can deal with grouped, nested or hierarchical structured data (Faraway, 2005). Furthermore, LMM are statistical models that incorporate both fixed-effects parameters and random effects (Bates, Mächler, Bolker, & Walker, 2015) and can separately treat the effects caused by the experimental manipulation (fixed effects) and those that were not (random effects) (Pinheiro & Bates, 2000). Importantly for the present experiment, LMM can handle missing data (Gelman & Hill, 2006). Here we used mixed-effect models to test how our dependent variables are influenced by the experimental manipulations (i.e., V/VA/VAV/VV) and by the load condition (i.e., Low and High). We adopted a different model for each dependent variable. The variables were: (a) performance; (b) subjective ratings of workload (i.e., Nasa-TLX), (c) sense of presence;(d) EEG-based workload (W_{EEG}); (e) skin conductance. As fixed effects, we used the load conditions (High vs. Low) and the sensory tasks (i.e. V for Visual, VA for Visual-Audio, VAV for Visual-Audio-Vibro, VV for Visual Vibro). The resulting models included the subjects as a random factor (i.e., random intercept) and the random slopes of load condition and sensory tasks over subjects. Type III Wald Anova function from the "car" package in R was used to determine the statistical significance of the fixed effects. Post hoc comparisons using the Tukey test were carried out from the "lsmeans" package in R.

2.4.2 Performance

As a first step, we focused on the effects produced by the multimodal integration and the cognitive load on the subject's performance. We obtained a significant main effect of the Cognitive Load ($\chi2$ (1) < 2.2e-16) explained by reduced performance in the High with respect to Low Load Condition. Also, we found a significant "ConditionXTask" interaction ($\chi2$ (1) < 0.001) explained by the increment of the subject's performance elicited by both bimodal and trimodal stimulation tasks compared to when only the visual signal is available (all p-values < 0.01), only in the High Load Condition (Figure 4A). No significant differences were obtained in the Low Load Condition (p>0.05).

2.4.3 Sense of Presence

The regression analysis showed the significant main effect of sensory tasks ($\chi2$ (1) =0.0015) explained by higher sense of presence in the task with higher sensory integration (VAV) with respect to when only the visual modality was available (V) ($p<0.007$) or when the auditory feedback was absent (VV) ($p=0.026$). The visual and audio integration (i.e., VA) produced a higher sense of presence with respect to the task with the only visual channel (V) was available ($p=0.048$) (Figure 4B). Also, we found the significance of the Load Condition ($\chi2(1)$ =0.0038) explained by a higher sense of presence in the High Condition, which increases the sense of presence by $\beta = 0.559$ ($p = 0.0118$).

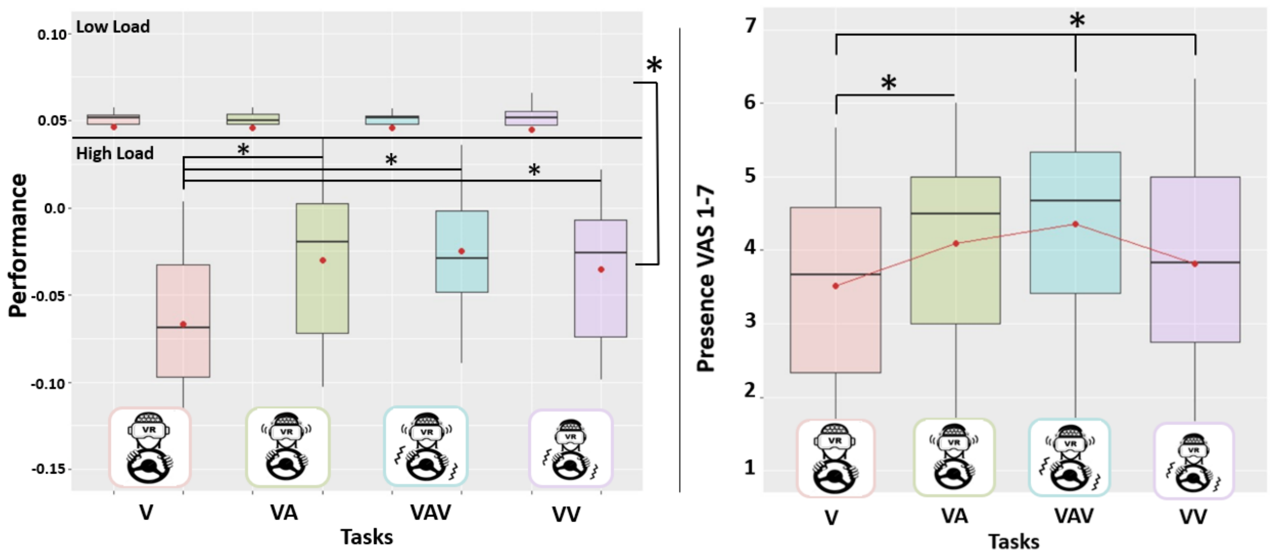

Figure 4 Bar graph representing behavioral correlates of the A) Performance Index for the Low Load Condition (top) and High Load Condition (bottom) for each sensory task. Values (y axis) are negative as they result from the normalization of the performance assess.

2.4.4 Workload

In terms of the EEG-based workload index (W_{EEG}), the regression analysis showed the significance of the main effect sensory task ($\chi 2$ (1) <0.001. Post-hoc tests revealed decrement of the workload in the multisensory tasks VAV and VV with respect to task in which the only visual channel (V) is available, respectively by β 0.333 (p =0.002) and by β 0.319 (p=0.006) (Figure 5A). No significant effect of the Load Condition neither of the interaction with the Sensory Tasks was found (p>0.05).

Similar effects were found for subjective measures of workload (NASA-TLX) where the regression showed the significant main effects of Sensory Tasks ($\chi 2(1)$ = 0.015), due again to the reduction of the perceived workload in the task in which the sensory integration is maximal (VAV) with respect to when the visual channel alone is available (V) (p=0.025) (Figure 5B). No other comparisons were significant (p>0.05). Also, we found the significance of the Load Condition ($\chi 2(1)$ < 0.001), explained by the higher perceived workload in the High with respect to the Low Condition.

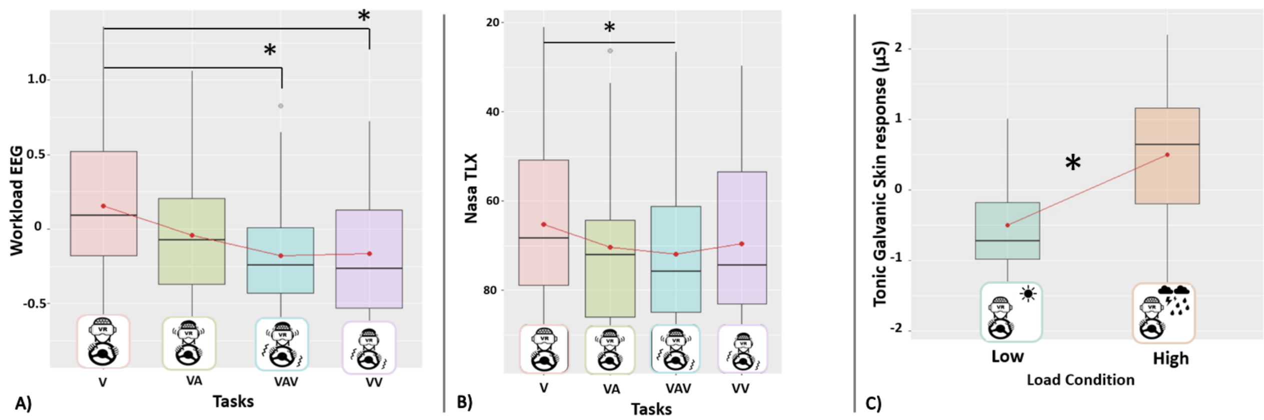

Figure 5 Bar graphs representing A) the EEG-based Workload index and B) the workload experienced by the subjects, both across the Sensory Tasks; C) the tonic GSR averaged across sensory tasks for each Load Condition.

2.4.5 Skin Conductance

Regarding skin conductance, both tonic and phasic components (i.e., average SCL and SCR peaks amplitude) revealed a significant main effect of the Load Condition (respectively $\chi2(1)$ = 0.001 for the tonic and $\chi2(1)$ <0.001 for the phasic component). Post-hoc test revealed higher arousal in the High Load Condition with respect to the Low Load Condition (β 0.998, p=0.004 and β 25, p<0.0001, respectively for tonic and phasic component).

2.4.6 Presence predicted by P300 latency

The previous results show an effect of the multisensory integration on the sense of presence and workload mostly during the Condition of High Cognitive Load. In the next analysis, we tested whether the participants' sense of presence experienced in the Low or High Cognitive Load Condition was predicted by the interaction between P300 latency and immersivity features, such as the amount and types of sensory stimulation provided to the subjects, and by the task load. To this aim, we ran a multilevel mixed log-linear regression analysis on the model with the participants' subjective ratings of the sense of presence as our dependent variable and P300 latency, Load condition (Low vs. High) and sensory tasks (i.e. V for Visual, VA for Visual-Audio, VAV for Visual-Audio-Vibro, VV for Visual Vibro) as our fixed effects. The resulting model included the subjects as a random factor (i.e., random intercept), the random slopes of Load condition and sensory tasks over subjects. Type III Wald Anova function from the car package in R was used to determine the statistical significance of the fixed effects.

Results showed a significant main effect of Load ($\chi2(1)$ =0.001), pointing that when presence increase, P300 Latency decreases more rapidly in the High condition. Additionally, we found significant interactions between P300 latency and Load condition $\chi2(1)$=0.012 and tasks X condition χ2(3)=0.003049 which were qualified by the significant P300 latency X Load Condition X tasks interaction χ2(3)= 0.003.

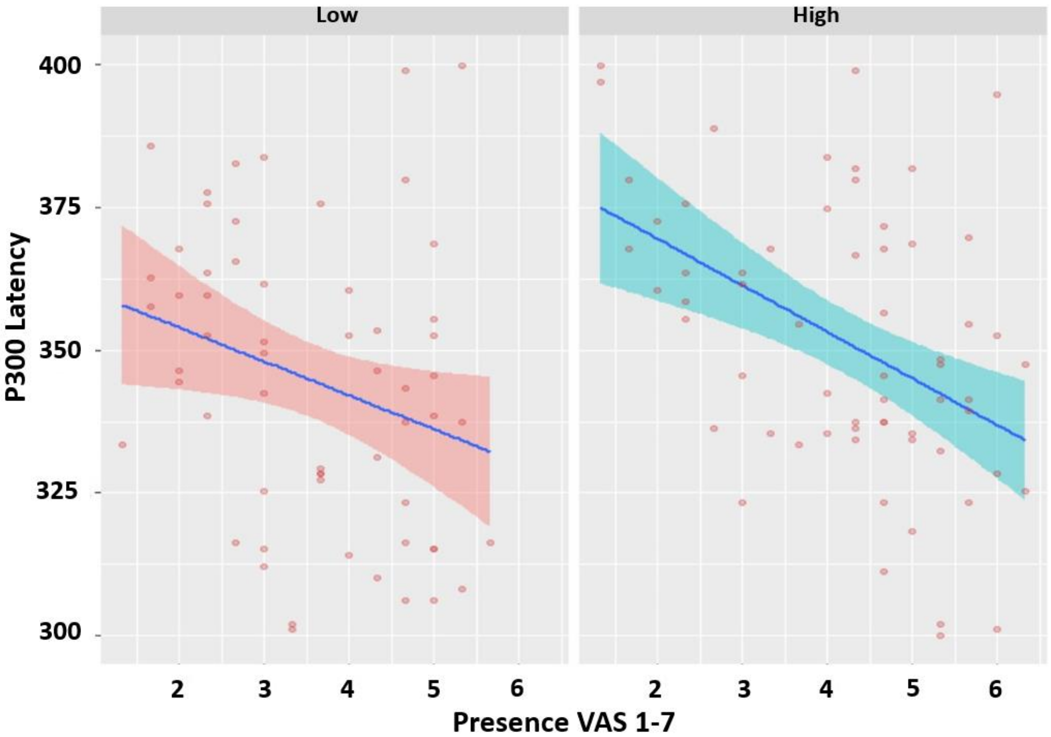

Figure 6. How Presence was predicted by P300 latency in the High task was significantly different from the Low task. An increase in the sense of presence in the High task in fact, caused the P300 latency to decrease more rapidly.

No further test using other measures as predictors of presence highlighted any significant behavior.

2.5 Discussion

In this study, we employed a target detection task in immersive virtual reality, neurophysiological and behavioral measures to devise a highly controlled naturalistic paradigm to investigate the facilitation effect of the multisensory cueing with respect to the unimodal visual stimulation, under different conditions of perceptual load (High vs Low). Our results showed an interaction effect between multisensory integration and perceptual load in enhancing the subjects' performance, which was higher each time the spatiotemporal integration (bimodal or trimodal) between sensory modalities occurred, but only in the condition of high load. As compared to the unimodal stimulation, the sense of presence increased when there was concordance between visual and auditory signals, both when the vibrotactile feedback was or not delivered (VAV and VA, respectively). In general, the illusion of being present in the virtual environment was higher in the condition of high, not low, cognitive load. By contrast, the haptic feedback was beneficial for the workload, as evident through the EEG-based index and subjective measures (NASA TLX).

2.5.1 Multisensory integration boosts performance under load condition

In line with previous reports on the facilitatory orienting effects by multisensory events in different load situations (Lunn et al., 2019; Santangelo & Spence, 2007), we found that the subjects' performance in the high perceptual load condition, was significantly enhanced by the multimodal cueing, compared to unisensory stimuli. Differently from classical visual search or spatial orienting tasks in which participants have to press a button or a pedal, here we asked participants to hit the perceived targets by steering a car in a virtual scenario. As can be seen from a reduction in performance, the High Load condition makes the task more difficult compared to the Low load condition, causing the subjects to miss more targets. Due to the higher mental load elicited by this condition, both the bimodal (i.e., visual-audio and visual-vibrotactile) and trimodal (VAV) stimulations improve significantly the performance compared to the mere visual condition. Given these results, our results converge with Santangelo and Spence(Santangelo & Spence, 2007) suggestion on the employment of multisensory stimuli in an applied context. As shown by Ho and colleagues in fact(C. Ho, Reed, & Spence, 2007; C. Ho, Santangelo, & Spence, 2009), multisensory stimulation could represent an important feature in the design of driver warning signals. Additionally, our data extend those results to high load conditions resulting from multisensory environmental noise. However, Lunn and colleagues(Lunn et al., 2019) demonstrated rather clearly that multisensory stimuli are not ‘immune’ to perceptual load effects as proposed by Santangelo and Spence (Santangelo & Spence, 2007), (see also Lavie’s work (Nilli Lavie, Ro, & Russell, 2003)on immune stimuli) but they still showed clear evidence of facilitatory attentional capture by multisensory stimuli.

2.5.2 The workload is decreased by vibrotactile cues

An important matter worth investigating is the interplay between multisensory stimulation and mental workload. Only a few studies investigated such relationship employing subjective questionnaires (Hancock, Mercado, Merlo, & Van Erp, 2013; Vitense, Jacko, & Emery, 2003) see (Kahol, French, Panchanathan, Davis, & Berka, 2006) for haptic feedback and EEG based attention peaks). To our knowledge, this is the first study exploring the effects of multimodal stimuli on workload through an EEG based index, especially with virtual reality. The subjective workload measure assessed by the Nasa TLX (Hart & Staveland, 1988) revealed that the trimodal condition was the only one perceived as less demanding compared to the visual one. Conversely, the EEG analysis showed a decrease in workload in both the Visual-Audio-Vibro and the Visual-Vibro condition, regardless of the load (of course, WEEG index over Low and High related conditions was not comparable among load levels, because of the different normalization). The reason for this facilitatory effect on the cognitive system could have multiple interpretations. For which concerns the trimodal solution it can be easily conjectured that signals from both audio and haptic feedback do really make the task more immersive and easier to perform as one could guess looking at participants performances (see also (C. Ho et al., 2007)), workload and presence questionnaires. Instead, it can be speculated that the absence of environmental sounds, while detrimental for the illusion of presence, for the performance it represents a cognitive relief for those subjects engaged in the task. The mental cost of filtering out the engine sounds and rain noise in the high cognitive load condition, in fact, is probably higher or more stressful (Szalma & Hancock, 2011) than the one needed to ignore the steering wheel vibrations, and this could explain why, despite the higher presence and performances scorings in the visual-auditory (i.e., VA) condition compared to the visual-vibrotactile one (i.e., VV), the latter happens to be less cognitive exhausting. However, it is worth considering that the different localization between the environmental vibrotactile source (steering wheel) and the cueing source (abdomen) could have contributed to reducing the cognitive effort by facilitating the discrimination of the cues from the distracting vibrations in the VV condition. Nevertheless, this finding suggests that in the designing of future workstations, solutions could exploit different sensory combinations to achieve different goals, depending on the task difficulty and duration, and taking into consideration whether prioritize performances on one side or user stress and mental fatigue on the other.

2.5.3 Multisensory stimulation and Perceptual Load modulate the sense of Presence

Exploring literature only a paucity of studies demonstrates an enhanced sense of presence concurrent with a multisensory exposition in virtual reality. For example, few studies investigated the effect of environmental sounds instead of silence in a virtual reality scenario (e.g., (Hendrix & Barfield, 1996); (Dinh et al., 1999). A more recent study (Cooper et al., 2018) carried out with projectors and 3d glasses, describes the relationship between the subjective sense of presence and the combination of sensory modalities (visual – audio-tactile) by reporting an increase of presence in the trimodal condition compared to the bimodal and in the latter compared to the unimodal one.

