Corso di Dottorato di Ricerca in
Morfologia e Funzione Normale e Patologica
di Cellule e Tessuti
Curriculum in Neuroscienze e Ambienti Virtuali

Exploring Cognitive Processes with Virtual Environments

Tesi di Dottorato di Ricerca
ING-INF/05

Candidato:
Ing. Federico Vanni

Relatori:
Prof. Massimo Bergamasco
Prof. Francesco Fornai

Università di Pisa

Dipartimento di Morfologia Umana e Biologia Applicata
Ciclo 2009-2011

Presidente del Programma Dottorale
Prof. Amelio Dolfi

Candidato
Ing. Federico Vanni

Relatori
Prof. Massimo Bergamasco
Prof. Francesco Fornai

Data dell'esame finale
03.12.2012

Contents

List of Figures

List of Tables

Summary

The scope of this thesis is the study of cognitive processes with, and within Virtual Environments (VEs). Specifically, the presented work has two main objectives: (1) to outline a framework for situating the applications of VEs to cognitive sciences, especially those interfacing with the medical domain; and (2) to empirically illustrate the potential of VEs for studying specific aspects of cognitive processes.

As for the first objective, the sought framework has been built by proposing classifications and discussing several examples of VEs used for assessing and treating disorders of attention, memory, executive functions, visual-spatial skills, and language. Virtual Reality Exposure Therapy was briefly discussed as well, and applications to autism spectrum disorders, schizophrenia, and pain control were touched on. These applications in fact underscore prerogatives that may extend to non-medical applications to cognitive sciences.

The second objective was sought by studying the time course of attention. Two experiments were undertaken, both relying on dual-target paradigms that cause an attentional blink (AB). The first experiment evaluated the effect of a 7-week Tibetan Yoga training on the performance of habitual meditators in an AB paradigm using letters as distractors, and single-digit numbers as targets. The results confirm the evidence that meditation improves the allocation of attentional resources, and extend this conclusion to Yoga, which incorporates also physical exercise. The second experiment compared the AB performance of adult participants using rapid serial presentations of road signs — hence less abstract stimuli — under three display conditions: as 2-D images on a computer screen, either with or without a concurrent auditory distraction, and appearing in a 3-D immersive virtual environment depicting a motorway junction. The results found a generally weak AB magnitude, which is maximal in the Virtual Environment, and minimal in the condition with the concurrent auditory distraction. However, no lag-1 sparing effect was observed.

Riassunto Analitico

La tesi attiene allo studio dei processi cognitivi in ambiente virtuale (AV). In particolare il lavoro presentato ha due obiettivi principali: (1) proporre un quadro di riferimento per le applicazioni degli AV nelle scienze cognitive, specialmente quando queste rientrano nel settore medico; e (2) illustrare empiricamente il potenziale degli AV per lo studio di specifici aspetti cognitivi.

Il quadro di riferimento del primo obiettivo è stato costruito discutendo classificazioni ed esempi di AV usati per valutare e trattare disturbi dell'attenzione, della memoria, esecutivi, delle abilità visuo-spaziali e del linguaggio. Sono stati brevemente discussi anche esempi di applicazioni di *Virtual Reality Exposure Therapy*, e per i disturbi dello spettro autistico, la schizofrenia, e l'analgesia. Esse sono infatti rappresentative delle prerogative degli AV trasferibili ad aspetti non medici delle scienze cognitive.

Per il secondo obiettivo del lavoro si è studiata l'attenzione nel dominio temporale. Sono stati realizzati due esperimenti, entrambi basati su un paradigma sperimentale di doppio-compito tale da indurre il fenomeno dell'*attentional blink* (AB). Il primo esperimento ha valutato l'effetto di 7 settimane di Yoga tibetano sull'AB di un gruppo di meditatori. La presentazione visiva seriale rapida comprendeva lettere come distrattori e numeri a una cifra come target. I risultati confermano che la meditazione riduce l'AB, ed estendono questa conclusione allo Yoga, che include anche l'esercizio fisico. Il secondo esperimento ha confrontato l'AB di un gruppo di adulti utilizzando presentazioni visive seriali rapide di segnali stradali — dunque stimoli meno astratti — in 3 condizioni: immagini 2-D sullo schermo di un computer, essendo simultaneamente presente o assente una distrazione uditiva, e presentazione 3-D in un AV immersivo che simula un incrocio autostradale. I risultati rilevano un AB lieve, che è massimo nell'AV, e minimo nella condizione 2-D con la distrazione uditiva. L'effetto *lag-1 sparing* non è presente.

Introduction

Virtual Environments are interactive, three-dimensional (3-D), computer-generated environments which may simulate an existing physical reality, or provide an alternative, completely new experience. What a participant experiences interacting with a Virtual Environment (VE) is named Virtual Reality (VR), and its applications span across many areas. Therefore, very often the design, development, and manipulation of a VE during its operating cycle pose multidisciplinary issues that require and bring together multidisciplinary teams. The nature of VR is twofold: that of an advanced technological human-computer interface, and that of a special medium, which should sustain an illusion of non-mediation in an individual who interacts with a multisensory substitute of physical reality. The possibilities originated from the potential of generating a compelling surrogate of physical reality are diverse, and an ever increasing number of them is turning actual. This is made possible by many concurring factors, the major being the advances in miniaturisation of the devices composing a VR system, their augmented computing power, the reduction in costs of those devices, and, importantly, the spreading of VR-related technologies in the society, mainly for entertainment purposes (e.g., 3-D cinema, and video-game consoles). Clearly, the first beneficiaries of the possibilities offered by advancements in VR technologies have been basic and applied researchers, who have been devising several ways of capitalising on VR. In this line VEs have been studied and tested for being used in medicine, as a pedagogical tool, but also as an aid for diagnostic and therapeutic aims. Thence, VEs were introduced also as sophisticated stimulation paradigms to study human performance, and human cognitive mechanisms, such as spatial cognition. The present dissertation stems from a research work aimed at investigating the use of virtual environments in cognitive sciences, but at the interface with medical conditions affecting the cognitive, and more generally, the mental domain.

Admittedly, the context of cognitive sciences is very broad, encompassing philosophy, psychology, neuroscience, language, and sprouting interdisciplinary connections with many diverse disciplines and application fields. The subset on which this dissertation focalises is the study of cognitive processes, and, more broadly, mental processes, in their relations with VEs. However, also the relations between VEs and cognitive sciences may be considered in several ways, and at different levels. Indeed, VEs involve technologies, thus one way of approaching the relations between cognitive sciences and VEs would have been to investigate how the technological factors and specifications of a VE impact on certain aspects of human cognition. Another level, different from the technological aspects, is the virtual experience enacted in a VE. At this level, one topic to study and measure would have been the psychological correlates of that experience, typically through studies of the sense of presence, and of how personality traits affect the experience in the VE as a function of different technologies. The choice of this research work is different, namely it is to start from a more applicative linkage between VEs and mental processes: the applications of VEs to medicine for neuropsychological assessment, cognitive rehabilitation, psychotherapy, and psychiatry. Two reasons determined this choice: first, the search of a framework into which VEs applications to cognitive sciences may be situated; second, the goal to produce an instantiation of this framework that demonstrates the potential of VEs for relevant medical applications. The proposition of this framework, instantiated for the applications to neuropsychological assessment, rehabilitation, and mental health is complemented by empirical case studies, which centre on the exploration of human attention with VEs, but not in explicit relation to medical conditions.

The cognitive process taken as the target of the empirical case studies is attention — albeit attention is not a unitary process —, specifically the study of the time course of attention with experiments based on the attentional blink (AB) paradigm. There are a number of reasons for this choice. The foremost reasons are the centrality of attention to control human information processing, and the fact that attention interacts with multiple perceptual and cognitive operations. This makes attention essential in human performance, because it controls the information flow so that perceptions and actions of an individual are kept coherent to one another, and in view of the individual's intentions and goals. Another relevant reason for empirically studying attention in VEs turns out from the review of VEs applications to cognitive assessment and rehabilitation, and to basic neuroscience in general. In fact, not many applications of VEs targeted attention, and none of them considered the time course of selective attention in the information processing flow, as indexed by the AB. The AB is an attentional deficit arising when two targets, T1 and T2, appear in a rapid stream of distractors, and the stimulus onset asynchrony between the two targets is between 100 ms and 400 ms.

As a result, this dissertation explores cognitive processes in VEs, as stated in its title, pursuing two series of research questions. Firstly, how VEs are applied for the assessment and therapy of mental disorders following neurological damage and psychological conditions, what are the grounds for this kind of applications, and if a framework built for categorising those applications is suitable to cover the wider range of VEs applications to cognitive sciences. Secondly, the research questions taken on about attention are split in two experimental studies. The first, Study 1, does not involve VEs, and investigates whether the the distribution of attentional resources as indexed by the AB performance can be changed following a 7-week Tibetan Yoga training undertook by habitual meditators. This kind of training combines a mild physical exercise associated with visual imagery, and a mindfulness-like meditation practice. The rationale for formulating this research question is supported by several findings. Large individual differences have been found in the general population in terms of shown AB magnitude (Martens et al., 2006), and the same magnitude has been alleviated by adding task-irrelevant distraction to the AB task (Arend et al., 2006; Olivers and Nieuwenhuis, 2005; Taatgen et al., 2009; Lapointe-Goupil et al., 2011), and following a training in action video games (Green et al., 2003). Furthermore, meditation as a pure mental training has been shown to produce a change in the distribution of limited brain resources as pointed by the AB performance (Slagter et al., 2007; van Leeuwen et al., 2009). Study 2 on its side, introduces a VE in the experiment, and attempts to answer two questions: whether rapid visual stimuli which are not as simple as letters or numbers, and which may be meaningfully contextualised in a VE, elicit an AB with similar characteristics as that arisen with abstract and simple visual stimuli; and whether there is an effect of the VE on the AB magnitude, similar to that of an auditory distraction. Road signs are used as visual stimuli for assessing the AB magnitude, and they have been presented on a 3-D virtual motorway junction to have them in a meaningful context. The rationale for this second experiment is partially in common with the first study, also following the observation that the kind of attention quality involved in mindfulness meditation implies a diffuse state of attention, with a less narrow attentional focus on the task at hand (Bishop et al., 2004), which is similar to the attentional state referred to for the explanation of the effect of a concurrent distraction alleviating the AB magnitude according to Olivers and Nieuwenhuis (2006). Concerning the background findings about AB evaluated with less abstract symbols as stimuli, words and natural pictures have been used in the past, including faces. Opposite effects of target categories, whether the same or different in a single rapid serial visual presentation, on the AB performance, have been found by Evans and Treisman (2005) and Einhäuser et al. (2007). Effects of emotional processing through the use of faces have been found as well on the AB by Stein et al. (2010), who demonstrated a trade off between perceptual load of T1 and fearfulness of T2 in determining the AB magnitude, and by de Jong et al. (2010), who noticed how an angry face used as T1 hampers the processing of a letter stimulus used as T2. This is to say that, although a robust phenomenon, the AB magnitude and duration may change depending on both individual differences, and featural or semantic characteristics of the stimuli used to assay it.

In view of the objectives and background laid out, the present dissertation is organised in 6 chapters. Chapter 1 illustrates the conceptual and technological fundamentals of VR in order to provide the basic background information underpinning VEs construction and characteristics, and motivating their applications. Some illustrative definitions of VR and VEs are considered to understand the technological components that create VEs, and allow to deliver them for engendering a compelling virtual experience. The functional components of a VR system are concisely described too, and connected to the important properties of level of immersion and interactive capabilities characterising VEs. The chapter terminates with a section introducing virtual presence as a major indicator of the human response to a VE. Its determinants, a framework to relate it to immersion and interaction of the VE

(Slater, 2009), and possible general methods to measure it are discussed in the end. This arrangement of Chapter 1 is intended to provide an overview of all the key elements playing a role in the human experience of a VE. This overview encompasses both the technology-oriented, and the experience-oriented approaches to VEs, as they are complementary for the applications of VR. Chapter 2 tackles the series of research questions formulated for defining a framework to situate applications of VEs to neuropsychological assessment, cognitive rehabilitation, psychotherapy, and psychiatry. Such questions are answered by discussing different criteria adopted to categorise applications of VEs to medicine, and in diverse fields. Consequent categorisations are separately proposed for applications to neuropsychological assessment and cognitive rehabilitation on one side, and to psychotherapy and psychiatry on the other. Prominent examples of VEs applications found from a scientific literature search are described to instance the proposed categorisations branches. In a unifying attempt, a framework for all the VEs applications to cognitive sciences is proposed and discussed by combining the two separate categorisations, and drawing on other non-medical applications of VEs. Chapter 2 terminates with the motivation for the choice of attention, and specifically of its temporal aspects, as the target of the empirical studies undertaken, and illustrated in Chapter 4 and Chapter 5. Thereby, Chapter 3 introduces the fundamental concepts about attention, its role in the human information processing, and taxonomies of attention mainly drawing on that proposed by Chun et al. (2010). Then, Chapter 3 concentrates on temporal attention, the attentional blink phenomenon, and the empirical evidence cumulated in the last two decades about it. Chapter 4 is devoted to the empirical Study 1. It illustrates in detail the rationale for investigating the effect on the AB of a training based on Tibetan Yoga. This illustration is based on a review of studies demonstrating training-related alterations of the AB in video game players, and in meditators. The remainder of Chapter 4 presents the experiment undertaken in terms of material and method, results, and conclusive discussion. Chapter 5 has a parallel structure for empirical Study 2. Thus, it discusses in detail its rationale, mainly by reviewing existing evidence about AB with natural pictures, faces, and words, and about concurrency benefits in AB performance generated by task-irrelevant distractions. Subsequently, material and method, results, and discussion for Study 2 are presented. Chapter 6 is for the dissertation conclusions. Here the conclusions are summarised and restated based on the previous chapters, and insights for future work on the lines considered in the present dissertation are suggested. There is also a section for Appendixes, which contain background material used for the experiments, such as a description of the Tibetan Yoga exercises used for the training in Study 1, and the questionnaire and experiment instructions employed in Study 2.

1. Virtual Environments Fundamentals

In this chapter the conceptual and technological fundamentals of Virtual Reality are illustrated in order to provide the basic background information underpinning Virtual Environments construction and characteristics, and motivating their applications.

Some illustrative definitions and the key aspects of Virtual Reality and Virtual Environments are considered to understand the technological components that create Virtual Environments, and allow to deliver them for engendering a compelling virtual experience. Then, the concept of immersive Virtual Environments is introduced, underscoring the important role of interactive capabilities. Finally, conceptualisations of the sense of presence, and its role in Virtual Environments are presented in order to complete the account about the virtual experience arising from the participation of an individual in a Virtual Environment.

1.1. Virtual Reality Definitions

A comprehensive definition of Virtual Reality (VR) is very difficult, and risks to offer too a simplistic view of the topic. Indeed, although some ideas and rudimentary components underpinning VR had been introduced since the late 1950s (see Alonso et al., 2008 for a brief review), the term VR is quite recent (Kelly et al., 1989) and several definitions have been proposed and discussed, especially in the 1990s, at the beginning of VR history as we conceive it nowadays.

Generally speaking, VR deals with building an interactive three-dimensional (3-D) computer-generated environment — the *Virtual Environment (VE)* — that might either simulate an existing physical reality, or provide an alternative, completely new experience. The experience of this VR is enacted by a human participant through a set of display and interface devices (Pan et al., 2006), that enable the participant to *feel* the VE and actively interact with and move within it. Thus, VR has two inherent natures: that of a collection of technologies enabling an advanced human-machine interaction; and that of a medium able to provide a novel kind of experience to the human body and mind. The conception of VR as a medium capable to immerse the human user in a VE, and engender a virtual experience of it *as if real*, is the fundamental idea guiding VR research and applications in relation to cognitive science. However, in order to get a complete understanding of VR and its applications, it is advantageous, and necessary to explore to some extent both the technological and the experiential nature of VR.

The term *Virtual Reality* was first coined by Jaron Lanier, the computer scientist and inventor CEO of VPL Research Inc., in the second half of the 1980s (Kelly et al., 1989). VPL Research was a company settled in California, and one of the first developer and marketer of VR technologies. It had several projects that used computer graphics, goggles for 3-D viewing of digital spaces and datagloves to interact with digital objects (see for instance Zimmerman et al., 1987; Stone, 1992): *Virtual Reality* was just the term chosen by Jaron Lanier as a common heading for such projects.

Then, the debate started about defining VR, and various definitions were proposed. The definitions essentially differ in the degree to which they emphasise either the technology or the human experience inherent in VR (Steuer et al., 1995), and the focus depends on the context where they are built, whether of engineering, psychology or communication science.

Some emblematic definitions that seem to clearly outline these two different dimensions of VR are the following:

D1. "Virtual reality (VR) is the use of computer graphics systems in combination with various display and interface devices to provide the effect of immersion in the interactive 3-D computer-generated environment. We call such an environment a virtual environment (VE)." (Pan et al., 2006)

D2. "[Virtual reality] is a technology that uses computerized clothing to synthesize reality." (Franchi, 1994)

D3. "Virtual Reality is electronic simulations of environments experienced via head mounted eye goggles and wired clothing enabling the end user to interact in realistic three-dimensional situations." (Coates, 1992)

D4. "[Virtual reality can be defined as] a class of computer-controlled multisensory communication technologies." (Biocca, 1992)

D5. "Virtual Reality uses computers to create 3D environments in which one can navigate and interact. [...] The main goal of VR is to create in the user the illusion of being in an environment that can be perceived as a believable place with enough interactivity to perform specific tasks in an efficient and comfortable way." (Alonso et al., 2008)

D6. "Virtual Reality is an alternate world filled with computer-generated images that respond to human movements. These simulated environments are usually visited with the aid of an expensive data suit which features stereophonic video goggles and fiber-optic gloves." (Greenbaum, 1992)

D7. "VR is a medium for the extension of body and mind." (Biocca and Delaney, 1995)

D8. “A *virtual reality* is defined as a real or simulated environment in which a perceiver experiences telepresence.” (Steuer et al., 1995)

Several basic aspects of VR and VEs emerge from these definitions:

1. VR and VEs represent a synthetic digital environment generated and controlled by a *computer* or a network of computers.
2. The environment is intended to be experienced by a *user* through special devices that serve to display and interact with the VE. These devices are often in physical contact with the user, who *wears* them.
3. The interaction with the environment may be *immersive* for the user, who should *perceive* the VE as believable, experiencing the sense of *presence* in it.
4. The VE is a 3-D *space*, perceived by the user as a *believable place.*
5. The interaction with the VE may be a *multisensory communication.*
6. The VE may either *simulate* an existing reality or provide the believable *illusion* of an *alternate world.*
7. The experience in the VE should allow the user to *perform tasks,* and the VE should *respond* to the user’s actions.
8. VR is a *medium* that affects *body* and *mind.*

Starting from these items, and looking at the comparatively short history of VR, it is clear that VEs lie between two *poles*: one is the core of the view of VR as a *machine*, i.e., a technology based on a collection of hardware and software components; and the other pole is the nucleus of the view of VR as a *medium*, capable of engendering in its users (multi-)sensory *experiences* of a substitute of the physical environment. These two poles, which form a technology-oriented view, and an experience-oriented view of VR, are not incompatible; rather, they represent two complementary constituents of VR, as elementarily the *virtual experience* is made possible by the *machine*. In contrast, the choice of one of the views against the other would expose to the risk of missing the factors related to the other view, which are instead important to understand and scientifically *measure* VR.

For this reason, in the following we will take both the experience-oriented and technology-oriented points of view not as one opposed to the other, but as complementary, in order to catch all the factors and dimensions useful to understand and scientifically approach VR. This is the approach we claim adequate for getting a single overall picture of VR, into which it is possible to situate also Virtual Environments applications to cognitive sciences.

1.2. Virtual Reality Components

Though taking into account its intrinsic multidimensional quality, in order to get a single view of VR and VEs a first step is to define what a VR system consists of.

The VR system represents in fact the technological essence of VR. It includes hardware and software components, and its goals are to generate the VE, to deliver it to the user, and to manage and maintain it according to the interactions occurring between the user and the VE. A schematic general view of the components of a VR system is depicted in Figure 1.1.

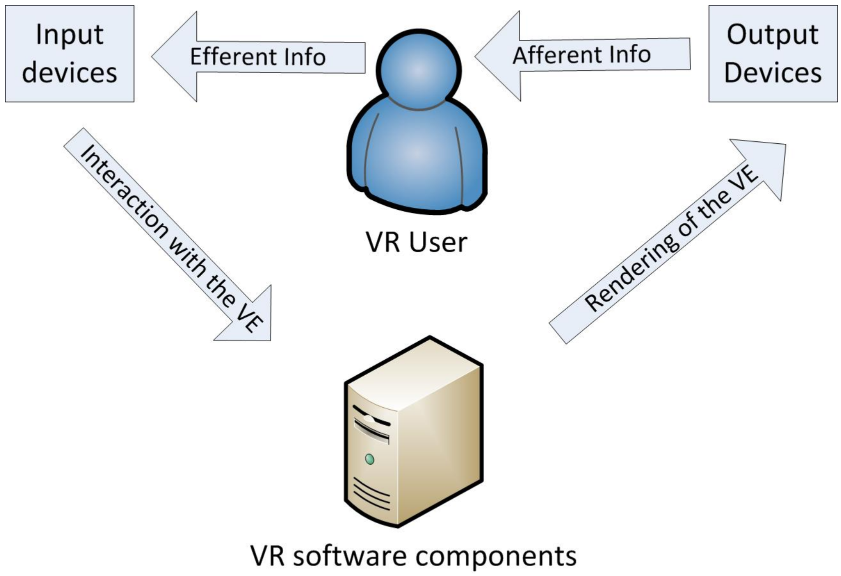


Figure 1.1.: Technology-oriented schematic representation of the components of a VR system: a software application manages a collection of input and output hardware devices which render the VE to the user, and enable her/his interaction with the VE. The management includes the continuous update of the VE properties and behaviours according to either the user's actions or a pre-programmed sequence.

1.2.1. Output and Input Devices

As referred to in definitions D5 and D8, the VE is *perceived* by the user: specifically, the simulated environment is experienced through the senses of the user, which thus require a stimulation from the VR system. Such a stimulation is provided by *output devices* that convey synthetic sensory data to the user's sensory systems. The output devices may in principle stimulate any of the human senses, but the most typically used are visual displays, aural displays, and, to a lesser but increasing extent, haptic displays. When the VR system generates data streams for different sensory modalities, e.g., a visual stream and a haptic stream, the VE is called *multimodal*, and the multisensory characteristic of VR technology is exploited as indicated for example in definition D4. The output devices bring to the user the afferent information coming from the VE. The most typical output devices are those for vision. As repeatedly mentioned in the listed definitions of VR, the VE is three-dimensional, so it often requires a stereoscopic display of the virtual scene. This is accomplished by delivering to the left and right eye of the participant a different view of the same scene, quickly refreshed by the computer according to the participant's head motion to simulate binocular disparity, and obtain depth perception. This may be achieved in several ways. One way is to display on special rear-projected screens two different views of the same scene, resembling the difference in perspective arising from the two eyes, and using special goggles that shut out the left and right lens in synchrony with the projection of the right and left views respectively. Another method is to provide the participant with an helmet (head-mounted display, HMD) with two built-in monitors, one in front of each eye, thus conveying the proper view to the corresponding eye. The first method is typical of CAVE-like systems. CAVE is an acronym standing for CAVE Automatic Virtual Environment (Cruz-Neira et al., 1993), which designates a room whose walls, and possibly the floor too, are rear-projected screens. Thus, in a CAVE, the participant in the virtual experience moves inside this room, and is surrounded by the stereoscopic graphic display of the VE projected onto the walls. The participant wears special eye-goggles, and her/his head motion is tracked. Other body parts may be tracked as well depending on the interaction expected with the VE.

There are also other means to achieve a 3-D visual display, but those just mentioned are particularly relevant for the work presented in this dissertation. Indeed, the "mini-CAVE" used for visual display in the empirical study presented in chapter 5 is a CAVE-like system. Its peculiarity lies in the fact that it covers part of the peri-personal space of the participant, but not the whole space around her/him (see chapter 5). On the other hand, the HMD importance is due to the wide use of this means to display VEs in most of the applications for cognitive sciences, and for mental health in general. As for the haptic display, devices acting as both sensors and actuators are used (Riva, 2006): the haptic feedback may be confined in the end-effector of an instrument handled by the user, or can use an exoskeleton frame to impart forces on the user according to the interaction with the VE (Bergamasco et al., 2007).

A further important characteristic of VEs is the possibility for the user to interact with the VE, which appropriately responds to the user's action (see definitions D3, D5, and D6). For this purpose additional interface devices are necessary, the *input devices*. Such devices have two main functions: first, they allow the user to effect changes in the VE; second, they track the position and orientation of the user's head, and other body parts if relevant, to update in real-time the displays of the VE, so as to avoid discrepancies in the sensorimotor loop of the user (Slater et al., 2009). Therefore, typical input devices are position trackers. Trackers include sensors and processing units, and are intended to track in real-time the spatial position and orientation of a body part to allow for an effective interaction with the VE and its objects. Thus, head is tracked for visual displays, the hand and arm are tracked for haptic interaction, and the whole body may be tracked for simulations involving locomotion. Motion capture, i.e., the real-time capture of positions and angles at body joints, may be used along with tracking devices, for example for animating a virtual character in the VE according to the user's actions. Various technologies exist for implementing tracking devices, and devices for motion capture, but their description is beyond the scope of this thesis.

Displays and tracking devices are often attached to, or worn by the user (e.g., eyegoggles for stereoscopic viewing or HMD, dataglove for tracking fingers flexion, tracking sensors fixed to the user's head, force-feedback exoskeletons for haptic interaction) as reported in definitions D2, D3, and D6. These "computerized clothing", "wired clothing" and "data suit" have been the stereotypical representation of VR for long in the first decade of its development. The equipment for an immersive VR experience was indeed bulky and expensive, and seemed to be designated to remain a niche application, confined in research laboratories. This situation substantially changed in the last decade, thanks to the tremendous progress in electronics in terms of miniaturization and availability of augmented computational power at low cost. The result has been to open up new ways for VEs applications spanning from entertainment to industry and healthcare, enabling as well the applications to cognitive sciences that will be discussed in chapter 2.

1.2.2. Software Components

Output and input devices allow the user to perceive the VE as they support the exchange of data streams between the VE and the user. But what about such data? And what about the *very* VE? The two questions are indeed the same. As a matter of fact, the VE does not physically exists, it is a digital environment, and as such it consists of digital data that represent models of what is intended to be delivered to the user, carrying specific properties and behaviours. A clear explanation of this aspect is provided by Sanchez-Vives and Slater (2005): the virtual scene the user should perceive is described in a computer database, that includes all its properties and behaviours (i.e., the model). Geometric, acoustic, radiant and physical characteristics of the scene are stored in the database, including the virtual objects present in the scene. As mentioned in almost all the VR definitions we listed, a computer generates and controls the VE. Particularly, the computer constructs and controls this database, and the output and input devices. The tracking devices send information in real-time to the computer, that based on such information *renders* the VE through the display devices, by using appropriate information in the database. In this way at any moment in time the display of the VE matches the user's position, orientation and action. For example, if the user turns the head, the head tracking device sends the new orientation to the computer, and this uses the information in the database to update, according to the new viewpoints, the scene images that are administered to the left and right eye of the user by a stereoscopic visual display. Also, if the user's body is represented in the VE, and the user's arm is tracked, when the user moves the tracked limb its virtual counterpart needs move accordingly, and this is enabled by the software running in the computer that renders the database based on the user's efferent information coming from the tracking devices in real time. Input and output devices and the whole VE rendering algorithms, as well as those governing the interaction with the user are controlled by modular computer programming scripts that compose the *VR software components* indicated in Figure 1.1. There are pieces of software for managing input and output devices, and there is the very *VR (software) application* that generates and sustains the VE.

1.3. Immersion and Interaction in Virtual Environments

In short, the VR system integrates the technological components of a VE: indeed, it includes a computer that stores the model of the VE, and controls interface and display devices to render it and allow the interaction with the user. Clearly, the user introduces a subjective component in the loop, as the response she/he provides in the interaction with the VE and to the stimulation of the VR system depends on subjective factors of the user. Here, we temporarily put them aside, and we consider instead the physical properties of the VR system. Such properties are objective in the sense that they can be objectively measured as they are a direct consequence of the characteristics of the adopted technologies. They play an important role because they determine the *level of immersion* of the VE, that is one of the main criteria to classify VEs.

Generally, an *immersive* VE is "one in which the user is perceptually surrounded by the VE" (Loomis et al., 1999), that is a system which stimulates multiple senses of the user in order to convince her/him of the simulated environment.

If we come to more operational terms, there are two ways to consider the level of immersion associated to a VE, and both are determined by the physics of the system, i.e., depend on its objective properties (Slater, 2009). The first approach defines the level of immersion in terms of displays and tracking information available (Slater and Wilbur, 1997): the more sensory modalities are supplied by the VE displays and the more the tracking enables a consistent rendering compared to the real world, the more the delivered VE is immersive. The alternative approach is to shift the focus on the interaction of the user with the VE. In this respect, Slater (2009) refers to the concept of *sensorimotor contingency* (SC) supported by an immersive VR system. SCs are part of a sensorimotor theory of perceptual experience that is posited in terms of ways of exploring the environment (O'Regan and Noë, 2001). In this framework, to become aware of an object, e.g. with vision, an active exploration of it is necessary (Maye and Engel, 2011). The object characterization is indeed based on the relationships between the actions of the perceiver and the resulting sensory stimulation. The sensorimotor contingencies are the structure of the rules that govern the sensory changes effected by different motor actions (O'Regan et al., 2001). For example, shifting the gaze and moving the head to see underneath something, or rotating the head to better hear a sound are actions associated to visual and auditory SCs respectively. Consequently, different SCs are associated to specific actions and to specific sensory modalities.

In the context of immersive VEs, depending on the VR system properties, there will be specific SCs supported by the system, and in turn these SCs will define a set of *valid actions* that a user can carry out in the VE to get a meaningful perception and interaction. Such valid actions might be categorised as *sensorimotor actions* when they are actions that the user can perform in the VE to meaningfully change perception, and as *effectual actions* when they refer to user's actions that change the VE (Slater, 2009).

In view of this theory, it is possible to characterise the level of immersion of a VE by means of the valid actions that are possible within that VE and to compare the "immersiveness" of two VR systems in terms of the set of valid actions they support. A system supporting only a subset of valid actions with respect to another is less immersive than the latter. Clearly, this approach does not change the fact that immersion depends on the physical, objective properties of the VR system generating the VE. The difference with respect to the first approach, which in practice is based on the number of sensory channels created and on the values of technical parameters measuring the fidelity of tracking and display devices, lies in the integration of the system interactive capabilities as determinants of immersion, in terms of the valid actions the user may take in the VR system to meaningfully perceive or effect changes in the VE. This establishes a link between the physical properties of the VR system, which determine the set of valid actions a user may take in a VE, and the plausibility of those actions, and of the system responses too. This plausibility may be affected by the subjectivity of the user.

Recapitulating, two key factors affecting the experience of the user in a VE are the levels of immersion and interaction that the VE affords. Nonetheless, these two factors do not depend on the user. Rather, they are the result of the physical properties of the VR system, and can be associated to a set of valid sensorimotor and effectual actions that can be taken in the VE. The response of the user to the VE is partly a consequence of these properties, but for the remaining part depends on subjective factors.

In the next section we examine a question that is a core in the experiential view of VR: the consequences on the user of immersion in and interaction with the VE or, differently posited, the contribute of the user's response, that is subjective, to the VE characterization.