Our study confirms this claim using a modern Head Mounted Display. Through a combination of visual, audio and vibrotactile stimulations, accounting for car vibrations, we show that the Visual-Audio (VA) feedback and the Visual-Audio-Vibro one (VAV) are capable to enhance the sense of presence compared to the visual channel alone. In the real-life experience, our brain is constantly engaged in the multimodal processing of the surrounding environment. The coordination of two or more sensory modalities facilitates the selection of relevant features from irrelevant stimuli. This explains why the combination of sensory stimuli elicits a higher sense of presence. But not all the combinations are equally effective. For instance, visual and tactile inputs alone are too far from a realistic situation to enhance the sense of presence. By contrast, the alignment between the visual and auditory sensory channels more closely mimic real-life conditions. We hypothesize that this effect occurs since the absence of sound represents a highly unconventional situation. Our result is consistent with previous studies reporting that a condition with the absence of sound elicits a lower sense of presence (Dinh et al., 1999; Hendrix & Barfield, 1996). Moreover, the lack of acoustical signals could even account for spatial disorientation (Gillingham, 1992). Finally, we speculate that the silence to which participants were exposed during the Visual-Vibrotactile task may have contributed to the higher awareness toward the haptic feedback produced by the steering wheel and the belt, which in turn may have introduced an evenly uncommon perceptual situation. Moreover, we found an increase of presence in the High Load Condition compared to the Low one. In a factor analysis run on the presence questionnaire, Witmer and colleagues (Witmer et al., 2005) reported that sensory fidelity, immersion, and interface quality accounted for the 20,3% of the variance, while involvement alone accounted for the 31.9%. From an immersive point of view (Mel Slater, 1999), which represents the degree of sensory fidelity achieved by the technological setup (Mel Slater et al., 2003) key to evoke a sense of being in the virtual world (see (Cummings & Bailenson, 2015) for review), the two task were equivalent. Nevertheless, being the High Load Condition more perceptually stimulating and demanding, according to Witmer (Witmer et al., 2005) one can speculate that elicited a greater sense of presence because it was more involving. Finally, there is evidence (Riva et al., 2007) supporting the possibility that the High Load Condition elicits more sense of presence because is more emotionally involving than the Easy one, which, with his fine weather and absence of particular sounds , results more neutral. In Riva's study, in fact, the level of presence was significantly higher in the anxious and in the relaxing conditions than in the neutral one.

2.5.4 Higher perceptual load elicits higher galvanic skin response

The sense of involvement in a given situation is a complex construct to measure since is a mental state that can involve situational and personal motivation (Celsi & Olson, 1988). An increase in skin conductance is considered a reliable bioindicator of human arousal variations (Boucsein, 2012) and in virtual reality studies is typically used to evaluate the anxiety produced by a situation or the stress related to a threat (Gonzalez-Franco et al., 2010; V. Petkova & Ehrsson, 2008; Mel Slater et al., 2010). Here we showed that both the phasic and tonic component of the electrodermal response was higher in the High Load Condition. Peripheral electrodermal responses (McCleary, 1950) are found concurrently with the activation of brain regions implicated in emotion, attention, and cognition (Critchley, 2002). Despite we did not directly assess the sense of involvement, in our study we can at least hypothesize that the High Load Condition was more arousing and stressful. Higher levels of arousal could have been necessary to the participants to achieve a proper level of performance in a more demanding task and this could have led to an increase in presence. It is worth mentioning however that GSR was not found to be a predictor of presence in the analysis.

2.5.5 The sense of presence is correlated with the mental workload

Whilst many authors are trying to find a correlation between presence and performance (Cooper et al., 2018; Ma & Kaber, 2006; Stevens & Kincaid, 2015), we are strongly convinced that this relationship is not always linear since arbitrary signals on one hand and functional but not realistic information on the other, could still both serve well their pragmatic purpose as performances enhancers even impairing the overall sense of presence experienced (Vitense et al., 2003). Here, in fact, we demonstrated that despite the Vibro-tactile condition elicited a lower sense of presence, its performances were still significantly higher than those obtained in the Visual condition.

On the contrary, we explored the possibility of a relationship between the sense of presence and the EEG based workload. Since the term "*presence*" was coined in fact, this construct has been mainly evaluated through self-report measures (Insko, 2003). This kind of analysis, however, did not apply directly to the Workload-EEG index, since for its nature, it is normalized between a minimum and a maximum for each subject, depending on the calibration runs (Easy, Hard). In other words, W_{EEG} values coming from different subjects cannot be comparable. For this reason, we tried to correlate presence with P300 Latency and P300 Amplitude. In particular, the latency of the P300 is often increased if the categorization of the eliciting stimulus becomes more difficult, thus, representing the timing of mental processing (Arico et al., 2014; Kathner et al., 2014). While we found that presence was not predicted by p300 amplitude and barely by our workload index, we showed for a correlation between the presence and P300 latency in the High load task. Nevertheless, it must also be noted that P300 amplitude and latency could be affected by several factors (e.g. attention, fatigue, age and gender) (Polich & Kok, 1995). The possibility to assess presence with ERP was previously investigated by Kober and Neuper (Kober & Neuper, 2012) but never so far using a head-mounted display. Our finding, while not easy to interpret, may suggest that the timing pressure and supposedly the increased involvement derived by a more demanding task, plays an important role in shaping the sense of presence. Future studies will explore more accurately this finding that for now, opens novel ways to study the dynamics between the sense of presence, workload, and ERPs.

2.6 Conclusion

We devised a novel virtual reality driving task to investigate the influence of multisensory stimulation and perceptual load on subjects' performance, sense of presence, arousal (i.e. GSR), and both perceived (i.e. questionnaire) and experienced (EEG-based, W_{EEG}) workload. Finally, we explored whether a modulation in the sense of presence was reflected in workload. Five main findings herein are discussed.

First, results highlighted that in virtual reality, in the High load task only, bimodal and trimodal stimulation were equally effective to significantly improve performance compared to the use of visual channel alone. Second, we showed that the most complete sensory combination (i.e. VAV) and the Visual-Vibro one (i.e., VV) were capable to decrease the experienced workload according to the EEG-based index. Instead, the perceived workload, according to the NASA-TLX questionnaire, was reduced only by the trimodal condition.

Third, as postulated by Sheridan (Sheridan, 1992) we found that task demand affects the sense of presence, which was higher in the High condition. Moreover, in line with previous studies (Cooper et al., 2018), we found that the higher level of multisensory stimulation (e.g., VAV) was more efficient in enhancing the sense of presence compared to the bimodal (VV) or unimodal stimulation (V).

Fourth, we extend previous findings (Lakie, 1967) showing that skin conductance level increases along with the task demand, proving that the same effect occurs in a virtual reality task.

Fifth, we show that in the High load task, the latency of P300 decreases as an effect of the increasing sense of presence, in other words, there is a linear dependency between P300 latency and sense of presence.

Overall, we extend to virtual reality previous evidence of the performance-enhancing properties of multisensory stimulation under a high perceptual load. Moreover, we provide new insights into the interplay between multimodal stimulation, workload, and presence. Understanding the factors that moderate the relationship between virtual reality and human response is crucial to areas of concern that range from general theories of perception and cognition to the pragmatic design of future applications. These findings may open novel ways to measure in a more naturalistic manner, effects only investigated through subjective and behavioral techniques since now.

2.6.1 STUDY 2 - Human-machine interaction assessment by neurophysiological measures: a study on professional air traffic controllers using Augmented reality

2.7 Abstract

This study aims at investigating the possibility to employ augmented reality systems along with neurophysiological measures to assess the human-machine interaction effectiveness. Such a measure can be used to compare new technologies or solutions, with the final purpose to enhance an operator's experience and increase safety. In the present work, two different interaction modalities (Normal and Augmented) related to Air Traffic Management field have been compared, by involving 10 professional air traffic controllers in a control tower simulated environment. The experimental task consisted of locating aircraft in different airspace positions by using the sense of hearing. In one modality (i.e. "Normal"), all the sound sources (aircraft) had the same amplification factor. In the "Augmented" modality, the amplification factor of the sound sources located along the participant head sagittal axis was increased, while the intensity of sound sources located outside this axis decreased. In other words, when the user-oriented his head toward the aircraft position, the related sound was amplified. Performance data, subjective questionnaires (i.e. NASA-TLX) and neurophysiological measures (i.e. EEG-based) related to the experienced workload have been collected. Results showed higher significant performance achieved by the users during the "Augmented" modality with respect to the "Normal" one, supported by a significant decrease in experienced workload, evaluated by using the EEG-based index. In addition,

Performance and EEG-based workload index showed a significant negative correlation. On the contrary, subjective workload analysis did not show any significant trend.

2.8 Introduction

In many operational environments (e.g. aircraft piloting, air-traffic control, industrial process control, robot-assisted surgery) operators have to face complex systems and machines to accomplish an operational activity. Improvements in such technology or even new solutions are often proposed, with the aim to enhance the human-machine interaction (HMI) and consequently increase the operator's performance and consequently overall safety. In this context, the most studied user's mental state is the *Mental Workload,* due to its strong relationship with the user's performance variations (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016). The mental workload can be assessed by using different approaches: i) *Performance* assessment e.g. by using a secondary task, provides an objective but indirect measure of the workload; ii) *subjective questionnaires* (e.g. NASA-TLX) provide a direct but subjective measure of the perceived workload; iii) *neurophysiological measure* provides both a direct and objective measure of the experienced workload (Gianluca Borghini et al., 2015). With respect to the former two techniques, the latter has the advantages to not impact on the main task, since it does not require any input by the user's side, and to be available even online, i.e. during the execution of the task. A lot of works demonstrated as electroencephalography (EEG) – based workload measures outperform the other kind of techniques (e.g. ECG, fNIRs) (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016). In particular, most of the studies showed that the brain electrical activities mainly involved in the mental workload analysis are the theta and alpha brain rhythms respectively gathered from the *Pre- Frontal Cortex* (PFC) and the *Posterior Parietal Cortex* (PPC) regions. Previous studies demonstrated as the EEG theta rhythm over the PFC presents a positive correlation with the mental workload (Gevins & Smith, 2003; Jausovec & Jausovec, 2012). Moreover, published literature stressed the inverse correlation between the EEG power in the alpha frequency band over the PPC and the mental workload (Jausovec & Jausovec, 2012). Depending on such evidence, theta EEG rhythms over frontal sites and alpha EEG rhythms over parietal sites have been used to define an EEG-based workload index, by using the ratio between frontal theta and parietal alpha rhythms. Nowadays, there are not many studies performed in real settings demonstrating the practical advantages of neurophysiological measures for HMI assessment. For example, Borghini et al. (Gianluca Borghini et al., 2015), performed a preliminary study on few helicopter pilots in which it has been investigated a neurophysiological workload measure to compare avionic technologies. The present work targets tower ATCOs working in small or medium-size towers. In this context, sounds emitted by airplanes play a very important role to accurately locate aircraft even without looking at them. Sound is also crucial with low visibility conditions where it provides important input information. In this context, ATCOs should be able to easily identify within the airfield any kind of stimuli (i.e. aircrafts engine sounds), but nowadays one of the encountered difficulties is to locate such stimuli precisely. In this regard, it has been developed at ENAC an interaction modality (the "*Augmented*" solution) by which operators could more easily retrieve information from the airfield. This solution has been compared with a "*Normal*" condition (i.e. without any augmentation) by using performances, subjective and neurophysiological measures.

2.9 Material and Methods

2.9.1 Experimental Subjects

Ten French professional ATCOs (6 males, Mean Age: ~41±4.6 years) coming from different airports and formations took part in the experiment. All the ATCOs were normal hearing. Their mean experience in hours was 7330 hours (SD = 5349.98). Their mean number of years in Control Tower was 12.8 (SD = 7.24). Informed consent was obtained from each participant after explanation of the study, which conformed to the revised Declaration of Helsinki and was approved by the local institutional ethics committee.

2.9.2 Augmented solution

The augmented solution has been implemented by retrieving the participant head orientation via a Microsoft HoloLens mixed-reality headset (we use here only its inertial measurement unit feature). Hence, we can detect in which location the participant's head is pointing at. This head axis (sagittal) will select sound sources (engine sounds) aligned with its position. All these sound sources are related to aircraft which are spatially placed in the airport vicinity, i.e. in the tower competency. In this way, the sounds that are heard by the participants are spatialized. Figure 7 shows the two experimental conditions. In particular, in the *Normal* modality, all the sound sources have the same amplification factor. In the *Augmented* solution, the amplification factor of the sound sources located along the head axis is immediately increased, while the gain of sound sources located outside this axis is decreased.

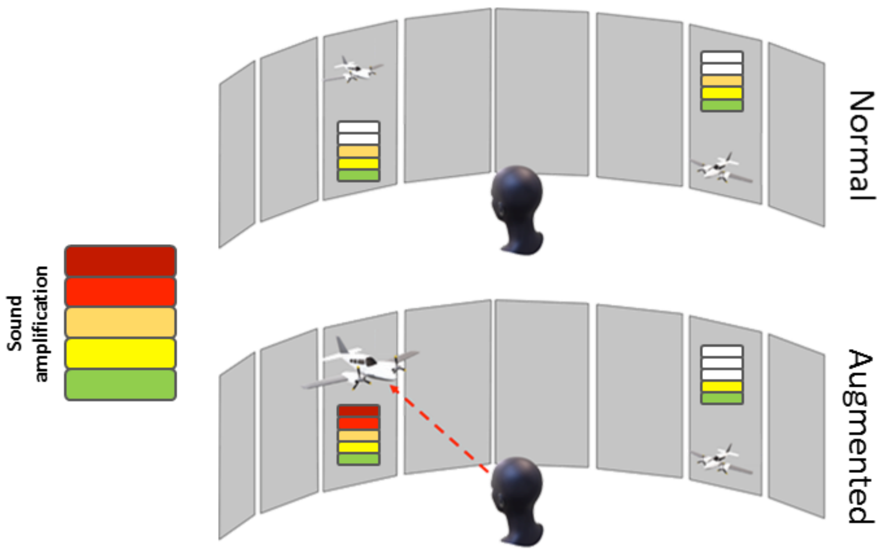

Figure 7 Schematic representation of the two compared interaction modalities (i.e. Normal vs Augmented). In particular, in the Normal one all the sound sources (aircraft) have the same amplification factor. In the Augmented solution, the amplification.

2.9.3 Experimental Protocol

The experiment was performed at École Nationale de l'Aviation Civile (ENAC, Toulouse, France). With respect to the aim of comparing the two different interaction modalities (i.e. Normal and Augmented); the experimental hypothesis was that the *Augmented* solution should be able to enhance ATCO's performance and/or decrease the experienced workload with respect to the *Normal* solution. To investigate this hypothesis, it has been reproduced a synthetic but realistic Control Tower environment, in particular, the *Muret* airport (France) by using 8 screens and Flight Gear (FG) open flight simulator (Figure 8).

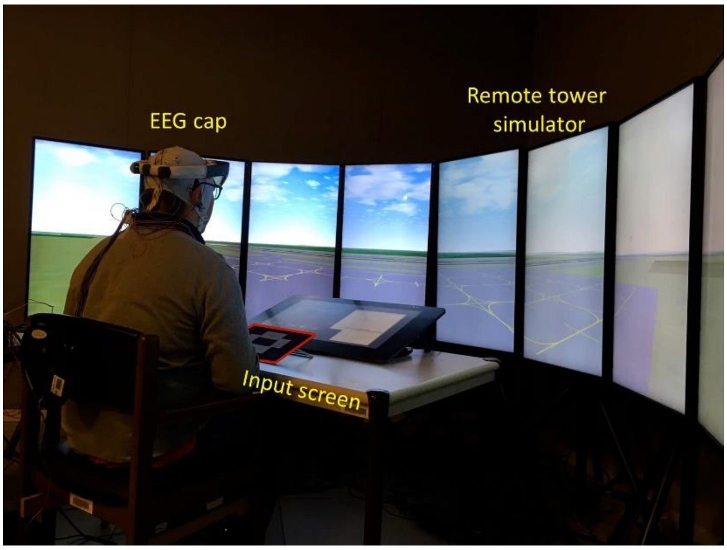

Figure 8. Overview of the experimental setup

The airport airfield has been divided into five distinct "Areas of Interest" (Figure 9). The ATCO has the point of view as the real Control Tower (located in front of the runway) and he had to discriminate the location of aircrafts in that airfield (each area could not contain more than one aircraft at a time) in the two different modalities (i.e. Normal and Augmented). In this regard, the ATCOs had an additional screen (named Input screen, Figure 8) in front of them by which to communicate the location of aircrafts within the five areas. In order to simulate different working scenarios, different trials have been realized as the combination of weather conditions (i.e. sunny or foggy) and difficulty levels (i.e. one or more aircraft at the same time), randomly changed along with the experiment. Participants have been trained to use the simulator in both the experimental conditions before the experiment started. Each participant has been asked to perform the experimental task four times for each interaction modality (i.e. Normal and Augmented), in two blocks of two repetitions each (with a resting pause in the middle). The presentation of each modality has been randomized for each subject.

2.9.4 Performed Analysis

This section describes the metrics (e.g. behavioral, subjective and neurophysiological indexes) that have been used to compare the effectiveness of the two modalities (Normal vs Augmented). In particular, to quantify *performance* achieved by the subjects, percentages of correct responses across experimental conditions have been computed and averaged over all the experimental trials. In order to evaluate the mental *workload perceived* by the ATCOs during the different phases of the experimental protocol, users have been asked to fill the six (Mental demand, Physical demand, Temporal demand, Performance, Effort, Frustration) 100-points range subscales NASA-Task Load indeX (Hart & Staveland, 1988). The global workload score from 0 to

100 was obtained for each modality (i.e. Normal vs Augmented) by averaging the individual dimension scale scores. Finally, a *neurophysiological* mental workload index has been computed from the EEG activity for each subject, as the ratio between frontal theta and parietal alpha frequency bands contributions (WL_{EEG}). In particular, for each subject, scalp EEG signals have been recorded by the digital monitoring beMicro amplifier (EBNeuro system) with a sampling frequency of 256 (Hz) by 13 Ag/AgCl passive wet electrodes covering the frontal and parietal sites (Fpz, AFz, AF3, AF4, Fz, F3, F4, Pz, P3, POz, PO3, PO4) referenced to both the earlobes and grounded to the left mastoid, according to the 10-20 standard (Jurcak, Tsuzuki, & Dan, 2007).

Figure 9 A satellite view of Muret airport, with the five "Areas of Interest" used for this experiment.

In order to compute the WL_{EEG} index, the following steps have been performed. First, EEG signals have been band- pass filtered with a fifth-order Butterworth filter [1-30Hz] and segmented in 2-seconds long epochs, shifted of 125ms (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016). Artifacts contributions that could affect the morphology of theta and alpha bands (e.g. eyes blinks and saccades, muscular artifacts, amplifiers saturations) have been removed by following specific procedures available in the EEGLAB toolbox (Delorme & Makeig, 2004). All the EEG epochs marked as artifact have been rejected from the EEG dataset with the aim to have an artifact-free EEG signal from which to estimate the brain variations along with the different modalities (i.e. Normal vs Augmented). At this point, the *Power Spectral Density* (PSD) was calculated for each EEG epoch using a Hanning window of the same length of the considered epoch (2 seconds). Then, the EEG frequency bands of interest have been defined for each participant by the estimation of the *Individual Alpha Frequency* (IAF) value (Klimesch, 1999). In order to have a precise estimation of the alpha peak and, hence of the IAF, each ATCO has been asked to keep the eyes closed for a minute before starting with the experiment. Finally, the theta rhythm (IAF-6 ÷ IAF-2), over the EEG frontal channels (Fpz, AFz, AF3, AF4, Fz, F3, and F4), and the alpha rhythm (IAF-2 ÷ IAF+2), over the

EEG parietal channels (Pz, P3, P4, POz, PO3 and PO4) have been divided to compute the WL_{EEG} index. In conclusion, the z-score method (Zhang et al., 1999) has been employed to compute a normalization of WL_{EEG} index distribution.

2.10 RESULTS

2.10.1 Behavioral data

A paired student's t-test (α=0.05) has been performed between the two modalities in terms of achieved performance. Statistical analysis showed a significantly higher performance (p=0.001) achieved during the Augmented modality with respect to the Normal condition (Figure 10).

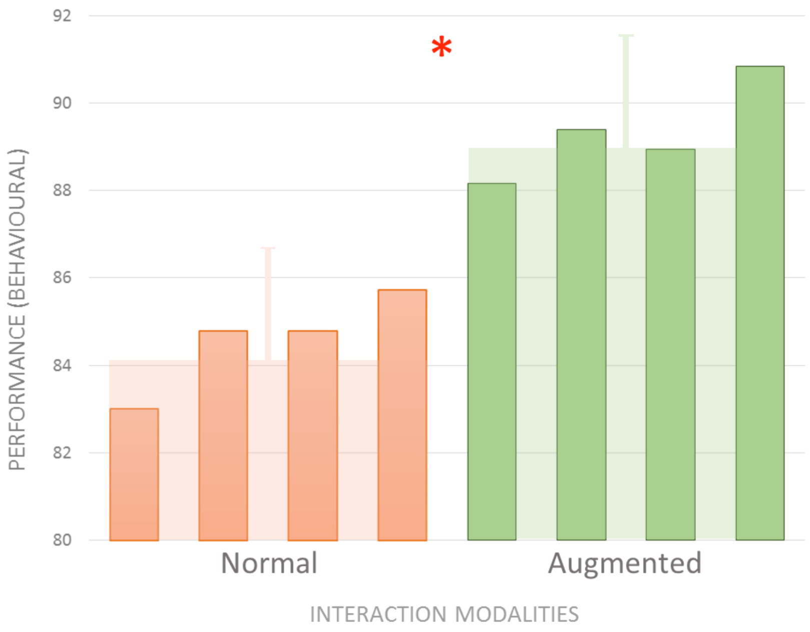

Figure 10 Performance and standard deviation values achieved by the users across the Normal and Augmented modalities for each repetition. The t-test highlighted a significant increment in performance during the Augmented modality.

2.10.2 Subjective data

A paired student's t-test ($\alpha=0.05$) has been performed between the NASA-TLX scores related to the two experimental conditions. NASA-TLX analysis showed a decrease in the perceived workload of ~5% during the Augmented modality, with respect to the Normal one. Anyhow, this difference was not significant ($p=0.214$).

2.10.3 Neurophysiological data

A paired student's t-test ($\alpha=0.05$) has been performed to investigate differences between modalities. The statistical analysis performed among the experimental conditions (Figure 11), revealed a significant decrease in the experienced workload during the Augmented solution with respect to the Normal one ($p=0.003$).

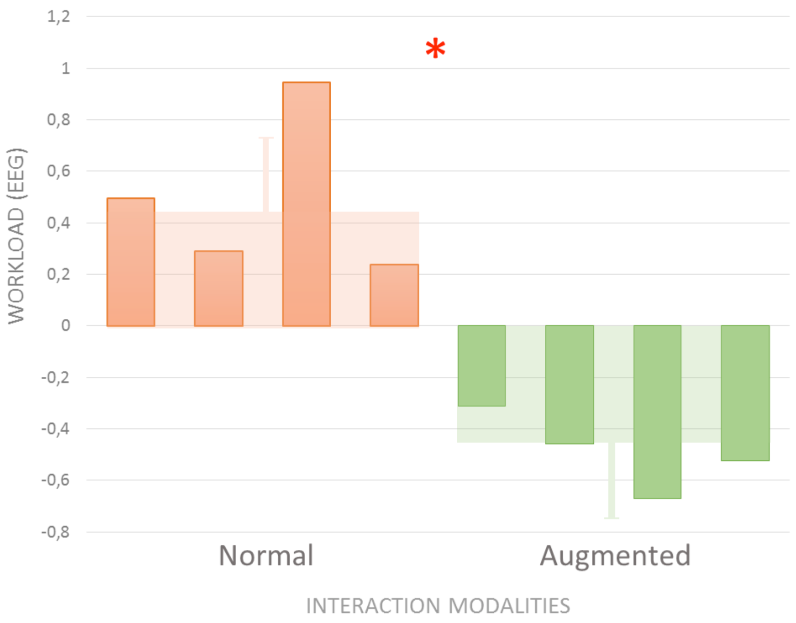

Figure 11 WLEEG scores and standard deviation related to the workload experienced by the ATCOs, for each modality and repetition. The statistical analysis highlighted a significant decrement in WLEEG scores during the Augmented modality.

2.10.4 Correlation analysis

A Pearson correlation analysis has been performed among performances and WL_{EEG} scores. Results showed a significant negative correlation between the two variables ($R=-0.9$; $p=0.0024$).

2.11 DISCUSSIONS

Results highlighted a clear advantage of the Augmented interaction modality with respect to the Normal one. More precisely, both performance and neurophysiological workload index analyses have confirmed such behavior. Furthermore, such indexes showed a significant negative correlation, confirming that higher performance is followed by a lower experienced workload and vice-versa. On the contrary, the subjective analysis did not reveal any significant trend, underlining the poor resolution (in terms of sample size) of such technique in comparison to neurophysiological measures (i.e. same trend, but statistically different). In addition, neurophysiological measures allow to measure the operator's mental states not only post-task (as for subjective analysis), but even during the task execution since the measure does not require any input from the user side and does not interfere with the task that he is performing. This feature could allow for better tuning of the technology the operator is interacting with, in order to optimize the human-machine interaction enhancing the performance of the whole system.

3. STUDY 3 -

3.1 Introduction

There are many small airports all around the world. For these airports, it is more and more difficult to maintain the cost-effectiveness of the operations. In order to improve cost-efficiency in air traffic provision, many countries are currently considering the remote tower air traffic control. The general idea is that the air traffic controller (ATCO) does not have to be located in the tower of the aerodrome that he is controlling, but he can operate from another location. With respect to standard control towers, this new solution will allow monitoring the traffic in small airports thanks to high-resolution cameras, advanced sensors and radio transmissions. The idea is that ATCOs should continuously interact with the tools that they are currently using but in a room that has a video wall that emulates the outside of a control tower. Anyhow, some implied information which could be crucial for ATCOs at a specific moment could be lost in a remote tower environment. For this reason, in the last year, many tools have been developed to try to enhance the performance of the operator, by providing synthetic feedback able to replicate, and even to augment real tower sensations (i.e., vibrations and/or sounds from the surroundings (Reynal et al., 2019). In this regard, Augmented Reality is nowadays one of the most important enabling technologies for innovation in a number of operational environments, since it offers the possibility to enhance the user interaction with the real environment by adding information to it in the form of synthetic overlays that enhance the user perception of the surrounding world (Masotti & Persiani, 2016). Dealing with an exponential increase of novel technologies and solutions able to enhance operator performance, it becomes essential to have reliable and precise tools to support the design phase.

New technologies or novel solutions in operational environments are often, among objective measurements, evaluated by using subjective assessment and/or judgment from operational experts (Papenfuss & Möhlenbrink, 2016; Van Schaik, Roessingh, Lindqvist, & Fält, 2010). However, it has been widely demonstrated how such measures could suffer from poor resolution due to a high intra and inter-operator variability depending on the nature of the measure itself (i.e., subjective). In addition, it is widely accepted in scientific literature the limit in using subjective measures alone, such as questionnaires and interviews, because of the impossibility to quantify "unconscious" phenomena underlying human behaviors (Gopher & Braune, 1984). Performance measure, when available, could provide just a part of the story, since an operator could achieve the same performance level by using different solutions, but experiencing different cognitive resources (e.g., experienced workload).

In this regard, in the last decade, it has been demonstrated that neurophysiological measures could be used to assess human mental states and that such kind of measures would achieve a higher resolution with respect to subjective assessing measures, providing additional information with respect to performance-based ones (Blankertz et al., 2016). With respect to subjective measures, the neurophysiological ones have the advantages to not impact on the task performed by the user, since they do not require any explicit input. Moreover, they can even be measured online, i.e., during the execution of the task (Gevins & Smith, 2003; Parasuraman, 2001). In this context, the mental workload is one of the most studied human mental states, because of its strong connection with the user's performance variation, affecting human error, system safety, productivity and operator satisfaction (Xie & Salvendy, 2000). The evaluation of mental workload by using (only) subjective measures is a quite disputed point in the scientific literature. de Winter (de Winter, 2014) stated that "*mental workload is an operational concept and not a representational concept, the idea that mental workload can be captured by the use of a questionnaire, and in particular by the use of the NASA-TLX alone, is too simplistic*." Mental workload is a more complex dynamic concept that needs to be assessed by more than just ratings on a subjective scale (de Waard & Lewis-

Evans, 2014). Mental workload reflects available cognitive resources (i.e., attentional resources and working memory capacity) during the execution of a task (de Waard and Lewis-Evans, 2014). It is widely demonstrated in literature as different biosignals features which are quite sensitive to workload variations. For example, eyes blinks, the rapid closing, and reopening of the eyelid are considered to be an indicator of both fatigue and workload. Recarte et al. (Recarte, Pérez, Conchillo, & Nunes, 2008) suggested that the specific amount of visual attention required by the task could lead to a blink inhibition but that the fatigue associated with long tasks would impair such inhibition. The monitoring of Heart Rate (HR) is commonly reported as being a measure sensitive to variations in mental workload. In particular, an increase in workload induces an increase in HR (Jorna, 1993). Regarding brain activity, electroencephalography (EEG) represents one of the most widely used techniques to infer relevant information regarding workload variations, because of its higher temporal resolution and relatively lower cost with respect to other brain imaging techniques (e.g., fMRI, MEG). One of the most studied features related to workload variation are frontal theta and parietal alpha bands. Activation of frontal and right parietal cerebral regions, reflected by synchronization in the theta band (4–8 Hz) and desynchronization in the alpha band (8–12 Hz), is sensitive to working memory load (Klimesch, 1999). In the same way, an increase in attentional resources has been linked to a desynchronization of the alpha band (Klimesch, Doppelmayr, Russegger, Pachinger, & Schwaiger, 1998) and theta band synchronization (Gevins & Smith, 2000). Both working memory and attention share the same cerebral regions and vary in the same way for numerous tasks. Anyhow, alpha-band synchronizations were also found during tasks soliciting frequent task switching (Pope, Bogart, & Bartolome, 1995). Other works highlighted an increase in alpha band power with an increase in task demand, especially as the number of tasks increased (Pope et al., 1995). Also, it was proposed that both alpha-band synchronization and desynchronization might be responsible for two different working memory maintenance mechanisms (Capilla, Schoffelen, Paterson, Thut, & Gross, 2012). As a result, alpha-band synchronization would support interfering item inhibition while alpha-band desynchronization would support relevant item maintenance

(Puma, Matton, Paubel, Raufaste, & El-Yagoubi, 2018). Moreover, the cognitive psychology literature demonstrates that the human psychophysiological activation, i.e., the arousal, has an "inverted U-shape" relationship with performance in that some levels of activation may help an individual to perform at a level that is higher than their baseline state (Yerkes & Dodson, 1908). Although this theory has been in some ways corrected and revised (Arent & Landers, 2003; Landers, 1980), the propaedeutic role of a proper psychophysiological activation to achieve the own best performance is a pillar of the behavioral researches (Eysenck, 2012). In this field, the skin sweating has been demonstrated to be one of the most sensitive physiological reactions to arousal variations (Bach, Friston, & Dolan, 2010), therefore the galvanic skin response (GSR) is considered the gold-standard bioindicator of human arousal (Boucsein, 2012).

Although many works have demonstrated the possibility to measure user's mental states in laboratory settings by using neurophysiological measures, just few works demonstrated the applicability of neurophysiological measures out of the labs i.e., real/realistic settings (Aloise et al., 2013; Pietro Aricò et al., 2017; P Aricò et al., 2016; Arico, Borghini, Di Flumeri, Sciaraffa, & Babiloni, 2018; Blankertz et al., 2016; G. Borghini et al., 2017; Cartocci et al., 2019; Gianluca Di Flumeri, Borghini, et al., 2019; Modica et al., 2018). Even less, the studies demonstrating the effectiveness of such measures in comparing different solutions in operational environments. In fact, although the recent technological achievements in wearable devices research towards the possibility to measure biosignals (e.g., brain, heart, and ocular activities) with a zero invasiveness for their employment in operational activities (Gianluca Di Flumeri, Aricò, et al., 2019), the technology, especially regarding EEG sensors, is not still ready for this kind of use (P. Aricò et al., 2018). On the contrary, pre-operational activities (e.g., designs of new solutions) could already benefit from this technology, since requirements in terms of invasiveness could be lower, benefiting instead from a powerful user's evaluation technique (i.e., neurophysiological measures).

Aricò et al. (P. Aricò et al., 2018) proposed a work aiming at investigating the possibility of employing neurophysiological-based measures to assess human-machine interaction effectiveness. In particular, two different interaction modalities (normal and augmented) related to the Air Traffic Management (ATM) field have been compared in terms of behavioral performance and an EEG-based workload index (frontal theta and parietal alpha ratio), by involving professional ATCOs in a control tower simulated environment at ENAC, Toulouse (France) in a laboratory-based task.

In the current study, we demonstrated: (i) the effectiveness of neurophysiological measures in comparison with subjective ones; and (ii) how the simultaneous employment of information coming from neurophysiological measures and behavioral ones could allow a holistic comparison of different solutions in remote tower pre-operational activities (i.e., design testing of new operational solutions).

In this regard, we employed specific neurophysiological, subjective and behavioral measures during: (i) the execution of two operational scenarios in remote tower operations; and (ii) specific critical events within each scenario, with the aim to compare the two investigated solutions.

3.2 Materials and Methods

3.2.1 Experimental Platform

The Simulation Platform used for the experiment has been developed by using ENAC facilities (Figure 14). The platform was designed to be able to simulate with high realism a Remote Tower environment. In this ecological perspective, the visuals on the airport vicinity had to be as realistic as possible. Therefore, we used a photorealistic view of Muret, an airport located in the South of France and under the supervision of the French Directorate General for Civil Aviation. The view was composed of several photographs of the airport previously stitched together and managed by an *ad hoc* software made with Unity and called RealTower.

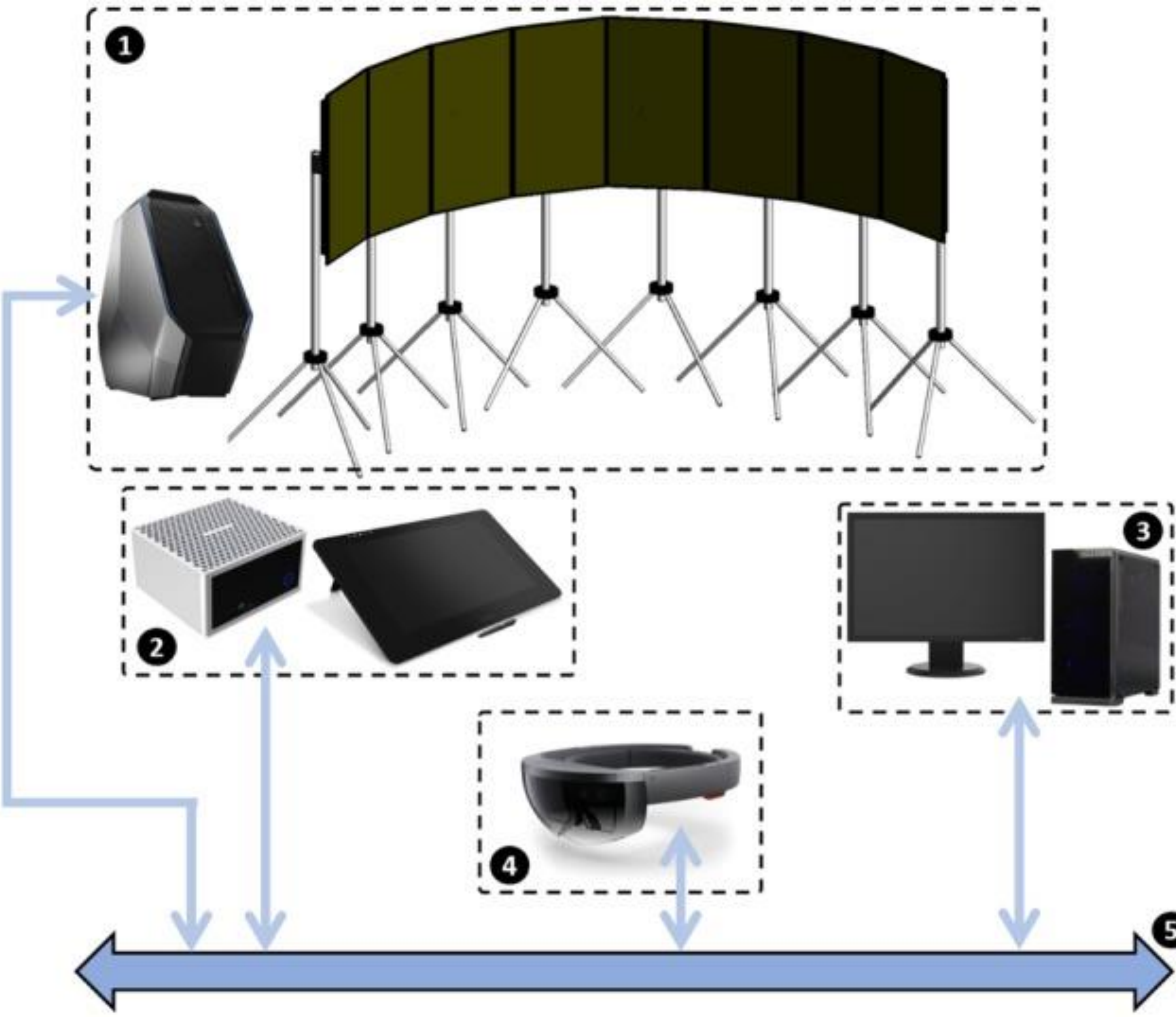

Figure 12 The general setup used for the experiment. (1) The panoramic view to display airport vicinity using eight UHD screens and the Alienware Area51 computer (2) and (3) Computers and screens to display respectively ground and air radars (4) Microsoft HoloLens device to retrieve user's head movements (5) ENAC Ivy bus software to make all these equipment communicate together.