1.4. Presence in Virtual Environments

Steuer defines VR as an environment where the perceiver experiences *telepresence* (Steuer et al., 1995, definition D8 in section 1.1). Telepresence (Minsky, 1980) refers to the sense a human operator feels of being *present* in a remote physical environment. Originally this definition had been coined for teleoperators: a teleoperator is a machine that allows a human operator to "move about, sense and mechanically manipulate objects at a distance" (Sheridan, 1995). A robotic system can be teleoperated by a human operator. The robot is located in a remote physical environment, and the operator, by distally manipulating the robot, might develop a sense of being in the remote environment where her/his actions effect changes through the robot, although s-/he is not physically there. This concept has been later extended to virtual environments, initially by designating it as *virtual presence* (Sheridan, 1992; Slater and Steed, 2000; IJsselsteijn and Reiner, 2004), and later simply as *presence* (Slater et al., 1994; Loomis, 1992).

The general idea underpinning presence in VEs is that the user *feels* to exist in the computer-generated environment rather than in the physical place where she/he is physically situated (Witmer and Singer, 1998). The issues of debate are about a more precise definition of presence, also outside VEs, about which are its determinants, and which is its place in the frameworks of psychology and neuroscience. Reviewing these arguments in detail is beyond the scope of this thesis, so in the following only some aspects of presence will be described, since presence is the main component of human response to the virtual experience.

As mentioned in some of the definitions of VR we reported previously in section 1.1, the virtual experience provides an illusion, that should result in a convincing substitute of physical reality. The sense of presence experienced in the VE is an indicator that such an illusion succeeded, and thereby the target for VR applications to be effective. The advantage of using presence for measuring whether a VE is functional is that it is not application-specific, rather it is a general concept that can arise in a human individual when participating in an experience. The critical point is how to precisely define presence, and consequently find objective and reliable measurements of it.

1.4.1. Presence Conceptualisations

There are two main research approaches to presence in general terms (Riva et al., 2011): presence as a function of the experience of a specific medium; and presence as a psychological phenomenon that is independent of the medium and acts to control the activity of the person in the environment.

The former approach is more related to the technological aspects of VR as a medium, and presence is often defined as the "perceptual illusion of non-mediation" (Lombard and Ditton, 1997). This definition of presence actually applies also to non-technological media like books or movies. In this context, presence is an all-or-none phenomenon occurring when the medium disappears from the conscious attention of its user (Coelho et al., 2006); and the perceived level of presence depends on how many instants over time have seen the occurrence of the illusion of non-mediation (Lombard and Ditton, 1997).

In this view presence in VEs is achieved when the technology components of the VR system are transparent to the user, and the sense of presence is the result of immersion and interaction levels offered by the VR system. Therefore presence is a subjective experience, resulting from the human response to VE, but it depends also on the medium characteristics.

On the other hand, the second approach takes presence as a psychological process that does not require a medium to occur: presence locates the self in an external space (either physical or cultural), and then serves as a feedback for the perceiver to tune her/his agency and control based on her/his intentions (Riva et al., 2011). This conception of presence is related to an ecological perspective of perception and action (Gibson, 1986). Indeed, the above definition of presence links perceiver's intentions and actions in the environment, with no need for a medium. Another definition of presence that draws on the ecological perspective, and specifically on the concept of affordances is that by Zahorik and Jenison (1998) who state that actions afforded by the environments lead the user to feel present in it. This reference to actions afforded by the environment in the general definition of presence, reminds us of the concept of "valid actions" supported by a VR system introduced by Slater (2009) for defining immersion in VEs. Nevertheless, the two situations are different in their starting point because Slater's position does not directly refers to presence, yet to the fact that sensorimotor and effectual actions supported by the VR system determine the level of immersion of the generated VE. In Slater's view, presence in the VE intended as the "sense of being there" (Heeter, 1992) (i.e., located inside the VE, despite the knowledge of actually being in the physical space) derives from an immersive system that supports valid actions similar to those found in physical reality. The more the approximation of the VR system to physical reality in terms of valid actions is rough, the more an additional creative mental processing is required to the user for achieving the illusion of being in the virtual place (Slater, 2009).

1.4.2. Presence Determinants in Virtual Environments

So far we have presented general approaches to presence, rather than operational characterizations that instantiate presence in VEs. Actually, the two approaches tend to converge when coming to what factors affect presence in VEs, and both consider presence as a subjective experience that represent the human user's response to the VE. The difference lies in how presence is determined: the media approach sees presence as the result of immersion in a technological medium, and the level of immersion depends on the technological properties of the VR system; in contrast for the ecological approach presence is still determined by immersion, but the latter is the result of the interaction between the user and the environment.

Anyway, the two approaches are to the opposite extremes of a line that includes other intermediate stances, which aim at standing more operational and specifically linked to VEs.

In particular, as noted by IJsselsteijn and Riva (2003), presence is a multidimensional perception, and it is determined by media characteristics and user's characteristics. Media characteristics comprise form factors and content factors, while user characteristics consist of psychological variables about the user.

Typical form factors affecting presence in VEs are the number of sensory channels generated by the VR system to stimulate user's senses; the extent of the field of view; the visual depth cues that impart pictorial realism in the VE; and the extent to which the user can exert control over the VE. Content factors instead refer to the content experienced by the user during the participation in the VE. In this context the virtual representation of the user's body and the presence in the VE of virtual characters seem to increase the sense of presence in the user. In particular, the presence of others in the VE is connected with some special kind of presence, i.e., co-presence and social presence, that often overlap, and refer to the sense of being in the VE with others (see Childs, 2010 for a review).

User's characteristics relate to the concentration of the user, her/his proneness to believe in illusion, the prior experience with VEs, the susceptibility to motion sickness (a side effect that can occur in a VE, but is far less frequent in current VR systems, thanks to their enhanced technological capabilities) (Ijsselsteijn et al., 2000), the value of the virtual experience to the user (Takatalo et al., 2008), and personality variables like empathy and imagination (Wallach et al., 2010).

Similarly, the determinants of presence are also referred to as external factors and internal factors (Slater et al., 1994). The former are those imparted by the VR system, and largely agree with media factors, whereas internal factors deal with how perceptions engendered by the VR system are mediated by mental models shaping the user's subjective experience.

In terms of presence multidimensional nature, in addition to co-presence and social presence, differing because the first requires co-location, i.e., the VE user needs to share the space with the other virtual characters (Zhao, 2003), other concepts as physical presence, spatial presence and self presence may be found in the scientific literature about virtual presence (see for instance Lee, 2004 for a review).

Physical presence and spatial presence point to the same dimension: that of feeling physically located in the virtual space (Biocca, 1997). Self presence is a particularly important component of presence when the user's body is represented in the VE: it is the effect of the VE on the mental models of the user, her/his identity and her/his emotional and physiological states (Biocca, 1997).

A further distinction of presence components that is very useful in operational terms is that by Slater (2009), which is based on the theory of sensorimotor contingencies introduced in section 1.3. Actually, Slater proposes a framework to solve the ambiguity arising from the multiple significations assigned to "presence". Such a framework is based on two concepts: *place illusion* (PI) and *plausibility illusion* (Psi), which are incorporated in an operational conception of presence as connected to how much realistically the human user responds to the VE.

PI is the "feeling of being there", i.e. the sense of being in the place created by the VE. Psi is the illusion that what is happening in the VE is real. PI depends on the physical characteristics of the VR system: if the level of immersion of the system ideally supports the sensorimotor contingencies typical of an equivalent real situation, then PI is a direct perceptual consequence of the physics of the VR system. However, in common immersive systems, which do not support all the sensorimotor contingencies of everyday reality, an additional role for PI to arise is played by the personality characteristics of the participant in the VE. PI as an illusion may break, thus quantitatively determining the level of presence in terms of "being there".

Psi is an orthogonal component of presence, and it causes the user to consider what is happening in the VR as really happening. Psi deals with the inference from perceptions of the "reality" and "credibility" of a situation. In order to arise and be maintained, Psi requires a correlation between actions and reactions in the VE, and between events happening in the VE not controlled by the user and the user sensations. It is just the correlational level rather than physical realism which determines Psi. Also Psi may break, but it is more difficult to be recovered as it deals with the credibility of the reality of the VE. This is also the reason why it is less directly dependent on the objective properties of the VR system: it is affected by plausibility of events and interactions occurring in the VE.

1.4.3. Presence Measurements

The successive step after defining presence in VEs is to cope with its measurement. There are three approaches to measuring presence in VEs: questionnaire-based methods, behavioural methods, and physiological response measurements.

Questionnaire-based methods are subjective, and they usually consist in asking the user of the VE to answer questions related to presence components and provide a rating for each component, following the performance of a task in the VE. Such methods have been often criticized as they can be confounded by many factors associated to the subjective interpretation of the questions, and the interest and expectations of the user (Freeman et al., 1999).

Behavioural methods use features of the VE to induce bodily responses that are measured and compared to responses expected in equivalent situations occurring in physical environments. In practice presence is assessed in terms of behavioural realism of the user during the virtual experience. For example, presence has been measured as a function of postural response in a moving VE (Freeman et al., 2000), and as the taking of typical behaviours exhibited around precipices when the user was immersed in a VE and had to walk on the edge of a precipice substituting a large part of the room floor (Meehan et al., 2002).

Physiological methods are more objective, as they measure user's physiological responses during the exposure to the VE, and use these measurements to assess presence. The most used physiological indicators are heart rate and electrodermal activity (Slater et al., 2009), respiration rates and skin temperature (Meehan et al., 2005), although recently also the use of other parameters has been tested, such as auditory event-related potentials (Kober and Neuper, 2012).

There are two directions for getting physiological measurements of presence. One is to compare the physiological measurement with a baseline consisting in the same physiological activity acquired when the user is in an equivalent situation in the physical reality: if the two physiological responses are similar, than this is a sign of presence. Alternatively, physiological responses may be used just as surrogates of presence, especially in stressful VEs, such as those eliciting height-fear (Meehan et al., 2005) or anxiety responses in phobics (Wiederhold et al., 2001).

A different approach to presence measurement is that related to the concept of break in presence (Slater and Steed, 2000). The idea is of inducing breaks in presence through subtle or gross events in the VE that makes the user aware that the VE is not "real". This can be done in various ways, e.g., by lowering the visual realism of virtual objects or by establishing contacts with the physical reality. The resulting breaks in presence can be counted and a probabilistic model of "presence in the VE" be constructed based on them (Slater and Steed, 2000), or physiological responses correlated with breaks in presence can be measured and used as presence surrogates (Sanchez-Vives and Slater, 2005).

1.5. The Virtual Experience

A synthetic account of the experience in an immersive VE may be drawn from the above sections. The VE consists in a digital environment rendered by a computer that uses the information in a database collecting representations of the virtual scene, its objects, and their physical properties and behaviours. The user experiences the VE through a series of interface devices. These devices create multiple channels to display the rendered VE to the sensory systems of the user, and to use information coming from the user (essentially position and orientation of body parts) to allow the interaction with the VE by taking actions, and effect changes on it.

The general goal of the VE is to provide an experience to which the user responds *as if real*. The extent to which the user realistically responds to the virtual experience mainly depends on the sense of presence elicited by the VE. Therefore the sense of presence is a way of measuring the functionality of the VE. This is true also when the VE does not simulate physical reality. Indeed, the illusion of being present in the VE rather than in the physical reality is important to respond to the virtual experience as if the virtual experience were real. The determinants and dimensions of presence are multiple, and somehow intertwined.

The physical properties of the VR system supporting the VE determine immersion and interactive capabilities of the VE, which in turn define the sensorimotor contingencies supported by the VE, and thereby a set of valid actions which are possible in the VE *as if real*. The set of valid actions, and personality characteristics such as the immersive tendency of the user determine the perceptual component of presence associated to the sense of being in the space created by the VE, despite the firm knowledge of actually being in the physical space. The orthogonal component of presence in the VE is that associated to plausibility illusion, which is more cognitive, because it is linked to the correlation between events occurring in the VE, and the sensations and expectations of the user. In this sense the user should feel a plausible correlation between external events happening in the VE and own sensations and expectations in order to feel present in the VE. A summary picture illustrative of this account is presented in Figure 1.2.

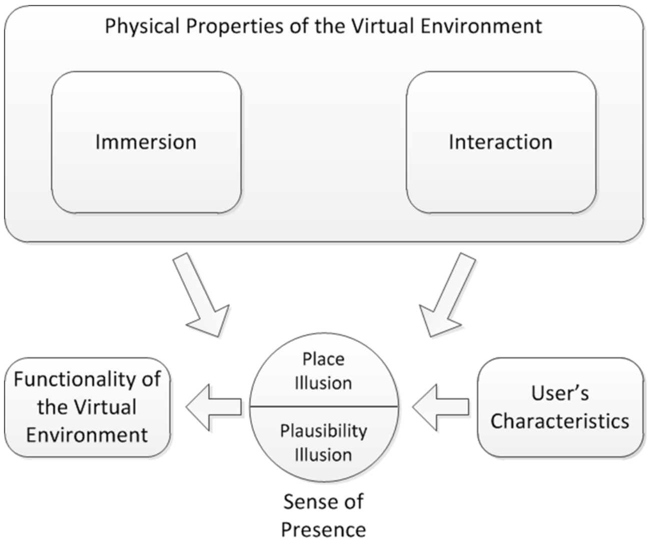


Figure 1.2.: Factors affecting the virtual experience.

Also the two perspectives mentioned at the beginning of this chapter, the technology-oriented one, and the experience-oriented one for VEs may be conciliated by remarking how the response of a human participant in a VE is determined by both objective, and subjective factors. The unitary element is provided by the experience engendered by the VE, which is primarily qualified by the experienced sense of presence. The operational conceptualisation of presence by Slater (2009) accommodates both objective and subjective properties, and thereby also technology-oriented, and experience-oriented views. Indeed, place illusion is largely a direct perceptual consequence of the physical properties of the VE, while plausibility illusion requires a correlation between actions and reactions in the VE, and between events happening in the VE not controlled by the user and the user sensations. So plausibility illusion relies less on the physical properties of the system, and more on the plausibility of events and interactions in the VE, thus on a subjective judgement.

In the next chapter several virtual environments will be considered as examples of applications to mental health. These VEs have different levels of immersion and different interactive capabilities. However, given their purpose, the measurement of presence is not frequent, and anyway not taken as the primary outcome to evaluate their functionality. Rather, correlation with measures taken in equivalent real settings, or behavioural and physiological patterns associated to different grades of a medical condition, are measured in connection to the virtual experience. These measures may be sometimes considered as a surrogate of presence, or even taken as a measure of it, provided that appropriate reference measurements are taken. Finally, it is not common in medical applications that the sense of presence is measured comparing level of immersion and interactive capabilities of the VE, and personality characteristics of the user.

2. Applications of Virtual Environments for Cognitive Sciences

This chapter provides an overview of how Virtual Environments (VE) are being applied in several fields, and focuses on their applications in connection with cognitive sciences.

In particular, the medical domain is taken as exemplary, because healthcare purposes represent the major driving forces for this kind of developments and applications, and because in that field the prerogatives of VEs are fully leveraged, at least at a potential level. Thus, the medical arena may be a starting point for outlining a framework to situate Virtual Environments applications in relation to cognitive sciences. Because the ultimate goal of this chapter is to propose such a framework, a criterion for classification of VEs applications is chosen from the applications to the medical field, and the prerogatives of VEs applications to medicine, which are generalisable, are discussed. The framework is gradually made up by considering categories of applications to neuropsychological assessment, cognitive rehabilitation, psychotherapy, and psychiatry. For each of these categories prominent examples are provided, and finally, a complete framework is proposed. Starting from considerations about the framework, and the examples discussed, the chapter ends by discussing the motivation for the study of temporal attention proposed in theoretical and empirical terms in the next chapters.

2.1. Categorisations of Virtual Environments Applications

VEs have been applied to numerous fields, covering very different areas. In the industrial context VEs are used for diverse purposes. One of these is virtual prototyping, a process that uses VR to build the virtual prototype of a product, and assess its performance (Seth et al., 2011). Virtual prototyping is usually combined with a computer-aided design (CAD) environment, and may include both the assembly part of the prototype make-up, and the interaction with the user of the final product. Virtual prototyping is used in diverse areas, such as automotive, material for biomedical applications and various manufacturing sectors. Other major industrial applications of VEs are based on simulators. These were also the first kind of applications devised and developed with VR technologies. Industrial simulators are mainly used for training and ergonomics purposes. A typical use is for training the personnel to complex maintenance tasks, possibly at geographically remote locations (Gutierrez et al., 2010; Tripicchio et al., 2010).

The potential of multimodal VEs for training applications has been studied in recent years with a novel approach, different from the traditional mainstream based on simulators. Such an approach relies on the setting up through VEs of training scenarios that replicate specific perceptual conditions and provide *ad hoc* feedback on performance in order to let the trainee completely understand the task to be learned (Bergamasco, 2012). In this kind of training scenarios, the emphasis is on the training of individual sensorimotor and cognitive skills necessary to achieve a skilful performance of the task, and not on a full, exact replication of the environment as done in traditional simulators. Useful training applications with multimodal VEs following this new approach involved, among the others, watercraft rowing in the sport domain (Ruffaldi et al., 2011), balls juggling as an entertainment application (Ruffaldi et al., 2011), and upper limb rehabilitation and maxillo-facial surgery training for the medical field (Frisoli et al., 2009; Gosselin et al., 2010).

VEs are also applied in the cultural heritage arena (Carrozzino and Bergamasco, 2010), where they can be used both as a valorisation tool and as an advanced human-computer interface for opening up new ways of interaction and fruition of the cultural contents. A typical example in this area is the Museum of Pure Form (Bergamasco et al., 2001), a virtual museum that allows the visitor to view in 3D and touch with haptic interface in an immersive CAVE-like VR system digital statues modelled from their physical counterparts preserved in museums spread worldwide.

Important and numerous applications of VEs refer to healthcare, but they will be considered in more detail in the remainder of this chapter.

In general, a categorisation of VEs applications may be formulated according to at least two criteria. An intuitive one is to define the categories by application domain; thus a possible, non exhaustive, categorisation includes industry, medicine, training, entertainment. And each of these categories could be then broken down in sub-groups such as automotive, surgery, education, gaming, cinema etc. However, this sort of categorisation is quite arbitrary, because many different categories and, especially, sub-groups can be defined; and often sub-groups are prone to overlap to one another: e.g., there are industrial training, as well as game for training, and in such cases it would be subjective the decision about whether to emphasise the industrial, training or entertainment aspect of the applications.

The second method of classification is to give more relevance to how VEs are applied, achieving in such a way categories that go across application domains, and emphasise the VEs peculiarities beneficial for a whole class of application fields.

Anyhow, it is hardly feasible and almost unfruitful to attempt a coverage of all VEs application domains, whatever the chosen categorisation criterion. Here we adopt the method based on how — and why — VEs technologies are applied, and we focus on VEs applications directed to cognitive assessment and rehabilitation, including psychotherapy and psychiatry. Indeed, albeit this thesis is not intended to validate healthcare VEs, our idea is that those examples offer the right substrate onto which an elaboration about more general applications of VEs to cognitive science may be built.

2.2. Simulation-Centred *vs* Interaction-Centred Applications in Medicine

As discussed in chapter 1, VEs are simulated environments that provide a powerful means of human-computer interaction. Given appropriate technological properties of the VR system hardware, with sufficiently elaborated models of the VE and its objects, it is possible to have highly-interactive and physically-realistic simulations of real situations, as well as very natural paradigms of interaction with such simulated situations. Therefore, depending on whether the VE accent for an application is on the *simulation for* or the *interaction with* the target users, two categories are recognizable concerning *how* VEs are applied: that of *simulation-centred* VEs applications, and that of *interaction-centred* VEs applications (Riva, 2006).

Simulation-centred VEs are those in which the primary goal is to have an accurate simulation of real objects, their appearance, and their behaviour in terms of physical interactions. Interaction-centred applications are instead those where the VE primary role is to open up enhanced ways of interaction for its user, in order to achieve the goal of the application itself, sometimes even by deliberately forcing the VE to respond unrealistically.

This categorisation, by distinguishing VEs as simulation tools from VEs as interaction tools, proves to especially fit the medicine application domain. When dealing with VEs applications in medicine, it is indeed possible to consider the following different sub-domains (see Riva, 2003; Adamovich et al., 2009; Gregg and Tarrier, 2007; Rizzo et al., 2002 for a review):

- Surgical simulation and planning
- Telepresence surgery
- Virtual endoscopy
- Medical education
- Medical Rehabilitation
- Psychotherapy and Psychiatry

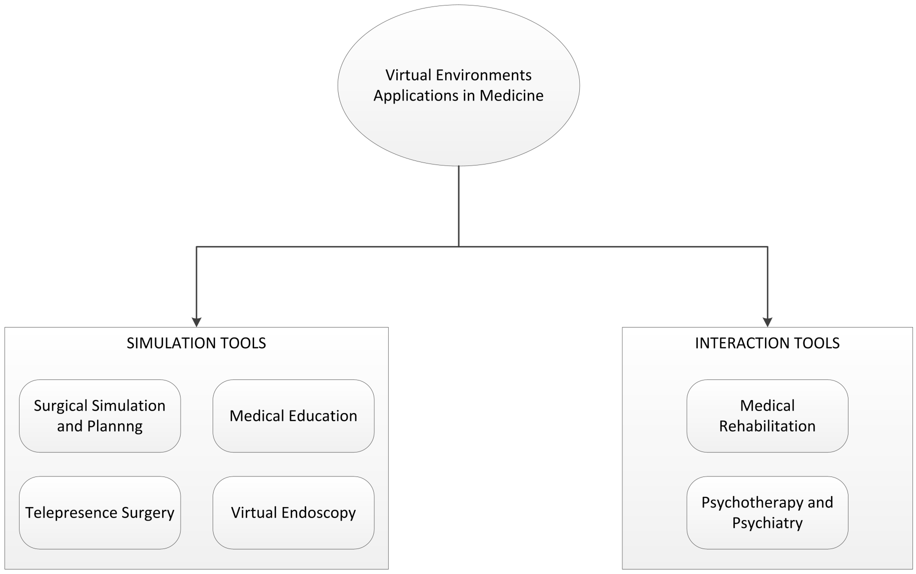


Figure 2.1.: A categorisation of VEs applications in medicine. Application sub-domains dealing with surgery, endoscopy and medical education chiefly rely on the potential of VEs as an accurate simulation tool. Real objects and their behaviours are simulated accurately, e.g., for training surgeons skills with real-like operative practice on virtual patients, or assessing non-invasively the airways of a real patient through a virtual endoscopy. On the other hand, applications in the field of medical rehabilitation and psychotherapy exploit the capabilities of VEs as natural and enhanced interaction tools. For instance, in neuropsychological rehabilitation VEs can selectively stimulate the neglected space and require response for it from a patient suffering from unilateral spatial neglect, or other VEs can provide a "safe" social place for a social phobic to meet virtual characters, and undergo an exposure therapy.

Whereas the first four sub-domains usually need applications that primarily exploit the simulation capabilities of VEs, medical rehabilitation, and psychotherapy do mainly rely on the interaction potential made available by VEs (Figure 2.1).

2.2.1. Simulation-Centred Applications in Medicine

Historically, the first applications of VEs to medicine were in surgery more than a decade ago (Krummel, 1998). At present, such applications cover different tasks, and partially overlap with medical education, specifically with training of surgical skills. Actually, VEs are used for surgical training, for diagnosis and pre-operative planning, and for intra-operative guidance and assistance (Pednekar and Kakadiaris, 2000).

The main reasons to use VEs for surgery are different according to the area of use. In terms of surgical training, VEs are unique as to the possibilities they offer to entirely simulate surgical procedures on different kinds of tissue, located in specific positions and orientations, and to repeat the tasks several times, in safe conditions, with the possibility to also play back what done. In addition, for training purposes it is important to replicate the multifarious and multimodal feedback potentially returned to the surgeon during the intervention, such as during drilling or while searching for anatomical landmarks. These are especially important requirements, in particular for the training of minimally invasive surgery (MIS, Gallagher et al., 2005), which requires particular psycho-motor skills and *per se* offers an impoverished interface and feedback for the surgeon (Tendick et al., 2000). Hence, simulators have been chiefly developed for MIS, also due to the restricted information set to be modelled, but they have anyway been developed to cover the training of numerous surgical tasks, such as vitro-retinal surgery (Rossi et al., 2004), knee arthroscopy (Lu et al., 2009), and other orthopaedics interventions (Mabrey et al., 2010). In addition, simulators for dental surgery (Thomas et al., 2001), and maxillo-facial simulators for training of complex tasks as the Epker osteotomy (Mégard et al., 2009), neurosurgery (Kockro and Hwang, 2009), and gastrointestinal surgery (Lauscher et al., 2010) have been set up too.

In short, VR surgical simulators are based on models of the human organ of interest, including deformable models specifying how the related tissues react to surgical operations, and on tracking and rendering algorithms to enable the presentation to the surgeon of the multimodal information and interaction possibilities required by the specific training scenario. In addition to visual feedback, tactile and force feedback are crucial for surgical training, as well as appropriate auditory feedback prompting the surgeon of the accuracy of a path-following or of a change in the nature of the tissue encountered while the surgical action (Pednekar and Kakadiaris, 2000).

More recently, also Augmented Reality (AR), i.e., the combination of virtual information with those coming from the physical world, has been used in surgical training (Botden and Jakimowicz, 2009). In this case virtual models of the organs to be operated are presented to the trainee in a hybrid setting often including physical instruments and a mannequin or a real video display used in laparoscopic interventions. One of the main advantages of AR simulators is the real haptic feedback they allow to experience, though in the most recent VR simulators haptic interfaces are included (Konietschke et al., 2010; Gosselin et al., 2010), and represent a key feature of these systems, as they also work in playback, that is they make possible to repeatedly experience a procedure, including in haptic terms.

VEs for diagnosis, pre-operative planning, and intra-operative assistance and guidance are patient-specific. This means that the model of the organs comes from medical images acquired on the specific patient (Pednekar and Kakadiaris, 2000). These images are typically Computed tomography (CT) scans, Magnetic Resonance (MR) images, Positron Emission Tomography (PET) scans and Ultrasound images. All these images need to be fused together for achieving a single model and providing multimodal information of the patient, and need to be registered, i.e., a common reference frame between the pre-operative model and the patient's anatomy needs to be established for subsequent congruent operations that are based on the model, but operated on the patient. Clearly, the main benefit of surgical planning is the possibility to anticipate a simulation of the intervention and its outcome, to practice it and rehearse in order to optimise surgery results.

Telepresence surgery (a.k.a. telesurgery) is a different surgical paradigm that uses robotics and VR technologies to practice surgery with no physical contact between the surgeon and the patient. The operations are physically executed on the patient by a robot, which is endowed with surgical tools and is teleoperated by a surgeon, who may be in a distance of a few meters from the patient or even very far (Marescaux et al., 2001).

The role of VR in telesurgery is essentially to immerse the surgeon in the operating field despite its physical distance. This is achieved by providing a stereoscopic visual display that offers a three-dimensional magnified view of the operating workspace, a haptic feedback from the interaction of the robotic surgical tools with the tissues of the patient, and the possibility to accurately scale the surgeon's movements and combine previously acquired medical information with those coming from the patient in real-time. A typical example of telesurgery platform is the DaVinci system (Intuitive Surgical, Sunnyvale, California).

Virtual Endoscopy starts again from medical imaging sources, and specifically CT and MR data, to construct three-dimensional models of a human organ. The endoscopy is then performed by flying through this virtual model, and is completely non-invasive for the patient. The main goal of virtual endoscopy is to provide information about the inner surface of anatomical structures in an interactive exploration (Vining, 1996). The use of virtual endoscopy is being evaluated for large airways diseases (Thomas et al., 2009), for the ear (Karhuketo et al., 2002), for the urinary and gastrointestinal tracts (Battista et al., 2009; Aschoff et al., 2008), and for angioscopy (Louis et al., 2009).

As for medical education, besides the training tools mentioned in the surgical field, and the possibility to use virtual endoscopy as a training aid, the contribution of VEs is mainly for providing anatomical knowledge in three dimensions, with the possibility to have photorealistic human atlases that can be navigated and touched by trainers and trainees. An important database in this framework is that from the Visible Human Project (Spitzer and Ackerman, 2008), which includes three-dimensional anatomical representations of the normal human male and female bodies, largely based on digital bioimages. Recently, also AR has been used to build synthetic environments for teaching anatomy (Thomas et al., 2010).

2.2.2. Interaction-Centred Applications in Medicine

Biocca and Delaney (1995) defined Virtual reality as a medium able to extend body and mind. A reason for this definition is that in the VE the user may be an active participant, and her/his interaction, encompassing both perceptual and cognitive levels, and covering multiple sensory systems, may go beyond the observation and, even, the manipulation of virtual objects that accurately resemble their physical counterparts. Moreover, the extension enabled by the VE consists of actions and behaviours that the user can take when participating in a virtual experience, actions and behaviours being impossible in the physical reality, e.g., due to the health condition of the same user.

The interaction paradigm that enables such a *going beyond* — that might be interpreted as the “extension of body and mind” referred to in the cited VR definition — is the cardinal nucleus holding the applications of VEs to medical rehabilitation and assessment, and to psychology. In addition, in the interaction with the VE this extension of body and mind can be controlled, because the VE is a computer-generated world consisting of input and output data streams that become afferent and efferent information for the user, and concretise the active interaction process, possibly even transforming user’s actions and behaviours. These data streams are fully controlled, hence offering a controllable interaction in a controllable environment, still maintaining quite an ample degree of flexibility.

Therefore, interaction-centred applications of VEs try to realise an extended form of interaction by seeking two major goals: to fully exploit the potential of an extended natural interaction, which overcomes the possibilities of interactions in the physical world, and still to maintain the ecological validity in the resulting virtual experience, which remains a controlled context. These are the prerogatives of VEs that mainly motivate interaction-centred applications in medicine, and, more broadly, in neuroscience research.

When considering the interaction-centred applications of VEs to medicine in more detail it is possible to further split the applications in medical rehabilitation, especially to keep up with the focus of the present work, that is to explore and shed light on the potential of VEs for cognitive science. Thus, Figure 2.2 proposes a refined categorisation, also adopting the heading of *Virtual Rehabilitation* (Burdea, 2003), which is often used to indicate all the diverse applications of VEs to medical rehabilitation.

Virtual rehabilitation is an umbrella which covers the applications of VEs to medical rehabilitation, and can be specified in various ways: a means of specification is the patients population addressed, another is the extent to which VEs and VR technologies are used to provide rehabilitation, and a third means is the protocol used, or the fact whether the therapist is physically present with the patient or not, in the latter case implementing a form of *telerehabilitation*.

In the present work it is adopted a classification of virtual rehabilitation based on the origin of the dysfunction to be remediated. In particular, neurorehabilitation is distinguished from general rehabilitation — i.e., rehabilitation for dysfunctions not originated by injuries to the central nervous system — and within neurorehabilitation, cognitive assessment and rehabilitation are separated from motor rehabilitation. However, it is worthwhile noticing that what Rizzo et al. (2004) call *the assets* for VR applications — from which we drew here for defining the prerogatives

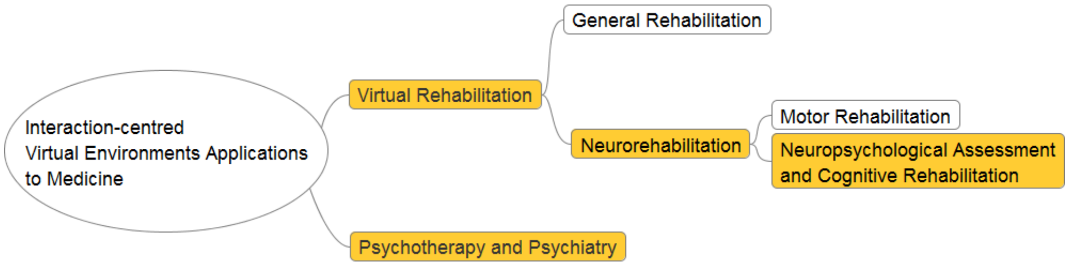


Figure 2.2.: A proposed categorisation for application fields of interaction-centred VEs applications in medicine. Virtual Rehabilitation is an umbrella term covering the use of VEs for medical rehabilitation. The primary scope of application is neurorehabilitation, both for motor training and for cognitive rehabilitation and neuropsychological assessment. Psychotherapy and Psychiatry, usually not included under the heading of virtual rehabilitation, are still part of the categorisation, with a number of mental disorders that may find in VEs a valuable therapeutic instrument. Alternatively, following a trend emerging from part of the scientific literature, it would be possible to change the heading "Psychotherapy and Psychiatry" with "Mental Health".

mentioned in section 2.3 — are similar for the diverse populations benefiting from virtual rehabilitation.