To display the out-of-the-window view on Muret airport (RealTower software), we used an Alienware Area51 computer equipped with an Intel Core i7 processor and two Nvidia GeForce GTX 1080 graphic cards coupled together with SLI technology. This computer was connected to eight UHD screens (Iiyama Prolite X4071). The spatial sound was relayed using the speakers from these Iiyama monitors. The choice to use the screen's speakers instead of binaural sound made with Head-Related Transfer Functions, for example, was justified with the fact that we had to have a simple way to promote the platform to several people through demo sessions. The air radar view was managed using a Dell Precision 5810 computer equipped with an Intel Xeon processor and an Nvidia Quadro M500 graphic card. The ground radar view was displayed using a Zotac Magnus EN980 computer equipped with an Intel Core i5 processor and an Nvidia GeForce GTX 980 graphic card.

A homemade version of a chair was also used to spread vibrotactile feedback to the user (see Figure 13). Two Clark Synthesis T239 Gold tactile transducers have been attached to it, one behind the back and another one under the sit. All this equipment communicated with each other *via* software, written in C#, Java, and Python, and connected through an *ad hoc* bus called Ivy (Buisson et al., 2002), which makes it possible to easily transmit network messages by means of mechanisms based on regular expressions.

Figure 13 The wooden chair on which two tactile transducers have been attached to spread vibrotactile feedback (see the two red circles).

Because of the electromagnetic field generated by the haptic chair's transducers, no inertial measurement unit (IMU) could be used. Hence, we chose to monitor participants' head movements using a Microsoft HoloLens headset device, (Figure 14). The inertial unit embedded inside the Hololens in fact, allowed us to record its movements without a flaw, compared to the bad acquisitions obtained with the IMU alone.

Figure 14 Experimental setup (left) and pseudo-pilots positions used during the experiment (right). Pseudo-pilots have given written informed consent for the publication of this image.

Besides, two pilot positions (called "pseudo-pilot" positions) have been designed to simulate real flight communications (see Figure 14). These facilities were composed of two ground radars and communication means (e.g., microphones working in network coupled with two Griffin Powermate push-to-talk buttons) to make the pseudo-pilots able to dialog in real-time with the subject ATCO during the experiment.

3.2.2 Experimental Task and Working Scenarios

The experimental task consisted of two different scenarios of 10 min long. In one case, the scenario was supported by novel augmented solutions (*augm-RT*). For the other scenario, it was not provided any augmentation solution, so it had to be performed in a standard operational way (*norm-RT*).

These scenarios were designed with the help of two professional ATCOs, with the goal to be as much realistic as possible. Specific operational events that ATCOs often encounter during their working time have been enclosed in such scenarios.

Therefore, the two scenarios were similar from an operational point of view (i.e., the same kind and number of operational events) and overall difficulty level, but not exactly the same, to avoid any habituation and/or expectation effect of the experimental subjects. To avoid any confounding due to eventual differences in the two scenarios, augmented solutions were applied for half of the experimental group on Scenario 1 (and Scenario 2 was without augmented solutions), and the other half on Scenario 2 (and Scenario 1 was without augmented solutions). Scenarios were run by two pseudo-pilots placed in a different room from the one where the experiment was conducted.

Here below they have been summarized the schedule of each scenario:

–*Scenario 1*: four aircraft are parked at the Muret—Lherm Aerodrome at the beginning of the scenario. Another aircraft is scripted to be in a downwind position 3 min after the start approximately.

–*Scenario 2*: four aircraft are parked at the Muret—Lherm Aerodrome at the beginning of the scenario. Another aircraft is scripted to be in a downwind position at 3 min 30 s approximately, and another one is scripted to be in downwind position at 9 min 30 s approximately.

In the two scenarios, one SPATIAL event and two RWY events are scripted to have a minimum number of recordings. However, pseudo-pilots were encouraged to raise more events during the experiment. Therefore, in the recordings, we always had more than one SPATIAL and two RWY events per participant. For the analysis, we always considered the same number of events to compare the two scenarios (*norm-RT* and *augm-RT*).

In addition to the operational scenarios, we designed two more scenarios at two different difficulty levels, to be used to calibrate the W_{EEG} algorithm (see "Neurophysiological

Measures" section for further details). These two scenarios were designed by using the same operational characteristics of the two experimental scenarios, e.g., the ATCO point of view, visibility conditions, type of task required (i.e., remote tower operations). They were simulated pilots asking to start their engine, then to reach the holding point, before finally requesting a take-off, while other ones were aligned for landing. One scenario was composed of a few numbers (i.e., two) of aircraft (*EASY*), and the other one integrated a high number of aircraft (i.e., eight) in order to increase the difficulty level of the required task (*HARD*). Also, within the *EASY* scenario, only three actions were required from pseudo-pilots, while within the HARD scenario, there were 17 different actions requested.

Once the experimental platform and experimental task were ready, a shakedown test with three professional ATCOs has been conducted in order to ensure that the components of the platform and the augmented solutions were working properly and that the experimental setup achieved the maximum level of operability and realism.

From an operational point of view, we expected that the proposed Augmented solutions should be able to enhance performance and/or induce a reduction in cognitive workload experienced by the operators with respect to standard operation (i.e., no augmented solutions activated).

3.2.3 Design and Rationale for Augmentation Modalities

Specific events that could occur at the airport are crucial in terms of safety and therefore are of particular interest in the ATM domain and have been taken into account during the design phase of the experimental scenarios. These events are reported below:

- *Unauthorized movements on the ground*: it occurs when an aircraft starts its engine and starts to move on the parking area or the taxi circuit without previously asking for it or simply warn the control tour;

- *Runway incursion*: it represents one of the most dangerous events that could occur on the runway, it consists of an aircraft entering the runway without permission while another one is about to land soon (i.e., on its final approach segments).

  These events were enclosed in both scenarios the same number of times. Two specific augmentation modalities (i.e., a distinct one for each of these two events) have been designed to give the ATCOs a way to solve these events faster while being more aware of the highlighted situations. The following sections describe these two interaction modalities and related feedback to the operator.

3.2.3.1 Spatial Sound Alert for Unauthorized Event on Ground

Spatial sound alert modality is used to warn the participant that an unauthorized event on the ground (unauthorized start of aircraft engines or taxiing procedure) has been detected. The goal is to keep the participant's attention toward the abnormal event. This related event is called "SPATIAL" in the present article. The head of the participant is constantly tracked using the Microsoft HoloLens device to monitor whether his/her gaze is actually aligned with the related event or not. Spatial sound alert modality uses an audio cue. A distinctive sound is played in order to unequivocally attract the participant. Basically, this sound is composed of three pure sine waves at 880 Hz (A) of 50 ms length separated by 50 ms of silence, looped continuously until the participant's head is aligned with the event. To avoid creating doubt in case the participant would have anticipated the situation and is already focused on, the alert is not activated if the gaze of the participant is already in the direction of the abnormal situation.

Spatial sound alert modality is expected to reduce the time taken by the participants to resolve an abnormal situation (e.g., an aircraft moving on the parking area without any authorization, Figure 15).

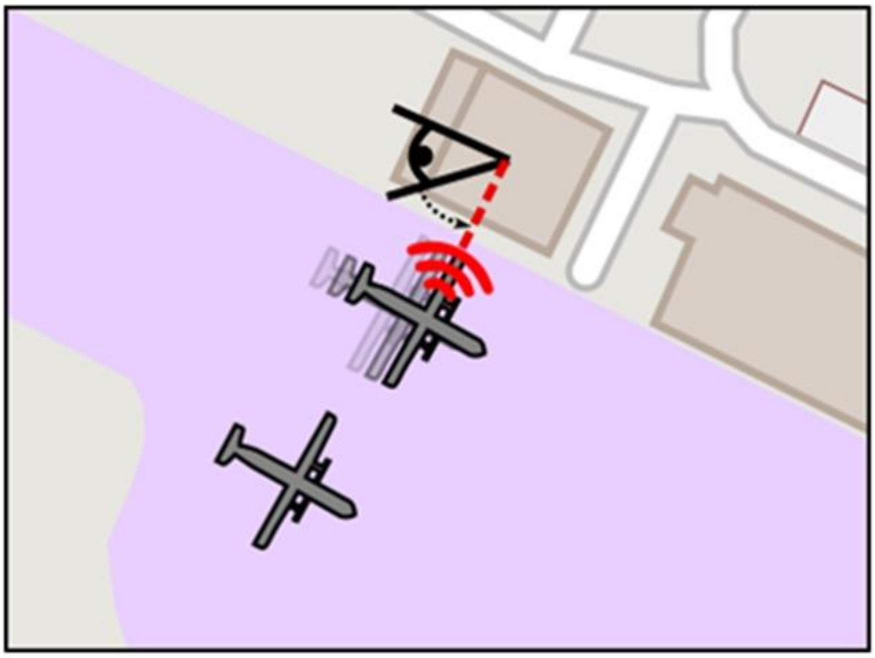

Figure 15 Alert risen by an unauthorized aircraft movement. The alert is spatialized, attracting the gaze of the user in the direction of the abnormal situation.

3.2.3.2 Runway Incursion Alert

A runway incursion event can be at the origin of the worst case of an incident: a collision with at least one vehicle with an important speed. The related event was here called "RWY." To support the ATCO in mitigating this risk, a highly disruptive alert has been designed. This alert combines Spatial sound alert (see the previous paragraph) and vibrotactile feedback. The combination of the vibrotactile input with a Spatial sound alert aims at giving cues concerning the event location. The vibrotactile feedback is calibrated to be clearly more highlighted compared to the spatial sound provided by Spatial sound alert modality. When an aircraft is in its final approach segment, all other aircraft being or entering on the runway (by crossing the holding point) will trigger the runway incursion alert. The alert is provided with a continuous vibration through the transducer located under the seat of the chair, made of an uncomfortable signal (i.e., modulated in frequency and amplitude). This signal does not allow any habituations. Moreover, to inform the participant of the localization of the incursion, the runway incursion's specific alert is completed by the generic Spatial sound alert, which is spatialized toward the holding point.

As for Spatial sound alert modality, Runway Incursion modality is expected to reduce the time taken by the participants to resolve a Runway Incursion situation (Figure 16).

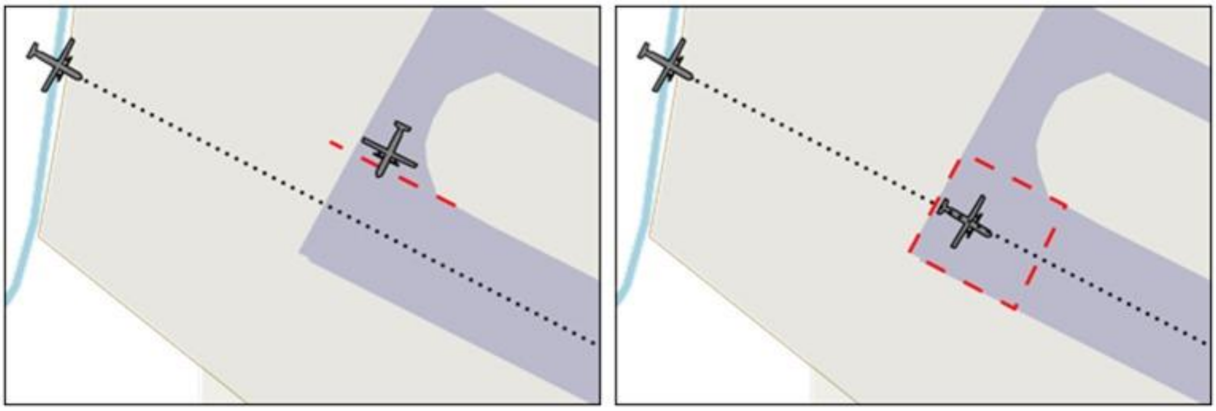

Figure 16 Left: a situation in which the aircraft located on the ground is stopped behind the holding point, generating no runway incursion situation. Right: the aircraft located on the ground entered the runway by crossing the holding; because of its location on the runway while another aircraft is about to land (on the final approach segment), this situation is a runway incursion.

3.2.4 Experimental Subjects

Sixteen professional ATCOs have been involved in the study. Each subject was asked to perform a specific training before starting with the experimental protocol, to get familiar with the simulation platform and employed technologies, and to avoid any learning effect during the scenario's execution. If the participant told us that he or she was not confident with experimental modalities, the training phase was repeated again. The group was composed of eight males and eight females, with a mean age of 39.4 years (*SD* = 7). None of them reported auditory problems (they are frequently checked due to their profession). Their main experience in the control tower was 10 years (*SD* = 6.8). Each participant signed a consent form before the experiment. The experiment was conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000. It received a favorable opinion from the local Ethical Committee.

3.2.5 Experimental Protocol

For each ATCO attending the protocol, EEG and ECG recordings were carried out using 13 Ag/AgCl passive wet electrodes (EEG, in particular, FPz, AFz, AF3, AF4, Fz, F3, F4, Pz, P3, P4, POz, PO3, PO4) referenced to both the earlobes and grounded to the left mastoid, according to the 10–20 standard (Jurcak et al., 2007) + 1 bipolar channel (ECG) placed on the chest of the subject. The device adopted in the experiment was the BEMicro (from EBNeuro Company) and the sampling rate was set to 256 Hz. In addition, GSR activity was monitored by means of the NeXus-10MKII system (MindMedia BV, Netherlands) and its dedicated software BioTrace+ with a sampling rate of 64 Hz. Biosignals have been synchronized with all the events coming from the simulation platform thanks to a specific device (Trigger Station, BrainTrends Srl). In particular, at the beginning of each experimental condition, the simulation platform sent a specific trigger to both the two biosignal amplifiers, in order to be able to synchronize all the information offline.

Once all the biosensors were placed and before starting with the execution of the experimental scenario, each subject was asked to perform short recordings, to be used as calibration for neurophysiological indexes computation. In particular, 1 min with eyes closed, to compute the Individual Alpha Frequency (IAF, Klimesch, 1999, see "Neurophysiological Measures" section for further details), and two recordings of 3 min each, of EASY and HARD difficulty levels scenarios, to be used as a calibration for the algorithm used to evaluate the EEG-based workload index (W_{EEG}, see "Neurophysiological Measures" section for further details). After these calibration recordings, the two experimental scenarios were run in a random order and the EEG, ECG and GSR signals have been recorded simultaneously during their execution. During the execution of the scenarios, a Subject Matter Expert (SME) provided his subjective judgment about participants' performance and mental workload. Immediately after each scenario, participants were asked to fill a specific questionnaire to evaluate their perceived mental workload (i.e., NASA Task Load Index NASA-TLX;(Hart & Staveland, 1988)).

3.2.6 Employed Measures

Subjective, behavioral and neurophysiological measures have been employed to compare the two operational solutions (i.e., norm-RT and augm-RT). Table 1 shows a summary of the used measures.

Table 1

Measures employed to compare the two operational solutions (i.e., norm-RT and augm-RT).

	Subjective measures	**Behavioral measures**	**Neurophysiological measures**
Mental workload	NASA-TLX		W_{EEG} index
Performance	SME evaluation	Reaction time	
Arousal			Skin conductance level

3.2.7 Subjective Measures

3.2.7.1.1 Cognitive Workload (NASA-Task Load Index)

The questionnaire used to estimate the mental workload was the NASA-TLX, that uses six 100-points range subscales to assess mental workload:

- 1. Mental demand: *How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?*
- 2. Physical demand: *How much physical activity was required? Was the task easy or demanding, slack or strenuous?*
- 3. Temporal demand: *How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?*
- 4. Performance: *How successful were you in performing the task? How satisfied were you with your performance?*

- 5. Effort: *How hard did you have to work (mentally and physically) to accomplish your level of performance?*
- 6. Frustration: *How irritated, stresses, and annoyed vs. content, relaxed, and complacent did you feel during the task?*

The six individual dimension ratings were linearly combined into a global score, ranging from 0 to 100.

3.2.7.1.2 Subject Matter Expert (SME) Evaluation

Direct and non-intrusive evaluation of ATCOs performance and experienced workload during the execution of the two operational scenarios with the augmented solutions activated (augm-RT) or not (norm-RT), was carried out by the SME, a professional ATCO with more than 25 years of experience in control tower and as ATM instructor.

The SME questionnaire consisted of two questions dealing with the overall performance achieved and the workload experienced by the participants. For each question, the SME had to provide a rate from 0 (Very Low) to 100 (Very High). In particular, the SME was asked to evaluate from one side the performance execution of the experimental subjects, as could he act like an instructor. On the other side, the SME should evaluate the workload level that in his opinion the ATCO had experienced overall in that specific scenario.

3.2.7.2 Behavioral Measures

Behavioral measurements (i.e., response times) were collected by automatically retrieving response times every time a specific operational event of interest occurred (i.e., SPATIAL or RWY events). With the goal to be able to compare these measurements during the behavioral data analysis, they were acquired during norm-RT *and* augm-RT scenarios. The only difference was that during norm-RT scenario, the participant was not helped with augmentation modalities because they were not activated, but SPATIAL and RWY events were still raised-up. Timers were used to retrieve response time (in milliseconds) for the two types of events:

- For the SPATIAL event, related to Spatial sound alert modality: the timer was started when the related unauthorized movement on ground was initiated by the pseudo-pilots and stopped when the participant successfully managed the induced situation;
- For the RWY event, related to Runway incursion alert modality, the timer was started until the moment that an aircraft entered the runway by crossing the holding point while another one was about to land very shortly. The timer was then stopped when the participant managed the dangerous situation by asking the pilot in short final to go for a go-around procedure.

This was true for norm-RT and augm-RT, but during augm-RT, specifically, the timer was started when the alarm started to sound (and for RWY event, also when the chair started to vibrate) and stopped when the alarm stopped (i.e., when the participant has aligned his head with the azimuth of the event).

Since, in the end, more than one measurement was collected for each type of event, response times were averaged for each participant and scenario. Unfortunately, the simulation interface that the operators had to interact with, was not fully accessible, then it was not possible to record another kind of behavioral measures apart ascribed reaction times. Anyhow, Reaction Times represents for sure a precious information (one of the most important ones), especially for the comparison of the proposed augmented solutions. In fact, from an operational point of view, it is very important being able to be aware as quickly as possible about possible emergencies, or abnormal situations. Anyhow, we tried to compensate for the lack or more behavioral measures by involving the SME and by asking him to rate this kind of overall performance measure.

3.2.8 Neurophysiological Measures

For each subject, we recorded EEG, ECG and GSR activity. In particular, EEG has been used to calculate the W_{EEG} index, the tonic component of the GSR signal was computed to quantify the level of arousal of the operator across the operational scenarios. Heart Rate estimation from ECG signal, and Blink Rate estimation from FPz channel, together with Frontal EEG Theta and Parietal EEG Alpha bands have been estimated to confirm that the two EASY and HARD calibration scenarios were actually able to induce two different workload levels.

3.2.8.1 Heart Rate Estimation

The ECG signal of the EASY and HARD conditions was first filtered by using a 5th-order Butterworth band-pass filter (High-Pass filter: cut-off frequency fc = 5 Hz; Low-Pass filter: cut-off frequency fc = 20 Hz), in order to reject the continuous component and the high-frequency interferences, such as that one related to the mains power source. At the same time, such filtering aims to emphasize the QRS process of the ECG signal. The following step consisted in measuring the distance between consecutive R-waves' peaks (RR distance) of the ECG signal (each R peak correspond to a heartbeat) in order to estimate the HR signal. In this regard, the Pan-Tompkins algorithm (J. Pan & Tompkins, 1985) has been employed, since it is the most used algorithm for the HR estimation for the ECG signal.

3.2.8.2 Blink Rate Estimation

The blink detection has been performed using a variant of BLINKER pipeline (Kleifges, Bigdely-Shamlo, Kerick, & Robbins, 2017). In its original version, BLINKER algorithm selects the best channel among an arbitrary number of EEG channels, allowing the optimal identification of blinks. However, we forced the BLINKER algorithm to use the FPz channel to detect the blinks. To proceed to the blink detection, the FPz signal has been bandpass filtered between 1 and 20 Hz. Potential blinks have amplitude 1.5, the standard deviation of the signal, duration higher than 100 ms and interblink interval higher than 50 ms. Among these only the blinks showing a correlation with the tent-like shape higher than 0.9 has been considered. The last check selects the blinks with a Positive Amplitude Ratio higher than 3 to remove saccades. For each condition, it has been computed the number of Blinks per minute (that represents the Blink Rate).

3.2.8.3 EEG Spectral Features Estimation and W_{EEG} Index Evaluation

As stated before, depending on the literature evidence, theta EEG rhythms over frontal sites, and alpha EEG rhythms over parietal sites have been taken into account for the computation of the EEG-based workload measure, i.e., the W_{EEG} index. The W_{EEG} index has already been validated in many operational settings, in which it was used to assess the workload experienced by the subjects across different difficulty levels. The W_{EEG} index is based on the application of a linear regression algorithm, the *automatic stop StepWise Linear Discriminant Analysis* (asSWLDA; Aricò et al., 2016b), able to extract EEG features sensitive to workload assessment (i.e., Frontal Theta Bands and Alpha Parietal Band). In the following, all the algorithm steps that allow computing the W_{EEG} index starting from the EEG signal have been reported (Figure 17). Different processing and artifact removing algorithms have been applied to the EEG signal in order to: (i) remove eye-blinks contribution, that could affect frequency bands related to workload and consequently introduce a bias in the measure; and (ii) remove all the other sources of artifacts (i.e., muscular artifact, saturation of amplifier). First of all, the EEG channels have been first band-pass filtered with a 5th-order Butterworth filter [low-pass filter cut-off frequency: 30 (Hz), high-pass filter cut-off frequency: 1 (Hz)]. The FPz channel has been used to remove eyes-blink artifacts from the EEG data by using the regression-based algorithm REBLINCA (Di Flumeri et al., 2016), that could affect the EEG frequency bands involved in workload assessment. With respect to other regressive algorithms (e.g., Gratton method, Gratton et al., 1983) the REBLINCA algorithm has the advantages to preserve EEG information in blink-free signal segments by using a specific threshold criterion that recognizes the occurrence of an eye-blink, and only in this case the method cleans the EEG signals. If there is not any blink, the method has not any effect on the EEG signal. The band-pass filtered [1–7 (Hz), 5th order Butterworth filter] FPz signal has then been used as a template to remove the eye-blinks contribution from the EEG signal. At this point, following the algorithm described in Aricò et al. (2016b), the EEG signals have been segmented into epochs of 2 s, shifted of 0.125 s. Specific procedures of the EEGLAB toolbox have been used (Delorme & Makeig, 2004) to remove artifacts generated by muscular activity and bio amplifier saturation. In particular, three criteria have been applied. *Threshold criterion*: if the EEG signal amplitude exceeds ±100 (μV), the corresponding epoch would be marked as an artifact. *Trend criterion*: each EEG epoch has been interpolated in order to check the slope of the trend within the considered epoch. If such slope was higher than 10 (μV/s) the considered epoch would be marked as an artifact. *Sample-to-sample difference criterion*: if the amplitude difference between consecutive EEG samples was higher than 25 (μV), it meant that an abrupt variation (no-physiological) happened and the EEG epoch would be marked as an artifact. In the end, all the EEG epochs marked as artifact have been rejected from the EEG dataset with the aim to have an artifact-free EEG signal from which estimate the brain variations along with the different conditions. All the previous mentioned numeric values have been chosen following the guidelines reported in Delorme and Makeig (2004).

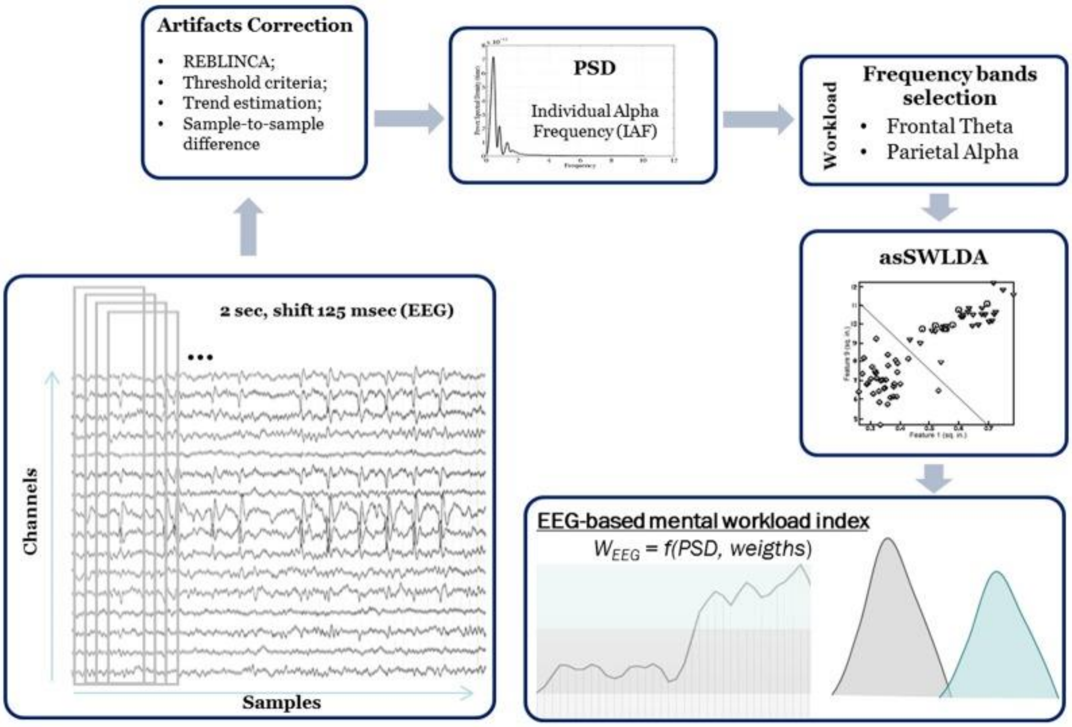

Figure 17 Electroencephalography (EEG) processing chain for the WEEG index computation.

At this point, *Power Spectral Density* (PSD) was calculated for each EEG epoch using a Hanning window of the same length of the considered epoch (2 s length, which means 0.5 Hz of frequency resolution). Then, the EEG frequency bands of interest have been defined for each subject by the estimation of the *IAF* value (Klimesch, 1999). In order to have a precise estimation of the alpha peak and, hence of the IAF, as stated before the subjects have been asked to keep the eyes closed for a minute before starting with the experiments. Finally, a spectral features matrix (EEG channels × Frequency bins) has been obtained in the frequency bands directly correlated to the mental workload. In particular, only the theta rhythm (IAF-6 ÷ IAF-2), over the EEG frontal channels (AFz, AF3, AF4, Fz, F3, F4), and the alpha rhythm (IAF-2 ÷ IAF+2), over the EEG parietal channels (Pz, P3, P4, POz, PO3, and PO4) have been considered as variables for the mental workload evaluation.

At this point, a linear classification algorithm (asSWLDA, Aricò et al., 2016b), an upgraded implementation of the well-known SWLDA algorithm has been used to select the subjective discriminant EEG spectral features related to the workload. With respect to the standard SWLDA approach, the asSWLDA algorithm embeds an automatic procedure to select the best number of relevant features to keep into the discrimination model. This property was demonstrated so far to increase the robustness to both the under and the overfitting phenomenon (Aricò et al., 2016b). Once trained with specific "calibration data," the algorithm can be used to compute a workload index (i.e., W_{EEG} index) on other data (i.e., experimental scenarios) by combining the selected EEG features with specific weights in output from the model itself. In particular, the algorithm has been calibrated by using the EASY and HARD related data already described.

In conclusion, *z*-score transformation (Zhang et al., 1999) has been used to compute a normalization of W_{EEG} index distribution.

3.2.8.4 Skin Conductance Level Estimation

The signal related to the Skin Conductance, named hereafter GSR, has been recorded with a sampling frequency of 64 Hz. A constant potential was applied in order to induce a skin electrical current. The variations of such current are a function of the skin conductance variations. The recorded signal was then entirely processed by using the Matworks MATLAB software. First of all, the signal was down-sampled to 16 Hz, in order to reduce the data amount. Also, the signal was then filtered through a 5th order Butterworth low-pass filter (cut-off frequency = 2 Hz) in order to remove all the higher frequency components that are not related to skin sweating activity. Finally, the signal was processed by using the Ledalab suite, a specific open-source toolbox implemented within MATLAB for the GSR processing. The Continuous Decomposition Analysis (Benedek & Kaernbach, 2010) has been applied in order to separate the Tonic (Skin Conductance Level—SCL) and the Phasic (Skin Conductance Response—SCR) components of the GSR. In the following analysis, the mean value of the SCL during the experimental scenarios has been taken into account.

3.3 Results

All the statistical comparisons have been performed through two-sided Wilcoxon signed-rank tests. In fact, data come from multiple observations on the same subjects, but it is not possible to assume or robustly assess that the distribution of the observations is Gaussian, therefore paired non-parametric tests have been used (Siegel & Castellan, 1956).

3.3.1 Calibration Conditions Analysis

HR, Blink Rate, Frontal Theta, and Parietal Alpha EEG bands values have been calculated on the data related to calibration conditions (EASY and HARD) in order to confirm that they were well designed (i.e., were able to induce two different workload levels, Figure 18). Statistical results suggested that all the considered neurophysiological indicators changed significantly among EASY and HARD conditions, confirming the suitability of the design of the EASY and HARD conditions in inducing two different levels of experienced workload. In particular, HR increased significantly (EASY = 72 ± 8; HARD = 77 ± 8; $p = 0.002$) and the Eyeblink rate decreasing significantly (EASY = 16 ± 7; HARD = 14 ± 8; $p = 0.03$) during the HARD condition. Regarding EEG frequency bands, Frontal Theta significantly increased (EASY = 264 ± 75; HARD = 357 ± 94; $p = 4.38 \times 10^{-4}$), and even Parietal Alpha increased (EASY = 80 ± 27; HARD = 91 ± 26; $p = 0.013$) with increasing difficulty level (HARD).

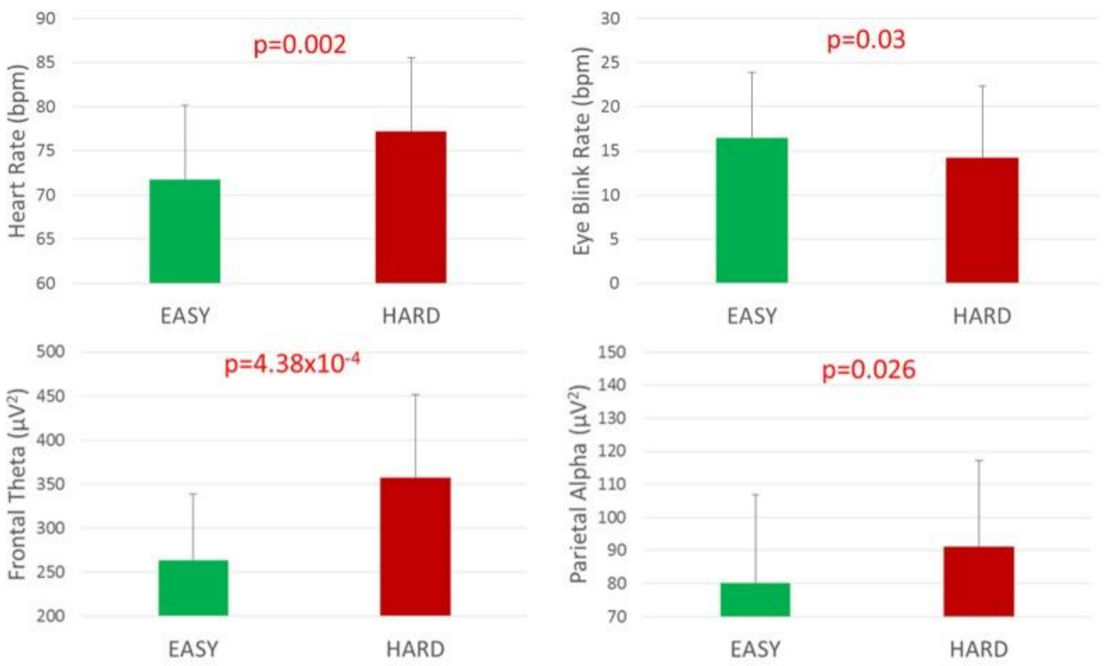

Figure 18 Neurophysiological measures between EASY and HARD conditions. Significant p-values have been marked in red. Values showed are mean and standard error, at a confidence interval of 0.95.

3.3.2 Subjective

The perceived workload was evaluated considering both the NASA-TLX workload score (i.e., directly filled from experimental subjects, $W_{NASA\text{-}TLX}$), and post-run SME assessment (W_{SME}). Statistical test exhibited no significant trends between the two conditions (norm-RT = 39 ± 14 and augm-RT = 42 ± 12, $p = 0.28$). On the contrary, scores provided by the SME highlighted a significant increase in overall workload during the augm-RT (70 ± 10) condition with respect to the norm-RT (79 ± 9) one ($p = 0.03$, Figure 19).

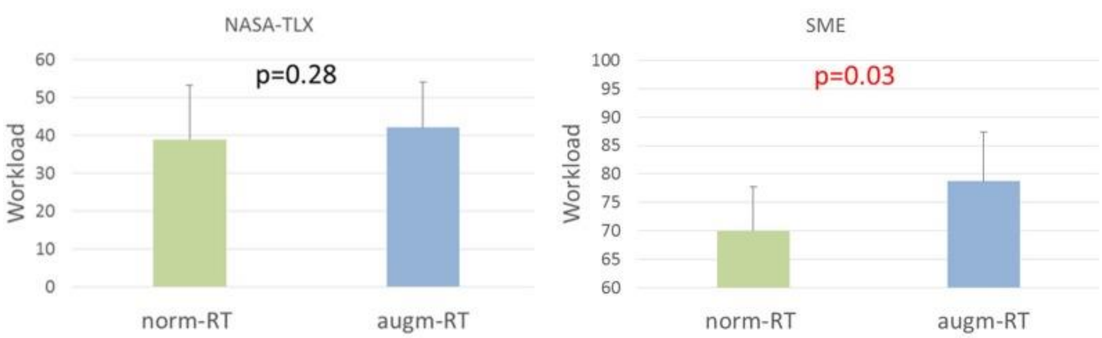

Figure 19 Subjective post-run workload scores assessed by experimental subjects (i.e., WNASA-TLX, on the left) and subject matter expert (SME; i.e., WSME, on the right) values for each condition. Significant p-values have been marked in red.

Even Performance values provided by the SME did not highlight any significant difference between the two experimental conditions (norm-RT = 76 ± 5; augm-RT = 78 ± 6; p = 0.29, Figure 20).

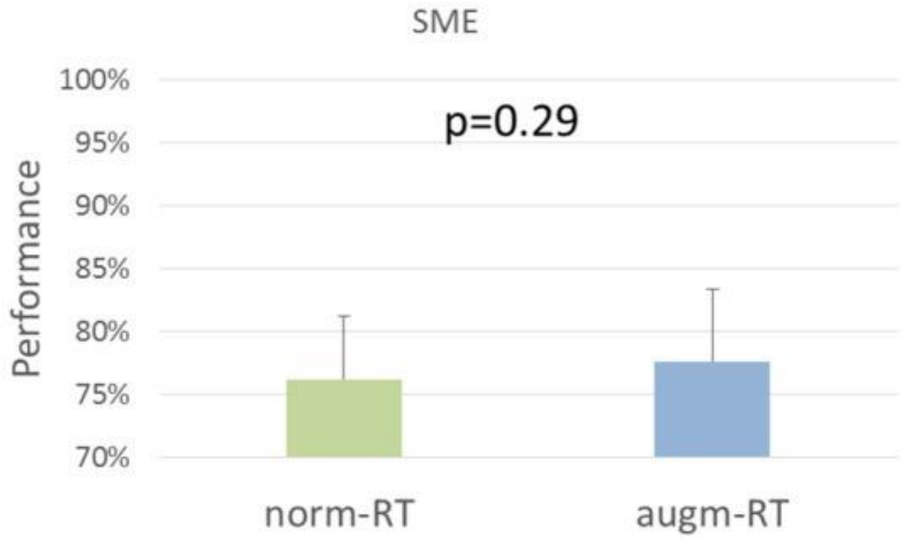

Figure 20 Performance values assessed by the SME for each experimental condition. Values showed are mean and standard deviation.

3.3.3 Neurophysiological Results

3.3.3.1 EEG-Based Workload Index (W_{EEG}) Evaluation

EEG features (i.e., Frontal Theta and Parietal Alpha) of EASY and HARD conditions have then been used to calibrate the asSWLDA algorithm for each subject, in order to compute the W_{EEG} index for each experimental scenario (i.e., norm-RT and augm-RT). Statistical analysis showed a significant overall increase (norm-RT = 0.004 ± 0.09; augm-RT = 0.1 ± 0.15; p = 0.04) of the index during the augm-RT scenario execution with respect to the norm-RT one (Figure 21, right). In addition, we performed a point by point statistical analysis (two-sided Wilcoxon signed-rank tests) between the W_{EEG} indexes of each scenario to highlight the reason for this increase in experienced workload. The test highlighted that the significance was reached (i.e., red rectangle) just around the appearance of the spatial sound events (Figure 21, left). On the contrary, during the appearance of Runway incursion events (i.e., gray bars), the workload increased in both the scenarios, in fact, statistics did not show any significant trend. In this regard, we reported in Figure 21 both the W_{EEG} indexes of the two scenarios second by second. In addition, we reported *p*-values for each time point and the events distribution for all the subjects.