The basic assumption underpinning neurorehabilitation is neuroplasticity, i.e., the capacity of the brain to change across the lifespan, and the fact that sensorimotor and cognitive training may induce such changes to counteract dysfunctions caused by brain lesions. Correspondingly, the stated goal that defines neurorehabilitation is to change brain function in order to remediate a disability due to an impairment caused by damage to the central nervous system, most often the brain (Robertson and Fitzpatrick, 2008; Rose et al., 2005).

There are different mechanisms involved in brain plasticity (Johansson, 2000), and some of them take advantage from VR-based or VR-augmented rehabilitation practices. VR-based practices completely substitute the conventional exercises with tasks that fully unfold by interacting with a VE. The distinctive points of this sort of therapy pertain to what sensory modalities are stimulated, and to the degree of interaction foreseen between the patient and the VE. Such aspects are very important for both motor and cognitive neurorehabilitation using VEs.

The primary goal of rehabilitation is to recover or — if not possible — compensate for a lost function. Both motor and cognitive dysfunctions may result from lesions to the central nervous system, especially the brain. There are multiple fundamental factors influencing the rehabilitation course under the aspects of motor and cognitive recovery (Sveistrup, 2004; Rose et al., 2005). One is early intervention, the others can be summarised as task-oriented training and repetition intensity, and for both of them VEs offer unique opportunities. The advantage of VEs concerning repetition intensity is straightforward: from the patient side, VEs may provide training scenarios that are appealing and motivating for the patient, reducing the "boring effect" that often hinders patient's motivation, and by this way also the rehabilitation outcome (Burdea, 2003). From the therapist point of view, the aid of VEs technologies, especially if combined with robotic devices for motor rehabilitation, allow for an increased repetition intensity, as technologies do not become tired and make possible for the therapist to follow more than one patient at a time.

Anyway, the crucial advantage featured by VEs is the kind of interactive environments provided (Rose et al., 1998, 2005). VEs can in fact implement task-oriented scenarios that provide sensorimotor training in an enriched and motivating environment. Environmental enrichment consists of an increased level of interaction with the environment, and it has proved to increase adult neurogenesis (Olson et al., 2006) in an ecological-like approach. Elaborating on the prerogatives of VEs described in section 2.3, VEs can provide an ecological context that engender enrichment, while adapting to specific patients' needs for rehabilitation, and maintaining a continuous, real-time control in the stimuli-response relation.

In the context of cognitive rehabilitation (CR), diagnosing a dysfunction, identifying cognitive weaknesses and strengths, and monitoring the course of the rehabilitation outcomes are achieved through neuropsychological assessment (NA). NA is based on psychometric tests, which should in principle measure simple, or complex cognitive functions of an individual in a reliable and standardised form, while predicting the level of functioning of the same person in the everyday life. Unfortunately, this has since long been identified as a major issue of psychometric tests, because they often miss the level of ecological validity necessary for generalising to lifelike situations the functioning measures they yield. This issue has been only partly mitigated by the shift from paper-and-pencil tests to computer-aided versions of the same tests, which are now very commonly and widely used for NA. VEs technology, with the capability to create a whole environment that is ecologically valid, and allows for naturalistic interactions, yet keeping a strict control on the stimulus-response function, has proved to be effective to further alleviate the problem originated by the lack of ecological validity (Rizzo et al., 2000).

The term rehabilitation is far less used when dealing with therapies, and assessment, of mental disorders pertaining to clinical psychology, and psychiatry. The two major lines of application of VEs in this context are for virtual exposure therapy, as an adjunctive intervention to standard cognitive-behavioural therapy, and for altering the perception/representation of own body when this is disturbed. In the first case, VEs are used as flexible and controlled environments where patients can engage in systematic, natural, and realistic exposures to virtual stimuli that are critical for arising and sustaining the symptoms of their mental disorder. In other words the VE can be conceived also as an environment that affect patients' emotional processing, and facilitates learning of their dysfunctioning thoughts, beliefs, and behaviours. In the second case, VEs are used to change the perception of own body representations and somatoperceptions, in order to correct body image disturbances, or to enhance control over pain.

Based on the proposed categorisation, the next sections of this chapter describe the potential of the interaction-centred applications of VEs with a focus on neuropsychological assessment and cognitive rehabilitation, and on psychotherapy and psychiatry. Specific general prerogatives are introduced, before presenting prominent examples in the specific domains of NA, CR, psychotherapy, and psychiatry. The applications to motor training in neurorehabilitation are not discussed because they go beyond the scope of this thesis, which is focused on cognitive aspects. Even so, it should be noted that the potential of VEs exploited in motor rehabilitation is largely the same as in cognitive rehabilitation, albeit robotic devices are more often used in connection with VEs when motor training is the primary goal. In addition, in many situations motor rehabilitation goes with therapies for enhancing basic cognitive abilities and executive functions of patients. For example, this is the case of stroke patients, for whom the value of VEs has been demonstrated for both motor recovery and cognitive rehabilitation (see Henderson et al., 2007; Weiss et al., 2005 for a review).

2.3. Prerogatives of Virtual Environments for Interaction-Centred Applications

The potential of VEs for neuropsychological assessment, cognitive rehabilitation, and, more broadly, for mental health have been discussed in several occasions since the late 1990s (Rose et al., 1998; Rizzo et al., 2000; Schultheis and Rizzo, 2001; Rizzo et al., 2002, 2004; Rose et al., 2005). The conclusions converge on a set of key prerogatives of VEs that are largely independent of the specific cognitive functions they apply to, and underpin also VEs value for motor rehabilitation, and for neuroscientific research. These prerogatives are listed, and briefly discussed in the following of this section.

Environmental enrichment is an important asset to the use of VEs for rehabilitation purposes, as it promotes neuroplasticity, which is clearly crucial for recovering the impaired function. The enrichment provided by a VE encompasses both integrated multisensory stimulation, and a highly-interactive setting for the patient. Hence, a VE may counteract the reduction in interaction suffered by patients with disabilities, and enhance their motor activity.

Improved ecological validity corresponds to the possibility to provide naturalistic scenarios, relevant for predicting performance in everyday functional environments, for achieving more generalisable rehabilitation outcome, and for studying brain activity in a real-world-like situation, which, however, maintains the controlled characteristics of a laboratory experimental activity. This asset of VEs is a clear example of the synergistic relationships connecting all the prerogatives towards a common resulting action beneficial for NA, and CR, but also for sensorimotor training of neurological patients, and for basic research about brain function. Indeed, the ecological validity of a VE consists of a naturalistic, and contextually rich, interactive environment, enabling naturalistic, and interactive behaviours of the participant in the virtual experience, even during brain activity imaging (Bohil et al., 2011). The lack of ecological validity has long been recognised as a major issue of psychometric tests in NA, and of mainstream interventions in rehabilitation. VE has the potential to overcome this issue, yielding scenarios and tasks for NA that provide results applicable also to everyday contexts. This is also favourable in respect to the recognition of the both the impairment, and the value of the aid offered by NA and CR, from the side of the cognitive patients and their families.

Systematic and controlled delivery of 3-D multimodal stimuli resolves, in combination with improved ecological validity, the dualism inherent to traditional NA, CR, and basic research on brain activity. The dualism lies in the difficulty to keep naturalistic a task and its context when probing brain activity, or psychometric properties of a patient. The same dualism can affect also rehabilitation procedures, in the motor, cognitive, and psychiatric domains. Beyond the continuous control of stimuli for reliable, and standardised deliveries, adaptation to patients' needs, both general and relative to their contingent response, is possible by tuning the parameters of VEs. The 3-D characteristic of feedback, and the multiple modalities, primarily visual and auditory, but recently also haptic, and in its infancy even involving taste and smell, enable a more fully sensorimotor engagement. This is beneficial in several ways, again in synergy with the other VEs prerogatives: psychometric and behavioural responses are more realistic, participants may move about inside the VE, increasing motor activations, and also the interaction level, in turns enhancing the environmental enrichment. Furthermore, multimodal stimuli promote the immersion of the participant in the VE, providing the ground for triggering a high sense of presence in the VE, hence achieving engaging and highly-motivating contexts. The sense of presence is a determinant factor for the functionality of a VE (Sanchez-Vives and Slater, 2005), and high motivation to practice is fundamental for rehabilitation purposes (Schuemie et al., 2001; Weiss et al., 2005).

Immediate multimodal feedback may be beneficial in several ways. Firstly, immediate is referred to real-time feedback about the performance of the participant in the virtual experience. Online feedback is important for every skills acquisition (Bergamasco et al., 2012), and in rehabilitative contexts appropriate feedback is useful for guiding the patient, and also for heightening her/his motivation to engage in the rehabilitative task (Adamovich et al., 2009; Weiss et al., 2005). Secondly, multimodal feedback can be used to dynamically test learning processes, such as strategies for acquiring spatial cognition, and navigation, which are important research topics *per se* (see Bohil et al., 2011 for a review), and in relation to cognitive impairments, like topographic disorientation (Livingstone and Skelton, 2007). Lastly, the feedback can be deliberately altered to treat specific impairments, as unilateral spatial neglect, or to test patients' abilities to adapt to novel situations Baheux et al. (2007).

Cueing stimuli to guide successful performance is a specific approach based on error-free learning, in contrast to the trial-and-error approach, that appeared to be beneficial, especially for procedural abilities, which can be spared in individuals suffering from declarative memory disabilities (Rizzo et al., 2004). Cues are presented in the VE prior to the participant's response, in order to guide her/him to a successful accomplishment of the task through an error-free performance. This is an example of how VE can go beyond a mere mimicry of the physical environment: on the contrary, the VE can provide stimuli fostering processes very difficult, or even impossible to achieve in a real environment. Furthermore, in the VE the cues may vanish as the patient progresses, and performance improves, and is retained.

Complete digital performance capture is a capability inherent in the the technologies composing VEs, and in those used in combination with them, like motion capturing systems for human motion analysis, and even brain activity recording systems (EEG, but recently also fMRI). In particular, immersive VEs include tracking systems affixed to participant's head, for example to the eyegoggles used for stereoscopic vision, and this information, normally used to update the graphic 3-D scene, may also serve to monitor and record the patient state, or to effect changes on the VE. This is important in rehabilitation, as patient's progresses may be precisely measured along several relevant dimensions of the performance, with the exact knowledge of what were the natures and levels of the stimuli eliciting the measured response. In psychotherapy, for example, the biofeedback of a patient may be connected with the VE, and effect changes on it according to the patient state to help regulate anxiety (Repetto et al., 2009).

Safety testing and training environments are a prerogative of VEs. The "virtual" nature, and the possibility to mimic physical reality, and to adapt it to the contingent situation by maintaining a naturalistic interaction, enable to test patients in otherwise risky tasks (for example testing and training the ability to cross a road, or to drive in patients with visual-spatial neglect, as in Katz et al., 2005). Moreover, some patients need undergoing therapies objectively difficult in real settings, as exposure therapy to alleviate fear of flying, and they might not accept exposure to real phobogenic stimuli, in contrast to exposure to a virtual counterpart, in the sheltered place offered by the VE (examples of this kind of applications are in VR exposure therapy for the fear of flying, as reported in Rothbaum et al., 2006, and for spider phobia as in Garcia-Palacios et al., 2002, and social anxiety disorder, as in Anderson et al., 2005).

In next sections the reader may find examples of how these prerogatives have been used in specific applications to virtual rehabilitation, and to mental disorders.

2.4. Applications to Neuropsychological Assessment and Cognitive Rehabilitation

The damage to the central nervous system that may cause dysfunctions are multiple, either acquired, or congenital. Thus, one possibility for categorising applications of VEs to NA and CR might be to use the origins of the dysfunction as categories: acquired brain injuries, neurological disorders, and developmental disorders. However, this does not seem to fit the purpose, because such categories are often the cause of a wider range of dysfunctions, not limited to the cognitive domain. Most commonly, the scientific literature about cognitive rehabilitation interventions progress by the cognitive abilities addressed. Here the same approach is taken, considering the cognitive domains presented in Figure 2.3, and reporting prominent examples of how VEs are being used for NA and CR within such domains.

CR and NA interventions are classified under five main cognitive domains: attention, memory, visuo-spatial skills, executive functions, and language. Perceptual abilities are implicitly included at different points in this categorisation (e.g., for visual and spatial skills, and for the most general fact that they are a preparatory processing step for the cognitive stage).

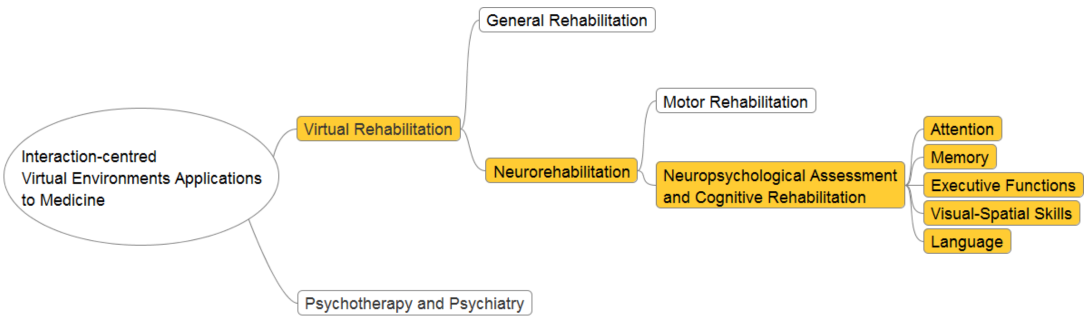


Figure 2.3.: Cognitive-domain-based categorisation of Virtual Environments applications to neuropsychological assessment and cognitive rehabilitation.

Examples of Applications to Attention Disorders

Attention is a multidimensional concept, and should not be considered as a unitary cognitive construct, or a single process. Rather, it is a property of multiple cognitive operations (Chun et al., 2010), and is almost ubiquitous in everyday life, because its central function is to select for processing relevant information, which would otherwise be dismissed, never entering the conscious awareness of an individual. A more accurate and detailed account of human attention, and of its role in the human information processing is presented in chapter 3. Thus, here the discussion is focused on the specific ways VEs have been applied to attention for NA and CR.

The attentional abilities that may be impaired and require CR are defined according to the skills required to perform one or more tasks, and several taxonomies have been proposed (see Johnstone and Stonnington, 2009 for a review). Below is a list that outlines what are the functional capacities implicated:

- *Arousal* is the level of alertness with respect to response to the environment. There are both a tonic and a phasic arousal: the former is the day-to-day alertness, independent of the immediate task demand, while the former is the ability to respond to changes in the environment, or in task demand.
- *Focused attention* is one's ability to focus on a specific object or stimulus, especially visual, auditory, and tactile stimuli, and discretely respond to them ignoring distractors.
- *Sustained attention* (a.k.a., *Vigilance*) is the ability to sustain a focused attention sate over an extended period of time, e.g., requiring a repetitive response.
- *Divided attention* is the ability to simultaneously attend to, respond to multiple tasks at a time. It includes also the concept of *alternating attention* as the ability to move between tasks, which have different cognitive requirements.
- *Executive attention* includes those attentional functions aimed at controlling one's thoughts and actions to produce a coherent behaviour (e.g., initiating a task, making a decision etc.). However, when dealing with rehabilitation, executive attention (often also referred to under the heading of executive control, cognitive control, or conflict monitoring) is usually not included in the taxonomy of attention. Rather, it is associated to the executive functions.

In terms of VEs aids for NA and CR of attention, the most outstanding application is for Attention-Deficit/Hyperactivity Disorder (ADHD), which is a behavioural disorder mostly affecting children, with symptoms of inattention and/or impulsivity or hyperactivity (Faraone et al., 2003). The Virtual Classroom designed and implemented by Rizzo et al. (2006) is an immersive VE chiefly used for the assessment of attention in children. The VE represent a school classroom, where the child is immersed, with a first-person perspective. The child is transported in the VE through an head-mounted display (HMD), and may interact using a joystick, or a response box when asked to accomplish tasks. These tasks are intended as psychometric tests of different attentional abilities. In the virtual classroom there are virtual classmates, who may chat, and act as distractors for the child. There is also a teacher, who provides the instructions for the tests. These are usually conducted using the virtual blackboard: for example the teacher asks the pupils to respond every time an animal drawing appears in a stream of non-animal drawings presented on the blackboard. Meanwhile, the school principal may knock on the door, and enter for a brief talk to the teacher; loud voices, may come from the open window, or a coloured car may pass by it; some classmates may chat. In this way sustained and focused attention of the child are assessed, in a naturalistic context, and contributions of different kind of distractors may be evaluated. The "Virtual Classroom" proved to be capable of effectively differentiating between participants with ADHD and normal control children, as it correlated with common tools used for ADHD assessment (Parsons et al., 2007). Moreover, this VE has been used to assess attention-related abilities in individuals with attention problems not descending from an ADHD (e.g., children with traumatic brain injuries, as illustrated in Nolin et al., 2009). In the scientific reports about studies using the Virtual Classroom as a NA tool a frequent conclusion is that much sensitivity is achieved compared to mainstream psychometric tools as the continuous performance test, because a major advantage of the VE is its ecological validity.

Another interesting aspect of VEs for NA and CR underscored by the Virtual Classroom is the possibility to adapt the technological platform underpinning the application for a cognitive domain to another cognitive domain. Indeed, a "Virtual Office" version of the Virtual Classroom is used for the assessment of memory impairments in adults (this application — see Rizzo et al., 2002 — is recapped in the following, when dealing with memory impairments).

Not many other applications of VEs for the assessment and rehabilitation of basic attention abilities have been proposed. One is that brought forward by Lengenfelder et al. (2002). This is an application for assessing the influence of divided attention on driving performance of patients with traumatic brain injuries. The VE is a 1.75-mile long road, with curves, presented on a computer desktop. The patient has a steering wheel, and gas/brake foot pedals to simulate car driving. Using these tools, and looking at the graphical VE on the screen as through the car windscreen, the patient has to drive the virtual car controlling its speed, and keeping it at the centre of the lane. In addition to this primary task, the patient's divided attention is demanded to accomplish a secondary task: speak aloud the 4-digit numbers appearing on the screen while driving, and continue driving. The numbers may appear in a fixed location or at different locations, and they may succeed one after another every either 2.4 seconds or 0.6 seconds. The performance of a group of patients was compared with the performance of a healthy group of matched individuals. The results are positive for the potential of VEs to be used as an assessment tool of divided attention. In fact, patients made more errors than healthy controls on the secondary task as expected based on previous studies, and performance resulted more influenced from the rate at which numbers appeared, than from the random localisation of the numbers.

Attention impairments associated with spatial neglect have been also addressed with virtual environments, and examples of these are briefly accounted in the one of succeeding paragraphs, which introduces applications targeted to spatial abilities.

Examples of Applications to Memory Impairments

Also for memory, several taxonomies have been proposed and adopted over the decades. Some of them focus on functional processes (encoding, consolidation, and retrieval of information or motor skills), others on the modality through which the information to be memorised was presented (auditory memory, visual memory, and motor memory), and on the short-/long-term of the storage, i.e., whether the memories are intended for immediate use, or they have to be permanently stored as knowledge available for successive retrieval. Besides, memory may be incidental or intentional, and retrospective or prospective, episodic or semantic, declarative or procedural. Traumatic brain injuries (TBI), cerebral vascular accident, Parkinson's disease, and Alzheimer's disease lead to memory impairments.

The "Virtual Office" evolved from the Virtual Classroom (Rizzo et al., 2002) has been successfully used to assess recall of non-typical targets in TBI patients. The VE again proved to correlate with measures taken through mainstream psychometric batteries, seemingly induce the emergence of specific unimpaired memory components not assessed by the usual test battery.

Brooks et al. (2002) used a VE simulating a four-room bungalow to assess prospective memory in a sample of stroke patients, compared to an age-matched group. Participants had to cluster furniture items to be removed for a move in a larger house, according to the destination room; to indicate for a "fragile" notice suitable objects, to grant the removal men access; to remember closing the kitchen door every time to avoid the cat going outside; and to hit a button next to a virtual clock exactly every 5 minutes. The results showed the usefulness of this VE to test prospective memory of stroke patients, being sensitive to particular problems of what and when to remember.

Pugnetti et al. (1998) compared different memory performance between multiple sclerosis patients, and healthy controls for targets presented during either passive or active navigation conditions in a VE. The VE was capable of revealing a different spatial memory performance of patients according to the navigation modality — either passive or active — utilised. This difference cannot be highlighted in standard psychometric tests.

Parsons and Rizzo (2008) undertook the preliminary validation of a battery of psychometric measures based on a VE presenting a city where a participant may immersively navigate, while accomplishing assessment tasks. The VE includes a memory assessment module, which measures correlated with standard psychometric memory tests, and not with non-memory measures. However, the authors admit the necessity to proceed with validation in respect to many aspects.

Examples of Applications to Impairments of Executive Functions

Executive functions are higher functions that majorly distinguish humans from other mammalians. They can be summarised as those functions that enable planning and organisation, initiation, self-regulation, monitoring and termination of activities, sequencing of actions and time management. In a word, executive functions make possible the effective performance. A major cause of executive dysfunction is brain damage to the frontal lobe.

The same 4-room virtual bungalow used in Brooks et al. (2002) for assessing prospective memory impairments has been used by Morris et al. (2002) to assess the ability in strategy formation and rule breaking of patients who had undergone frontal lobe neurosurgery, in the same furniture removal task. Rehabilitation practices for sequencing of actions, and multitasking were implemented within a virtual kitchen (Zhang et al., 2001) where the task was to prepare a meal going through a specific series of steps, and in a VE simulating a super-market (Rand et al., 2009) where a virtual shopping was undertaken. These same VEs, and similar ones, have been extended and used to assess more generally activities of daily living, and they demonstrated success in small groups of patients, although standardisation was lacking. Also virtual equivalents of psychometric tests have been designed and studied, although they still need extensive validation before achieving the status of a standard tool, usable in clinical practice. One example is a virtual equivalent of the Wisconsin Card Sorting Test (WCST, Elkind et al., 2001). The setting is a virtual beach scene, where virtual bathers are at umbrellas with different with particular objects on each of them. Patients are requested to deliver beach-related objects to the virtual characters according to a matching pattern alike in the standard WCST. Patients may receive verbal feedback from virtual bathers. The results of the study were encouraging, as the measures for executive dysfunction correlated, although the virtual version resulted more difficult than the traditional one. Besides, an equivalent of the multiple errands test (Shallice and Burgess, 1991) for the assessment of executive functions has been proposed by Raspelli et al. (2009). It is a virtual shopping mall where the patients have to select and buy products, also obtaining information, according to specific rules.

Examples of Applications to Impairments of Visual-Spatial Skills

Visual-Spatial skills are complex because they rely on many parts of the human brain, and encompass perception, processing, and interpretation of visual-spatial stimuli. Hence, many functions may be affected by impairments of visual-spatial abilities. The visual spatial system may be seen as having two main functional components (Johnstone and Stonnington, 2009): an input component, aimed at the accurate perception of visual information, and an output component, aimed at integrating this information with contextual cues, and with action. The perceptual component includes receptive visual functions (visual acuity, depth perception, visual fields), visual-spatial attention, and spatial perception, whereas the integration component covers spatial orientation and body schema. Thus, it is clear that some deficits of visual-spatial skills may be assigned to other cognitive domains, such as attention, in different taxonomies. Virtual Environments have been applied to the assessment, and rehabilitation of impairments affecting visual-spatial attention, spatial orientation, and body image.

An important disorder of visual-spatial attention is the unilateral spatial neglect. This neuropsychological condition often impairs stroke patients with a lesion in the right brain hemisphere. These patients are impaired at responding to stimuli occurring in the hemispace contralateral to the brain lesion. VEs have been developed for both the assessment, and the rehabilitation of unilateral spatial neglect. The fundamental idea is to track the head and eye movements of the patient while viewing a VE — typically wearing a head-mounted display that delivers the VE shutting out the physical environment — and to use moving cues, and possibly distortions biassing the unimpaired hemispace, to measure and assist patient's attention to the neglected space. A review of these applications has been conducted by Tsirlin et al. (2009). The main advantages of VR-based methods over traditional ones are in tracking and multimodal capabilities of VEs, and in their potential for providing lifelike tasks and contexts. Thereby, several parameters of the patients can be captured and tracked in real-time, revealing their inactivity during both the assessment and rehabilitation phases, and prompting them to counteract the deficit. Several sensory modalities may be used either for stimulating the patient, or investigating the consequences of the neglect disorder. Finally, a VE may enable to assess and intervene on the spatial neglect not only in the peripersonal space. The applications are still to be led to a standardised profile, so that they can be applied in everyday clinical protocols, but they have shown to be able to augment both assessment, and rehabilitation of unilateral spatial neglect. Some concerns remain on the issue of spatial perception in VEs: indeed, evidence exists about an impaired estimation of spatial relations in stereoscopic VEs (Plumert et al., 2005). This might have some effect on the applications for unilateral spatial neglect, and requires further investigation.

Other applications for visual-spatial skills were for the neuropsychological assessment of spatial perception in patients with Parkinson's disease and closed head injuries (Stirk and Foreman, 2005), and for training navigation and spatial orientation in patients with brain damage, and in the elderly population (Morganti et al., 2007, 2009). The assessment of spatial abilities in patients with Parkinson's disease involved mental rotation and manipulation of objects seen on a movable virtual tray, and the memory and adoption of different viewpoints with respect to a target object in a virtual golf playground where the patient can move about. The tool for assessing topographic disorientation is instead based on a virtual maze, and a virtual road through which participants have to move in order to reach a specific location, which was previously located in a paper-version of the test. All these applications were successful as correlating with traditional psychometric assessment, adding more capability in tracking patients' behaviours, and in doing it in more situated environments compared to mainstream test batteries. Also for patients with early Alzheimer's disease, or ageing people with a mild cognitive impairment, Cushman et al. (2008) proposed a VR-based test for assessing navigational skills. The VE was a detailed 3-D representation of the hospital lobby, and was presented on a laptop PC. Patients could navigate in the VE, and several navigational subsets were administered to them (e.g., route learning, free recall, route drawing, self-orientation, landmark recall, etc.). The test using the VE proved to yield close correlations with an equivalent real-world navigational test, although with a trend for lower scores.

Body schema refers to the sensorimotor representation of own body, and is often distinguished from body image, which is a body representation more dealt with visual-spatial and semantic type of body awareness (de Vignemont, 2010). Both these types of body representations may be disrupted by psychological or neurological disorders. In some taxonomies body representation impairments are classified under visual-spatial deficits, in other under the more general heading of neuropsychological disorders. Here, for the sake of unity, VEs applications to patients with disruptions of body awareness are all mentioned in the section about applications to psychological and psychiatric disorders (section 2.5).

Examples of Applications to Language Disorders

Two main kinds of language disorders have seen the application of a VE: aphasia, and deficit of productive language and communication skills, especially in children. For aphasic patients Geiszt et al. (2006) developed a virtual therapy room, presented to the patient via an head-mounted display. There are a single-patient, and a group session version. In the single-patient session the patient interacts with the therapist in the VE, while in the group session there is a virtual community of avatars, and a virtual teacher, who comment on the patient's performance, and in turn may be called to respond themselves to questions. The questions, taken from standard language rehabilitation instruments, or based on environments created purposely by the therapist, appear on a virtual blackboard. The patient normally views the VE in a first-person perspective, and the virtual group is sat around a virtual table. However, in order to enhance motivational effect, the point of view of the patient may be shifted in a third-person perspective after every correct response, so as to let the patient see herself/himself while providing the correct response.

Especially for children suffering from autism spectrum disorders, VEs may be useful to train social communication skills, for example by creating interactive narratives with virtual characters that respond to children gestures. This is believed to facilitate therapy for all those children who have poor productive language and communication skills. One such VE has been used also to analyse the virtual character gestures and gaze-behaviours necessary for completing a joint attention task (Alcorn et al., 2011), that is a task in the VE targeting joint attention, a crucial skill to be trained in autistic children, as it induces many benefits, also on language performance. Joint attention is accomplished when the gaze or gesture of an individual can be followed by the patient, who is able, when appropriate, also to respond. In the VE a virtual character tries to attract the attention of the child to some objects by using either the gaze or a pointing gesture. Furthermore, the environment may be created starting from gestures of, or elements chosen by, the same child.

2.5. Applications to Psychotherapy and Psychiatry

As depicted in Figure 2.4, the applications of VEs to clinical psychology and psychiatry cover a wide range of disorders, and are more oriented to therapy than assessment. The bulk of the applications are for the Virtual Reality Exposure Therapy (VRET), which is a VR-based version of exposure therapy. This therapy is chiefly adopted for treating anxiety disorders, like specific phobias (as acrophobia, spider and cockroach phobia, fear of flying), social anxiety disorder, and post-traumatic stress disorder (PTSD). Applications to body image disturbances are less widespread, and mainly address patients with eating disorders. Pain control therapies postulate that transportation in a distracting VE may help to control pain, and this idea has been used primarily with burn victims. A more recent application of VEs for analgesia exploits the possibility to trigger the illusion of ownership for a virtual body part, and preliminary studies investigate the instantiation of this concept in amputees to reduce phantom limb pain. This application attempts to achieve this goal by affecting the body schema of the patient, but is aimed at controlling pain, so we refer it in the examples of applications to pain control. Autism spectrum disorders and Schizophrenia may benefit from social VEs, which foster attention to virtual partners and social communication skills in autistic children, while assessing and training schizophrenic patients about how hallucinations are dependent on the social environment. The application to Obsessive-compulsive disorders is quite limited at present, and is oriented to both assessment and therapy.

In the following some examples are reported to illustrate how VEs are applied to mental disorders.

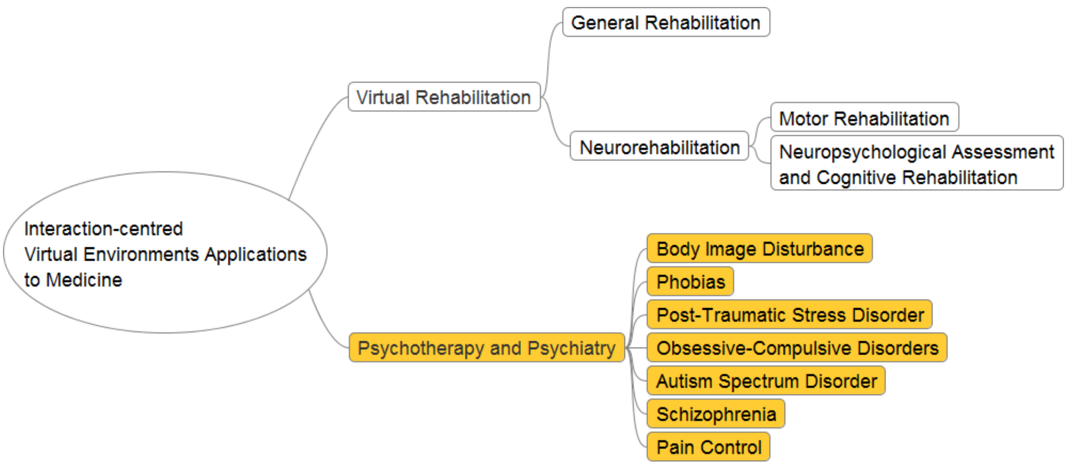


Figure 2.4.: Mental-disorder-based categorisation of Virtual Environments applications to psychotherapy and psychiatry.

Examples of Applications to Body Image Disturbances

The typical application of a VE for the assessment and treatment of body image disturbances is to eating disorders. Body image distortion and dissatisfaction are addressed by VE-based treatments, as a support to cognitive-behavioural therapy. In a literature review Ferrer-García and Gutiérrez-Maldonado (2012) discussed how VEs are useful to study, assess, and treat body image disturbances in patients with eating disorders, because VR can represent a subjective concept as body image, and can provide patients objective, solid, and therapist-independent information about their erroneous mental representation of body image. In this line, the VEs used to assess and treat body image disturbance usually consist of multiple virtual rooms where patients are required to adjust their virtual counterpart body and body parts size according to their perception of own body image in the assessment phase, and to help to modify distorted thoughts about body size and shape. For example, in the assessment phase it is evaluated how the body image of the patient changes in terms of distortion and dissatisfaction before and after either having had a low-calorie or high-calorie virtual meal in a virtual kitchen. In the therapeutic phase the VE is exploited to facilitate the emergence of discrepancies between the biased representation of body image, and the proprioceptive signals deriving from the interaction with the VE. The initial studies of VEs applications to eating disorders date from more than 10 years ago, and positive evidence has been collected on their potential as a tool assessment and treatment of body image disturbances associated to eating disorders. However, it has been underscored the need for further research, and for more controlled studies with larger samples Ferrer-García and Gutiérrez-Maldonado (2012), in order to stabilise the potential and the clinical use of this tool.

Examples of Applications to Phobias and Post-Traumatic Stress Disorder

These disorders are considered together here because the related applications of VEs are all based on Virtual Reality Exposure Therapy (VRET). Exposure therapy consists of systematic graded exposure to anxiety-provoking stimuli in order to induce desensitisation to the phobogenic stimuli, reducing the fear, and the consequent symptoms (Foa and Kozak, 1986). The traditional exposure therapy is *in vivo*, as the anxiety-provoking stimulus is physically delivered to the patient. Clearly, this is not always possible, for inherent and cost-effectiveness reasons, and the patient herself/himself may not accept this form of exposure. One option is then to use *imaginal* exposure, with the patient instructed to imagine anxiety-provoking situations. This alternative has its own issues: it is difficult to control what exactly the patient is imagining, and for some patients it is difficult even to imagine the stressful situations they normally avoid. VEs may be an option for overcoming these issues, and have a more natural and controlled exposure to the feared stimuli than imaginal therapy, while delivering the treatment in a protected, and cost-effective environment.

Initial studies of VRET for specific phobias dates back to two decades ago, and several specific phobias, as well as more general anxiety disorders have seen VRET applications (see Gregg and Tarrier, 2007 for a review). For example, Krijn et al. (2007) compared VRET, cognitive behavioural therapy (CBT), and bibliotherapy with no contact with the therapist for patients suffering from fear of flying. The VRET was delivered in 4 1-hour weekly sessions using a head-mounted display. The VEs for exposure to fear of lying were two: a virtual airport where patients could freely move in a 1 m^2area, and a seat inside an aircraft, where they experienced vibrations typical of take-off, turbulence, and landing. As usual in exposure therapy, the exposure to the anxiety-provoking situations was graded, starting from the less feared, based on patient's self-rating of any proposed situation in terms of a 0–10 scale of subjective units of distress. Results shown that bibliotherapy alone was not effective, and that neither the sole VRET nor the sole individual CBT were able to induce a sufficient decrease in fear of flying. However, group CBT, and VRET were able to reduce negative cognition on flying. This is in favour of the view that VRET may be an important adjunct to CBT, especially for those cases when *in vivo* exposure is not feasible, but it is not yet as effective as necessary for replacing mainstream CBT. In the same study by Krijn et al. (2007), also underlying phobias were treated using VRET. They were claustrophobia and acrophobia. The first was treated in 3 VEs: two elevators (a bigger and a smaller one), a hallway becoming narrower as the patient walked towards its end, and a closet room which could be made as narrow as a 1 m^2. The VEs used for exposure acrophobia-related anxiety were a fire escape with 6 floors in open space, a garden on the roof of a building, and virtual building with 8 floors. Both claustrophobia and acrophobia resulted comparably effective with CBT, and capable of reducing anxiety. In other studies VRET effectiveness on fear of flying was more substantial than in this study. This is the case for example of the VRET for fear of flying presented in Rothbaum et al. (2006). In this article a controlled comparison between VRET and standard *in vivo* exposure therapy is presented. The results support a comparable effectiveness of VRET and standard *in vivo* exposure, also at 6-month and 12-month follow up. A reasonable explanation may lies in the difference between the VEs used, which can affect the capability of the VE to actually elicit a real-like anxiety. In fact, *a posteriori*, a number of patients included in the study of Krijn et al. (2007) were not sufficiently feared in the VE.

Also for less specific anxiety disorders like social anxiety disorder (SAD) VEs proved to be effective in reducing the scores in the respective anxiety scales, and in some case studies also reducing avoidance of the feared situations at a 3-month follow-up (Anderson et al., 2005). When a social anxiety component has to be treated, the VEs need to include social stimuli, and these are essentially obtained by using virtual characters who display, mainly through facial expressions and gestures, different attitudes that may provoke anxiety in a viewing patient, to different degrees. Robillard et al. (2010) carried out a randomised controlled study on generalised SAD patients comparing the effect of CBT with *in vivo* exposure, and the effect of

VRET. Both treatments were superior to the no-treatment condition on the waiting-list group, and effective at reducing the score of anxiety according to the Liebowitz Social Anxiety Scale and on other measure variables. The authors claim this is a first step to determine whether CBT combined with *in vivo* exposure, and CBT combined with VRET, can be considered a substantially equivalent treatment for SAD. Different VEs were used to present typical anxiety-provoking social situations: a virtual bar, where the patient has to greet and take a seat at a virtual friends' table, a virtual dinner in a friends' home, a presentation in a virtual job meeting, and an assertiveness requirement with a virtual shop assistant. These applications of VEs are particularly interesting *per se*, because they involve human-like virtual characters (commonly referred to as *avatars*), which bring both technical and behavioural challenges. They have been used also for the treatment of fear of public speaking, where typically VEs represent a meeting room or a virtual auditorium, including small groups of attending avatars, or a larger conference audience (see Vanni et al. (ress) for a review). The patient has to read a brief text, or introduce her-/himself in front of the virtual audience, which shows different positive, encouraging, or negative, disagreement-like facial expressions, gestures, and comments. The successive sessions of VRET are again graded starting from less fearful situations, and they move to the more feared when the patient's anxiety score in the previous exposure has reduced. It has been shown that a relatively low visual fidelity is necessary for eliciting the anxiety typical of fear of public speaking (Slater et al., 1999), whereas it is more important that the virtual characters are not static, and display facial expressions and seemingly autonomous behaviours able to trigger relatively marked positive or negative attitudes. The evidence cumulated for these applications of VEs are positive, in the sense that the virtual social scenes were able to elicit anxiety similar to what happens in real *in vivo* situations. Moreover, the scores in scales as the Personal Report of Confidence as a Speaker (PRCS) were generally improved after VRET. The limitations of the current available studies in this field are about the patients population (they are mainly University students with low scores in PRCS rather than patients with severe fear of public speaking or generalised SAD), and the lack of controlled studies comparing on large clinical samples standard CBT therapy and pharmacotherapy, with the outcome of VRET.

Post-Traumatic Stress Disorder (PTSD) is a disabling anxiety disorder, which may result from the exposure to trauma, especially involving an actual or threatened injury to themselves or to others. Consequently, combat-related PTSD exists among soldiers who took part in wars, as well as PTSD may affect victims of a terrorist attack, or of a natural catastrophe, like a particularly destructive earthquake. Imaginal exposure therapy and CBT with an exposure therapy component are therapeutic options effectively used for PTSD patients (Rizzo et al., 2009). Prominent examples of VEs applied as a tool for implementing VRET of PTSD involved the treatment of combat-related PTSD (Rizzo et al., 2009), and of PTSD in the aftermath of the September 11, 2001 terrorist act (Difede et al., 2007). The VE for combat-related PTSD presented by Rizzo et al. (2009) is a *Virtual Iraq* scenario, where desert roads, a city and marketplace, and a humvee patrolling the war area can be presented. The patient — a soldier diagnosed with combat-related PTSD — wears a head-mounted display helmet, and can travel on the humvee as a single vehicle, or in a convoy, and virtual soldiers and pedestrians can be present around. 3-D directional sound, vibrotactile, and olfactory stimuli may be added, as well as other specific virtual objects, in order to resemble the characteristic of the setting and events related to the trauma, and to modulate patient's anxiety in a way functional to the therapy. Preliminary results are available from an open clinical trial, and case reports using the *Virtual Iraq* environment, but they are not sufficient to draw general conclusions. However, most of the participants who completed the treatment (2 weekly sessions for 5 weeks, about 1.5–2 hours per session) showed a significant reduction of the symptoms severity based on the PTSD check-list, also at a 3-month follow-up. The benefit of a user-centred design of the VE, based on the feedback of the patients about the resemblance of the virtual scene with a realistic situation related to their actual traumatic setting came up from the studies conducted by Rizzo et al. (2009). Difede et al. (2007) presented a study comparing the effect of a VRET integrated in a CBT comprising cognitive restructuring, relaxation training, and psycho-education, to a waiting-list group. Participants were patients diagnosed with a severe PTSD following exposure to the World Trade Center attack as fire-fighters, or civilian personnel. The VE presented the World Trade Center area, showing streets, buildings, and the sky. The patients wore a head-mounted display, and could walk and look around, with the virtual scenery accordingly updated in real time. Different 3-D sequences were used for graded exposure: a jet flies over the towers, it may or may not crash them, the sound of the explosion may be either presented or not, streets sounds may be altered with screams, the towers may or may not collapse. It is the therapist that decides which sequences are delivered to the patient. The results on the group of 10 patients found a significant improvement of the Clinician-Administered PTSD scale compared to the waiting-list group, also for patients who had undertaken with no improvements imaginal exposure therapy prior to the VRET. also at 6-month follow-up 9 patients out of 10 had maintained their improvements. Difede et al. (2007) remark that an important evidence of this study, with its inherent limitations as it is carried out on a restricted group, and comparing with a waiting-list group rather than a group undertaking standard therapy, is the effectiveness of a VE that presented the same situations, albeit graded, to patients who came from quite different exposures to the traumatic event.

Examples of Applications to Obsessive-Compulsive Disorder

Obsessive-compulsive disorder (OCD) is characterised by the occurrence of obsessions and/or compulsive rituals, and the empirically-based psychotherapy is CBT involving exposure and response prevention (Abramowitz et al., 2009). In this case exposure is based on confronting with stimuli provoking the compulsive rituals and anxiety. In addition, response prevention is intended to refrain from performing the rituals. The application of VEs for OCD is still minimal. Kim et al. (2008) used an anxiety-provoking VE on patients diagnosed with OCD, and compared the outcome of exposure to the VE with a group of healthy participants. The VE was a typical home, with several rooms the patient could move about by using a joystick, and wearing a head-mounted display. Participants were initially instructed to familiarise with the VE, performing actions like opening the door, turning the gas valve on or off, and switching the light. Then they had to take away specific objects. Finally, they were allowed to freely check (the gas valve, the light switches, a water supply), a behaviour that is typical for OCD patients. The results showed a longer checking time than the control group, higher anxiety scores, but also a more rapid decrease of anxiety in the VE. According to Kim et al. (2008) these results are a clear indication that appropriately designed VEs have the capability to trigger anxiety and typical behaviours in OCD patients, as required for their use as a tool for exposure and response prevention. Kim et al. (2010) then extended their previous work by adding an office-like VE, and proposed to use indices based on frequencies and durations of checking behaviour in the VE as a novel behavioural measure for the compulsive checking behaviour of patients diagnosed with OCD. The results were generally positive, as the indexes calculated as a result of the task in the VE positively correlated with those based on self-reports and standard clinical measures. Although the same authors underscored some discrepancies emerged from the results, and remarked the need for further studies that corroborate this initial evidence, this study demonstrates that VEs are a promising tool also for psychometric assessment of OCD, as VR is able to deliver situations that elicit behaviours as-if-real in the patients.

Examples of Application to Autism Spectrum Disorders and Schizophrenia

As already pointed out in a previous paragraph about VEs application for language disorders (section 2.4), children on the autism spectrum may benefit from VEs. Two main kinds of use of the VE showed benefits for autistic patients, both exploiting the potential of VEs as an enhanced learning environment: as a tool for enhancing social skills, and as a tool for more general habilitation in autism (Standen and Brown, 2005; Bellani et al., 2011 for a review). For example, Mitchell et al. (2007) showed that a group of adolescents diagnosed with autism spectrum disorder improved in judgements and reasoning about whether and where to sit in videos of real café and bus, following practise in a VE presenting them with a virtual café, filled with virtual people. Participants had to sit down with their food tray either at a free table, if available, or ask for permission to sit at a table already partially occupied, if no free table was available, in a virtual café. Moore et al. (2005) used a collaborative VE to evaluate the ability of autistic children to operate avatars and recognise emotions presented by the avatars facial expression, and the virtual context. Participants were successful at both operating the avatars, recognising the emotions expressed by them, and identifying the appropriate context for those emotions. Cheng and Ye (2010) used another collaborative VE, where autistic children participated in a virtual classroom with their avatar, and the avatar of a teacher. Social competence learning was evaluated in terms of the ability to listen to others, eye-contact, and understanding of others' expressions and behaviours. The results were encouraging, with performance in such abilities improved. Overall, the available evidence is positive for the use of VEs with autism spectrum disorders, and a key advantage seems to be also the high acceptance of VR-based interaction of autistic patients, who also might enable their social interaction following its deployment through avatars, and in a protected environment as the VE is, with respect to the physical world. This consideration, however, again leads to the notable fact that studies about the generalisability of findings collected within the VE are fundamental.

As for application of VEs to schizophrenia, they are relatively more recent, and mainly address the social interaction difficulties associated to this psychiatric disorder. For instance, Park et al. (2011) used a VE for a social skills training based on role-play. Patients diagnosed with schizophrenia trained their conversational, assertive, and emotional expression skills through role-playing either in a real environment, or a VE. Only the use of real actors and videos differentiated the social training intervention in the real environment with respect to the VE. In the latter, the social interaction and role play occurred through avatars. The participants in the VR-based training reported higher motivation for the treatment, while overall social skills were similar after the standard and the virtual interventions. Interestingly, the participants in the VE-based role playing for social training showed improved conversational and assertiveness skills over those who took the standard role-play training. In contrast, the latter resulted better for training non-verbal skills. Freeman (2008) highlighted also different purposes for VE applications to schizophrenia. Social perception of others has been assessed on schizophrenic patients with a VE by Kim et al. (2006). They were successful at finding the social impairments present in these patients in terms of interpretation of relevant cues in the VE, and emotion recognition of avatars. Concerning the development VR-based treatments for schizophrenia, their potential is discussed by Freeman (2008), but their actual implementation and trial are still lacking. Freeman (2008) suggest three possible ways of employing VEs for developing treatments for schizophrenia: how symptoms are modulated by specific factors, e.g., by simulated hallucinations in a VE after specific emotional manipulations; a kind of VRET to persecutory fears; and training about how to maintain engagement in a social situation, when the symptoms arise.

Examples of Applications to Pain Control

Applications of VEs to analgesia are based on the idea that the multimodal stimulation provided by an immersive VE is able to alter the perceptual representation of one's body, thus shifting attention away from the pain perception, and induce an analgesic effect. Furthermore, a higher sense of presence in the VE (Malloy and Milling, 2010), and a more active participation in it, i.e., maximising interactivity with the VE Wender et al. (2009), proved to be factors influencing the pain relieving effect of the virtual experience. The most studied VE application for pain attenuation is to burn wound care. Hoffman et al. (2004) presented a water-friendly VR system capable of delivering an immersive, three-dimensional VE to a patient while undergoing wound care procedures in a hydrotank. The system is water-friendly as electron-based images are converted to photon-based images, and conveyed to a head-mounted display wore by the patient in the hydrotank via optic fibers, preventing from any electrical shock risks. The VE represents an icy three-dimensional canyon with a river and waterfalls, and the patient may shoot snowballs at snowmen, igloos, robots, and penguins in the VE by pointing the gaze and hitting a button on a joystick. Animations and spatial sound effects are associated to the snowballs impact. The study compared a 3-minute wound care procedure without the VE distraction, with a 3-minute procedure while experiencing the VE. It is a single case study: on 0–10 scales, the patient rated a sensory pain decrease from 7 to 2, and an affective pain decrease from 6 to 3 passing from the no-VE condition to the VE condition. Also the amount of time spent thinking of the pain dropped significantly. The sense of presence was rated as 7, qualified as moderate. Hoffman et al. (2008) extended the research with a controlled study, which confirmed the reduction of the worst pain severity, and highlighted a correlation between the sense of presence in the VE, and how much the pain dropped.

A more recent application of VE to pain control capitalises on the plasticity of body schema. In this framework VR seems capable to induce the sense of ownership for a virtual limb (Perez-Marcos et al., 2009), and even to trigger out-of-body experiences (Lenggenhager et al., 2007). In this context, research is probably still too initial for considering clinical applications, however, Murray et al. (2010) presented positive findings in an exploratory study that used VEs to harness visual imagery of upper-limb amputees in order to reduce the phantom limb pain. VEs where the patients can complete tasks with a virtual limb, gaining a more robust illusion of owning and controlling the virtual limb, showed benefits when compared with the mainstream approach of the mirror-box (Ramachandran and Rogers-Ramachandran, 1996), where instead the patient sees the image of the anatomical remaining limb on a mirror in a location compatible with the amputated limb, and is instructed to move the remaining one. The visual feedback actually reduces the phantom limb pain, but the reduction is rapidly lost after removal of this visual feedback. VEs might be useful to alleviate this problem. This may also be considered as an application to pain control.

2.6. Summary Discussion

The applications of Virtual Environments to cognitive science considered in the previous sections focused on the medical domain, for two main reasons. First, we believe the categorisation proposed for medical applications is especially capable of underscoring and widely encompassing all the prerogatives of VEs useful for cognitive science and beyond. Second, the applications to medicine are more numerous compared to the non-medical ones in the field of cognitive neuroscience, which nonetheless are growing in the recent years, because VR systems are becoming compatible with neuroimaging techniques, and because of the advances in brain-computer interfaces that open up new ways of research (see Bohil et al., 2011 for an up-to-date review on applications of VR to neuroscience). Overall, we propose a summary framework for VEs applications for cognitive science as depicted in Figure 2.5.

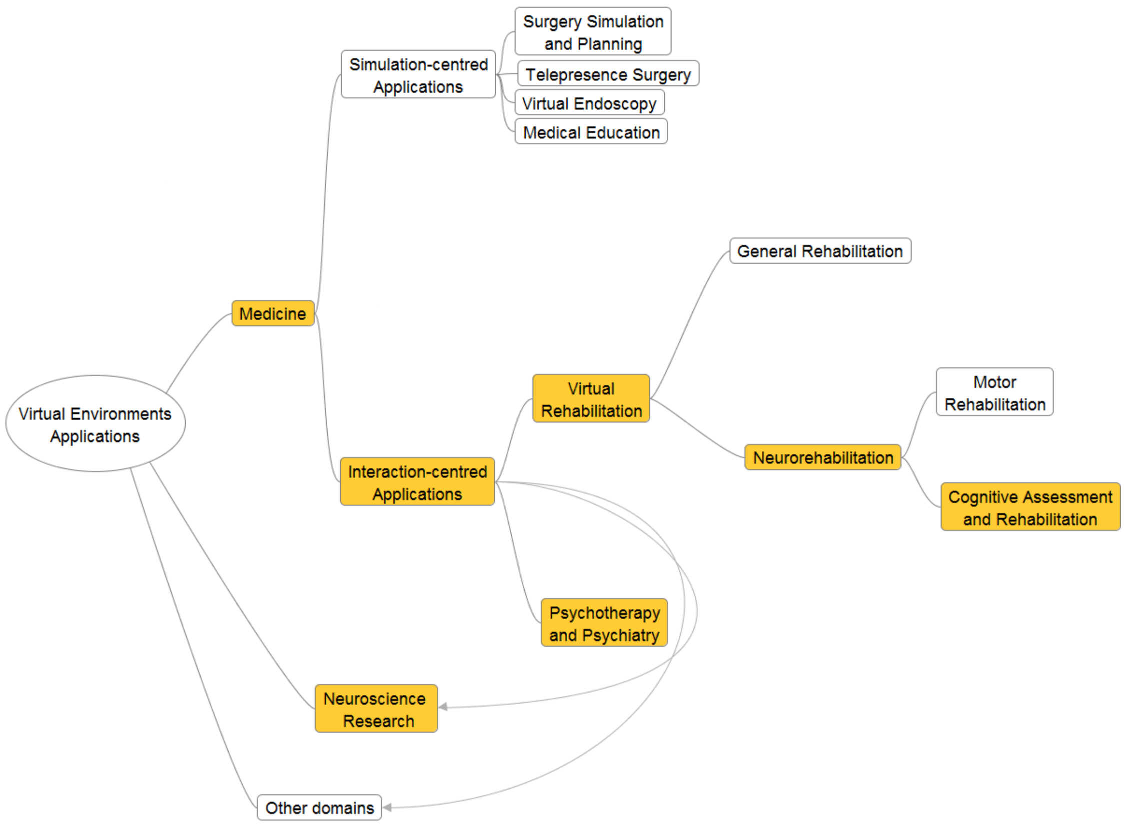


Figure 2.5.: Summary chart to graphically outline the categories of virtual environments applications for cognitive sciences. The background-shaded bubbles in the chart indicate the categories that include applications for cognitive science. The links between "Interaction-centred Applications", "Neuroscience Research", and "Other domains" remark the possibility that interaction-centred applications of virtual environments are not exclusively found in medicine.

Non-medical applications of VR to neuroscience research, which are not considered in detail here, but are included in the framework, focus on spatial cognition and navigation, and, more recently, on the ability of VR to induce body-transfer illusions (Bohil et al., 2011), as briefly mentioned when introducing the most recent applications to pain control.

In cognitive assessment and rehabilitation many applications have been proposed, and, at least to some extent, evaluated. In particular, VEs have been proposed to address impairments of attention, memory, spatial abilities, executive functions, and language. Generally, all the applications yielded positive results when evaluated, demonstrating that the potential of VEs for NA and CR is concrete. In fact, studies on psychometric tests in VE usually correlate with standard tests, and, when the point has been addressed, also with functioning in the real-world. This correspondence is one of the crucial advantages of VE applications in the field. Also VR-based rehabilitation interventions were associated with positive outcomes, although a large number of VEs applications are oriented to assessment. However, some limitations are still present, which require additional research, and also design principles for the VR systems which foster their spreading among clinicians and professional therapists. First, there is a lack of large studies comparing the effect of VEs with those of standard assessment procedures and therapies, and of standardisation of VR-based protocols. This can be achieved over time (often the history of a VE application hardly exceeds a decade), but the process might be hindered. The hindrance is the second limiting factor of VEs applications in this field: the requirement of special competences for handling the VR equipment, both during the intervention itself, and for preparing the virtual tasks. This limitation can be overcome with the adjunct of user-friendly interfaces for the therapists, who can directly and dynamically operate on the objects and the behaviours of a VE at a high abstraction level to adapt the task for the patient, being understood the essential requirement that medical, professional, engineering, and computer science components must cooperate throughout the whole process from the design of the VE application for a well defined target population and intervention, to its implementation, pilot validation, and standard use. A general remark emerging from the discussions of the different VEs applications, is that they are not intended and suitable to replace the current tools. Rather, they deploy themselves as valuable adjuncts which solve limitations of current standard interventions, and offer additional capabilities not conceivable with mainstream paradigms.

Looking at the specific application sub-domains, the area of language disorders seems to have been the less attractive for VEs applications, likely because contexts like executive dysfunctions, and memory impairments more evidently and directly took advantage from the impact of ecologically valid environments. Moreover, the possibility to move within a 3-D VE is a key benefit, which has anyway less impact on language disorders. In the attention sub-domain, the number of applications is not as large as the ubiquitous nature of attention would suggest. Indeed, the major application is the “Virtual Classroom” (Rizzo et al., 2006), which mainly addresses ADHD assessment. Also a VR-based assessment of divided attention during car driving is proposed (Lengenfelder et al., 2002), but no applications specifically dedicated to the other abilities of attention are available. A final remark concerns the applications of VE for assessment and rehabilitation of cognitive impairments associated to prevailing age-related diseases, like Alzheimer and Parkinson’s diseases. Considering the pace at which the prevalence of such diseases is increasing, and the effect this has on the society, it is expected in the near future that a burst of research and translational effort will be put on applications of VEs for the assessment and treatment of these kinds of prevailing age-related impairments.

The applications in psychotherapy seem to be somewhat more advanced — also due to a relatively longer history —, at least for the virtual reality exposure therapy in the treatment of anxiety disorders (specific phobias, social anxiety disorder, and post-traumatic stress disorder). Also these applications are not yet become standard clinical protocols, but a significantly greater number of controlled studies have been carried out to compare them with standard psychotherapies. Furthermore, the psychological connotation of the disorders targeted within this category, in contrast to the neurological prominence of the disorders considered in the NA and CR categories, contributes to boost part of the research related to VEs applications to mental health, because such research lies on the same line of emerging research challenges about more basic neuroscience, which are studying the transformational power of VR, i.e., how VR can affect body perception and representation, the sense of self, and the sense of agency. The ability to leverage this power of VR would constitute an extraordinary breakthrough towards a brand new generation of interventions. A similar argument is valid for applications involving autonomous virtual characters, very useful for social skills training, and exposure therapy to treat social anxiety disorder . Indeed, devising applications of new technologies in order to promote social inclusion, especially of those impaired due to ageing, is a hot topic in the current research agenda of information and communication technologies for health.

Lastly, a kind of application of VEs which did not emerge from the examples presented in the previous sections is the use of serious games for rehabilitation purposes. The issue is worthy the mention, notwithstanding it is not a VE application in a strict sense. Serious gaming applications are the implementation of rehabilitation tasks as game-like tasks, in the attempt to strengthen the patient's motivation, and focus her/his attention on the rehabilitation process, rather than on the pain, also avoiding mind wandering during the exercise. These applications, which may be used for both sensorimotor training and cognitive rehabilitation (Rego et al., 2010), differ from those described previously, because they are computer games, not necessarily VEs, but they use technological components of VR systems. Technologies capturing the patient's movement and gestures are used, like structured 3-D laser scanners and accelerometer-based motion sensing systems used in commercial game consoles. These technologies enable a natural interaction paradigm with the computer presenting the game environment, and allow for the digital capture of patient's performance. Given the motion component usually inherent in these games, most applications are for motor rehabilitation (e.g., balance and upper limb rehabilitation for patients with spinal cord injuries, and post-stroke respectively). For cognitive rehabilitation and psychotherapy, only a few studies have been undertaken, and they offered a complementary therapy of impulse-related disorders (Fernández-Aranda et al., 2012), and a training means for spatial and verbal memory rehabilitation in patients with traumatic brain injuries (Caglio et al., 2009). These 3-D video games may associate sensorimotor and cognitive training in the same environment. Thus, when adopting VR technologies they are transversal within the framework proposed in Figure 2.5, as they may apply to either cognitive or mental disorders, and may even encompass both sensorimotor and cognitive training in the same application.

2.7. Motivation for Virtual Environments for Studying Attention

All the cognitive domains considered in the proposed categorisation (see Figure 2.3) are important for human adaptive functioning in daily living. However, compared to the other domains attention is rather special, as it is a multidimensional construct, which affects multiple perceptual and cognitive operations. Indeed, attention selects incoming information and modulates its processing, so as to privilege relevant information, and block out the *noise*, which is determined dynamically, according to the current goals of the individual, and to the state of the environment. This influences also other cognitive domains: for example visual attention is necessary for an appropriate spatial behaviour, and selective attention biases what is stored in memory. The different attentional abilities referred to in section 2.4 are the macroscopic evidence of this process, and they are often assessed in task contexts which are abstract, making it difficult to motivate the patient, and to yield results generalisable to the daily living functioning. As discussed above, this problem can be partly tackled by leveraging on the VEs prerogatives, but a common characteristic shared among non-medical and medical applications of VEs to cognitive science is that they both somewhat neglect the study of attention. Some attention abilities are assessed in children suffering from ADHD(Rizzo et al., 2006) and in patients with TBI (Lengenfelder et al., 2002), while VEs applications to basic neuroscience do not consider attention as a primary focus.

Additionally, the studies of applications found in the scientific literature generally evaluate attention at a macroscopic time scale. The time course of attention is chiefly considered for assessing vigilance, i.e., over an extended period of time — several seconds, or minutes — while the basic time course of attention influencing which stimuli are brought to an individual's awareness, and which are not, operates on the time scale of a few hundred milliseconds, or less. At the best of our knowledge these temporal aspects have not yet been studied within a virtual environment, leaving two questions open. Firstly, whether the time course of selective attention in a VE is similar to that in a real environment, or to that expressed when interacting with a computer, like in a video game. Secondly, whether stimuli presented in a VE, which may be more natural *per se* and included in a meaningful context, are able to elicit an attentional response different than the simpler, abstract stimuli commonly used in laboratory tests. The answers to these two questions are useful also in view of possible applications to assess the time course of selective attention in a neuropsychological assessment perspective, as well as in a perspective assessing attention biases in mental disorders. Indeed, biases of the time course of attention have been postulated, and partially demonstrated for emotional processing in general, and also studied in relation to anxiety disorders (de Jong et al., 2010, 2009). Therefore, before thinking about the potential incorporation of assessment tools of the time course of selective attention in a VE application, it is necessary to investigate the more basic research questions formulated above.

In the present thesis the two questions are instantiated in two experiments. In both instantiations the empirical probing task is the typical experimental paradigm used for studying the time course of selective attention, that is the attentional blink. This is an attentional deficit that may arise when an individual is requested to report two targets appearing in a rapid serial presentation of distractors, and the two targets onset asynchrony is shorter than 400–500 ms. A similar experimental paradigm has been used in two empirical studies reported in the following chapters (chapter 4 and chapter 5). In the first study the time course of attention has been assessed with simple, abstract stimuli, in the real environment. In this case the capability of a relatively short and mild integrative body-mind training to make more efficient the temporal distribution of attentional resources has been investigated. The modified attention distribution is reflected by a quicker time course of attention processing. In the second study, the two research questions mentioned above in relation to VEs are considered more directly. In fact, the attentional blink paradigm is explored in the VE compared to its classical administration on a computer screen, and the visual stimuli used are less abstract and meaningfully related to the used VE.

Chapter 3 presents the theoretical basis about attention, and the attentional blink phenomenon, while chapter 4 and chapter 5 introduce the specific rationale for, describe, and discuss the two empirical studies undertaken.

3. The Study of Temporal Attention

This chapter introduces the *cognitive component* of the thesis. Indeed, the fundamentals about human attention and its temporal aspects, with a special emphasis on the attentional blink phenomenon are presented. These aspects are in fact explored in the empirical case studies described and discussed in next chapters. Several taxonomies are possible for attention, reflecting its multifarious nature as well as the multiplicity of processes and behaviours it prominently influences. Attention has been widely studied in virtually all of these aspects, but its time course has been addressed more recently. Rapid Serial Visual Presentations (RSVP) are the typical experimental paradigms used to tap temporal attention, and the attentional blink (AB) is probably the most studied phenomenon arising from dual-target RSVP. In the following sections, after the provision of background information about attention and its major characteristics and properties, the AB is introduced with the associated empirical findings. Formal and informal models of AB are summarised, and some open issues are proposed that will serve as the rationale for the empirical case studies presented in next chapters.