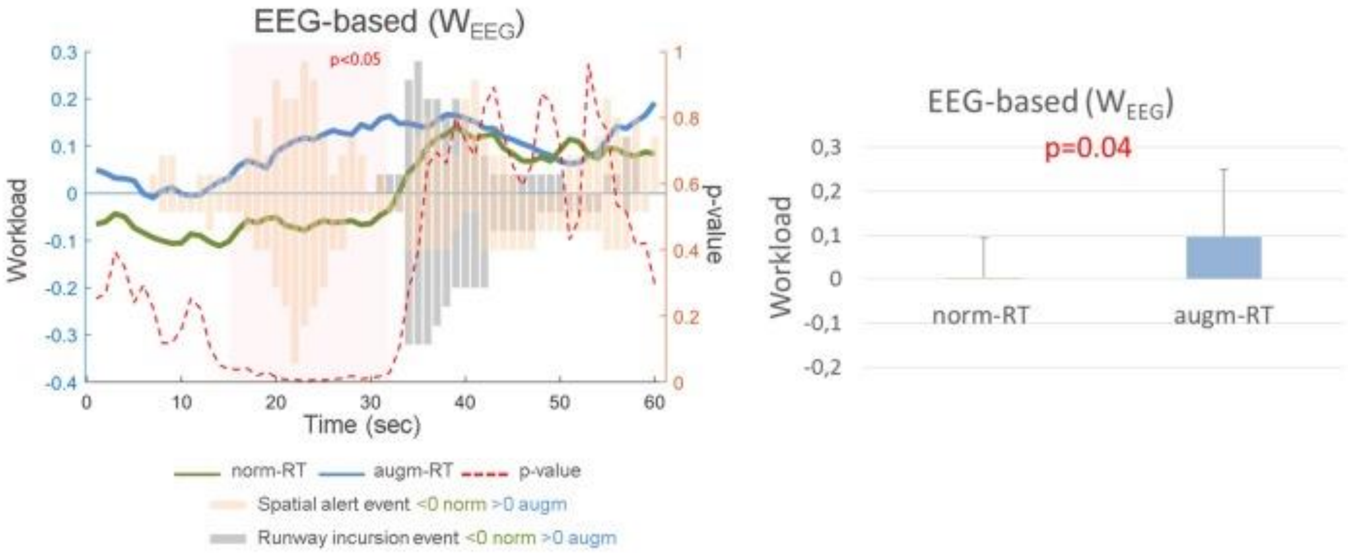

Figure 21 WEEG score exhibited a significant increase in the experienced workload during the augm-RT condition with respect to the norm-RT one, which is consistent with the results provided by the SME workload assessment. Bars higher than 0 are referred to augm-RT scenario events, while bars lower than 0 are referred to norm-RT scenario events.

3.3.3.2 Skin Conductance Level

We computed the SCL of the GSR signal for each experimental scenario, in order to investigate the emotional involvement (i.e., arousal level) of operators during the execution of the two scenarios.

The statistical analysis showed a significant increase (norm-RT = 2.34 ± 0.84; augm-RT = 2.42 ± 0.82; $p = 0.049$) of the Tonic component during the augm-RT scenario execution with respect to the norm-RT one (Figure 22).

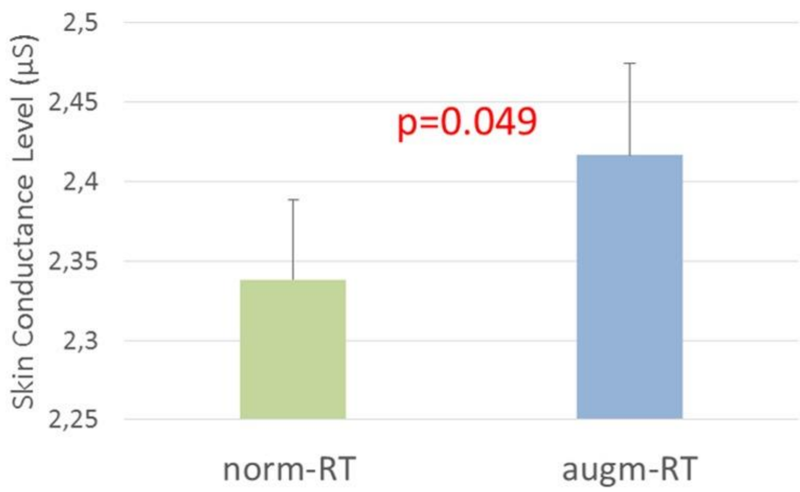

Figure 22 The analysis of Skin Conductance Level (SCL) of the galvanic skin response (GSR) signal revealed an increase in the arousal during the augm-RT condition with respect to the norm-RT one. Significant p-values have been marked in red.

3.3.4 Single Event Analysis

The simulation platform developed at ENAC allowed recording the reaction time needed to the experimental subjects to reply to specific events under which the augmented solutions could be activated (i.e., augm-RT scenario) or not (i.e., norm-RT scenario). In the following, the reaction times and the related W_{EEG} index recorded during these events (i.e., starting from the onset of the event, until the ATCO addressed the issues highlighted within the specific event) have been analyzed.

3.3.4.1.1 Behavioral Results

Statistical results highlighted a significant decrease in reaction times needed from the experimental subjects to identify (and react) to the specific events (Runway Incursion norm-RT = 41 ± 35; augm-RT = 20 ± 30; $p = 0.03$; Spatial Alert norm-RT = 58 ± 57; augm-RT = 10 ± 5; $p = 6 \times 10^{-5}$) if the augmented solutions were activated (i.e., augm-RT) in comparison to the scenarios without any augmented solution activated (norm-RT, Figure 23).

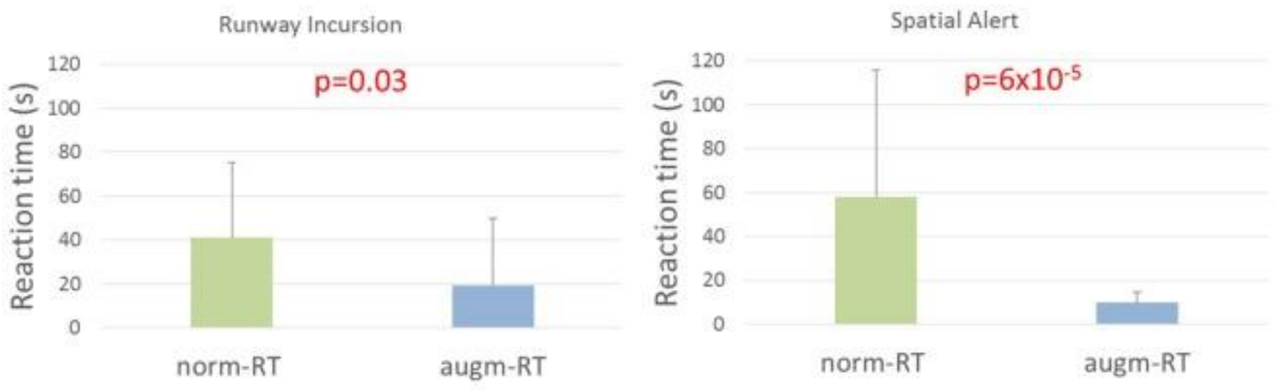

Figure 23 Reaction times recorded by the simulation interface exhibited a significant decrease in the time needed by the subjects to identify and react to the specific event, highlighting a local increase in performance induced by the specific augmented solutions.

3.3.4.1.2 W_{EEG} Scores

Thanks to the possibility to calculate the EEG-based workload index along with the execution of the two scenarios, over time, it was possible to measure the experienced workload of the subject during the occurrence of the specific operational events where augmentation could (augm-RT) or not (norm-RT) be applied. Statistical analysis did not highlight any statistical difference in the workload experienced by the experimental subjects during the events related to Runway incursion alert (norm-RT = 0.46 ± 0.03; augm-RT = 0.46 ± 0.06; $p = 0.79$). In addition, if we consider norm-RT scenario from an operational point of view, it can be seen that the Runway incursion event induced a significantly higher workload than the Spatial alert event (1.7×10^{-6}). On the contrary, during the event related to Spatial sound alert, it was highlighted with a significant increase in the experienced workload (norm-RT = 0.01 ± 0.28; augm-RT = 0.34 ± 0.54; $p = 0.04$) if the augmented solutions were activated, with respect to the condition without any augmentation activated (Figure 24).

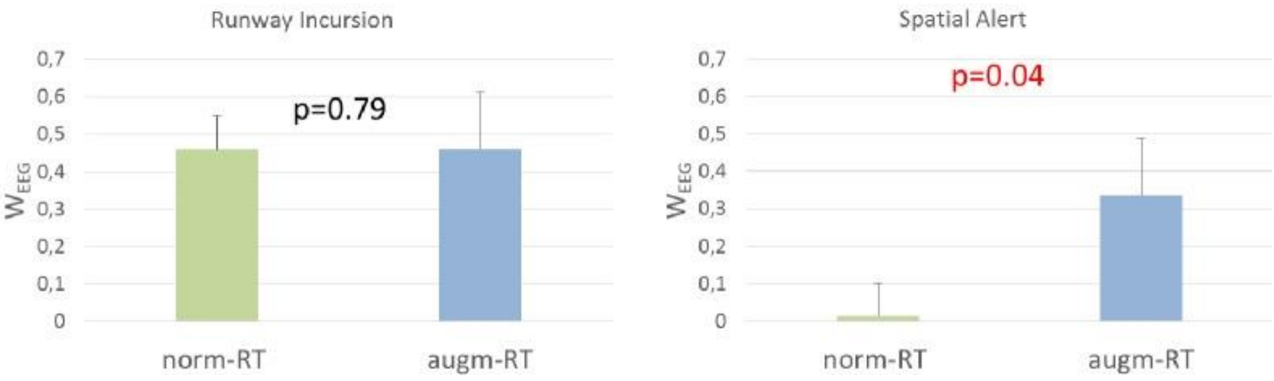

Figure 24 WEEG values calculated during the specific events highlighted no significant difference during the Runway incursion event, and a significant increase during the Spatial Alert event. Significant p-values have been marked in red. Values showed are mean and standard deviation.

Figure 25 shows for each of the 16 experimental subjects involved in the study-specific WEEG behavior second by second and related events appearance.

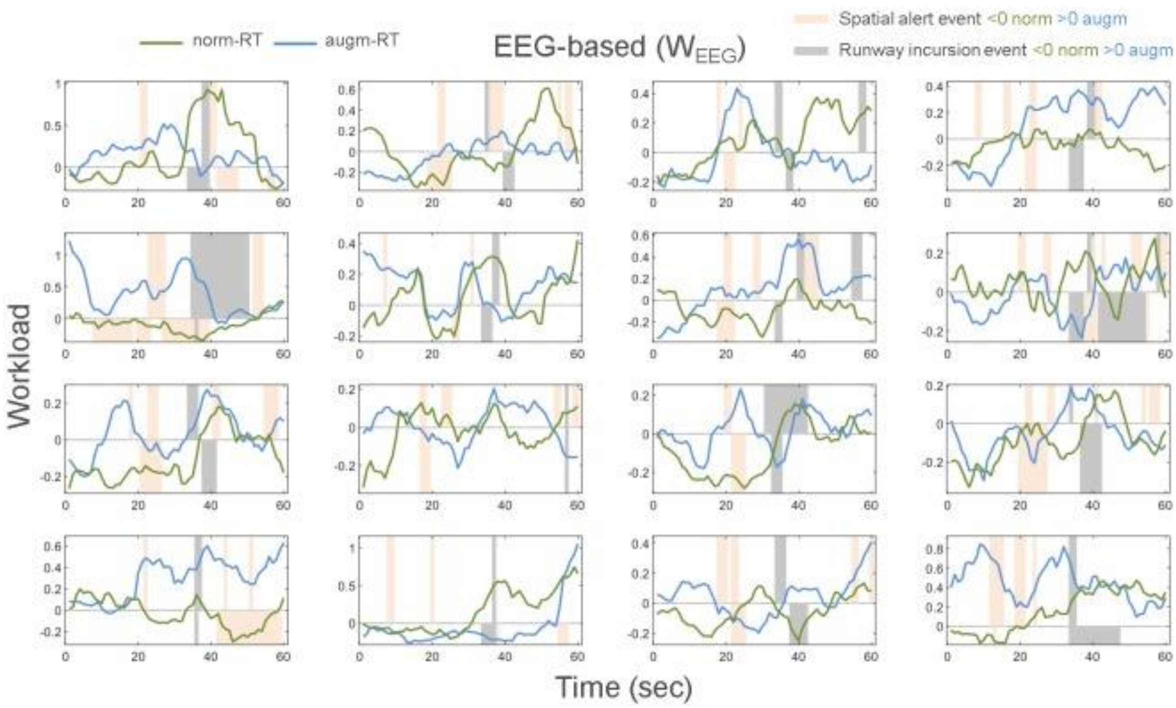

Figure 25 WEEG index shape for each of the 16 experimental subjects and related operational events appearance (i.e., Spatial Alert and Runway Incursion) have been reported. In particular, the time duration of each event has been reported for each of the two scenarios (i.e., norm-RT and augm-RT). Bars higher than 0 are referred to augm-RT scenario events, while bars lower than 0 are referred to norm-RT scenario events.

3.4 Discussion

The present work aimed to demonstrate that the combination of information coming from neurophysiological measures, together with behavioral ones, could provide a more comprehensive assessment in the design process of new technologies and solutions in operational environments with respect to the evaluation of only behavioral and/or subjective measures. In this regard, the unimodal assessment could not be enough informative to investigate if new technology could be effective or not in order to enhance operator performance. In fact, considering the subjective measures alone, an incongruence emerges: the NASA-TLX could suggest that the two solutions did not induce any decrease in perceived workload. On the other side, the SME glimpsed an overall increase in the operator workload, but any improvement in the overall performance. This point highlights also the low resolution of the NASA-TLX questionnaire in assessing perceived workload, as demonstrated in de Waard and Lewis-Evans (2014). In this regard, taking a look to neurophysiological measures, in particular to the W_{EEG} index, it highlights an increase in the experienced cognitive workload if the augmented solutions were activated (consistently with the SME workload assessment), and in particular, a statistically significant increase during one specific event (i.e., the one related to Spatial sound alert). On the contrary, during the Runway incursion event both the solutions (norm and augm-RT) induced a similar increase in workload and no differences were highlighted between the two modalities in this case. Anyhow, this information alone could not be enough if we do not look at the same time at the performance achieved by the experimental subjects. In this regard, the overall performance filled by the SME did not reveal any difference between the two experimental conditions (i.e., norm-RT and augm-RT), but the subject's reaction times significantly decreased during the experimental events (RWY and SPATIAL) if the related augmented solutions were activated. If we consider also the significant increase in the tonic component of the GSR signal during the augm-RT related scenario, that implies a higher psychophysiological involvement (i.e., arousal) of the experimental subjects. Therefore, the overall story can suggest the following insights:

- 1. The proposed augmented solutions induced a local increase in subject performance on those operational events where the augmentation was applied. The overall performance seems to not be affected by the proposed solutions.
- 2. The augmented solutions induced an increase in the workload experienced by the participants. Anyhow, this increase is still acceptable and also fruitful, since it did not negatively impact the performance (that indeed locally increased) and has to be intended only as a consequence of the higher engagement of the controller. This behavioral effect is totally in line with physiological results obtained in terms of arousal (following point).
- 3. Augmented solutions induced an increase in the arousal level of the subjects, representing a positive aspect for the overall performance assessment. This result corroborates the workload highlights and supports the overall interpretation of the propaedeutic effect of augmented solutions in enhancing the controller engagement, and consequently its performance. In fact, according to the inverted U-shape relationship theory between human performance and psychophysiological activation (conjugated with the various concepts of arousal, stress, workload;(Janelle, 2002)), in general, a proper increase in workload and arousal is not disruptive, but necessary for humans to achieve their best performance.

Finally, taking again a deeper look at the single event assessment, it is clear that augmented solutions induce a decrease in the time needed by the operator to catch the emergency and react to it. In other words, although the overall performance level of the operator seemed to not change significantly, the augmented solutions induced a local increase in performance. Anyhow, the related events during the augm-RT scenario, in particular during the event related to the Spatial sound alert, induced a higher experienced workload, identified by the W_{EEG} index. This effect could be due to the information that the ATCOs would not be aware of if they would not have been alerted. Even if in this specific experiment workload increase did not induce a decrement in performance, it has to be stressed on the possible negative effects of increased workload over time, such as fatigue. Increased workload and resulting fatigue have in fact continued to play a major role in many transport accidents to the present day (European Agency for Safety and Health at Work, 2011). In this regard, training of operators could have a significant effect on the experienced workload. Even if the use of the two sensory channels of hearing and touch could be seen as natural, the novelty for them of these kinds of interaction techniques and the information provided consequently could need a longer appropriation time, implying specific care in learning sessions. In other words, although ATCOs have been well trained to use in a correct way the proposed augmented solutions before the experiment and they could be useful in specific operational situations (i.e., Runway Incursion and Spatial sound alert), they could become in general too much distracting, inducing in some way an increase in experienced workload. However, novelty and lack of familiarity can be modulated by learning, which could mitigate this effect in the long term by decreasing workload, and eventually related induced fatigue effects or potential lack in performance.

3.4.1 Highlights for Multimodal Interaction in Remote Tower

The present work allowed us to provide a few suggestions about the implementation of multi-modal interaction in future remote towers platforms. In particular:

- **Auditory and Vibrotactile feedback shall be used to improve the salience of relevant remote operational events or highlight changes in operational status**. The outcomes of this experiment highlighted a possible advantage of implemented Augmented solutions when applied to specific operational events (i.e., Spatial Alert and Runway Incursion Alert) since the related reaction time values were significantly shortened once Augmented solutions were activated.

- **Multimodal remote tower platform implementation shall take into account the ATCO control tower soundproofing background**. Results showed that providing ATCOs with visual, audio and vibrotactile feedback seemed to increase the ATCOs experienced workload level. ATCO feedback collected during the validation activities suggests that this effect may be influenced by the controller's familiarity with the specific tower environment; namely if ATCO used to work in a soundproofed tower or not. The constant provision of stimuli that are not necessarily associated with relevant operational events, seems to be perceived as disturbing in a control tower soundproofed from the outside aerodrome environment. The provision of multimodal feedback in a multimodal remote tower platform shall be modulated according to ATCO familiarity soundproofed tower environment; in order to minimize the distracting effect that may induce a decrease in performance and an increase in experienced workload.

- **The implementation of augmented multimodal feedback in remote tower requires an extended familiarization period**. Results showed an increase in cognitive workload while using augmented solutions because of the novelty of the information and/or by the novelty of the modalities used (e.g., audio and vibration). Nowadays, ATCOs use the audio and haptic modalities unconsciously since many control towers are soundproof, and the vibration is generally not felt. ATCOs will need more intensive training and familiarization period to understand the real effect of the provided augmentations on performance."