3.1. Human Attention

The word *attention* is a very common one, widely used in diverse contexts. However, the specific meaning it takes is not always the same, although a shared idea underpins all such meanings.

For example, the first meaning reported by the online version of the Oxford Dictionaries Online (2012) is: “notice taken of someone or something; the regarding of someone or something as interesting or important”. The first meaning indicated by the Cambridge Dictionaries Online (2012) is even more general: “notice, thought or interest”, but it explicitly introduces a *mental* connotation of attention. This connotation is still more clear in the definitions of attention found on the Merriam-Webster Dictionaries Online (2012): “the act or state of applying the mind to something” and “a condition of readiness for such attention involving especially a selective narrowing or focusing of consciousness and receptivity”. Another interesting observation is that often attention is defined in reference to an action. In this respect, the other definitions of attention from the Cambridge Dictionaries Online (2012) report of “catch somebody’s attention”, “pay attention to something”, “turn your attention to something” and “the centre of attention”.

These definitions, albeit not so specific, still provide important distinctive elements of attention. First, rather than defining what attention *is*, they suggest what attention *does*. Indeed, according to all the definitions, attention is some unspecified act, state or condition that notices, regards as interesting or thinks of something or somebody. Second, attention is connected to the human mind: it works by applying the *mind*. Third, attention can be *directed* to a centre of attention, as attention can be caught, payed and turned to. Lastly, attention is also connected to readiness, and, especially, to *selection* and *focus* of consciousness and receptivity.

These distinctive elements actually reflect important characteristics of human attention, and show, as already pointed out in 1890 by James (1890) — and quoted in almost every textbook including a chapter about attention — that "everyone knows what attention is.". Indeed, everyone knows what attention does, at least at an overt macroscopic level: attention refers to the ability to attend to stimuli, but this is only part of the story about attention; and the wide use and intuitive, almost innate, nature of the concept of attention has contributed to a twofold situation. On one side the amount of scientific research works conducted on attention over the last century and more is huge, addressing many different aspects of attention; on the other side, attention is often treated as a catch-all term, lacking the details of the different mechanisms and functions that attention moderates or contribute to. Next sections of this chapter cannot of course be exhaustive about attention. Rather, they attempt to provide an overview about its characteristics and functions that should prove useful to understand the rationale and results of the empirical case studies presented in chapter 4 and chapter 5.

3.2. Attention in Human Information Processing

Attention has a crucial role in performance. In fact it is through attention that we are able to achieve our goals, essentially because we are able to maintain a coherence between our perceptions and our actions in view of our intentions and goals.

Several accounts have been proposed to explain attention and the functions it serves in performance, but it is beyond the scope of this chapter to present an extensive compendium of these theories and their differences. Instead, it is possible and productive for this work to outline the basic characteristics of attention emerged from different research works. These basic characteristics can be considered as the fundamentals of human attention.

A first step to understand the role of attention in performance, and also what are some related research questions, is to consider how information is processed by the human mind. A basic information processing model is shown in Figure 3.1.

This three-stage model is an adaptation of what proposed by Proctor and Van Zandt (2008) and Johnson and Proctor (2004). Their accounts are not the uniqe general frameworks proposed for describing the flow of information in the human from the

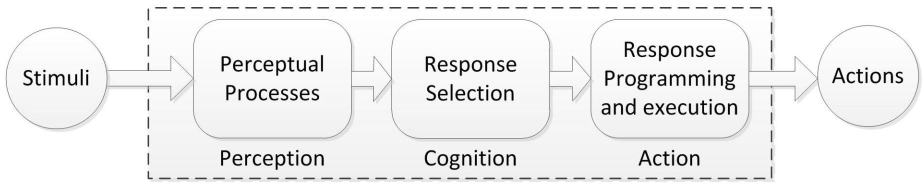


Figure 3.1.: A basic model of the human information processing. The major stages are reported: usually in this kind of models each stage represents a set of processes or functions.

perception of an environmental stimulus up to the execution of a response. Several models have been discussed, also for specific sources of information, e.g., visual or auditory information, and each stage in the model need to be further developed to a greater detail for in-depth analysis and understanding of information processing. However, the three-stage model allows to catch the fundamental elements of human information processing, to highlight the different roles of perception and cognition — although the boundaries between the two are quite elusive — and the role of selective attention.

Each stage of the nodel indicates a major function or process accomplished on the information coming from various sources. Actually, as already mentioned, the processes are composed of sub-processes, but they are grouped in a single stage because they can be mapped onto the three general domains of perception, cognition and action.

In the first stage perceptual processing occurs: the stimuli coming from the environment reach with their physical properties the sensory systems. The different physical properties elicit responses from different sensory modalities (e.g., visual, aural, somatosensory), therefore the information associated to a modality uses an input channel to the brain that is dedicated to that specific modality. Clearly, meaningful stimulations are often multimodal, i.e., they are composed of stimuli coming from sources with physical properties characteristic of different sensory modalities: the information processing must be able to “recompose” the component stimuli, which are unimodal, and interpret the resulting pattern to execute in the end an appropriate response, i.e., to show an appropriate behaviour, part of a consistent performance. This is achieved in the subsequent stages. Perceptual processing is the first step towards the “recomposition”, and its goal is to detect, discriminate, and identify stimuli. This stage is mainly a *bottom-up* processing, driven by the sensory inputs.

The *top-down* processing intervenes in the second stage, the *cognitive* stage, that resorts to comparisons, arithmetic operations and decision-making in order to determine the appropriate response to the stimulation. Important characteristics of this stage are the potential recourse to knowledge stored in memory, and the comparison of the stimuli among them or with the retrieved memories. The appropriate response is selected among different alternatives according to the circumstances.

The final stage is aimed at executing, if necessary, an overt response. In some accounts the actual response selection is shifted from the cognitive stage to this stage. Then the response is programmed, i.e., the movement is planned, and translated into neuromuscular commands used to control the effectors, e.g. the hand or limbs, which execute the response.

It is worthy noticing that there is no direct matching between a single stage in the information processing model, and a single structure in the brain: multiple structures contribute to the three processing functions, and there is no exact mapping of the model functions in the cortical topography.

The role of attention in the framework of human information processing is to control the flow of information. Such a control can be exerted at different levels and with different mechanisms. This general role of attention is also a reason for the impossibility to isolate attention as a single process, because it encompasses both the perceptual and cognitive domains. Attention cannot be regarded as a unitary construct. An interesting characterisation is proposed from Chun et al. (2010), who consider attention as a fundamental property of multiple perceptual and cognitive operations.

From a general point of view, it should be considered that every stage of the information processing can be affected by different types of limitations, and that there is an overall limited capacity to the amount of information that can be processed moment-to-moment. At any moment in our life, we are bombarded with an uncountable bunch of stimuli coming from the environment and stimulating our sensory systems. In fact, our behaviours are appropriate only if we can efficiently cope with all this information. Since our information processing resources are limited, some sort of filtering is then indefeasible. Attention just comes at this point, i.e., to enable a selection process of the incoming information. This fundamental function of attention is referred to as *selective attention* (Posner and Petersen, 1990; Desimone and Duncan, 1995). The limited capacity at processing the information relevant for achieving intended goals has been postulated as the necessity out of which attention evolved (Pashler et al., 2001).

Another function of attention is *modulation*. The difference between selection and modulation is that attention biases the competition among competing options in favour of one of them through selection; then it influences the processing of the selected item. The influence exerted once selection is accomplished is modulation, which cannot occur without selection (Levin and Simons, 1997). Attention with modulation determines the extent to which information is processed, but also speed and accuracy of a response, and if the associated event will be remembered. Therefore attention affects various stages of the information processing. Vigilance, i.e. attention sustained over time, which is very important for many everyday tasks, is related to the modulatory effects of attention: modulation is the immediate effect of attention on selected stimuli and is necessary to sustain attention over extended periods of time.

As to the relationship of attention with information processing, a still controversial issue is where does selection occur in the flow. Two views are possible: an early-selection view and a late-selection view. The former considers the locus of selection at the perceptual stage, in a way that some stimuli are almost completely ignored, without undergoing any semantic processing. In contrast, in the late-selection view information is selected after the perceptual stage, when some dgree of semantic processing has been carried out. Therefore, the question is about the processing of unattended stimuli. A recent approach to this problem proposed by Lavie et al. (2004) considers that the two views are not mutually exclusive, rather, they might be both applicable depending on how difficult is to process the target stimuli. Specifically, if the attended target is easy, i.e., it has a low perceptual load, then there is an excess of attentional resources that go to the processing of the unattended stimuli, entailing a late selection. On the contrary, when the attended target (possibly a task) is difficult, it draws all attention, and distractors go less processed, as in early-selection views.

Perceptual load is thus a determinant of task difficulty. Typically, perceptual load is manipulated in the visual modality, and it includes the number of competing items in a display and the similarity between targets and non-targets (Torralbo and Beck, 2008).

3.3. Attention Metaphors

Due to the lacking of a simple defintion of what attention is, and in view of the importance of attention in human performance, also some useful metaphors of attention have been used in the history of attention studies to operationally deal with attentional experiments.

The Spotlight of Attention This is a metaphor that could be also named the focus of attention, and reminds to the dictionary definition mentioning "the centre of attention". This metaphor has an inherent spatial connotation, but it can be enunciated in general terms. Attention selects the information that is relevant for the task at hand, and modulates its processing. Within the spotlight of attention the processing is more efficient: the reaction times are faster, the features composing the stimuli are bounded in a single object that can be recognised as a whole, and discrimination is firmer (Shipp, 2004). The *focus* of attention identifies the core of the spotlight, and it reminds of another property of attention according to the dictionary definitions: the possibility to direct attention. The focus of attention, that can be thought of as anchored to the spotlight, can be directed to different sources of information or stimuli.

Spatial attention is the term used to indicate the property of attention to be allocated to different locations across space. For example, one can visually track a moving object: in this case the spotlgiht, or the focus of attention is moved across space, and the stimuli occurring outside such a spotlight are unattended — unless they are particularly salient. When spatial attention is guided by eye movements it is said to be *overt*, whereas it is *covert* when a location is attended without acccompanying eye movements.

A metaphor similar to the spotlight of attention is that of a gradient of attention in space: there is still a focus, but as the distance from the focus increases, the perceptual efficiency decreases.

Attentional Resources This metaphor refers to a limited resource model of attention, which can be allocated to the different tasks under the control of an executive system (Fernandez-Duque and Johnson, 2002). Attention is considered as a resource available up to a limited amount. This amount has to be shared among the multiple tasks to be accomplished in parallel. When the requirements of attentional resources exceed the limit, performance degrades, and prioritisation of tasks execution may be activated (McDowd, 2007). This metaphor originally foresaw a single pool of attentional resources (Kahneman, 1973), and it was useful to explain dual-task interference. However, this view of the metaphor has then been criticised, based on evidence that increasing the difficulty of a primary task, does not necessarily entail a degradation of the performance of a secondary effortful task. Consequently, models foreseeing multiple reservoirs of attentional resources have been proposed (Navon and Gopher, 1979) and (Pashler, 1998), postulating that the interference occurs only when two tasks spill resources from the same reservoir.

The advantage of this limited capacity metaphor over the spotlight one is that the latter does not account for a graded allocation of resources to different tasks, since even the gradient version of the metaphor does not allow for a reduction in the “light intensity” at the primary focus when a secondary task comes at hand.

Attentional Effort Here the effort refers to the experience of “paying more attention” when confronted with a challenging task, also based on the motivation to maintain attentional performance (Sarter et al., 2006; McDowd, 2007).

In general the metaphor of attentional effort is less central in the scientific literature about attention than the capacity-limited-based metaphors described in previous paragraphs. Anyway, there are two conceptions of this attentional metaphor. The more general one regards the attentional effort as a function of task difficulty, and is incorporated in the limited capacity models of attentional processing. In this conception the attentional effort essentially depends on the task demands, and specifically it increases as a function of tasks demands. Another conception proposed in the work by Sarter et al. (2006) posits the attentional effort is a cognitive incentive on attentional performance. Particularly, the attentional effort is “activated” by top-down mechanisms to counteract the decline in performance. The top-down activation depends on the motivational contingencies of the performer.

3.4. Attention Taxonomies

Assuming that attention is a ubiquitous concept as it is crucially involved in virtually all human behaviours, that no unitary construct of attention exists, and that the general role of attention is to control information processing, several taxonomies can be built for attention.

One of these taxonomies, commonly used when dealing with rehabilitation of attention impairments, is based on attentional abilities to be deployed in the context of specific tasks:

- *Arousal* is the level of alertness with respect to response to the environment. There are both a tonic and a phasic arousal: the former is the day-to-day alertness, independent of the immediate task demand, while the latter is the ability to respond to changes in the environment, or in task demand.
- *Focused attetnion* is one's ability to focus on a specific object or stimulus, especially visual, auditory, and tactile stimuli, and discretely respond to them ignoring distractors.
- *Sustained attention* (a.k.a. *Vigilance*) is the ability to sustain a focused attention sate over an extended period of time, e.g., requiring a repetitive response.
- *Divided attention* is the ability to simultaneously attend to, and respond to multiple tasks at a time. It includes also the concept of *alternating attention* as the ability to move between tasks, which have differrent cognitive requirements.
- *Executive attention* includes those attentional functions aimed at controlling one's thoughts and actions to produce a coherent behaviour (e.g., initiating a task, making a decision, etc.). However, when dealing with rehabilitation, executive attention (often also referred to under the heading of executive control, cognitive control, or conflict monitoring) is usually not included in the taxonomy of attention. Rather, it is associated to the executive functions.

Another interesting taxonomy differentiates external attention from internal attention (Chun et al., 2010). This taxonomy is particularly valuable because it incorporates several other distinctions used to classify attention, and highlights its linkage to information processing. Specifically, external attention selects and modulates information coming from the sensory systems, whereas internal attention selects and modulates internally-generated information, such as the content of working memory, or a task set.

Thereby, *exogenous attention*, also referred to as *bottom-up* or *stimulus-driven* attention, which is captured for example by salient, modality-specific stimuli in the environment, falls under external attention. On the other hand, the complementary form of attention, that is *endogenous attention,* or *top-down* attention is goal-directed, and falls under internal attention. In fact, stimulus-driven attention may have as a target spatial locations, time points, sensory modalities, elementary features, and entire objects. Instead, goal-directed attention has as targets cognitive representations, such as the representations in working memory, and long-term memory, the rules, and the reponse selection appropriate for achieving the goals set out for the task at hand. From this view a distinction between *spatial attention* and *temporal attention* emerges, which is of special importance for the empirical work presented in this dissertation.

Spatial attention prioritises spatial locations in the environment, and may be overt, or covert with respect to the use of eye movements to guide attention to attend a spatial location. In contrast, temporal attention deploys over time, by attending objects all appearing at the same spatial location, but at different time points. Both of these attention types have limited capacity: there is a limited number of objects which can be attended across space, and a limited number of objects that can be attended over time. Specifically, this capacity-limited aspect of temporal attention is a key point in the study of the time course of attention, which is less studied than spatial aspects of attention, and is the target of the experiments presented in chapter 4 and chapter 5 of this dissertation. The role of attention in information processing entails that the time course of attention, or rather the limited number of objects that may be attended over time, is a constraint to the rate of information processing. This is linked also to the limited capacity of working memory, so that temporal attention manifestations are explained in frameworks that implicate not only attention, but also perceptual stages of processing, and encoding/consolidation of information in working memory. Next section (section 3.5) delves into this aspect, by presenting the attentional blink phenomenon, widely adopted to tap temporal attention, the associated empirical evidence, and the models proposed to explain it.

A complementary view for a framework of attention is based on functionally and/or anatomically distinct brain networks underpinning attention-related mechanisms. Two frameworks have been proposed, and are reported here, because although less directly linked to temporal attention, they introduce dissociable systems of attention mentioned in scientific works about the relations between attention and meditation (see for example Jha et al., 2007).

One framework is that of Attention Netwroks, proposed by Posner and Petersen (1990). Here there are three brain networks corresponding to three dissociable functions of attenion: alerting, orienting, and executive attention. *Alerting* is the achievement and maintenance of a vigilant and alert state of high sensitivity to incoming stimuli. *Orienting* is the selection of information from sensory input, especially as a result of moving the attention focus across space, and engaging it on targets. *Executive attention* encompasses mechanisms for monitoring and resolution of conflicts among competing options (thoughts, feelings, responses). These networks have correspondent anatomical structures in the brain (Posner and Rothbart, 2007): the locus coeruleus, right frontal and parietal cortices for alerting; superior parietal cortex and temporal parietal junction, frontal eye fields and superior colliculus for orienting; and anterior cingulate, lateral ventral and prefrontal cortices, and basal ganglia for executive attention. Norepinephrine, Acetylcholine, and Dopamine are the respective chemical modulators involved in the attention networks. There is substantive behavioural and neuroimaging evidence about predictions formulated relying on the attention networks framework. Those about meditation and attention are reported in section 4.1 for corroborating the rationale of our empirical Study 1.

In the same rationale for our Study 1 it is mentioned another framework that individuates two dissociable systems in the brain: the *dorsal system of attention*, and the *ventral system of attention* (Corbetta and Shulman, 2002). The dorsal system is implicated in preparing and applying goal-directed selection of stimuli and responses. It is modulated by stimuli detection, and includes intraparietal and superior frontal cortices. The ventral system works as a circuit breaker for the dorsal system, directing attention to salient or unexpected events. The ventral fronto-parietal system is right-lateralised. In terms of the relationship between attention networks and dorsal/ventral systems of attention, there is evidence of activation of dorsal sub-regions for input-level and response-level selection in correspondence to orienting and conflicts monitoring respectively (Jha et al., 2007).

3.5. Temporal Attention and the Attentional Blink (AB)

A fundamental role of attention is to select and modulate the processing of information relevant for performance, to the detriment of irrelevant information. This selection is required to cope with the limitations inherent in human information processing. As briefly mentioned in section 3.2, perceptual, cognitive and action stages of human information processing may all be affected by three kinds of limitations: data-limited processing — when the input information is degraded; resource-limited processing — when the processing system does not have the ability to efficiently handle a piece of information if certain requirements are exceeded; and structurally-limited processing — when the system is unable to accomplish several operations at once (Proctor and Van Zandt, 2008). A limitation that is critical in the information processing flow is that imposed by working memory, which may contain only a limited number of items at a time. When attending to several concurring stimuli, the ability to explicitly report them depends on their encoding in working memory. Therefore the interaction between attention and working memory is important. This particularly emerges in temporal attention, i.e., when attention is allocated to a stream of stimuli appearing at different time-points, but in the same location in space, and this stream is characterised by a high rate, e.g., 8–10 stimului/s. In this situation there may be interference between the presentation of the stimuli to the senses, their selection by attention, and their encoding and retrieval from working memory.

3.5.1. Attention and Working Memory

Along with attention, memory plays an important role in human information processing. A typical distinction when dealing with memory is between long-term memory and short-term memory. The attributes long-term and short-term refer to how long — or short — is information maintained in the respective memory storages. Working memory is a system that has a short-term storage as its component, but also includes the ability to manipulate the information temporarily maintained in such a storage (Hitch and Baddeley, 2010).

One of the most influential models of Working Memory (WM) is the Baddeley and Hitch's one (Baddeley et al., 1974). It posits a system comprising at least three interactive subsystems: an attentional control system, and two short-term storages for visual and acoustic material respectively. In a successive evolution (Baddeley, 2000) the model included a third temporary store, the episodic buffer, which holds multidimensional *episodes* combining multimodal information. Therefore, the episodic buffer is the component of WM where information encoded from different sensory modalities interact, and this interaction is also extended to the content of long-term memory (Baddeley, 2010). In the light of the three-stage model introduced for human information processing in section 3.2 and depicted in Figure 3.1, WM is part of the cognitive stage, and it has important interactions with attention, especially in relation to the attentional blink deficit, which will be introduced in subsection 3.5.2.

An important characteristic of WM is that it is capacity-limited, i.e., at any moment in time it can maintain a limited amount of information. In particular, the episodic buffer is assumed to have a maximum capacity of four episodes (Baddeley, 2010), and there is evidence that visual WM capacity is of about four objects, while the maximum capacity of WM for verbal information is about seven chunks (Chun et al., 2010).

As for the Baddeley's model of WM, attention selects what perceptual information enters into WM, and this function is especially critical just for the capacity-limited nature of WM. Perceptual characteristics of the material to be encoded may affect the capacity limit for maintenance in WM: for example binding multiple features in a single object increases the capacity of visual WM (Luck et al., 1997), whereas more copmlex features may reduce WM capacity (Alvarez and Cavanagh, 2004). However, it should be noted that there is also a position that explain such findings by assuming a fixed number of items maintained in visual WM, independent of their complexity (Awh et al., 2007). Errors found in change-detection tasks when the items complexity changes would be attributable to high similarity between the items, without any change in the number of objects currently maintained in WM.

Electrophysiological and neuroimaging studies have shown that attention modulates both the encoding of sensory information, and the encoding at a post-perceptual stage (see Awh et al., 2006 for a review). In particular, the competitive advantage granted by selective attention for post-perceptual encoding of attended items has been interpreted as a clear interaction between attention and WM.

Encoding/consolidation processes in WM are implicated in the models accounting the attentional blink phenomenon, an attentional deficit which arises when in a rapid stream of stimuli all appear at the same spatial location. The carachteristics of this phenomenon makes it especially suitable for studying temporal aspects of attention.

3.5.2. Attentional Blink

The phenomenon most often singled out for studying temporal aspects of attention is the *attentional blink* (AB). AB is not a fresh effect since it was first observed in 1987 (Broadbent and Broadbent, 1987) during an experiment differentiating detection and identification processes of a target within a rapid serial visual presentation (RSVP). This effect was then named *attentional blink* by Raymond et al. (1992), as alike eye blink it hampered reporting of a target that had to be *seen* in a serial visual display.

The visual AB appears when the second (T2) of two targets embedded in a RSVP of distractors is presented within 200–500 ms after the first target (T1). The RSVP entails that all stimuli, both targets and distractors, appear serially in the same location of a display for a fraction of a second (typically 100 ms or less). Specifically, the subjects are often not able to report the identity of T2 at the end of the stream, even though they correctly reported that of T1. When subjects are instructed to ignore T1 in the same condition — i.e., same perceptual characteristics of the RSVP stimuli — T2 is usually correctly reported. This means that the AB is not a perceptual deficit, rather it is attentional. The ususal outcome measure to quantify the AB is the *conditional accuracy* of T2 reports, calculated by taking only trials with a proper report of T1 — referred to as the T2|T1 conditional accuracy.

The AB deficit is experimentally elicited through the AB paradigm that in its classical form manipulates the Stimulus Onset Asynchrony (SOA) between targets (a.k.a. Target Onset Asynchrony, TOA), i.e., the period of time elapsing between the onset of the two targets T1 and T2 in a RSVP. T1–T2 SOA is manipulated in successive different RSVPs. Participants are asked after every RSVP to make an unspeeded report of the targets in their order of appearance. Usually, targerts and non-target items have an equal presentation duration, and, if present, also the inter-stimulus interval (ISI) is constant throughout the RSVP. T1–T2 SOA is often measured in terms of *lags* since T1 onset: lag-1 means that T1–T2 SOA is equal to the duration of $T1 + ISI$ if present, that is no items intervene between T1 onset and T2 onset, while lag-3 means that T1–T2 SOA equates the duration of $3\,\text{items} + 3 \star \text{ISI}$, so 2 targets intervene between T1 onset and T2 onset, etc. In other terms T1 is in position '0', the post-T1 item in the series is in position 'lag 1', the second follower is in position 'Lag 2', and so on. A pitcture illustrating the paradigm and terminology is presented in Figure 3.2.

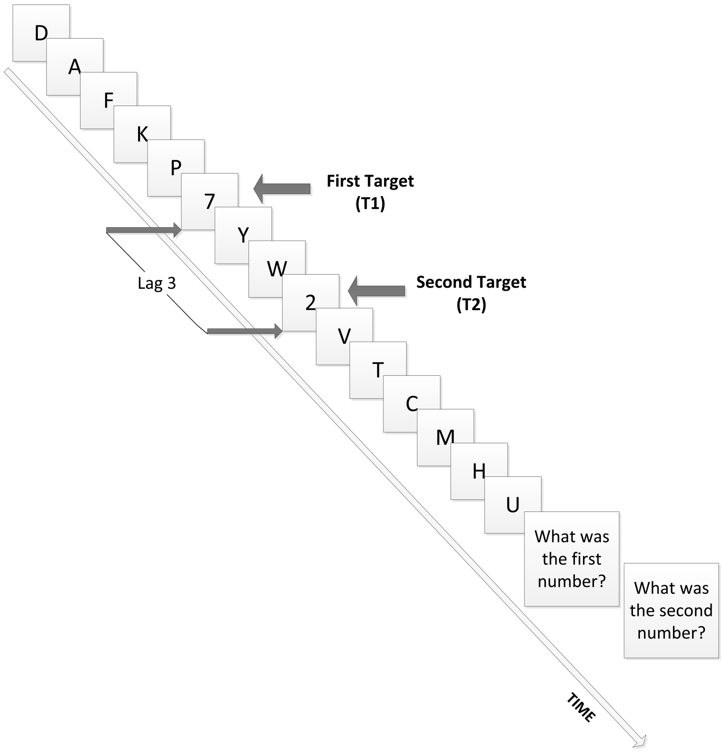


Figure 3.2.: Schematic illustration of a typical AB paradigm with T1–T2 lag = 3. The distractors are letters, while the targets T1 and T2 are the numbers 7 and 2 respectively.

A bunch of studies have undertaken experiments with the classical AB paradigm, and a typical characteristic of the conditional T2|T1 accuracy compared with the T1 accuracy has been established (Figure 3.3).

In the last two decades many studies were carried out to elucidate mechanisms and causes underlying the AB. Although the final word on these issues has not yet been written, a clear change of perspective may be followed in the scientific literature about AB (Martens and Wyble, 2010). During the first decade of studies since the AB discovery, the leading view held that AB derived from a central hard-wired bottleneck due to a capacity-limited post-perceptual stage of the human information processing. In particular, in this framework all items presented in a RSVP, targets and non-targets, are processed up to the formation of semantic representations. However, in order to be reported after viewing the RSVP, a target requires to reach the level of conscious awareness, or, in other words, to be encoded and consolidated in WM; this is achieved by means of a central stage of information processing that is capacity-limited. Consequently, the consolidation of T1 in a reportable WM content saturates the capacity of the central stage, and prevents the semantic representaion of T2 to be consolited in WM, hence precluding the participant in an AB experiment to report T2 when temporally too close to T1. In other words, it is a resource-depletion account that prevailed during the first decade of AB studies, attributing

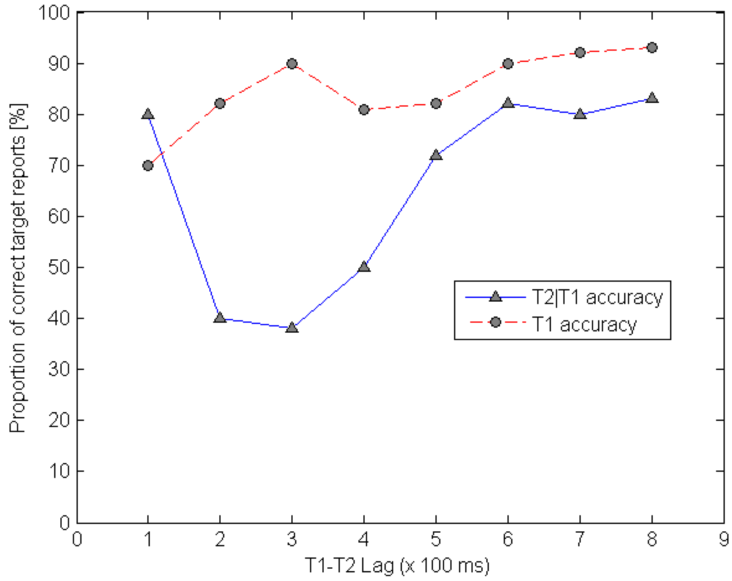


Figure 3.3.: Typical characteristics of the T2|T1 conditional reporting accuracy and of the T1 reporting accuracy as a function of T1–T2 lag, for lag $1 = 100$ ms, measured in an AB experimental paradigm.

to the depletion of limited processing resources operated during T1 processing the subsequent miss of T2.

More recently, the capacity-limited central stage of human information processing as the unique mechanism underlying the AB has been questioned (see Martens and Wyble, 2010; Dux and Marois, 2009 for a review). Specifically, several findings converge to the conclusion that the structural bottleneck of human information processing may be a contributing factor underpinning the AB, but not the only or fundamental one. Some form of attentional control seems instead to play an important role, and the AB appears to arise from the separate or interacting action of mechanisms intended to enhance targets or suppress distractors for an efficient encoding in WM and retrieval for their unspeeded report.

3.5.2.1. Empirical Evidence about the Attentional Blink

The different theories proposed to explain the AB are assessed on the basis of the extent to which they can account for the diverse empirical findings collected by means of the AB paradigm and some specific variants. Thence, before outlining the different theories it is important to describe at least the most prominent findings connected to the AB.

A first important consideration is the robustness of AB. The effect has been elicited with diverse types of stimuli, including alphanumerical characters (this is the case presented in the largest number of works, including the seminal ones), symbols

(Chun and Potter, 1995), words (Barnard et al., 2004; Luck et al., 1996; Sergent et al., 2005), pictures (Evans and Treisman, 2005; Potter et al., 2010) and faces (Einhäuser et al., 2007; Awh et al., 2004), as well as auditory tones (Tremblay et al., 2005; Duncan et al., 1997) and tactile stimuli (Dell'Acqua et al., 2006). In all these manifestations, the AB induces an interval lasting for several hundred ms during which the conscious awareness of a stimulus is impaired.

The AB has not been studied only in a within-modality scheme, with both targets in the rapid streak belonging to the same modality, being it visual, aural, or tactile. A fairly debated issue regards indeed whether the AB is a cross-modal deficit or if it is confined within a specific modality. Controversial results emerged from these studies. For example Haroush et al. (2011) found that auditory processing is enhanced during a visual AB: still, their interpretation is not in terms of a visual AB caused by attentional resources engaged by the auditory T2; rather, they argue that the occurrence of the visual AB releases resources, which become available for the auditory target, hence benefitting of an attentional enhancement. On the contrary, Jolicoeur (1999) found that a speeded two-choice discrimination task for a sound caused an AB for detecting the presence of a visual target in a RSVP. Differently, no impairment at identifying a second visual or auditory target was found by Duncan et al. (1997) when attending to a concurrent target presented in an independent stream and in another modality (vision or audition). Cross-modal effects are in favour of views that assumes the existence of an amodal pool of attentional resources, which are accessed for identifying targets in dual-task serial presentations independent of the targets modality. Conversely, experiments showing no between-modality effect on the AB support the existence of modality-specific resources, which pose restrictions on attention. A recent study exploiting the individual differences in the magnitude of observed AB provides important evidence of a cross-modal AB (Martens et al., 2010). A caveat that is often discussed when challenging the cross-modal AB issue is that of task switching. There are in fact also results going in the direction of the necessity of a task switching between the responses required by the two targets for eliciting a cross-modal AB, e.g., in an audiovisual case (Potter et al., 1998), while a simple switching between modalities would be not sufficient for an AB (Soto-Faraco and Spence, 2002).

Another interesting aspect of AB is that its magnitude significantly varies across individuals, up to an extreme that is represented by *non-blinkers*, who even do not show any AB, and are estimated to be the 5 % of the general population (Martens et al., 2006). This inter-individual variability of AB magnitude counteracts the idea of a structural nature of the AB, opening an avenue for studies that hypothesise a plasticity of the AB (Slagter et al., 2007; Green et al., 2003). In addition, non-blinkers have been found better at ignoring distractors (Martens and Valchev, 2009), with no differences in working memory capacities (Martens and Johnson, 2009). Individual variability of AB magnitude could therefore depend on attention allocation policies that an individual adopts from trial to trial, determining which objects deserve further processing and consolidation in memory (Martens and Wyble, 2010).

The fact that AB occurrence might depend on individual strategies of attention allocation is in agreement with another surprising result about the AB: the addition of concurrent task-irrelevant activity attenuates the AB magnitude (see for example Olivers and Nieuwenhuis, 2005, 2006; Lapointe-Goupil et al., 2011). The AB attenuation has been found with the addition of different kinds of task-irrelevant activity: listening to task-irrelevant music or free task-irrelevant mental association (Olivers and Nieuwenhuis, 2005), inward/outward visual motion and flickering superimposed on the RSVP (Arend et al., 2006), change in task instructions (Ferlazzo et al., 2007; Olivers and Nieuwenhuis, 2006), and a concurrent secondary task (Olivers and Nieuwenhuis, 2006; Taatgen et al., 2009; Lapointe-Goupil et al., 2011). The evidence of an alleviated AB has been interpreted in different ways: as the result of a more diffuse attentional state (Olivers and Nieuwenhuis, 2005), as a reduced interference thanks to less focused attention, which limits the number of attended distractors, or as an enhanced cognitive flexibility fostered by the concurrent task (Olivers and Nieuwenhuis, 2006). In Arend et al. (Arend et al., 2006) AB attenuation roots in the decreased attentional allocation to T1 caused by the concurrent moving visual background.

Actually, the manipulation of resources allocation to T1 and T2 in the RSVP has proved to affect AB magnitude. When T1 identification is more difficult, a larger AB can be found (Taatgen et al., 2009). Also perceptual, spatial, and temporal cues displayed an effect on AB. If T2 is preceded by a non-target sharing the same colour in a classical dual-target RSVP, or spatially pre-cued in a dwell-time paradigm, the AB induced by T1 may be alleviated (Nieuwenstein et al., 2005). Besides, the AB is postponed, i.e., it appears at longer lags since T1 onset, as long as dostractors sharing a target-defining feature are used to mask T1 (Olivers and Meeter, 2008). As for temporal target cueing, Martens and Johnson (2005) found that providing explicit knowledge about the temporal lag between T1 and T2 for the different trials, either symbolically or directly showing a couple of cueing symbols with an inter-stimulus interval corresponding to the temporal information to be conveyed, correspondend to a reduction of the AB magnitude. The authors' proposed explanation was in favor of some degree of top-down attentional control, governed by participants' intentions, as a strategy that may in part overcome the restrictions causing the AB.

Lag-1 Sparing The main outcome of an AB paradigm is an impairment at reporting T2 in a dual-target RSVP when T2 appears between 200 ms and 500 ms since T1 onset. However, there is another outcome that somehow deviates from this trend, and which has been found in more than a half of the experiments studying the AB: it arises when T2 appears at lag 1 in the stimuli stream. Indeed, looking at Figure 3.3, the T2|T1 report accuracy is often unimpaired, i.e., no AB is observed, when T1 and T2 are consecutive. This effect is named "lag-1 sparing", and is reliably found in many experiments (Visser et al., 1999). Lag-1 sparing is explained in different ways by the different theories about AB roots and causes. A major distinction across all the accounts is whether lag-1 sparing depends on the temporal separation between T1 and T2 or on the nature of the stimulus appearing at lag-1 position. The lag-1 sparing phenomenon has been studied also in paradigms that add a spatial displacement between targets to the temporal separation typical of the AB, and in paradigms assessing a possible multisensory AB, by using targets eliciting different sensory modalities, e.g., a visual T1 and an auditory T2. When the spatial location of T2 changed relative to T1 location, no lag-1 sparing was generally found (Visser et al., 1999). In addition, the lack of lag-1 sparing is commonly found when there is a multidimensional change between the defining attributes of T1 and T2 (Visser et al., 1999). Finally, lag-1 sparing observation depends on how correct reports are counted: if the reporting order of T1 and T2 is taken into consideration for determining the correctness of atrial, the lag-1 sparing effect often disappears (Hommel and Akyürek, 2003).

In the following an overview of the main models proposed to account for the AB and related empirical evidence is provided. The overview follows a distinction between informal and formal theories proposed in a review by Dux and Marois (2009). The section closes with a summary of the main points that can be drawn from the comparison of the different models, and their analysis in the light of the evidence cumulated by means of AB-related experiments.

3.5.2.2. Informal Theories

Inhibition Model One of the first informal theories proposed to explain the origin of the AB was the *Inhibition Model* (Raymond et al., 1992). The model derived from a paradigm where T1 was a white letter in a stream of black letters, and T2 was a black 'X'. According to this model, the detection of T1 physical features (i.e., the white colour) at an early perceptual stage triggers an attentional response intended to facilitate T1 identification. Attention is considered to be allocated episodically, therefore the attentional response corresponds to the initiation of an attentional episode. Such initiation is metaphorically depicted as a gate that opens to allow the processing of T1 until its identification is completed. However, if the series item at lag-1 — indicated as the T1+1 item — is presented before T1 identification completion, T1 features and T1+1 features would be available together for the identification process, hence providing a high potential for confusion and identification errors. For this reason a suppressive mechanism is triggered by T1+1 occurrence, which inhibits post-perceptual processing of the items trailing T1. This is likened to the gate closing. Thus, the AB appears because T2 is suppressed until T1 identification is completed. The exception of lag-1 sparing is motivated by a slow dynamic of the suppression activation, i.e., the gate closure is slow, and lets T2 pass to the identification process when it appears as the T1+1 item. Therefore, lag-1 sparing depends on the temporal proximity between T2 and T1.

Interference Theory The AB occurs also when detection and not identification of T1 is required (Shapiro et al., 1994). This evidence challenges the inhibition model hypothesis that the root of the AB is a mechanism intended to avoid a features conjunction error during target identification. Specifically, the interference model posits that the AB deficit occurrence depends on the result of matching items representations with featural and identification object templates in visual WM (the terms visual short-term memory, VSTM, is used in the original paper). All items undergo perceptual processing, and typically T1, T1+1, T2, and T2+1 items gain access to the VSTM, where they are assigned with weights and interfere for being retrieved as required by the reporting task. The assigned weighting depends on the limited capacity of VSTM, and on the degree of similarity between the items. The AB arises because T2 has a lower weight since VSTM is already close to saturation of its limited capacity. In contrast, lag-1 sparing occurs because when T2 is in lag-1 position, one less item occupies VSTM (the T1+1 and T2 items coincide in this case), hence diminishing the interference that causes T2 representation decay in the AB condition. The explanation of no AB for T1-T2 SOAs longer than 400 ms is not very specific, mentioning that VSTM is periodically flushed.

Two-Stage Theory An influential model for the AB, which can be considered the prototype of the so-called bottleneck accounts is the two-stage theory presented by Chun and Potter (1995). According to such a theory, when a RSVP is presented, almost all items access a stage 1 of processing, they are recognised, and their conceptual representations activated. However, in stage 1 each item representation is volatile, subject to decay or to be overwritten by the trailing items. Therefore, in order to be later reported, target representations require to undergo a stage 2 of information processing, which encodes and consolidates them in WM. Specifically, when in stage 1 a likely target is detected, a transient attentional response is triggered, which enhances the target and first post-target item for selection to enter stage 2. Nonetheless, stage 2 is capacity-limited and its latency exceeds the item presentaion duration of RSVP. Hence, T2 has to wait for the complete processing of T1 at stage 1, before entering stage 2 for consolidation in WM and successive report. The AB arises when T2 appears during T1 processing in stage 2, and its representaion is lost during the waiting. In this theory, lag-1 sparing is explained by the fact that the dynamic of the attentional enhancement is not rapid, enhancing T1 and the item at lag-1 postion: so if T2 appears at lag-1 position, and particularly within 100 ms from T1 onset, the probability of an AB is dramatically reduced.

The two-stage model derives from the observation that also categorically defined targets, as black letters in a stream of black digits, may trigger an AB. This is not in agreement with the inhibition model (Raymond et al., 1992), that argued the origin of the AB as an error in the conjunction of features with target identity.

Central Interference Theory In this framework, Jolicoeur (1998) extended the two-stage model with a theory that incorporates both the AB and the psychological refractory period (PRP). PRP occurs in dual-task paradigms, when paricipants have to make separate responses to a primary and a secondary task. The latter is sequentially presented with respect to the primary task, and partially overlapped, as it stimulates the participant before she/he has responded to the primary task. PRP consists in slowing down responses to both tasks (Pashler, 1984). The central interference theory posits that there is a single locus of interference between the response to the secondary task and the concurrent processing of the primary task: this locus is the encoding of information in short-term memory (STM) for T2 in the AB, also termed consolidation in short-term memory, and the response selection for the secondary task in PRP. Such processes indeed requires central mechanisms that are capacity-limited, that is they require certain operations to be performed serially. For the AB the consolidation of T2 is subject to interference, but the interfering operations may be multiple, including the encoding of T1 in short-term meory, but also response selection for T1, retrieval from long-term memory, and the switching between the primary and secondary task. In this view, the two-stage model (Chun and Potter, 1995) may be considered as a special case of the central interference theory, as the former considers the consolidation of T1 in STM the unique source of AB, whereas the former considers multiple possible causes, as response selection for T1 (Jolicoeur, 1998).

Attention Dwell Hypothesis This account is against the view of visual attention as a serial system, capable of rapidly shifting among objects. Specifically, Ward et al. (1996) formulated the attentional dwell time hypothesis following a series of experiments manipulating the SOA between targets but also targets spatial locations. The proposed framework foresees that objects compete in parallel, according to their matching to a target template, for the allocation of visual processing resources that are limited. The competition process lasts for several hundreds of ms, and the winning objects undergo extended processing at the expenses of the other objects, forming representations that contribute to a sustained state in which these selected representations and their properties are available to guide behaviour. In this context, AB arises because T1 wins the competition over T2, as T1 enters first the competition process, and T2 is subject to masking and interference. A greater number of competing objects, like in a typical RSVP, induces a greater demand on the capacity-limited visual resources, thus contributing to the AB. No explanation for lag-1 sparing is offered by the attentional dwell model, since this effect was not observed in Ward et al. (1996) experiments. This might be due to the spatial displacement among targets adopted for their experiments: indeed, lag-1 sparing is rarely found when target stimuli appear at different spatial locations (Visser et al., 1999).

Two-Stage Competition Model This is an extension and modification of the two-stage theory by Chun and Potter (1995), and it has been proposed in Potter et al. (2002). The model relies on empirical findings derived from experiments that study very short T1-T2 SOAs in a variant of the classical AB paradigm. The two stages of Chun and Potter's model are maintained: the first for detection of a potential target, and identification of the target as its end, and the second, initiated by such identification resulting from stage 1 and subject to limited capacity, to consolidate the target in short-term memory. The new claim is that if T2 appears before T1 identification, as attention in stage 1 is labile, T2 attracts attention, accruing resources faster thanks to the attentional response triggered by T1 detection. This is the reason of the lag-1 sparing effect, and of an accuracy superiority of T2 over T1, with a relevant proportion of reversals in T1-T2 reporting, at very short SOAs (e.g., shorter than 53 ms in Potter et al., 2002 experiments). Especially, for lag-1 sparing, what might happen according to this model is that T1-T2 SOA is shorter than the time needed to identify T1 such that T2 enters stage 2 for consolidation and later report, but enough long for T1 to be briefly retained without stage 2 processing, and be correctly reported with T2. The temporal position of T1 in the RSVP becomes determinant for a privileged access to stage 2 as long as the SOA with T2 increases.

Other Bottleneck Models Further accounts of the AB deficit either extend the two-stage theory or share some important points with it. The key one is that the major reason for the AB to arise is a central bottleneck due to some capacity limitation of the information processing. Another debate is about the amodal nature of such a central bottleneck. Actually, AB has been found also for auditory and tactile presentations (Duncan et al., 1997; Dell'Acqua et al., 2006), and always induced an attentional failure over a few hundreds milliseconds. In addition, also cross-modal AB has been studied (Haroush et al., 2011). In this case, the question is whether attentional resources allocated to one modality deteriorate attention to another modality, i.e., if the central bottleneck is amodal, or capacity-limited sepecific-modality resources may be separated independently. A prevalent evidence is in favour of an amodal central bottleneck, although AB might arise from both a limitation at the amodal stage of processing, and at the visual processing stage (see Dux and Marois, 2009 for references to this line of debate). An alternative account for the AB (Awh et al., 2004) proposes a multi-channel model, where separate channels process feature-based and configural information respectively, against a unique central processing bottleneck. But this account has been criticised as it uses faces as target stimuli, and faces salience might have been a major determinant of these experiments findings, biasing their processing.

Temporary Loss of Control Model This model, proposed by Di Lollo et al. (2005), questions the assumption of almost all the above described models, that AB arises due to the depletion of limited resources caused by T1 processing. In their experiments the AB occured when T1 and T2 belonged to different categories, whereas no AB was found when both T1 and T2 belong to the same category. This evidence is obviously against an account that based on depletion of resources allocated to the target leading the stream. Specifically, Di Lollo et al. (2005) found that when three targets T1, T2 and T3, are presented sequentially in a RSVP, T3 suffers no report impairment when sharing the same category with T1 and T2. This effect was also found in successive experiments, and it has been named “Spreading the [lag 1] sparing” by Olivers et al. (2007). The crucial claim of the temporary loss of control (TLC) account is that an input filter configured to pass target items and reject non-target items governs the initial processing of incoming stimuli. The filter is maintained by a central processor that is capacity-limited: indeed this processor can perform one function at a time. When a target is detected and initially identified the central processor switches from monitoring to consolidation operation, and the input filter passes from the endogenous control of the central processor to the exogenous control of the trailing items. If the trailing item is intra-categorical with the target currently under consolidation, it matches the filter and it is efficiently processed, being its reporting accuracy limited only by the short-term memory span. Conversely, if the trailing item does not belong to the same category, it will take longer to be processed as it does not match the input filter configuration, becoming susceptible of masking from the following items, and the input filter configuration will be disrupted, and requires time to be reconfigured for the new category. These two consequences of target category change determine the AB. According to this model, lag-1 sparing and its spreading are possible as long as the consecutive targets belong to the same category.

Delayed Attentional Reengagement Account The finding that a blank gap separating two targets elicits an AB questions the TLC hypothesis of a major role of distractors interference in inducing an AB (Nieuwenstein et al., 2009). In addition, Nieuwenstein (2006) found that the AB may be mitigated if the target is pre-cued by a distractor sharing a target-defining feature. In order to account for these new findings Nieuwenstein and colleagues proposed the delayed attentional reengagement account (Nieuwenstein, 2006; Nieuwenstein and Potter, 2006). This working hypothesis argues that T1 presentation triggers the deployment of top-down (i.e., goal-directed) attentional resources: when T1 information is no longer available due to a post-target replacement, the resources are disengaged and allocation of attention is delayed. Such a disengagement particularly occurs when T1+1 is either a distractor or a blank item. AB arises when the participant in the paradigm is not efficient at rapidly reengage top-down attention to T2 shortly after disengagement. Lag-1 sparing is possible because attention is sustained for T2, as it bears goal-relevant information. Both the nature of T1+1 and the duration of T1 enhancement (i.e., engagement of top-down attention) contribute to lag-1 sparing. An empirical finding compatible with this theory is the result of the experiments carried out in Nieuwenstein and Potter (2006). Here a superiority in the accuracy of a “whole report” condition — i.e., the report of all the letters in a RSVP — was found with respect to a “partial report” condition — when only coloured or specific letters of the stream had to be reported. The proposed account explains the whole report superiority as fostering a state that sustains attentional engagement throughout the entire task. On the other hand, the target pre-cuing in Nieuwenstein (2006) reduces the AB by helping re-engagement of attention.

3.5.2.3. Formal Theories

There have been both incorporation of the AB phenomenon in more general neuro-commmputational models, and the proposal of new dedicated models. In the following only some of them are considered, which account for most of the different approaches.

Locus Coeruleus Model Locus coeruleus (LC) is a nucleus located in the brainstem, which is thought to influence attentional control (Aston-Jones et al., 1999). LC activity leads to the release of norepinephrine, which is projected to many cortical areas, especially in regions thought to play a role to direct attention. Specifically, Aston-Jones et al. (1999) developed a neurocomputational model that links phasic and tonic modes of LC activity to the facilitation of focused attention and scanning with more labile attention, respectively. A focused attention state should lead to an increased accuracy at responding to task-relevant stimuli, while scanning should increase response to task-irrelevant stimuli. In monkeys the LC-mediated noradrenergic innervation increases the reponsivity of target neurons when a target stimulus is visually presented, and this is interpreted as a facilitation of processing in response to a task-relevant stimulus. Following the rapid, phasic response to the task-relevant or salient stimulus, the local release of norepinephrine within the LC causes a sort of refractory period, during which the facilitating effect of LC-norepinephrine system is not available. This refractory period has a duration that is similar to that of the AB (Nieuwenhuis et al., 2005). Therefore, the LC model for the AB postulates that the AB and the LC-norepinephrine system share the same dynamics: in particular, when T1 appears in a RSVP, LC neurons should fire and induce a noradrenergic modulation that enhances attention thus enabling T1 encoding in WM. By 200 ms since T1 onset the LC refractory period starts, and no further attentional enhancement may be triggered for several hundred ms — the AB interval. Lag-1 sparing is instead possible if T2 appears within the attentional enhancement window triggered by T1, i.e., before the refractory period of the LC starts.

Although it is not demonstrated that the administration of a noradrenergic agonist may affect the time course of AB, LC model contributes to the evidence that lag-1 sparing depends on the temporal sepration between T1 and T2, despite a non-target intervenes or not between prior to T2.

Global Workspace Model This is a formal theory that emphasises the relationship of AB with consciousness. The global workspace model has been proposed by Dehaene et al. (2003) to account for the processing of stimuli from their early sensory representations to conscious events, available for multiple processes, e.g., for delayed report as in the AB. The global worspace model has proved able to simulate an AB. In general, according to the model, a stimulus has to trigger the coordinated activation of a large neuronal population, comprising sensory and high-level areas interconnected by long-distance axons, in order to access conscious awareness. When a stimulus activates a sufficient sub-set of neurons in the workspace the neuronal activity becomes self-sustained, and the information associated to the stimulus becomes available and can be projected to higher-order processing areas. However, the neurons of the workspace activated by the stimulus inhibits their neighbours, preventing a subsequent stimulus to trigger the activation of the global workspace. This is the reason for the AB: sensory processing of T1 and T2 proceeds in parallel, as no inhibition is foreseen, then, when T1 invades the global neuronal workspace the processing of T2 at the same level is inhibited, and T2 may be blinked. Lag-1 sparing effect is not explicitly incorporated into this model, although a possible interpretation of it would be that of a delay in the T1-locked inhibition mechanism concurrently with a T2 appearing at short SOA.

Episodic Simultaneous Type/Serial Token Model The episodic simultaneous type/serial token model (Wyble et al., 2009), also named eSTST, is the extension of a theory about temporal attention and WM known as simultaneous type/serial token, STST (Bowman and Wyble, 2007). The two models are neuronal networks describing the extraction of information and encoding of targets from a serial presentation, by preserving temporal order. According to the STST model, visual WM uses types and tokens, operating in two stages, similarly to the two-stage theories like that of Chun and Potter (1995). Targets and non-targets of a RSVP are processed up to the activation of their semantic representations (i.e., their *type*), but episodic information, like the temporal order and repetition of the item in the series, is required to report the targets in their order. Episodic information is stored in WM as *tokens*. In order to report an item, the latter must be encoded in WM, and a link between its token and its type must be established for retrieval. In the RSVP, T1 detection triggers a brief transient attentional that enhances also the post-T1 item, originating the lag-1 sparing effect if such item is T2. In contrast, if T2 appears during T1 encoding in WM, the AB arises, as attention to new inputs is suppressed to protect T1 type-token binding and consolidation, and T2 type cannot be linked to a token in WM for later retrieval. The STST model does not account for spreading the sparing findings. This is one main reason for the extension into the eSTST model. The key difference is that eSTST foresees an interaction between excitation of attention triggered by a new target, and the suppressive effect provided by WM encoded. As long as consecutive targets arrive, the suppressive effect is not triggered, and the non-specific attentional enhancement allows the binding of the types to the tokens in WM. This explains the spreading of lag-1 sparing, but comes at a cost: a decrease in T1 accuracy, a high proportion of T1-T2 reversals when reportingthem, and inaccuracies at detecting stimuli repetition.

An important characteristic of this model is that it hypothesises a function for the AB: it is a self limitation to visual encoding brought forth to parse visual input in distinct episodes. The AB works as a cognitive strategy that is flexible, because sparing occurs when targets are presented uninterruptedly. However, this comes at a cost, indeed items may be encoded in parallel in WM, but with no assurance that their episodic distinctiveness is maintained.

Boost and Bounce Theory The Boost and Bounce theory by Olivers and Meeter (2008) essentially proposes two interacting stages of processing, but capacity limits are not the determinants of the AB. During sensory processing perceptual representations, but also semantic and categorical representations of items are activated. However, these representations are susceptible to decay, and their activation is influenced by the masking effect of the preceding and following items, depending on their similarity to the current item, i.e., on their saliency (a stimulus is salient to the extent that it differs from its preceding and trailing stimuli). WM is the second stage of this model: it implements an attentional set according to task instructions, it stores encoded target representations linked to a response, and employs the attentional set as an input filter implementing a gating mechanism for enhancing stimuli matching the attentional set, and blocking out those whch do not match it. The attentional set enhances task-relevant sensory representations (the *boost*), while irrelevant sensory representations are inhibited (they are *bounced*). In such a way, enhanced items access WM and can be consolidated for being later reported. In the model the strength of the modulation depends on the sensory evidence strength, therefore it does not depend only on the stimulus triggering the modulation, but also on the similarity between targets and distractors, and on the current state of attention (i.e., the current amount of excitatory and inhibitory feedback is affected by, but also affects sensory evidence).

In this framework, when a classical AB paradigm is administered, initially a stable bouncing state prevent distractors from entering WM. Then T1 is represented at the sensory stage, and its sensory signals trigger a boost, which enables T1 identity to enter WM. However, in a RSVP the boost peak arrives after T1 onset: it actually benefits the T1+1 item. Therefore, when T1+1 is a non-target, it receives the maximum enhancement intended for T1, and this strong signal of an item not matching the attentional set enters WM and reaches the gating neurons thus inducing a strong bounce. This bounce has maximum amplitude on the post T1+1 item: if T2 falls under such bounce it cannot access WM, even though it triggers per se a boost, as the bounce inhibitory activity is too strong compared to the presentation time of T2. Consequently, according to the boost and bounce theory of temporal attention, the AB is not caused by a central capacity-limited bottleneck and the processing load of T1; it instead depends on the timing of inhibitory and excitatory feedback elicited by T1 and its trailing stimuli in the RSVP.

As for lag-1 sparing, the boost and bounce theory explains it in the same way as the AB: the attentional boost peak coincides with T1+1 item arrival, therefore it is spared by the blink. Also the spreading-the-sparing effect has this cause: as long as post-T1 items are task-relevant, they benefit of the attentional boost and may be identified and consolidated in WM (which anyway has a maximum storage capacity of 5 items in this model). The lag-1 sparing effect is thus time-based, and not item-based as other models propose.

Threaded Cognition Model This model of the AB proposed by (Taatgen et al., 2009), combines the viewpoints attributing the AB to control issues (as the TLC and Boost and Bounce models) and two-stage theories attributing the AB deficit to a conflict between fast target detection and capacity-limited, slow WM consolidation. The model assumption is that during RSVP target detection and memory consolidation operate in parallel, being in fact sub-tasks of a single task, which is the RSVP task. Thus, these sub-tasks compete for cognitive resources in a multi-tasking context. Multi-tasking is modeled by threaded cognition: a set of cognitive resources may operate in parallel, but a single cognitive resource can serve a single task at a time. Without an additional control, for a default task, a resource is assigned to the first task needing it, and released when the task no longer requires it. In the case of RSVP a control production rule is used in order to protect consolidation from conflict with target detection. This overexertion of control suppresses target detection, hence originating the blink of T2 when it appears during T1 consolidation. However, (Taatgen et al., 2009) emphasises that the conflict between target detection and memory consolidation is only apparent, and it would not require the protection rule causing the AB, since the two processes only share procedural resources. Therefore, in this view the AB would derive from an unnecessary overzealous control mechanism. Lag-1 sparing and spreading are explained by this threaded cognition model as the result of recognition by the control that the items following T1 are targets, thus not applying the protective rule. Threaded cognition model is interesting also in further respects because it accounts for some additional important evidence connected to the AB. One explanation it offers is for non-blinkers. Several studies found that, although AB is a fairly robust phenomenon, individuals exist that do not show any AB when confronted with RSVP (Martens et al., 2006; Martens and Valchev, 2009). In the threaded cognition model account (Taatgen et al., 2009) propose that non-blinkers behaviour depends on not applying the control production rule suppressing target detection during consolidation of T1. The threaded cognition model might also explain the finding that the addition of distraction concurrently to the dual-target RSVP task can attenuate the AB. Indeed, additional distraction corresponds to additional demand on cognitive resources. In default situations this would slow down pwerformance, since concurrent tasks should wait for their chance to access the requred resources. In contrast, when a RSVP is running, there is not enough time for waiting, so production rules activation is dropped, and the protective control production rule that blocks target detection has the lowest utility being thus the first to be left out. Therefore, AB is reduced, not completel eliminated, because only concurernt tasks that provide enough prduction activations, i.e., are not successfully ignored, set the conditions to supersede the control production rule eliciting the AB (Taatgen et al., 2009).

3.6. Introduction to the Empirical Studies

One of the chief objectives of this thesis is to provide an empirical case study of using VEs to investigate cognitive processes. For this purpose, attention has been chosen as the target process, essentially for its importance in human performance, its interaction with multiple cognitive operations, and the limited number of studies considering attention in VEs, at least on the time-scale typical of attentional blink. In this framework, the following two chapters present two empirical studies — Study 1, and Study 2 — addressing temporal attention, and using the performance in AB-like paradigms as the main outcome measure.

Study 1 does not involve a VE, but serves as a starting point for elaborating the rationale of Study 2. It evaluates whether a short-term mindfulness practice may influence the allocation of attentional resources over time as indexed by changes in the AB magnitude. The link with Study 2 is the attention allocation policy (Taatgen et al., 2009) or diffuse attentional state (Olivers and Nieuwenhuis, 2006, 2005) referred to as a possible cause of AB modulation, for example when concurrent task-irrelevant activity is added to the dual-target RSVP task. On the one hand, a diffuse attentional state as opposed to a focused attentional state is one apsect of the typical definitions of mindfulness (Bishop et al., 2004). On the other hand, a concurrent sound distraction, and a background including visual motion, or flickering, have been proved capable of alleviating the AB deficit, possibly inducing a similar diffuse attentional state. Consequently, a concurrent sound presented in a dual-target desktop-displayed RSVP and a three-dimensional VE meaningfully related to the dual-target RSVP presented therein may be considered as two different forms of concurrent distractions, and they might show a similar modulatory effect on the AB magnitude.

Thereby, Study 2 explores AB performance in an immersive three-dimensional VE, and constrasts the results with AB performance scored using bidimensional versions of the same stimuli in a classic desktop-display version of the RSVP, with and without a concurrent sound distraction. The prediction that a concurrent distraction alleviates the AB magnitude is tested, and a potential relationship between the sound distraction and the distraction possibly introduced by the three-dimensional stimuli in the VE and the virtual landscape is discussed too. Furthermore, while in Study 1 letters and digits are used as visual stimuli, road signs are used as stimuli in the RSVP of Study 2, allowing for a discussion of the effects of less abstract symbols, and their inclusion in a meaningful context (a virtual motorway) in the immersive three-dimensional VE display condition.

4. Study 1 – AB Modulation After a Short-Term Mindfulness Yoga Training

In Study 1 we explored whether a 7-week training programme based on an integrative body-mind approach relying on Tibetan Yoga including mindfulness meditation can improve the distribution of attentional resources in a group of adult meditators, as indexed by the magnitude of attentional blink (AB) measured across dual-target rapid serial visual presentations (RSVPs) of letters (the to-be-ignored distractors) and digits (the to-be-reported targets). The objective is to contribute to the empirical evidence about the possibility to modulate the AB through mental training, adding the specificity of *Tsa Lung* Tibetan Yoga as a means of training that includes a bodily component along with the mental practice. The main result of the study is that this Yoga practice enables a better allocation of attentional resources, as the AB magnitude is reduced after the training. This chapter describes the rationale for the study, and presents the details of the conducted experiment in terms of procedure, methods, results, and conclusive discussion.

4.1. Study 1 Rationale

As illustrated in chapter 3, attentional blink (AB) is a robust phenomenon that arises when an unspeeded dual-target identification task is required for two targets embedded in a rapid stream of distractors, all appearing at the same spatial location of a display. Actually, the robustness of the AB goes beyond the visual modality, since also auditory and tactile AB have been observed, but in the experiments presented herein only the visual AB is considered.

The exact causes of the AB are still not clear, but it appears to originate from the combination of depletion of capacity-limited processing resources by the first target (T1) in the RSVP, with attentional control mechanisms (see subsection 3.5.2 for a more detailed discussion of the topic). At any event, the original hypothesis that AB reflects a structural bottleneck in central processing capacity is no longer accepted, also because evidence about a large individual variability of AB magnitude is now demonstrated, and several conditions proved to be capable of significantly alleviating or modifying the duration of an AB (Arend et al., 2006; Olivers and Nieuwenhuis, 2005; Stein et al., 2010; Einhäuser et al., 2007; Martens et al., 2006; Green et al., 2003).

According to some recent theories about AB causes, the large individual differences found for this attentional deficit might be the consequence of different allocation policies these individuals adopt for the deployment of attention(Taatgen et al., 2009; Olivers and Nieuwenhuis, 2006). Consequently, if AB is not hard-wired in our brain structures and reflects some cognitive strategy for the allocation of attention, it might be possible to modify the AB performance of an individual through some sort of mental training, possibly by producing a long-lasting change in the adopted attention allocation policy. Therefore, the AB paradigm is appropriate as an index of allocation of attentional resources.

4.1.1. Training-Related Alterations of the AB in Video-Game Players and Meditators

Several studies evaluated the effect of training on AB performance. Green et al. (2003) compared habitual video-game players with non-video-game players in different aspects of visual attention, and evaluated the change of attentional abilities following the training of non-players on a an action video game. Firstly, the results displayed enhanced attentional capacity, and enhanced allocation of spatial attention to the visual field, even outside the game-playing zone, for video-game players compared to non-players. Secondly, they specifically examined the temporal characteristics of video-game players visual attention by using an AB paradigm. The RSVP included a white letter as T1, and a white 'X' as T2, among black letters serving as distractors. The task was to identify T1, and detect the presence of T2, which actually appeared only in 50% of the trials. Letters duration was 15 ms, separated by 85 ms blanks. The video-game players featured a reduced AB, and general higher report accuracies at every T1–T2 SOA. Finally, the last experiment in Green et al. (2003) tested the AB performance — as well as other attentional capacities — of non-video-game players prior to and after 10 days of 1-hour training per day. The participants were split in two groups: one training on an action video-game that required attention to be switched/distributed around the visual field, and another training on a video game that required to focus on one object at a time. Participants trained on the action video game featured a faster recovery from the AB, and a marginal correlation of AB performance improvement with video-game playing improvement. This is an evidence of the ability of action-video game training to alter visual attention processing, and namely to overcome a limit in the distribution of attentional resources over time, as measured by the AB.