3.5 Conclusion

The present work demonstrated the suitability of neurophysiological measures during the design and validation phase of new solutions/technologies/tools in operational environments, in particular for remote tower operations using augmented reality. More specifically:

- The computed W_{EEG} index can be used for the evaluation of mental workload in operational settings and could provide a measure at a higher resolution if compared with the classical assessing methods (i.e., questionnaires). In this regard, it is able to provide information over time (i.e., at a level of few seconds), opening up the possibility to assess the workload level experienced by the operators during specific operational activities or particular events, allowing to identify critical points during the operational activity or along the design process and to optimize the human-machine interaction, even in real-time. As an example, highlighted in this study, the W_{EEG} index revealed that the Runway incursion events induced a significantly higher workload than Spatial alert events. By using subjective measures, it was not possible to highlight this kind of behavior, and this information could instead be relevant in operational activities.

- GSR was confirmed to be quite sensitive in showing the psychophysiological activation of subjects during operational activities, therefore appearing as a powerful and less invasive tool for evaluating human engagement.
- The evaluation of neurophysiological measures is able to provide additional information with respect to behavioral and subjective measures, and its employment during the pre-operational activities (e.g., design process) could allow a more holistic and accurate assessment of new technology/solutions.

4. GENERAL CONCLUSION

In this thesis, immersive virtual reality and augmented reality contributions to both laboratory and applied research projects have been explored. In Study 1 the impact of multimodal cueing and multisensory load modulation on performance, mental workload and presence were investigated recording behavioral and electrophysiological responses in virtual reality. In Study 2 and 3 instead, we investigated whether neurophysiological measures could be used to assess the human-machine interaction effectiveness. In particular, an augmented reality system was employed in a remote control tower to enhance Air traffic controllers performance.

In the first Study described, in the second chapter, a novel paradigm was devised. The task consisted in a naturalistic visual search task in virtual reality, in which participants drove a car on a virtual racetrack with the aim to hit the highest number of spheres (i.e., the targets)

Results confirmed that multisensory integration boosts performance under a high load condition, suggested that vibrotactile cues can be used to decrease mental workload and made clear that the sense of presence is modulated by both multisensory events and perceptual load. This paradigm indicates that virtual reality not only can be used to study attentive phenomena and mental workload but that maybe is, in particular, adapt to investigate brain responses in a more naturalistic way.

In Study 2 and 3, we explored whether a hearing augmentation system based on the gaze direction of professional air traffic controllers could boost their performance and reduce their workload. The results showed a clear advantage of the augmented interaction mode over the normal one as both the performance analysis and the analysis of the index of the neurophysiological workload have shown. In addition, these indices demonstrated a significant negative correlation, confirming that higher performance is followed by less experienced workload and vice versa. Conversely, the subjective analysis did not reveal any significant trends, underlining the poor resolution (in terms of sample size) of this technique compared to neurophysiological measures (i.e. the same trend, but statistically different). Moreover, neurophysiological measurements allow measuring the mental states of the operator not only after the task but also during the execution, since the measurement does not require any input from the user and does not interfere with the task he is performing. This feature could allow better tuning of the technology with which the operator interacts, in order to optimize the human-machine interaction improving the performance of the entire system.

4.1 Future Directions

In the last decade, virtual reality has established itself as a valid tool for the empirical investigation of embodied cognition theories (Banakou, Groten, & Slater, 2013; Maister, Slater, Sanchez-Vives, & Tsakiris, 2015; V. I. Petkova, Khoshnevis, & Ehrsson, 2011) and this trend is set to increase. In the last decade, we have seen not only the renewed accessibility of the instrument in economic terms (Bohil et al., 2011), which has gone from tens of thousands of dollars to a few hundred in eight years but also a stable progress in technology that promises to involve our senses in an increasingly convincing way. We have been witnessing the multiplication on the market of devices, such as gloves and sensors, compatible with virtual reality helmets that allow you to make a faithful real-time tracking of the body in space. Although muscle movement tracking is not new in the field of research, the ability to map in real time and at a low cost your actions on a virtual avatar in an environment that meets the laws of physics provide an ideal testbed for naturalistic cognitive studies and *embodied* theories (X. Pan & Hamilton, 2018). Additionally, the photorealism achieved to date by the graphics engines available for free to anyone (e.g., Unity3d and Unreal) has nothing to envy to the quality of computer graphics that have made the fortune of many colossal movies. Some studies have already pointed out some years ago how it is possible to investigate using virtual avatars social interaction mechanisms (Hamilton, Pan, Forbes, & Hale, 2016). However, with the advent of increasingly realistic avatars that no longer fall into the "uncanny valley" (Figure21)(Mori, MacDorman, & Kageki, 2012), it will be also possible to conduct experiments of social interaction with agents almost indistinguishable from a human being, but at the same time controlling each expressive variable of the face, body or voice.

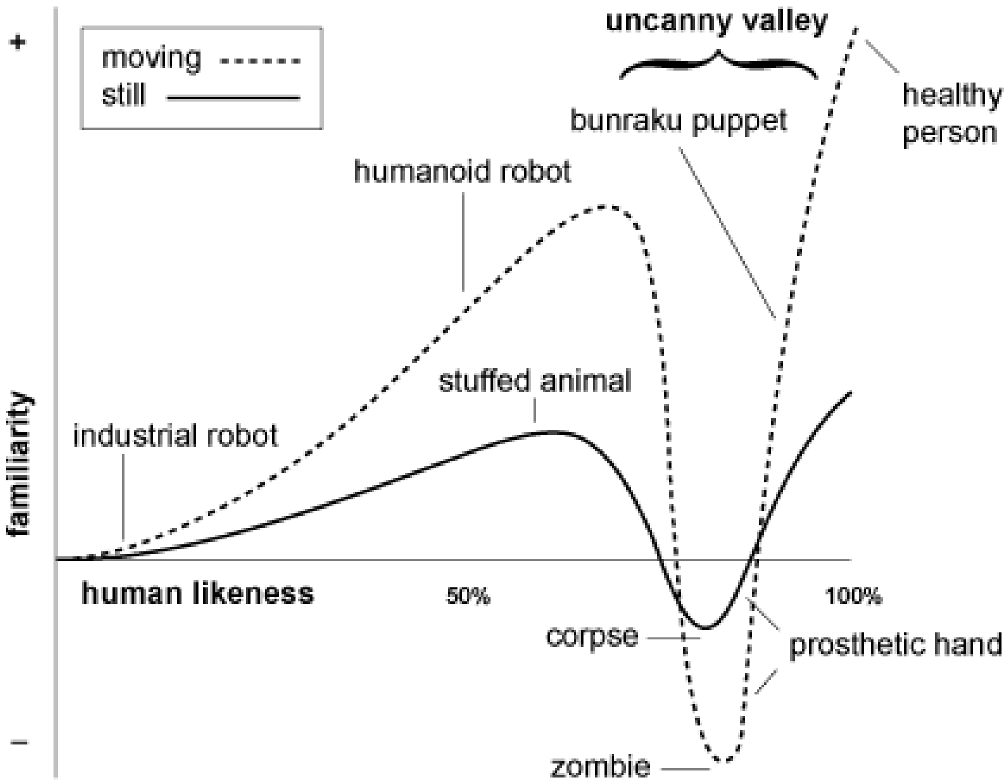

Figure 26 illustrates the Uncanny Valley effect, where the feeling of familiarity aroused by virtual avatars, robots and anthropomorphic automata grows as their visual resemblance to the human figure increases until, however, the extreme representative realism generates a sudden collapse, arousing unpleasant sensations such as repulsion and restlessness.

In the future, we will see how the use of these latest generation graphics engines will be used as a test bench for the development of "embodied" artificial intelligence. According to Barsalou (Barsalou, 2010) in fact, robotics will be an excellent testbed for testing hypotheses of brain architectures based on grounded cognition. Rodney Brooks, former director of the Massachusetts Institute of Technology's computer science and artificial intelligence lab, claimed the same twenty years ago. According to the author, intelligence always requires a body, and in the field of artificial intelligence we should stop thinking about complex internal representations and processes of reasoning instead focusing on the interaction between the system and the environment (Brooks, 1991).

More recently, a call to the development of deep learning algorithms more inspired by biology has also been submitted by influential names in the field of artificial intelligence (Bartunov et al., 2018) including Geoffrey Hinton, a recent Turing Prize winner and director of Google Brain. Certainly, embedded neural models inspired by biology represent a growing interest in robotics (Pfeifer & Bongard, 2006) and are an excellent inspiration in the development of autonomous adaptive systems. However, perhaps the implementation of new deep learning algorithms could find in graphic and physical engines (e.g., Unity, Mujuco, Pybullet) the ideal ground for their development. In fact, adopting the same reasoning that led psychologists and neuroscientists to invest in virtual reality, one could equip an artificial intelligence with a virtual body to be tested in a 3D environment that responds perfectly to the laws of physics. This would allow not only to conduct simulations of the interaction with the environment more quickly and in a less expensive way but would also make it possible to equip the AI with a body that changes over time and senses that are refined. This could be useful on the one hand to test hypotheses of brain architectures hypothesized by neurosciences but also, for example using genetic(Kramer, 2017) or evolutionary (Bäck, Fogel, & Michalewicz, 2018) algorithms to investigate the evolution of a neural system with a given body in a given context from generation to generation. Finally, by immersing both a human (through VR) and an Artificial Agent in the same virtual environment it will be possible to study their interaction or train the agent by imitation (J. Ho & Ermon, 2016; Schaal, 1999) without the risks of using a real robot and while they share the same body.

In particular, we are designing: 1) A VR study that leverages machine learning-based gesture recognition to animate a virtual human arm using electromyography (Fig. 27) for both amputees and the normal population. 2) A VR study that will explore the dangerous affordances elicited in a human by an artificial agent that shares with the former the same virtual space.

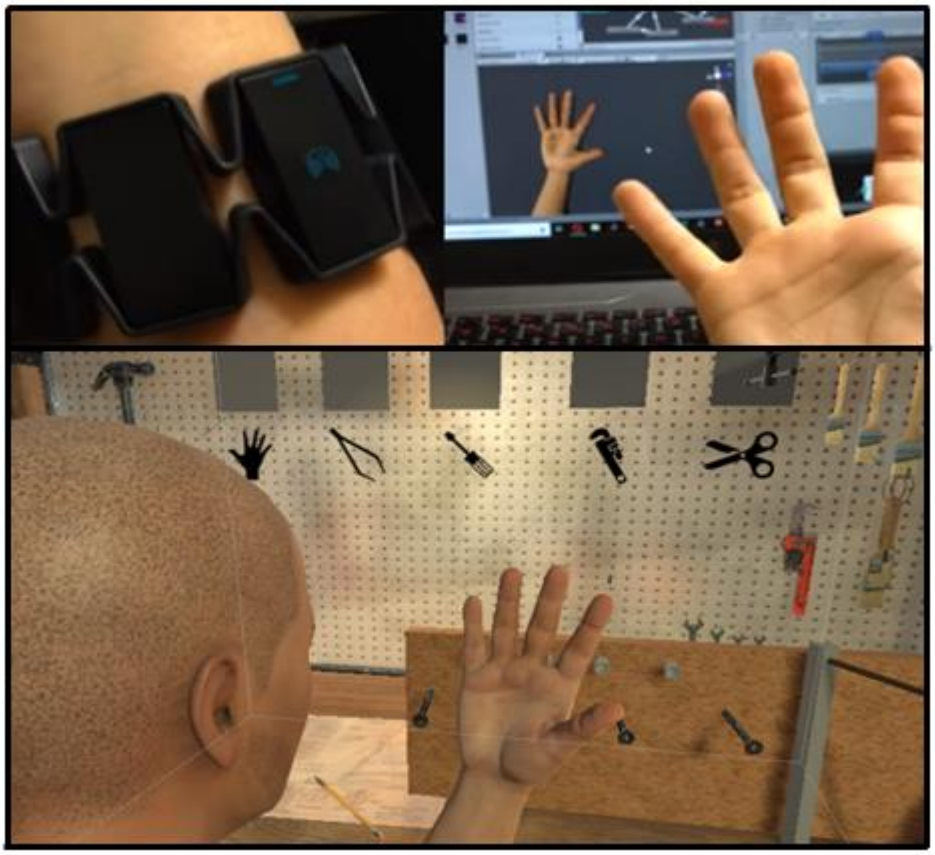

Figure 27 shows the Myo Armband EMG system, a Unity3d-based interface and the virtual scenario under construction.

5. Acknowledgments

I would like to express my deepest gratitude to my supervisors Enea and Viviana for their patience and invaluable guidance in this journey. I would also thank all my colleagues from both the Aglioti lab and the Applied Neuroscience Lab for the funny moments we shared and the help they provided me. I would also thank my partner Roberta and my pets Saki, Hiroshi, Kirikù , and Skippy for making my life special.

6. BIBLIOGRAPHY

Aloise, F., Aricò, P., Schettini, F., Salinari, S., Mattia, D., & Cincotti, F. (2013). Asynchronous gaze-independent event-related potential-based brain–computer interface. *Artificial intelligence in medicine, 59*(2), 61-69.

Arent, S. M., & Landers, D. M. (2003). Arousal, anxiety, and performance: A reexamination of the inverted-U hypothesis. *Research quarterly for exercise and sport, 74*(4), 436-444.

Arico, P., Aloise, F., Schettini, F., Salinari, S., Mattia, D., & Cincotti, F. (2014). Influence of P300 latency jitter on event related potential-based brain-computer interface performance. *J Neural Eng, 11*(3), 035008. doi:10.1088/1741-2560/11/3/035008

Aricò, P., Borghini, G., Di Flumeri, G., Bonelli, S., Golfetti, A., Graziani, I., . . . Benhacene, R. (2017). Human factors and neurophysiological metrics in air traffic control: a critical review. *IEEE reviews in biomedical engineering, 10*, 250-263.

Arico, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., . . . Babiloni, F. (2016). Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Frontiers in human neuroscience, 10*, 539. doi:10.3389/fnhum.2016.00539

Arico, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., & Babiloni, F. (2016). A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Prog Brain Res, 228*, 295-328. doi:10.1016/bs.pbr.2016.04.021

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., & Babiloni, F. (2016). A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks *Progress in brain research* (Vol. 228, pp. 295-328): Elsevier.

Arico, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., & Babiloni, F. (2018). Passive BCI beyond the lab: current trends and future directions. *Physiol Meas, 39*(8), 08TR02. doi:10.1088/1361-6579/aad57e

Aricò, P., Reynal, M., Imbert, J.-P., Hurter, C., Borghini, G., Di Flumeri, G., . . . Ferreira, A. (2018). *Human-machine interaction assessment by neurophysiological measures: a study on professional air traffic controllers.* Paper presented at the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE computer graphics and applications, 21*(6), 34-47.

Bach, D. R., Friston, K. J., & Dolan, R. J. (2010). Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *Int J Psychophysiol, 76*(1), 52-55. doi:10.1016/j.ijpsycho.2010.01.011

Bäck, T., Fogel, D. B., & Michalewicz, Z. (2018). *Evolutionary computation 1: Basic algorithms and operators*: CRC press.

Bagassi, S., De Crescenzio, F., Lucchi, F., & Masotti, N. (2016). Augmented and virtual reality in the airport control tower.

Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences, 110*(31), 12846-12851.

Barsalou, L. W. (2008). Grounded cognition. *Annu Rev Psychol, 59*, 617-645. doi:10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in cognitive science, 2*(4), 716-724.

Bartunov, S., Santoro, A., Richards, B., Marris, L., Hinton, G. E., & Lillicrap, T. (2018). *Assessing the scalability of biologically-motivated deep learning algorithms and architectures.* Paper presented at the Advances in Neural Information Processing Systems.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *Journal of Statistical Software, 67*(1). doi:10.18637/jss.v067.i01

Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, 47*(4), 647-658. doi:10.1111/j.1469-8986.2009.00972.x

Bimber, O., & Raskar, R. (2006). *Modern approaches to augmented reality.* Paper presented at the ACM SIGGRAPH 2006 Courses.

Blankertz, B., Acqualagna, L., Dähne, S., Haufe, S., Schultze-Kraft, M., Sturm, I., . . . Müller, K.-R. (2016). The Berlin brain-computer interface: progress beyond communication and control. *Frontiers in Neuroscience, 10*, 530.

Blascovich, J., Loomis, J., 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, 13*(2), 103-124.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat Rev Neurosci, 12*(12), 752-762. doi:10.1038/nrn3122

Borghini, G., Arico, P., Di Flumeri, G., Cartocci, G., Colosimo, A., Bonelli, S., . . . Babiloni, F. (2017). EEG-Based Cognitive Control Behaviour Assessment: an Ecological study with Professional Air Traffic Controllers. *Sci Rep, 7*(1), 547. doi:10.1038/s41598-017-00633-7

Borghini, G., Aricò, P., Di Flumeri, G., Salinari, S., Colosimo, A., Bonelli, S., . . . Babiloni, F. (2015). *Avionic technology testing by using a cognitive neurometric index: a study with professional helicopter pilots.* Paper presented at the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature, 391*(6669), 756-756. doi:Doi 10.1038/35784

Boucsein, W. (2012). *Electrodermal activity*: Springer Science & Business Media.

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence, 47*(1-3), 139-159.

Bruder, G., Wieland, P., Bolte, B., Lappe, M., & Steinicke, F. (2013). *Going with the flow: Modifying self-motion perception with computer-mediated optic flow.* Paper presented at the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).

Capilla, A., Schoffelen, J.-M., Paterson, G., Thut, G., & Gross, J. (2012). Dissociated α-band modulations in the dorsal and ventral visual pathways in visuospatial attention and perception. *Cerebral Cortex, 24*(2), 550-561.

Cartocci, G., Modica, E., Rossi, D., Inguscio, B. M. S., Arico, P., Martinez Levy, A. C., . . . Babiloni, F. (2019). Antismoking Campaigns' Perception and Gender Differences: A Comparison among EEG Indices. *Comput Intell Neurosci, 2019*, 7348795. doi:10.1155/2019/7348795

Celsi, R. L., & Olson, J. C. (1988). The role of involvement in attention and comprehension processes. *Journal of consumer research, 15*(2), 210-224.

Cipresso, P., Chicchi Giglioli, I. A., Alcañiz Raya, 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, 9*, 2086.

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis, 58*(1), 7-19.

Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., & Meyer, G. (2018). The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PLoS One, 13*(2), e0191846.

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *Neuroscientist, 8*(2), 132-142. doi:10.1177/107385840200800209

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). *Surround-screen projection-based virtual reality: the design and implementation of the CAVE.* Paper presented at the Proceedings of the 20th annual conference on Computer graphics and interactive techniques.