Another form of training correlated with improvement of the performance in AB tasks is meditation. Many different styles of meditation exist, and univocal definitions for the specific types are not readily available. However, the different meditation practices may be regarded as "a family of complex emotional and attentional regulatory practices, in which mental and related somatic events are affected by engaging a specific attentional set."(Raffone and Srinivasan, 2010). A useful scientific framework that attempts to conceptualise a number of meditation types has been proposed by Lutz et al. (2008). In this framework the role of attention is crucial, and serves to define two classes of meditation styles: Focused Attention meditation (a.k.a. FA meditation), and Open Monitoring meditation (a.k.a. OM meditation). While FA meditation requires to sustain and monitor an explicit focus (e.g., the own breath, a specific location in the environment, or a certain thought), OM meditation implies the non-reactive monitoring of the ongoing experience, but without the selection of any explicit focus. Importantly, FA and OM meditation categories have been able to predict relevant neurofunctional aspects consistent with the dissociable systems of attention identified in the human brain (Posner and Petersen, 1990; Corbetta and Shulman, 2002).

As seen in chapter 3, attentional capacities have been associated with dissociable systems in the brain: the three attentional networks identified by Posner and Petersen (1990) dissociate alerting, orienting, and executive control; instead, Corbetta and Shulman (2002) proposed a model dissociating between a dorsal system, and a ventral system of attention. Alerting entails the achievement and maintenance of a vigilant and alert state, orienting means the selection of information from sensory input, and executive control is designated to monitor and resolve conflicts among responses. On the other hand, in the dorsal/ventral model of attention, the former system prepares and applies goal-directed (top-down) selection, being modulated by stimulus detection, whereas the ventral system is stimulus-driven, directing attention to behaviourally-relevant stimuli, especially those salient, or unexpected.

In FA meditation alerting is used to sustain the explicit focus, orienting to select it, and executive control to detect distraction, and re-engage the focus of attention. The dorsal system is expected to be more implicated with FA meditation, as it underlies the selection of stimuli. OM meditation, which also initially uses focused attention to reduce distraction, in its mature form relies on brain regions implicated in sustaining attention for monitoring the ongoing experience, as the executive control network, and the ventral system, which is stimulus-driven.

Thus, the expectation is that both FA meditation, and OM meditation practices may have a training effect on attentional processes. Neurophysiological and behavioural evidence exist to corroborate this general prediction. In fact, Carter et al. (2005) used a binocular rivalry task administered with an head-mounted display to Tibetan Monks during and after either an FA meditation practice, or a form of "compassion" meditation that does not require to sustain attention on an explicit object. The result was that the group performing the experimental task during and after FA meditation was able to sustain for a significantly longer duration the stable image of one of the two competing percepts, underscoring the ability of the FA meditation state to improve the dorsal processing of attention. MacLean et al. (2010) demonstrated an improvement in vigilance (vigilance decrement was reduced) during sustained visual attention in a group of meditators trained in an FA style entailing

5 hours of daily meditation for 3 months. Brefczynski-Lewis et al. (2007) found that expert FA meditators showed stronger activations, measured in fMRI scans, than novice meditators in brain areas implicated in monitoring, attention engagement, and selective orienting. However, less activation was necessary in certain regions of expert meditators' brains, showing a correlation between the activation decrease and the increase of amount of meditation experience (hours of practice). This latter evidence has been associated with the reduction of the effort necessary to sustain the attentional focus as long as the meditation experience grows. Also OM meditation training showed consistent effects on attentional processes. Specifically, many studies addressing OM meditation centred on *mindfulness* training programmes, which incorporated OM meditation as one of their components.

The debate about the definition of mindfulness is still open (Baer et al., 2009), but in its various conceptualisations it deals with particular qualities of attention and awareness (Kabat-Zinn, 2003). Specifically, Bishop et al. (2004) proposes an operational definition, which sees mindfulness as a state consisting of two components: a self-regulation of attention, and an open, non-judgemental orientation to the ongoing experience. Self-regulation of attention, and orientation to experience achieved in a mindfulness state imply sustained attention, attention orienting, and non-reactive awareness of the stream of experience. These skills are trained through forms of meditation, also OM meditation, and are being applied for several years as clinical tools in psychotherapies. Especially, the main reference about mindfulness-based treatments for mental health is mindfulness-based stress reduction (MBSR) (Kabat-Zinn, 2003), which includes a meditation practice ascribable to OM meditation. Jha et al. (2007) assessed with attention network test (ANT) two groups of participants: a group was naïve to mindfulness, and participated in an 8-week MBSR training on concentrative, i.e., an FA meditation type; another group was composed of expert meditators who participated in a 1-month retreat of mindfulness meditation, i.e., an OM meditation type. A waiting list group was used as a control. Improvements in conflict monitoring, and orienting were found in expert meditators prior to the retreat, and in novices after the MBSR training. In addition, expert meditators showed increased alerting scores after the mindfulness retreat. Tang et al. (2007) found that even a short-term training program including OM meditation components induced improvements in conflict monitoring as measured by the ANT (and the reduction of stress-related indexes). These findings together point to the fact that early stages of mindfulness training may improve the function of the dorsal system of attention (orienting, and executive control), whereas intensive medium-to-long-term mindfulness training as in the 1-month retreat condition of Jha et al. (2007) may improve the ventral system function in terms of alerting (see Hölzel et al., 2011, and Chiesa et al., 2011 for a review on mindfulness impact on cognitive abilities).

As anticipated, also the AB has been used to assay the effects of meditation training on attention processing (Slagter et al., 2007; van Leeuwen et al., 2009). The idea generally underpinning these studies is that meditation practice may train attentional abilities that lead to a reduced AB because they allow the release of limited brain resources.

Slagter et al. (2007) conducted a longitudinal study to evaluate the effect on the AB of a 3-month intensive meditation training. Two groups participated in the study: a practitioners group, and a control group. Practitioners had previous experience in several different meditation styles, and participated in a 3-month retreat, which foresaw 10–12 hours of meditation per day. The meditation practised in the retreat was an OM meditation, supposed to ultimately broaden the attentional focus, and cultivate a non-reactive form of attention that does not judge, or affectively respond to sensory and mental stimuli. These attributes of attention are similar to those included in the definition of mindfulness (Bishop et al., 2004). The control group participants did not have any prior experience with meditation, and took a single 1-hour meditation class. Then, they were instructed to meditate 20 minutes per day in the week preceding the experimental session. The experimental task was a classical AB paradigm: the targets were two black numbers embedded in a stream of 15–19 letters. Two target onset asynchronies were tested: a short one, 336 ms (lag 4), and a long one, 672 ms (lag 8), and stimuli were presented for 50 ms each, with a 34 ms inter-stimulus interval. Experimental trials might be either single-target or dual-target, in order to accurately evaluate the effect of meditation training on T1 processing. Both behavioural, and electrophysiological data were collected and analysed: namely, T1 and T2|T1 report accuracies, and the T1-elicited P3b component of scalp-recorded event-related potentials (ERP) acquired during the streams presentation.

The results in Slagter et al. (2007) were a smaller AB, and a reduced amplitude of T1-elicited P3b potential, with a positive correlation between the size of AB reduction, and the extent to which P3b decreased. These findings corroborate the argument that AB depends upon the deployment of attentional resources to T1, and show that a pure mental training consisting of 3 months of intensive OM meditation practice can allow for an increased control over the allocation of attentional resources to the processing of T1. Specifically, these conclusions agree with those in Green et al. (2003), and extend them because based on a pure mental training rather than an action video game. In addition, they are also compatible with the hypothesis that an over-investment of attentional resources on T1 may cause the AB, as opposed to a diffuse attentional state, which would counteract the consequences imposed by the limited attentional resources when they need be shared between rapidly succeeding targets as in a RSVP (see Olivers and Nieuwenhuis, 2006, and the section 5.1 for more details about the over-investment hypothesis). According to Olivers and Nieuwenhuis (2006), the diffuse attentional state would distribute attentional resources without allocating too a strong focus on the task, and this would prevent non-target items from accidentally accessing the capacity-limited stage of processing, and interfering with targets in that stage. In Slagter et al. (2007) the diffuse attentional state could be induced by the OM meditation retreat, and no over-investment of attention to T1 would be demonstrated by the reduced amplitude of T1-elicited P3b potentials.

The cross-sectional study of van Leeuwen et al. (2009) assessed age-related effects on the AB in meditators, starting from the consideration that attentional abilities, and also the performance in the AB, drop with age. Specifically, they conducted a study on three groups of subjects: two older, age-matched groups (mean age around 50 years old), one composed of long-term meditation practitioners, and one with no prior meditation experience serving as a control; and a third control group of younger participants who never engaged in meditation (mean age around 24 years old). The meditation practitioners regularly practised both FA and OM meditation, with a wide range of experience, spanning from 1 to 29 years. All the groups participated in two blocks of trials, and their task was to identify two numbers embedded in a stream of black letters. T1 was red, while T2 was black. Each visual item in the stream was presented for 100 ms, and lag 1–lag 7 were tested.

The results of this study by van Leeuwen et al. (2009) confirm the authors' prediction of an aid offered by long-term meditation practice to overcome age-related decline in the AB performance. In fact, the meditators group showed a smaller, and shorter AB than the age-matched control group on lag 2, lag 6, and lag 7, and even a better performance on lag 2 than the younger control group. The better performance on long lags may be related to an increased ability in sustaining attention.

In Slagter et al. (2007) the meditation was based on the OM meditation practice, although participants had previously practised several types of meditation. In van Leeuwen et al. (2009) the participants in the experimental group practised both OM meditation and FA meditation. As discussed by van Leeuwen et al. (2009), the comparison between the two studies seems to suggest that OM meditation prevents over-investment of resources, and a quick disengagement of attention from T1, whereas FA meditation might be responsible for the increased ability of expert meditators to sustain attention, leading to a better performance on longer lags. However, it has not been possible neither to directly connect the attentional improvements to a specific meditation style, nor to correlate the improvements to the amount of meditation practice. This is due to the fact that practitioners came from different meditation experiences in Slagter et al. (2007), and did meditate for quite different durations at different times in van Leeuwen et al. (2009).

4.1.2. The Present Study

The findings summarised in previous paragraphs may be globally interpreted as an indication that meditation is a form of mental training that is able to train attentional processes. In particular, it can induce changes in how attentional resources are shared over time, and across targets and distractors, as measured by the performance in an AB task. A possible explanation for this evidence is that a more distributed attentional focus is induced by OM meditation practice, which reduces the investment of attentional resources to the first target in a dual-target RSVP.

Mindfulness is a kind of OM meditation, and *Tsa Lung* is a particular Tibetan Yoga practice, which combines body exercises typical of other Yoga traditions, with mindfulness meditation (see section A.1). Our Study 1, presented in next sections of this chapter, is a longitudinal study evaluating whether a 7-week Tsa Lung training programme on a group of adult meditators induces changes in AB performance. The prediction to be tested was that AB performance at post-training should be better than at pre-training. The peculiarity of the study compared to those of Slagter et al. (2007), and van Leeuwen et al. (2009), and, in general, to those assessing the effects of meditation on attentional processes (see Chiesa et al., 2011) is that, along with a mindfulness meditation practice, the training included also a mild physical exercise. Namely, in Tibetan Yoga, the mindfulness state is achieved through body exercise, and OM meditation. Hence, these are the components of the training evaluated in the present study.

A previous study by Tang et al. (2007) found that executive control was improved as measured by ANT through an integrative body-mind treatment. This treatment, however, although including relaxation, mental imagery, and mindfulness training, was different from Tibetan Yoga, because the latter includes also postures to be maintained, associated with visualisations and controlled breathing. Tibetan Yoga has been studied on patients suffering from lymphoma, and it showed to reduce sleep-related disturbances (Cohen et al., 2004). However, to our knowledge, Tibetan Yoga has not been studied in terms of its potential effects on attention processes.

The objective of the present study was just to explore whether a 7-week Tibetan Yoga training delivered to habitual meditators had some effect on the AB performance. In the affirmative case, this result would add to the evidence that a mindfulness training is able to change attention allocation over time, with the peculiarity that this particular training — Tibetan Yoga — also incorporates a mild physical exercise among its components.

4.2. Study 1 Material and Methods

Participants

25 adult meditators participated in the present study, but only 21 complied with the training requirements, and underwent both experimental sessions (see the description in *Procedure*).

The 21 participants who completed the Tibetan Yoga training were 9 males, and 12 females (median age 44 years old, ranging from 26 to 64 years old). They all had either normal or corrected-to-normal vision. They all came from the Lama Tzong Khapa Institute in Pomaia (Italy), where the two experimental sessions, prior to and after the training programme respectively, and the classes for explaining the training programme requirements and operations took place. These participants were all familiar with meditation, and they habitually practised different styles, categorised as both FA and OM meditation practices.

The participants provided self-reports of demographic information, and about the amount of training (physical and meditative) undertaken during the training period. They took part in two experimental sessions: one used as the baseline, immediately prior to starting the Tibetan Yoga training, and another just after the end of the 7-week training programme. Both sessions lasted for two experimental days. No control group was used in this study. The duration of an experimental session for a single participant was approximately 30 minutes.

Tibetan Yoga Training Programme

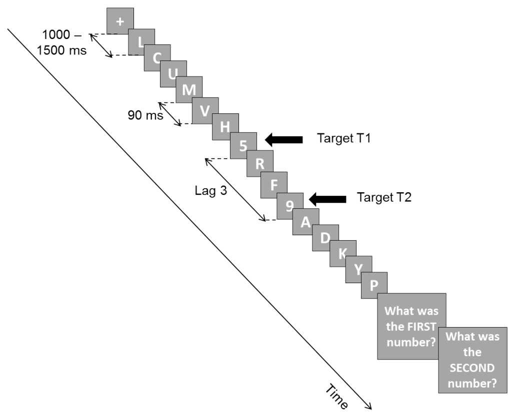
All participants performed different amounts of Tibetan Yoga exercises and meditation, but all for a minimum of 2 hours per week, over a period of 7 consecutive weeks. The physical component included 5 exercises, while the mental component required to engage in a kind of OM meditation, credited to the rubric of mindfulness. In fact, the result of this practice should develop the ability to achieve a mindfulness state as described in previous paragraphs of this section. A summary description of the Yoga exercises resembling how they were explained to the participants is included in section A.1.

Stimuli, Task, and Apparatus

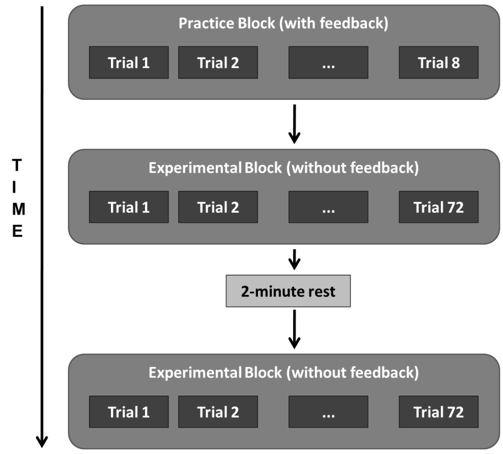
Participants' task in the two sessions was to identify, with no time pressure, two single-digit white numbers (the T1 and T2 targets) appearing in a stream of white letters (the distractors). Two types of trials were used: short-lag trials, in which the second target digit (T2) appeared in serial position 3 (i.e., at lag 3) after the first target digit (T1), and long-lag trials, in which the second target digit (T2) appeared in serial position 7 (i.e., at lag 7) after T1. All the visual stimuli in a trial appeared sequentially for 90 ms each, with no inter-stimulus interval, and at the same location, in the centre of a 24-inch LCD computer screen, against a grey background (Fig. 4.1a). The whole experiment, including visual stimuli presentation, and participant's responses collection, was implemented in Matlab®(MATLAB R2011b, The MathWorks Inc., Natick, MA, USA), using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Procedure

Each participant took part in the 7-week Tibetan Yoga training programme, and in two experimental sessions. Besides, participants provided information about the hours of meditation practised at baseline, and during the training period. The pre-training session was to measure the baseline AB performance, whereas the post-training session served to collect the data for comparison with those at the baseline, in a within-subject design.



(a) Graphic representation of a typical trial in Study 1.



(b) Graphic representation of an experimental session in Study 1.

Figure 4.1.: Study 1 – Graphical description of experimental task, and sessions structure.

The two sessions were composed of three blocks of trials: the first was a practice block, including 8 trials with feedback about the correctness of the responses, used to familiarise participants with the experimental equipment, and task. The second and third blocks included 72 trials each, and each trial consisted of an RSVP of 2 single-digit numbers drawn without replacement from the range 2–9, embedded among 13–17 different letters. In every block half of the trials were long-lag, and half were short-lag, and they were randomly intermixed within each block. The participant's task was, at the end of each trial, to answer to the two questions "What was the FIRST number?", and "What was the SECOND number?" by typing the respective target on the numeric keypad of a computer keyboard. Participants had no time limit to make their responses, and were instructed to type a '0' only when they were sure to have not seen a target.

Data Analyses

Only participants' responses in the two experimental blocks of the two sessions were collected and analysed in order to answer to the following research questions:

1. Is an AB present in the tested population at baseline?
2. Does the meditation experience of participants at baseline affect the AB performance?
3. Does the body-mind training based on Tibetan Yoga modify the AB performance?

For questions 1. and 3. separate general linear models (GLM) were implemented to conduct repeated measures ANOVA, using "Lag" as the single within-subject fixed factor for addressing the first question, and "Lag" and "Session" as the fixed within-subject factors for addressing the third question. Both “Lag” and “Session” had 2 levels (lag 3/lag 7, and Pre-/Post-training). The primary outcome measure was the proportion of correct reports for the second target (T2), conditional upon the correct reports of the first target (T1), averaged across the trials of the two experimental blocks for each participant. This variable is referred to in the following as P(T2|T1). The mean proportion of T1 correct reports averaged across the trials of the two experimental blocks for each participant is considered as well, and referred to as P(T1). Indeed, the AB consists of a decrease of P(T2|T1) relative to P(T1) at short lags.

For question 2. a separate 2-way mixed ANOVA was carried out. Participants were split in two groups based on the average amount of meditation practised per week in the 4 weeks prior to starting the Tibetan Yoga training (t_{M0}). One group comprised participants for whom $t_{M0} > 20\,hrs$, and another group included the remaining of them. Such groups yielded the 2-level between-subject grouping factor “Meditation experience at baseline”, which was included in the ANOVA model along with the “Lag” factor. The response variable in the ANOVA was again the T2|T1 report accuracy resulting from the first experimental session, i.e., P(T2|T1) at baseline.

4.3. Study 1 Results

Is an AB present in the tested population at baseline?

The results of the one-way repeated measures ANOVA with “Lag” as a single fixed factor revealed a statistically significant main effect of “Lag” (p-value < 0,001) for both T2|T1 accuracy and T1 accuracy. As depicted in Figure 4.2, T2|T1 accuracy is lower than T1 accuracy at lag 3 (270 ms), the short SOA, whereas it is similar to T1 report accuracy at lag 7 (630 ms). Therefore, the population of adult meditators recruited for the Tibetan Yoga training in fact shows an AB deficit prior to the training when confronted with the RSVPs delivered in the present experiment.

Does the meditation experience of participants at baseline affect the AB performance?

However, these meditators came from different experiences, and amounts of practice. Especially, it is interesting to test whether their performance in the AB task at baseline is affected by the amount of meditation practised in the 4 weeks immediately prior to the first session day. The results of the separate ANOVA analysis adopting “Meditation experience at baseline” as a further between-subject factor showed no significant effect (p-value = 0,711) of the latter on T2|T1 mean accuracy, and no significant interaction with “Lag” (p-value = 0,604), which instead had a statistically significant effect (p-value < 0,001).

Does the body-mind training based on Tibetan Yoga modify the AB performance?

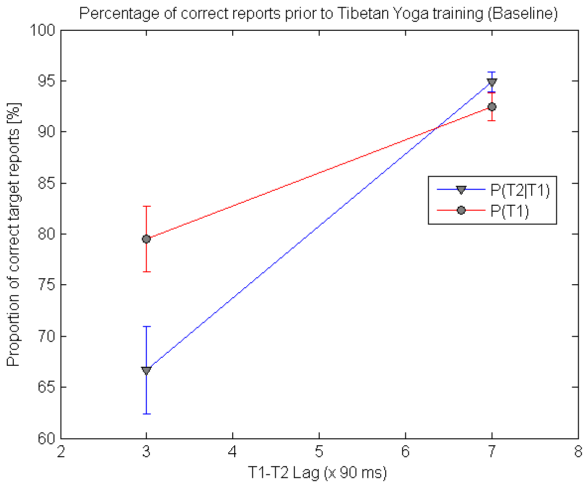


Figure 4.2.: Study 1 – Mean proportions of T2|T1, and T1 correct reports versus T1–T2 lag level. T2|T1 accuracy at lag 3 (66,662 %) is significantly lower than at lag 7 (94,895 %), and at lag 7 it is similar to T1 accuracy. The latter is greater also at lag 3 (79,497 %). Error bars represent standard errors of the mean.

The most important analysis for the present experiment is to test whether a difference exists in the AB performance of participants between baseline, and the end of the Tibetan Yoga training programme. The 2×2 "Lag"×"Session" repeated measures ANOVA yielded the results reported in Table 4.2, based on the descriptive statistics in Table 4.1.

P(T2|T1) – Cell means, Marginal means, and Grand mean			
	Session		
Lag	Baseline	Post-training	Marginal Means
3	66,662 %	77,086 %	71,874 %
7	94,895 %	95,377 %	95,136 %
Marginal Means	80,779 %	86,232 %	$\mu_G = 83,505\,\%$

Table 4.1.: Study 1 – Summary of the means of proportions of correct trials (T2|T1) in the different experimental conditions (μ_G denotes the grand mean).

Both "Lag" and "Session" resulted to have significant main effects on T2|T1 report accuracy, and also a significant interaction of these factors was found. The main effect of "Session" on T2|T1 was detected, following an increase of the mean T2|T1 accuracy across lag levels from 80,78 % to 86,23 % (grand mean $\mu_g = 83,51\,\%$). As for the main effect of "Lag", T2|T1 mean accuracy across baseline, and post-training sessions, spanned between 71,87 % at lag 3, and 95,14 % at lag 7.

The interaction plot is shown in Figure 4.3a. A substantial invariance is measured at lag 7 for T2|T1 accuracy between baseline session and post-training session (94,90 % at baseline, and 95,38 % at post-training), whereas a larger difference is found for lag 3: 66,66 % at baseline versus 77,09 % at post-training. Actually, *post-hoc* tests

(a) Summary statistical results of Lag×Session repeated measures ANOVA on T2|T1 accuracy.

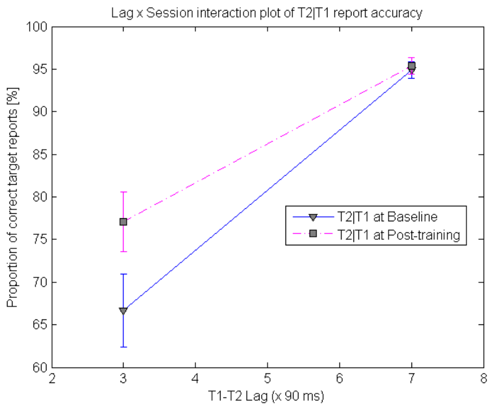
Factors	F-ratio	p-value
Lag	$F_{1,20} = 18,16$	<0,001
Session	$F_{1,20} = 13,83$	0,001
Lag × Session	$F_{1,104} = 14,88$	<0,001

(b) Summary statistical results of Lag×Session repeated measures ANOVA on T1 accuracy.

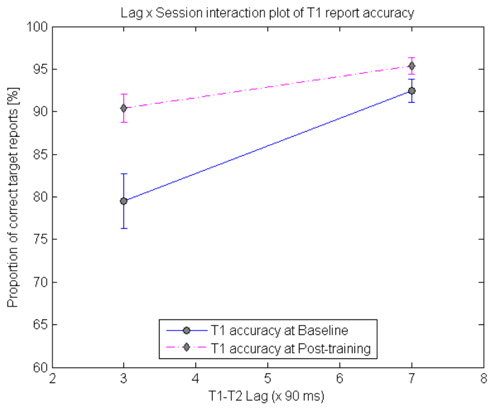
Factors	F-ratio	p-value
Lag	$F_{1,20} = 29,12$	<0,001
Session	$F_{1,20} = 16,29$	0,001
Lag × Session	$F_{1,104} = 15,50$	<0,001

Table 4.2.: Study1 – Results summary of 2×2 Lag×Session within-subject ANOVA on T2|T1 and T1 accuracies.

according to the Tukey's method showed that the difference between sessions at lag 7 is not significant, whereas at lag 3 it is statistically significant. Thus, there is a statistically significant increase in T2|T1 report accuracy at lag 3 following the Tibetan Yoga training, or, putting it in different terms, there is a reduction in the AB magnitude following the Tibetan Yoga training. The *post-hoc* analysis on T1 report accuracy proved a significant difference between T1 accuracy at baseline at lag 3 (79,497 %), and T1 accuracy after the 7-week Tibetan Yoga training completion, when T1 report accuracy at lag 3 is 90,41 %.



(a) Study 1 – Lag×Session Interaction Plot of mean T2|T1 report accuracy. Error bars represent standard errors of the mean.



(b) Study 1 – Lag×Session Interaction Plot of mean T1 report accuracy. Error bars represent standard errors of the mean.

Figure 4.3.: Study 1 – Lag×Session interaction plots for means of T2|T1 report accuracy, and T1 report accuracy.

4.4. Study 1 Discussion

The objective of Study 1 can be summarised as to evaluate changes in the allocation of attentional resources over time in adult habitual meditators, following a 7-week

Tibetan Yoga training. How attentional resources were allocated was measured through the performance in terms of AB magnitude in dual-target RSVPs, namely looking at the mean report accuracy of T2|T1. The expected outcome was that the Tibetan Yoga training improves the AB performance of participants, i.e., their mean T2|T1 report accuracy should increase after the training completion, in the condition where T2 is in serial position 3 after T1 (lag 3, 270 ms). The peculiarities of the Tibetan Yoga training underpinning this prediction were essentially two: (1) an OM meditation component, which can be categorised as a mindfulness practice, was part of the training programme, and as such it was a kind of mental training that had already proven to be capable of changing the AB performance (Slagter et al., 2007; van Leeuwen et al., 2009); (2) The mild physical exercise, performed as a preparatory activity to meditation, and accompanied by mental visualisations, should contribute to produce a similar effect on the distribution of attentional resources. However, the amount of meditation practised in the weeks prior to starting the Tibetan Yoga training might have affected participants performance in the AB task at baseline.

The results of the statistical analyses carried out on the collected data showed that an AB is indeed present prior to the Tibetan Yoga training start, but that it was not affected by an High vs Low "meditation experience" factor based on the average amount of hours of meditation practised per week in the last 4 weeks preceding the baseline experimental session. This is in agreement with the findings that AB is a robust deficit, and with the prediction that it can be produced by a dual-target RSVP, where letters are to be ignored, and two numbers are to be reported, all flashing at a rate of about 11 items/s.

Also the prediction that a 7-week Tibetan Yoga training including a mild physical exercise preparatory for the mindfulness meditation is able to change the AB performance was confirmed. This result adds to the evidence that mindfulness practices are capable of alleviating — but not eliminating — the AB, reflecting a more efficient distribution of limited resources over time, and among targets and distractors. The behavioural measure adopted in this study, however, is not able to determine *per se* the exact temporal dynamic of the processing connected to the two targets, as an event-related potential study could do. Nonetheless, another finding arising from the present study was that also on T1 accuracy there is a main effect of Lag, and Session, as well as an interaction effect of those factors. Specifically, T1 accuracy at training completion for lag 3, i.e., for a T1–T2 SOA causing the AB, increased as it had happened in Slagter et al. (2007). This pattern, with both T1 and T2|T1 report accuracies increasing, indicates that T2|T1 accuracy improvement did not come at the expense of a T1 report impairment. Consequently, these results are compatible with accounts of the AB that do not entirely ascribe T2 impairment to depletion of limited resources on T1 identification, and consider the possibility that an over-exertion of control (Taatgen et al., 2009), or an over-investment of attentional resources on the stimuli (Olivers and Nieuwenhuis, 2006) might be the causes of the AB. Besides, it seems unlikely that after 7 weeks the better performance in the dual-target task was determined by some form of perceptual learning, or even by participants' ability to predict the lag level in different trials, which had been randomly distributed in each experimental block.

One interesting observation spontaneously made by several participants was that if they actively tried to lessen the focus on the RSVP, while "opening" their attention, just as explained in the descriptions of OM meditation, and mindfulness, they felt to achieve a higher reporting accuracy. This fact has two consequences: firstly, it seems to go in the same direction of the hypothesis put forward in Olivers and Nieuwenhuis (2006), according to which a more diffuse attentional state, corresponding to a lessen focus on the RSVP, should lead to an alleviation of the AB, independently of how such a diffuse attentional state has been achieved. Secondly, it should be taken into account that unfortunately some participants' effort to adopt an OM state might itself have confounded the final outcome. Indeed, the resulting improvement of the AB performance might depend on both the Tibetan Yoga training, and the OM meditation-like state possibly practised during the experiment. The hypothesis that a diffuse attentional state may benefit the AB performance is at the root of part of our Study 2, presented in chapter 5, where it is assessed whether an auditory distraction, which is supposed to foster a more diffuse attentional state, has on the AB magnitude a similar effect as an immersive 3-D virtual environment meaningfully connected with the visual stimuli used in the AB paradigm. Details are presented in chapter 5.

A final remark is about the distinct contributions, if any, of the physical exercise component and the meditation component of the training, to the effect on the AB magnitude. A further step for extending this study would indeed be to try isolating the contributions to AB alleviation of the physical exercise component, and the mental component of Tsa Lung Yoga. In fact the physical exercise is a peculiar component of this kind of training, and of other integrative body-mind training, and it might represent a specificity of these training protocols compared to pure mental training.

5. Study 2 – AB with Road Signs: an Exploration in Virtual Environments

In Study 2 we explored the Attentional Blink (AB) with two specificities: first, the visual stimuli presented in the dual-target RSVP are road signs, showing arrows pictograms as distractors, and pictograms of animals and vehicles as targets; second, one of the three tested display conditions presents the RSVP in a 3-D VE representing a motorway junction with the rapid stream of stimuli appearing on a virtual portal-like signpost. The other display conditions are classical 2-D presentations of the road signs RSVPs on a desktop computer screen, either with or without the concurrent delivery of an auditory distraction. The auditory distraction is the sound of a car engine, which can be either idle or accelerated. The AB performance in the different conditions is evaluated considering three SOA between the two targets, including when they are consecutive (the lag-1 level). This chapter describes the rationale for the study, and presents the detail of the conducted experiment in terms of procedure, methods, results, and conclusive discussion.

5.1. Study 2 Rationale

A recent interesting finding of the AB research is a concurrency benefit alleviating the AB magnitude found when one of different kinds of concurrent task-irrelevant activity is added to the primary RSVP task. Visual motion superimposed on the visual stream, a concurrent sound or music, and free mental association have been tested and generated an attenuated AB magnitude (Olivers and Nieuwenhuis, 2005; Arend et al., 2006; Taatgen et al., 2009; Lapointe-Goupil et al., 2011). These effects have been explained in terms of either cognitive flexibility or a more distributed form of attention, which overcome the limitations that usually induce the AB.

The presentation of stimuli in a 3-D immersive VE that represents a meaningful context for the visual items used to elicit the AB seems to be a special sort of distraction whose effect is worthwhile investigating and comparing with a more standard distraction, like that provided by a concurrent sound in a 2-D display condition.

Besides, as discussed in Study 1, an attentional distribution similar to that trained with Tibetan Yoga including mindfulness meditation has shown to alleviate AB after a short-term practice of a few weeks. We argue this is another finding compatible with the view of an important role on AB modulation of personal cognitive strategies applied to attention distribution.

Moreover, it is interesting to start answering the more general questions about whether the time course of attention in a virtual environment is similar to that targeted by a classical AB experiment, which is undertaken looking on 2-D symbols on a computer screen.