Cummings, J. J., & Bailenson, J. N. (2015). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology, 19*(2), 272-309. doi:10.1080/15213269.2015.1015740

de Waard, D., & Lewis-Evans, B. (2014). Self-report scales alone cannot capture mental workload. *Cognition, technology & work, 16*(3), 303-305.

de Winter, J. C. (2014). Controversy in human factors constructs and the explosive use of the NASA-TLX: a measurement perspective. *Cognition, technology & work, 16*(3), 289-297.

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

Di Flumeri, G., Aricò, P., Borghini, G., Sciaraffa, N., Di Florio, A., & Babiloni, F. (2019). The dry revolution: Evaluation of three different EEG dry electrode types in terms of signal spectral features, mental states classification and usability. *Sensors, 19*(6), 1365.

Di Flumeri, G., Borghini, G., Aricò, P., Sciaraffa, N., Lanzi, P., Pozzi, S., . . . Simone, A. (2019). EEG-Based Mental Workload Assessment During Real Driving: A Taxonomic Tool for Neuroergonomics in Highly Automated Environments *Neuroergonomics* (pp. 121-126): Elsevier.

Di Flumeri, G., Borghini, G., Arico, P., Sciaraffa, N., Lanzi, P., Pozzi, S., . . . Babiloni, F. (2018). EEG-Based Mental Workload Neurometric to Evaluate the Impact of Different Traffic and Road Conditions in Real Driving Settings. *Frontiers in human neuroscience, 12*, 509. doi:10.3389/fnhum.2018.00509

Dinh, H. Q., Walker, N., Hodges, L. F., Chang, S., & Kobayashi, A. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. 222-228. doi:10.1109/vr.1999.756955

Eysenck, M. (2012). *Attention and arousal: Cognition and performance*: Springer Science & Business Media.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advances in psychophysiology, 2*(S 1), 78.

Faraway, J. J. (2005). Extending the linear model with r (texts in statistical science).

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *8*(12), 1643-1646. doi:10.1038/nn1608

Fodor, J. A. (1983). *The modularity of mind*: MIT press.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science, 308*(5722), 662-667.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences, 4*(1), 14-21.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the cognitive sciences, 4*(1), 23-48.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*(2), 593-609.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*: Cambridge university press.

Gevins, A., & Smith, M. (2005). Neurocognitive function EEG measurement method and system: Google Patents.

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex, 10*(9), 829-839.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4*(1-2), 113-131.

Gillingham, K. K. (1992). The spatial disorientation problem in the United States Air Force. *Journal of Vestibular Research, 2*(4), 297.

Gillmeister, H., & Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. *Brain Res, 1160*, 58-68. doi:10.1016/j.brainres.2007.03.041

Gonzalez-Franco, M., Perez-Marcos, D., Spanlang, B., & Slater, M. (2010). *The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment.* Paper presented at the 2010 IEEE virtual reality conference (VR).

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26*(5), 519-532.

Ha, H.-G., & Hong, J. (2016). Augmented reality in medicine. *Hanyang Medical Reviews, 36*(4), 242-247.

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology, 22*(10), R390-R392.

Hamilton, A., Pan, X. S., Forbes, P., & Hale, J. (2016). *Using virtual characters to study human social cognition.* Paper presented at the International Conference on Intelligent Virtual Agents.

Hancock, P. A., Mercado, J. E., Merlo, J., & Van Erp, J. B. F. (2013). Improving target detection in visual search through the augmenting multi-sensory cues. *Ergonomics, 56*(5), 729-738. doi:10.1080/00140139.2013.771219

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research *Advances in psychology* (Vol. 52, pp. 139-183): Elsevier.

Hasson, U., Nusbaum, H. C., & Small, S. L. (2009). Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences, 106*(26), 10841-10846. doi:10.1073/pnas.0903253106

Hendrix, C., & Barfield, W. (1996). Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators & Virtual Environments, 5*(3), 274-289.

Ho, C., Reed, N., & Spence, C. (2007). Multisensory in-car warning signals for collision avoidance. *Hum Factors, 49*(6), 1107-1114. doi:10.1518/001872007X249965

Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals: when spatial correspondence matters. *Exp Brain Res, 195*(2), 261-272. doi:10.1007/s00221-009-1778-5

Ho, J., & Ermon, S. (2016). *Generative adversarial imitation learning.* Paper presented at the Advances in neural information processing systems.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS biology, 3*(3), e79.

IJsselsteijn, W. A., de Kort, Y. A. W., & Haans, A. (2006). Is this my hand I see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleoperators and Virtual Environments, 15*(4), 455-464.

Insko, B. E. (2003). Measuring presence: Subjective, behavioral and physiological methods.

Janelle, C. M. (2002). Anxiety, arousal and visual attention: A mechanistic account of performance variability. *Journal of sports sciences, 20*(3), 237-251.

Jausovec, N., & Jausovec, K. (2012). Working memory training: improving intelligence--changing brain activity. *Brain Cogn, 79*(2), 96-106. doi:10.1016/j.bandc.2012.02.007

Jorna, P. (1993). Heart rate and workload variations in actual and simulated flight. *Ergonomics, 36*(9), 1043-1054.

Juan, M. C., Baños, R., Botella, C., Pérez, D., Alcaníiz, M., & Monserrat, C. (2006). An augmented reality system for the treatment of acrophobia: the sense of presence using immersive photography. *Presence: Teleoperators and Virtual Environments, 15*(4), 393-402.

Jurcak, V., Tsuzuki, D., & Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage, 34*(4), 1600-1611. doi:10.1016/j.neuroimage.2006.09.024

Kahol, K., French, J., Panchanathan, S., Davis, G., & Berka, C. (2006). Evaluating the role of visio-haptic feedback in multimodal interfaces through EEG analysis. *Augmented Cognition: Past, Present and Future*, 290.

Kalawsky, R. S. (2000). *The validity of presence as a reliable human performance metric in immersive environments.* Paper presented at the 3rd International Workshop on Presence.

Kathner, I., Wriessnegger, S. C., Muller-Putz, G. R., Kubler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. *Biol Psychol, 102*, 118-129. doi:10.1016/j.biopsycho.2014.07.014

Kleifges, K., Bigdely-Shamlo, N., Kerick, S. E., & Robbins, K. A. (2017). BLINKER: automated extraction of ocular indices from EEG enabling large-scale analysis. *Frontiers in Neuroscience, 11*, 12.
Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews, 29*(2-3), 169-195.
Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., & Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neuroscience letters, 244*(2), 73-76.
Kober, S. E., & Neuper, C. (2012). Using auditory event-related EEG potentials to assess presence in virtual reality. *International Journal of Human-Computer Studies, 70*(9), 577-587. doi:10.1016/j.ijhcs.2012.03.004
Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297*(5582), 846-848.
Kramer, O. (2017). *Genetic algorithm essentials* (Vol. 679): Springer.
Lakie, W. L. (1967). Relationship of Galvanic Skin Response to Task Difficulty, Personality Traits, and Motivation. *Research Quarterly. American Association for Health, Physical Education and Recreation, 38*(1), 58-63. doi:10.1080/10671188.1967.10614803
Landers, D. M. (1980). The arousal-performance relationship revisited. *Research quarterly for exercise and sport, 51*(1), 77-90.
Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *J Exp Psychol Gen, 133*(3), 339-354. doi:10.1037/0096-3445.133.3.339
Lavie, N., Ro, T., & Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychological science, 14*(5), 510-515.
Lawless, J. (2003). Event history analysis and longitudinal surveys. *Analysis of Survey data*, 221-243.
Lee, T.-W., Girolami, M., Bell, A. J., & Sejnowski, T. J. (2000). A unifying information-theoretic framework for independent component analysis. *Computers & Mathematics with Applications, 39*(11), 1-21.
Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of computer-mediated communication, 3*(2), JCMC321.
Lunn, J., Sjoblom, A., Ward, J., Soto-Faraco, S., & Forster, S. (2019). Multisensory enhancement of attention depends on whether you are already paying attention. *Cognition, 187*, 38-49. doi:10.1016/j.cognition.2019.02.008
Ma, R., & Kaber, D. B. (2006). Presence, workload and performance effects of synthetic environment design factors. *International Journal of Human-Computer Studies, 64*(6), 541-552. doi:10.1016/j.ijhcs.2005.12.003
Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends in cognitive sciences, 19*(1), 6-12.
Masotti, N., & Persiani, F. (2016). On the history and prospects of three-dimensional human–computer interfaces for the provision of air traffic control services. *CEAS Aeronautical Journal, 7*(2), 149-166.
Mast, F. W., & Oman, C. M. (2004). Top-down processing and visual reorientation illusions in a virtual reality environment. *Swiss journal of psychology, 63*(3), 143-149.
Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are We Ready for Real-world Neuroscience? *J Cogn Neurosci, 31*(3), 327-338. doi:10.1162/jocn_e_01276
Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychon Bull Rev, 18*(5), 904-909. doi:10.3758/s13423-011-0131-8
McCleary, R. A. (1950). The nature of the galvanic skin response. *Psychological Bulletin, 47*(2), 97.
Meehan, M., Insko, B., Whitton, M., & Brooks Jr, F. P. (2002). *Physiological measures of presence in stressful virtual environments.* Paper presented at the Acm transactions on graphics (tog).
Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389-391.
Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). *Augmented reality: A class of displays on the reality-virtuality continuum.* Paper presented at the Telemanipulator and telepresence technologies.

Miller, L. E., Longo, M. R., & Saygin, A. P. (2017). Visual illusion of tool use recalibrates tactile perception. *Cognition, 162*, 32-40. doi:10.1016/j.cognition.2017.01.022

Miller, R. L., Pluta, S. R., Stein, B. E., & Rowland, B. A. (2015). Relative Unisensory Strength and Timing Predict Their Multisensory Product. *Journal of Neuroscience, 35*(13), 5213-5220. doi:10.1523/jneurosci.4771-14.2015

Modica, E., Cartocci, G., Rossi, D., Martinez Levy, A. C., Cherubino, P., Maglione, A. G., . . . Babiloni, F. (2018). Neurophysiological Responses to Different Product Experiences. *Comput Intell Neurosci, 2018*, 9616301. doi:10.1155/2018/9616301

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98-100.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of neurophysiology, 78*(4), 2226-2230.

Nash, E. B., Edwards, G. W., Thompson, J. A., & Barfield, W. (2000). A Review of Presence and Performance in Virtual Environments. *International Journal of Human-Computer Interaction, 12*(1), 1-41. doi:10.1207/s15327590ijhc1201_1

Ngo, M. K., & Spence, C. (2010). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Atten Percept Psychophys, 72*(6), 1654-1665. doi:10.3758/APP.72.6.1654

Noë, A., & Noë, A. (2004). *Action in perception*: MIT press.

Olk, B., Dinu, A., Zielinski, D. J., & Kopper, R. (2018). Measuring visual search and distraction in immersive virtual reality. *R Soc Open Sci, 5*(5), 172331. doi:10.1098/rsos.172331

Palmarini, R., Erkoyuncu, J. A., Roy, R., & Torabmostaedi, H. (2018). A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing, 49*, 215-228.

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng, 32*(3), 230-236.

Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology, 109*(3), 395-417.

Pannunzi, M., Perez-Bellido, A., Pereda-Banos, A., Lopez-Moliner, J., Deco, G., & Soto-Faraco, S. (2015). Deconstructing multisensory enhancement in detection. *J Neurophysiol, 113*(6), 1800-1818. doi:10.1152/jn.00341.2014

Papenfuss, A., & Möhlenbrink, C. (2016). Assessing Operational Validity of Remote Tower Control in High-Fidelity Simulation *Virtual and Remote Control Tower* (pp. 87-113): Springer.

Parasuraman, R. (2001). *Adaptive automation: From theory to practice.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Parsons, T., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain sciences, 7*(4), 42.

Pavone, E. F., Tieri, G., Rizza, G., Tidoni, E., Grisoni, L., & Aglioti, S. M. (2016). Embodying Others in Immersive Virtual Reality: Electro-Cortical Signatures of Monitoring the Errors in the Actions of an Avatar Seen from a First-Person Perspective. *Journal of Neuroscience, 36*(2), 268-279. doi:10.1523/jneurosci.0494-15.2016

Petkova, V., & Ehrsson, H. (2008). If I Were You: Perceptual Illusion of Body Swapping. *PLoS One, 3*(12), e3832.

Petkova, V. I., Khoshnevis, M., & Ehrsson, H. H. (2011). The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. *Frontiers in psychology, 2*, 35.

Pfeifer, R., & Bongard, J. (2006). *How the body shapes the way we think: a new view of intelligence*: MIT press.

Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R., . . . Rugg, M. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology, 37*(2), 127-152.

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, 3-56.
Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological psychology, 41*(2), 103-146.
Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology, 40*(1-2), 187-195.
Puma, S., Matton, N., Paubel, P.-V., Raufaste, É., & El-Yagoubi, R. (2018). Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology, 123*, 111-120.
Pylyshyn, Z. W. (1984). *Computation and cognition*: MIT press Cambridge, MA.
Recarte, M. Á., Pérez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish journal of psychology, 11*(2), 374-385.
Reynal, M., Imbert, J.-P., Aricò, P., Toupillier, J., Borghini, G., & Hurter, C. (2019). Audio Focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility. *International Journal of Human-Computer Studies, 129*, 116-128.
Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., . . . Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior, 10*(1), 45-56.
Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci., 27*, 169-192.
Rizzolatti, G., & Craighero, L. (2007). Language and mirror neurons. *Oxford Handbook of Psycholinguistics. Oxford University Press, Oxford*.
Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive brain research, 3*(2), 131-141.
Rochat, M. J., Caruana, F., Jezzini, A., Intskirveli, I., Grammont, F., Gallese, V., . . . Umiltà, M. A. (2010). Responses of mirror neurons in area F5 to hand and tool grasping observation. *Experimental brain research, 204*(4), 605-616.
Sakata, H., Taira, M., Murata, A., & Mine, S. (1995). Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cerebral Cortex, 5*(5), 429-438.
Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience, 6*(4), 332.
Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1311.
Schaal, S. (1999). Is imitation learning the route to humanoid robots? *Trends in cognitive sciences, 3*(6), 233-242.
Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments, 1*(1), 120-126.
Siegel, S., & Castellan, N. J. (1956). *Nonparametric statistics for the behavioral sciences* (Vol. 7): McGraw-hill New York.
Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2015). afex: Analysis of factorial experiments. *R package version 0.13–145*.
Skarbez, R., Brooks, J. F. P., & Whitton, M. C. (2017). A Survey of Presence and Related Concepts. *ACM Computing Surveys, 50*(6), 1-39. doi:10.1145/3134301
Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence, 8*(5), 560-565.
Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *Br J Psychol, 109*(3), 431-433. doi:10.1111/bjop.12305
Slater, M., Brogni, A., & Steed, A. (2003). *Physiological responses to breaks in presence: A pilot study.* Paper presented at the Presence 2003: The 6th Annual International Workshop on Presence.
Slater, M., Linakis, V., Usoh, M., & Kooper, R. (1996). *Immersion, presence and performance in virtual environments: An experiment with tri-dimensional chess.* Paper presented at the Proceedings of the ACM symposium on virtual reality software and technology.

Slater, M., Pérez Marcos, D., Ehrsson, H., & Sanchez-Vives, M. V. (2009). Inducing illusory ownership of a virtual body. *Frontiers in Neuroscience, 3*, 29.

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS One, 5*(5), e10564.

Soranzo, A., Lugrin, J.-L., & Wilson, C. J. (2013). The effects of belongingness on the simultaneous lightness contrast: a virtual reality study. *Vision research, 86*, 97-106.

Spence, C., Pavani, F., & Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 26*(4), 1298-1319. doi:10.1037/0096-1523.26.4.1298

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: a review. *Hear Res, 258*(1-2), 134-142. doi:10.1016/j.heares.2009.04.015

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis. *Journal of Cognitive Neuroscience, 8*(6), 497-506. doi:10.1162/jocn.1996.8.6.497

Stein, B. E., & Meredith, M. A. (1993). The merging of the senses. Cognitive Neuroscience series. *MIT Press, Cambridge Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. J Cogn Neurosci, 1*(1), 12-24.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience, 15*(8), 520-535. doi:10.1038/nrn3742

Stein, W. (2012). *On the Problem of Empathy: The Collected Works of Edith Stein Sister Teresa Bendicta of the Cross Discalced Carmelite Volume Three*: Springer Science & Business Media.

Stevens, J. A., & Kincaid, J. P. (2015). The Relationship between Presence and Performance in Virtual Simulation Training. *Open Journal of Modelling and Simulation, 03*(02), 41-48. doi:10.4236/ojmsi.2015.32005

Stevenson, R. A., Geoghegan, M. L., & James, T. W. (2007). Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Exp Brain Res, 179*(1), 85-95. doi:10.1007/s00221-006-0770-6

Suma, E. A., Clark, S., Finkelstein, S. L., & Wartell, Z. (2010). *Exploiting change blindness to expand walkable space in a virtual environment.* Paper presented at the 2010 IEEE Virtual Reality Conference (VR).

Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: a meta-analytic synthesis. *Psychological Bulletin, 137*(4), 682.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb Cortex, 17*(3), 679-690. doi:10.1093/cercor/bhk016

Thomas, D. J. (2016). Augmented reality in surgery: the computer-aided medicine revolution. *Int J Surg, 36*(Pt A), 25.

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum Percept Perform, 34*(5), 1053-1065. doi:10.1037/0096-1523.34.5.1053

Van Schaik, F., Roessingh, J., Lindqvist, G., & Fält, K. (2010). Assessment of visual cues by tower controllers, with implications for a Remote Tower Control Centre. *IFAC Proceedings Volumes, 43*(13), 123-128.

Vecchiato, G., Maglione, A. G., Cherubino, P., Wasikowska, B., Wawrzyniak, A., Latuszynska, A., . . . Leucci, M. R. (2014). Neurophysiological tools to investigate consumer's gender differences during the observation of TV commercials. *Computational and mathematical methods in medicine, 2014*.

Vitense, H. S., Jacko, J. A., & Emery, V. K. (2003). Multimodal feedback: an assessment of performance and mental workload. *Ergonomics, 46*(1-3), 68-87. doi:10.1080/00140130303534

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625-636.
Witmer, B. G., Jerome, C. J., & Singer, M. J. (2005). The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments, 14*(3), 298-312.
Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple tasks environments. *International journal of cognitive ergonomics, 4*(3), 213-242.
Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology, 18*(5), 459-482.
Zhang, J. H., Chung, T. D., & Oldenburg, K. R. (1999). A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen, 4*(2), 67-73. doi:10.1177/108705719900400206

7. APPENDIX- Overview of publication status of chapters in the thesis

Chapter number and title	**Status**
1.Introduction- Grounded Cognition and VR	Submitted
2.Study 1	In preparation
3.Study2	Published
4.Study3	Published