Finally, the investigation of AB performance by using less abstract stimuli, even incorporated in a meaningful related context, as the use of a VE makes possible, is interesting *per se*. Indeed, this investigation may add to research works connecting natural scenes perception and AB (Evans and Treisman, 2005; Einhäuser et al., 2007; Potter et al., 2010), studies elaborating on possible functions of the AB (see Martens and Wyble, 2010 for a brief review), and future hypotheses to use VEs for the assessment of temporal aspects of attention in more ecologically valid settings.

5.1.1. AB with Less Abstract Stimuli

Visual AB arises with a variety of stimulus types, including alphanumerical symbols, words, and pictures. However, different characteristics have been found when using abstract symbols, words and pictures of natural scenes faces in the RSVPs. Effects on AB magnitude and duration of targets category in a single stream, and of whether targets identification or simple detection is required, have been found by Evans and Treisman (2005) and Einhäuser et al. (2007); and also emotional stimuli affected performance in an AB paradigm (Stein et al., 2010; Arnell et al., 2007). In order to understand if the AB has a specific function, and if its suppression, either inherent or induced, comes at some cost, it would be helpful to capture the AB in everyday life. One way to stride in this direction is to generalise the abstract stimuli usually adopted in laboratories to elicit the AB. VEs can be very useful in this sense, since one of their distinctive properties is the possibility they offer to provide rich stimuli, which maintain a rigorous control of the stimulus-response function, yet enabling for ecologically valid stimulations, more inclined to be generalised to daily living performance. Nonetheless, at the best of our knowledge no study of the effect on AB of a 3-D immersive presentation in a VE of the RSVP has been conducted until now.

Evans and Treisman (2005) studied the perception of objects in natural scenes by means of RSVP paradigms, starting from an apparent conflict. On the one hand, humans are able to grasp the gist of a complex everyday scene, and assess some of its high-level properties, especially if they are meaningful or threatening, in a fraction of a second (less than 100 ms). On the other hand, a marked impairment — they name it a capacity limitation — of visual attention emerges when two targets as simple as alphanumeric characters appear a few hundred ms from one another in a RSVP. Specifically, Evans and Treisman (2005) explored whether an AB arises when the task is to detect and/or identify two targets embedded in natural scenes appearing in close temporal proximity (220 ms – 880 ms). The two targets might belong to the same category (either two animals or two means of transportation), or to different categories (one animal and one means of transportation). The results of Evans and Treisman's series of experiments were that the performance in online detection of targets was good, independently of whether target categories were congruent or not. Instead, an AB arose for identification, and it was more marked in magnitude and duration when the two targets belonged to different categories, and participants did not have prior knowledge about such categories. No lag 1 condition was tested, but they found an AB up to a SOA of 880 ms, which was more profound for targets belonging to different categories, than in a same-category condition. Conditional T2 accuracy rates varied approximately between 60 % and slightly more than 90 % over the SOA span used. Another series of experiments, which used RSVPs of natural scenes at higher rates (about 13 items/s) than those used to study the AB (about 9 items/s), showed that natural scenes categorisation can occur at a pre-attentive stage. The conclusions of Evans and Treisman (2005) were in favour of a view according to which detection is possible at a pre-attentive stage, based on the presence of a set of disjunctive features detected in parallel, without the binding process, and the object representation that requires the selection from attention. Thus, it is possible to semantically identify a relatively complex object in a natural scene even only with some of its features detected in parallel; however, if the target object is no longer available when attentional resources can be drawn to bind its features and make its representation available to conscious awareness, the identification is more prone to errors, because the features binding stage is serial. In terms of AB, one prediction supported by these experiments is that when T1 and T2 belong to the same category, there is a priming effect of T1 features on T2 detection that is more pronounced when participants know in advance targets categories, but that still provides a sort of passive advantage when participants ignore T2 category. Thus, the same-category condition triggers in any case a less deep AB than the different-category condition.

Einhäuser et al. (2007) presented RSVPs of natural photographic stimuli at different rates (6 Hz – 40Hz), including two to four targets. Targets identification was then tested at the end of the RSVP for two of the targets with a 2-alternative forced choice between similar exemplars belonging to the same category. Target category might be either "faces" or "watches", and the questions used for requesting the 2-alternative forced choice — "Which watch?", "Which face?" — did not reflect the order of appearance of the two categories in the RSVP. As to the within-category AB, this experiment found it for both faces and watches as a dip at short T1–T2 SOAs (around 300 ms), but with the watches category displaying a longer and slightly delayed AB (the AB impairment lasted until about 590 ms for "watches", against about a SOA of 410 ms for "faces"), suggesting a category-dependence of the phenomenon. The results across categories attested an effect on the AB of the target category, and of the difference in category between the two targets. Namely, T2 recognition performance was better when T1 belonged to a different category. This finding is not aligned with the results in Evans and Treisman (2005), where the AB increased in depth and duration when T1 and T2 were drawn from different category sets. Einhäuser et al. (2007) does not specifically address this difference in the results of the two studies, however, they highlight the differences between them: the use of different categories (faces and watches vs animals and means of transportation); the use of a single, quite low (8 Hz) presentation rate in Evans and Treisman (2005) versus a plurality of (higher) presentation rates in Einhäuser et al. (2007); the forced choice decision for target identification in Einhäuser et al. (2007) against a free sub-category identification in the other experiments series; and finally, the different method used to analyse the data and draw conclusions.

Dux and Harris (2007) used line drawings of familiar objects as visual stimuli in RSVPs to study whether viewpoint costs arise during an initial identification stage, or during consolidation. To this end, either distractors or T1 were rotated in separate experiments. The results revealed a robust AB, unaffected by distractors orientation, but augmented by a 90° rotation of T1. Distractors orientation did not affect the AB magnitude for both colour-defined and semantically-defined targets. This evidence was interpreted as an indication that preliminary recognition of familiar objects is view-point invariant, and view-point costs arise in the consolidation stage. Interestingly, no lag-1 sparing effect was found, and no explanation is offered for this observation.

Potter et al. (2010) tested the ability of identifying the specific category of two target pictures appearing in an RSVP among pictures of single objects and scenes, when the superordinate category is communicated to the experiment participant immediately prior to every trial. Lags 1, 2, and 4 were tested, corresponding to 110 ms, 213 ms, and 427 ms respectively. Colour pictures were used. The objective was to test whether the lag-1 sparing effect is present when the targets are pictures never seen before, and only their superordinate category (e.g., a fruit, a vehicle, etc.) is known just a little in advance of the trial by the participant in the experiment. The two targets of a trial belong to the same superordinate category. The results of these experiments showed that an AB is present at lag 2 and lag 4, when the recovery from the attentional deficit is started. Furthermore, a lag-1 sparing effect is present, and the authors consider this evidence against the theory proposed by Evans and Treisman (2005), and in favour of a transient attention account for lag-1 sparing, as proposed by Wyble et al. (2009). Actually, Evans and Treisman (2005) did not test the lag 1 condition, but as the identification stage is serial, no lag 1 sparing should occur in their model, as T1 and T2 cannot be identified simultaneously. This is against the finding in Potter et al. (2010), which instead agrees with the hypothesis that transient attention triggered by T1 "opens" an attentional window such that if T2 appears within this window, i.e., 150 ms since T1 onset, it can be identified as it is the case of lag-1 sparing. The same transient burst of attention is responsible of the AB, in that it is suppressed around 150 ms after T1 onset in order to separate the encoding of two different targets, and preserve episodic information. When T2 appears during the suppression, an AB arises.

These studies, although starting from differently conceived experimental designs, converge on two conclusions: (i) the generalisation of AB results found with simple abstract stimuli to the AB manifestations with natural stimuli is possible, yet not straightforward; and (ii) targets category congruency, or even target category type, affect the magnitude and duration of the AB.

The AB performance with complex stimuli has been also studied using faces as targets. Marois et al. (2004) found an AB for the detection of pictures of indoor/outdoor scenes used as T2, when a face appeared as T1, and scrambled scenes were used as distractors. T2|T1 accuracy was tested for T1–T2 SOAs of 200 ms, 400 ms, and 800 ms, with the impairment that was particularly prominent for T2 appearing at 200 ms and 400 ms within T1 onset. Awh et al. (2004) challenged the view of a central amodal bottleneck at the root of the AB by investigating the effects of digits as T1 on faces and letters as T2 in RSVPs. The main finding was that faces discrimination as T2 was not impaired with the same digits discrimination task at T2, which in contrast impaired letters discrimination at T2. These results were proposed as the grounds to propose the multi-channel model account for the AB. The latter assumes the existence of separate channels for processing feature-based and configural information respectively, hence predicting that AB interference arises only when T1 processing occupies every channel available for processing T2. In the case of faces as T2 and digits as T1, since digits are discriminating using only the feature-based channel then the T2 faces have the configural channel available during the AB period, and do not suffer from the interference as in the T2 letters case.

Faces were also used in RSVP paradigms for studying the effect of emotional-laden stimuli on attentional resources by using the AB.Stein et al. (2010) used an emotionally neutral face as T1, and either a single happy or fearful face as T2. Scrambled faces were used as distracting fillers in the RSVP. The perceptual load of T1 was manipulated by using two intact faces flankers of T1 face (which was presented centrally at fixation). In the low-load condition the flankers and target were the same face, whereas in the high-load condition the two T1 flankers, and T1 itself were three different neutral faces. The task was to identify T1 gender, and detect T2 presence, without any prior knowledge on its emotional connotation. The results of the experiment demonstrated an AB attenuation for fearful expressions relative to happy faces in the T1 low-load condition, while in the T1 high-load condition the advantage of fearful faces disappeared, but the AB effect on the face detection accuracy remained, and affected similarly both happy and fearful facial expressions. The proposed explanation was that the extraction of emotional meaning from visual stimuli requires attentional resources, which are made unavailable when T1 perceptual load becomes high.

Arnell et al. (2007) found that an involuntary AB can be initiated by sexual or taboo words used as a critical distractor in a RSVP where the task is to remember a single different target. In these situations the high-arousal distractor word acted as a T1 in the RSVP, resembling an AB interference on T2. A similar pattern emerged when using negative pictures as distractors Most et al. (2005). However, current evidence is rather in favour of a role for arousing quality of stimuli in capturing attention, and inducing an AB.

Overall, in light of the studies conducted so far on the AB using natural scenes, faces, and words as stimuli, it seems clear that the AB can be induced also by such kinds of stimuli. Nonetheless, there are factors that may modulate its magnitude, postpone its manifestation, and affect its duration. Such factors are essentially related to the task requirement, either the detection or identification of a target, to whether the two targets belong to different categories, to the similarity of the features composing the categories, to the arousal in emotional terms of the stimuli, and to the perceptual load associated to T1. A role for the elicitation and characteristics of an AB with natural stimuli might also be played by the familiarity of the targets, the variety of the experimental set, and the level of specificity of the category required to define the targets, as briefly discussed in Potter et al. (2010).

5.1.2. Concurrency Benefits in AB Performance

A beneficial effect of concurrent task-irrelevant activity has been found on the AB in works conducted by separate groups, and with several types of concurrent tasks. Olivers and Nieuwenhuis (2005) found that T2 detection accuracy improved relative to a group performing a standard AB paradigm in groups whose participants were instructed either to think of their holidays or listen to a rhythmic tune during the RSVPs. The hypotheses provided by those authors to explain their results were three: an effect of overall arousal induced by the free-association thoughts about holidays and the rhythmic tune, a positive affective state, and an inherent widening of attention up to the inclusion of T2 in selective processing, triggered by the multi-tasking requirement itself. The same group of researchers in a successive study Olivers and Nieuwenhuis (2006) tried to elucidate the mechanisms underlying the beneficial effect of distraction on the AB magnitude. For this purpose, they investigated the effect on AB of a concurrent task, a concurrent stimulus, and the voluntary lessening of concentration during the RSVP. The concurrent task was to retain in memory, and recognise at the end of the RSVP a simple lines pattern. The concurrent stimulus consisted in a positive affective picture briefly displayed before the RSVP. Finally, the third experiment was to compare a group of participants undertaking the usual AB paradigm, with another performing the same, but with the instruction to reduce the concentration on the task during the RSVP. In all of these conditions an improvement was obtained in T2 detection performance.Olivers and Nieuwenhuis (2006) explained these findings with two hypotheses: the overinvestment hypothesis, and the positive-affect hypothesis, which both ascribe the improved T2 detection to a more diffuse attentional state. According to the over-investment hypothesis the probability of an AB is increased when attentional resources are *over-invested* on the RSVP. In such a case indeed, more distractors gain access to a capacity-limited stage of processing, where they interfere with targets encoding and may impede their consolidation for subsequent report. On the contrary, when attentional resources are drained by either the additional task or the explicit instruction to reduce concentration on the detection task, fewer distractors reach the capacity-limited processing stage, and targets are less exposed to interference. Additionally, the positive-affect hypothesis suggests that positive affect induce an increased cognitive flexibility, and putatively also short-term memory capacities, so as to reduce the AB because of a more flexible orienting towards multiple and/or novel stimuli. The most important obstacle to the reconciliation of these theories with the bottleneck theories proposed for explaining the origin of the AB, is the claim of AB as the consequence of a fundamental limitation in human cognition. As we have seen in chapter 3, this notion of a fundamental cognitive limitation is now criticised by further evidence, and eliminated in some AB accounts (see for example Martens et al. (2006); Olivers and Meeter (2008); Taatgen et al. (2009)).

Further evidence about the decrease in AB magnitude induced by concurrent task-irrelevant activity has been collected by several other research groups. Arend et al. (2006) studied whether a spatial manipulation of the visual background for the RSVP can induce an effect on AB magnitude similar to that revealed by Olivers and Nieuwenhuis (2005). To this aim, Arend et al. (2006) introduced a task-irrelevant background consisting of either a starfield simulation with dots illusory moving in the depth plane away from the centre of the screen where the RSVP streamed, a starfield moving in the opposite direction, i.e., towards the centre of the screen, or a flickering starfield, creating a distracting background as well, but without any illusion of motion. The results showed again an attenuation in the AB relative to the control condition with static visual background for all the backgrounds, with a more successful attenuation for the outward motion condition. With respect to the study by Olivers and Nieuwenhuis (2005), the authors claim the demonstration that the effect arises also from the same modality to which the AB is presented, and that it definitely depends on decreased, and not on increased attention. In particular, the postulated explanation for the AB attenuation is the reduction in the allocation of attention to T1 as a result of reducing spatial attention to the location of the RSVP.

Another means found successful at reducing the AB was the peripheral dot detection task, added as a secondary task to the dual-target RSVP by Taatgen et al. (2009). In this condition, along with reporting the two targets appearing in the RSVP, participants had to respond to a red dot moving in their peripheral visual field. Also this manipulation attenuated the AB magnitude, and the authors used this evidence to support their model of the AB — the threaded cognition model —, which regards the AB as the manifestation of an overexertion of cognitive control. The secondary task with the moving dot to be detected is able to relax control.

Finally, Lapointe-Goupil et al. (2011) investigated whether AB is attenuated by a concurrent task demanding central resources throughout the entire RSVP. They also examined a possible effect of decision criteria on the AB performance. The concurrent central demanding task was a timing task, which required participants to reproduce for its actual duration, a tone heard during the whole extent of the RSVP. The results demonstrated the ability of such a central demanding concurrent task to alleviate the AB, no modulation of T1 report accuracy was revealed, and a role of shift in decision criterion was detected as participants who first performed the single-AB condition became more liberal at identifying T2 in the concurrent-AB condition. The authors remark these findings are in agreement with both the over-investment hypothesis of Olivers and Nieuwenhuis (2006), and the threaded cognition model of Taatgen et al. (2009). Essentially, it is a demonstration against the inflexible structural bottleneck at the roots of the AB, and in favour of some flexibility in attention limitations.

5.1.3. The Present Study

We undertook the present study, often mentioned as Study 2 in the following, with two main objectives: (1) to test whether an AB deficit is elicited by a dual-target RSVP consisting of road signs as both distractors and category-defined targets, and (2) to test the hypothesis that an immersive 3-D VE has an effect on AB performance compared to a 2-D display condition and to the effect of a concurrent auditory distraction.

The prediction for objective (1) was to find an AB, likely with some difference with respect to the classical findings observed with letters and digits. The main predicted sources of variability in the performance are the categorical definition of the targets with respect to distractors, and the different categories (animal and vehicle) found in every target couples, following the considerations in Evans and Treisman (2005), Einhäuser et al. (2007), and Potter et al. (2010). The prediction for objective (2) was that the AB magnitude in the 3-D VE and in the condition with a concurrent auditory distraction are different from the AB magnitude in the 2-D Desktop visualisation. A specific goal of Study 2 is to explore whether the sign of this difference is the same for distraction-free 3-D stimulation and auditory-distraction 2-D stimulation, which is in turn expected to show an AB alleviation as reliably found in the studies mentioned above (Arend et al., 2006; Olivers and Nieuwenhuis, 2005; Lapointe-Goupil et al., 2011).

5.2. Study 2 Material and Methods

Participants

12 adult participants (2 females and 10 males between 25 and 40 years old, median age 32 years old) participated in the study. They all had normal or corrected-to-normal vision, and were recruited among the personnel, PhD students, and faculty members of the Perceptual Robotics Laboratory (Scuola Superiore Sant'Anna, Pisa, Italy). All participants undertook 3 separate experimental sessions in three different days, each including two blocks of actual experimental trials along with initial training and practice phases to ensure the understanding of the experimental task and procedure (see the *Procedure* paragraph). Every session corresponded to one of three possible display conditions for the experimental task. Participants had to respond to each task trial as specified below. The order of participation in the three sessions was counterbalanced across participants, yielding 2 participants for each of the 6 possible orders of participation in the 3 display conditions (Table 5.1). The duration of an experimental session for a single participant was approximately 40 minutes.

Display Conditions Orders			No. of Participants
"Desktop"	"Distraction"	"mini-CAVE"	2
"Distraction"	"Desktop"	"mini-CAVE"	2
"Desktop"	"mini-CAVE"	"Distraction"	2
"Distraction"	"mini-CAVE"	"Desktop"	2
"mini-CAVE"	"Desktop"	"Distraction"	2
"mini-CAVE"	"Distraction"	"Desktop"	2

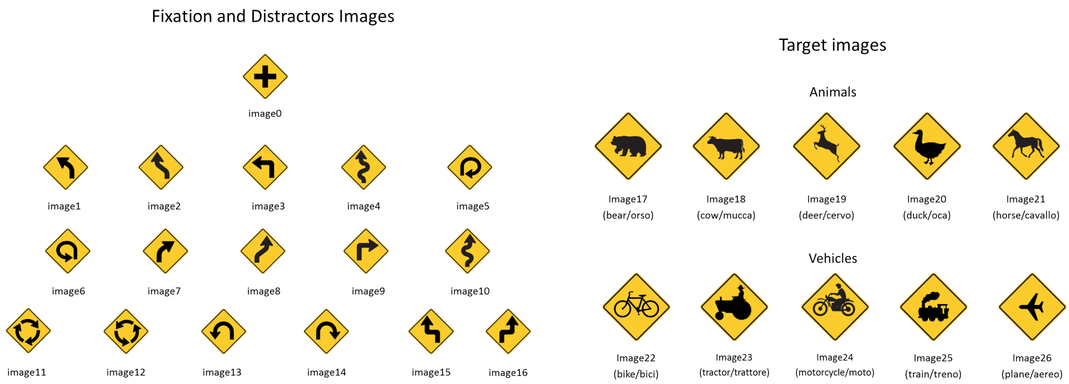
Table 5.1.: Study 2 – Possible display conditions administration orders. Administration order was counterbalanced across participants, so as to have 2 participants per display condition order.

Stimuli and Apparatus

During the experiment, except in the training phase undergone prior to the practice and actual experimental trials of each session, participants watched rapid serial visual presentations composed of 16 non-target stimuli to be ignored (a.k.a. distractors) and 2 target stimuli (a.k.a. targets) to be later reported. The two targets, indicated in the following as T1 and T2 based on their order of appearance in the RSVP, were embedded in the distractors stream and might be presented at a different SOA from one another (lag 1, lag 3, or lag 7, with lag $1 = 90\,\text{ms}$). T1 might occupy either position 5, 6, or 7 in the series. A fixation symbol was displayed at the beginning of each trial. All the mentioned visual stimuli were square road signs, composed of two elements, a yellow background, and a black pictogram (see Figure 5.1). The pictogram was the distinctive element of each stimulus. The complete sets of targets and distractors, and the fixation symbol are depicted in Figure 5.1.

The targets were drawn from two sets including 5 vehicle-targets and 5 animal-targets. In the streams of practice and experimental trials each of the two targets belonged to either the vehicle or animal set, and no two animals or two vehicles were allowed in the same stream. The 16 distractors of each stream were a random permutation of a 16-element set of directional road signs, i.e., road signs showing only arrows pointing at different directions. The fixation symbol was a cross-road sign.

In the training phase practised at the beginning of each session, images of all the targets were displayed at a low pace (presentation duration of 1.5–2 s for every target image) with or without the respective target name, according to the specifications presented in the *Procedure* description.



(a) Fixation (the cross sign) and distractors stimuli used in the RSVP trials. The fixation was always the first symbol in the RSVP trials, while all the distractors appeared in every trial, but in randomised order.

(b) Targets stimuli used in the RSVP trials, and in the training phase. Two different targets, one from the animals set, and another from the vehicles set, drawn randomly, and presented in a random order, appeared in each trial. The SOA between the two targets might equate lag 1, lag 3, or lag 7, i.e., no distractor, two distractors, or six distractors could intervene between the presentation of the first and second target. In the training phase, the target stimuli first appeared one after another with their name at the bottom, then the name was not presented, and the participant had to choose the target name from a list.

Figure 5.1.: Study 2 – Visual stimuli used as fixation, distractors, and targets in the RSVP trials.

For the “Desktop” and “Distraction” conditions, stimuli were presented on a 24-inch LCD computer display (60 Hz refresh rate). The participants’ viewing distance was approximately 70 cm, yielding a viewing angle on the display for the bounding yellow background of 5,2°×5,2° and for the pictogram of 2,5°×2,5°. In the “mini-CAVE” condition, visual stimuli were presented in a 3-D VE projected in a mini-

CAVE visualisation system (Figure 5.2). The mini-CAVE is a system composed of 4 rear-projected screens, surrounding part of a participant's peri-personal space. One screen is right in front of the sat participant, two screens are on the left and right side of the participant field of view, and a fourth screen serves as a plane for the upper limbs reaching plane of the participants. The VE is projected onto these 4 screens. Shutterglasses enabling the stereoscopic visualisation of the VE in the mini-CAVE, and mounting the head-tracking sensors, were wore by the participant for providing in real-time an adequate viewpoint of the 3-D scene. Participants sat at a viewing distance of approximately 120 cm from the central rear-projected surface of the mini-CAVE, staring at visual stimuli centrally appearing and covering a viewing angle of 2,7°×2,7°. In the mini-CAVE the stereoscopic virtual scene represented a motorway junction extending to the horizon in front of the viewer, with a portal-like signpost indicating the different directions through the junction and overarching the motorway. The experimental RSVPs and the visual stimuli of the training phase appeared onto the portal.

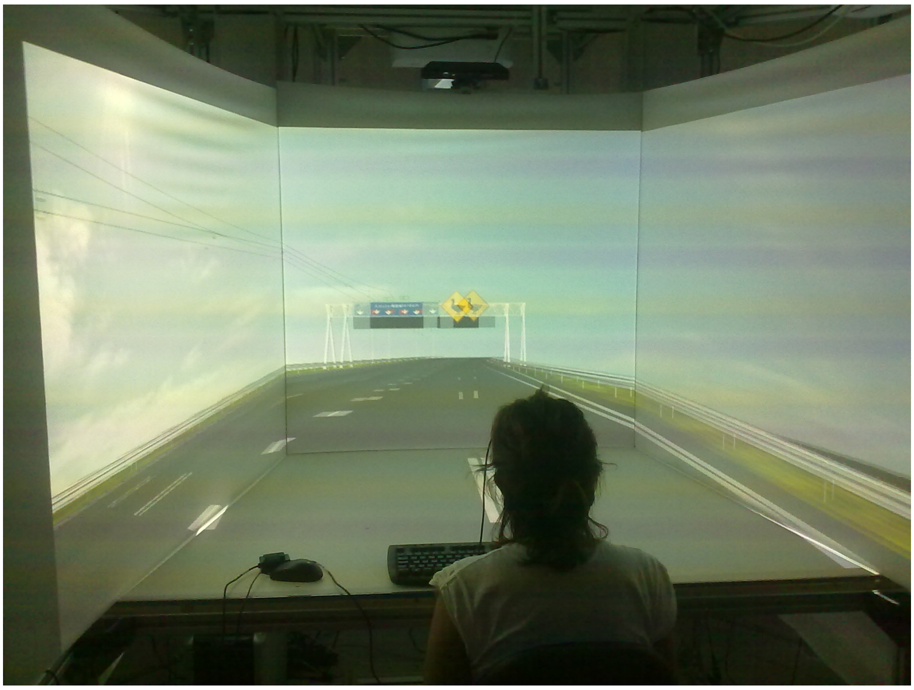


Figure 5.2.: Study 2 – A participant in the "mini-CAVE" display condition.

The visualisation of the RSVPs in the training, practice, and experimental phases of all the experimental conditions, and the rendering of the VE in the "mini-CAVE" condition were realised in XVR (Carrozzino et al., 2005; Tecchia et al., 2010).

Procedure

Each participant took part in three experimental sessions, and each session, corresponding to a different display condition, was carried out in a different day:

- **"Desktop" session**: 2-D presentation on a desktop computer screen of RSVPs made of coloured road signs against a black screen background. No VE and no concurrent auditory distractions were presented throughout this session. Participants' task in the experimental trials was to report T1 and T2 in the order they appeared in the trial.
- **"Distraction" session**: 2-D presentation on a desktop computer screen of RSVPs made of coloured road signs against a black screen background, and of a concurrent stereophonic sound of a car engine synchronised with the RSVP duration. No VE was presented throughout this session. Participants' primary task in all experimental trials was to report T1 and T2 in the order they appeared in the trial. In addition, for one block of trials, participants also had to indicate whether the concurrent sound was regular or not.
- **"mini-CAVE" session**: 3-D VE depicting a motorway junction with a portal-like signpost (Figure 5.3) showing the RSVPs made of coloured road signs. No concurrent auditory distraction was presented throughout this session. Participants' task in the experimental trials was the same as in the "Desktop" session, i.e., to report T1 and T2 in the order they appeared in the trial.

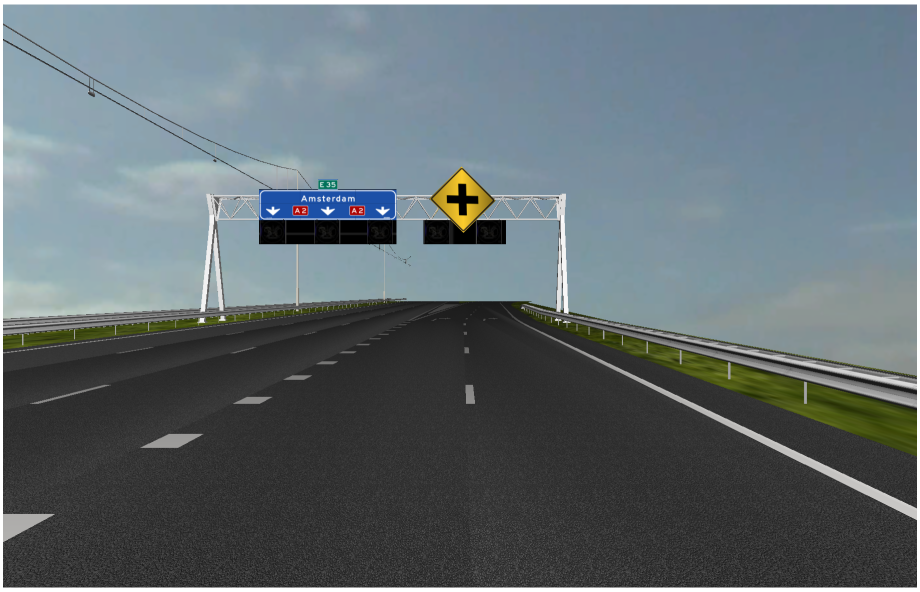


Figure 5.3.: Study 2 – A screenshot depicting the VE used for the "mini-CAVE" display condition. The road signs RSVP replaced the cross fixation symbol in the rightmost signpost on the portal.

Specifically, every session was organised in four successive phases as depicted in Figure 5.4:

1. A **preparation phase** to give instructions about the session task and procedure to the participant, and familiarise her/him with the experimental equipment. A brief questionnaire with demographic information and a self-rating of previous experience with computers, video games and VEs was administered to the participant in this phase at the first session they underwent. A sample copy of the questionnaire is reported in section A.4.

2. A **training phase** intended to ensure that the participant was aware of appearances and names of the targets to be looked for in the RSVP trials of the subsequent phases. This phase consisted of a preliminary passive observation of all the possible targets, each appearing for 2 seconds above its name, and of a target-name association test during which each target appeared without its name for 1,5 s, and the participant was asked to choose from a comprehensive list the name of the viewed target. Besides, in the "Distraction" session also the difference between the "regular" and the "irregular" sounds used as concurrent distractions in the practice and experimental trials was aurally presented to the participant. No data were collected from participants' responses in this phase.
3. A **practice phase** consisting of a block of 9 trials, each resembling a typical trial of the subsequent experimental phase. All the three possible lag levels were present in the 9 trials in a random order. In the practice phase of the "Distraction" session all trials were accompanied by the concurrent sound, but only half of them included also the question about sound regularity after the usual question of reporting the two targets. No data were collected from participants' responses in this phase.
4. An **experimental phase**, the only one collecting data then used in the analysis. This phase comprised two blocks, each composed of 54 trials. A 2-minute rest interval was foreseen between the two blocks. 18 lag-1, 18 lag-3, and 18 lag-7 trials constituted a block, but in random order. Every trial consisted of an RSVP including the 16 distractors and 2 targets. In the "Desktop" and "mini-CAVE" sessions participants' task was to unspeededly report, at the end of each trial, the two targets in the order they appeared. The reporting consisted in the selection of the two targets names from a list containing all possible targets names, using the arrow keys of a computer keyboard to browse the list, and selecting the intended name with a spacebar press. T1 was selected first and T2 just after T1. In addition, only in one block of the "Distraction" session, after reporting the two targets, participants were asked to indicate if the sound heard in concurrence with the RSVP trial was regular or not.

Participants received instructions about their tasks in the different phases at the beginning of every session. In the "Distraction" session they were instructed to attend also to the sound regularity during the trials of both experimental blocks, since they would be required in some trials to judge about the regularity of the heard sound.

In all sessions the participant sits in front of the computer screen or mini-CAVE visualisation system (Figure 5.2). Responses are made by choosing with arrow keys and the spacebar of a computer keyboard a target name from a comprehensive list including all the 10 target names.

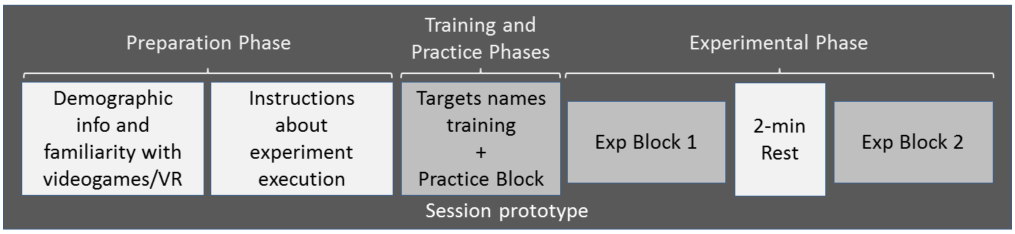


Figure 5.4.: Study 2 – Prototype of an experimental session. Each session corresponded to one of the 3 display conditions: "Desktop", "Distraction", and "mini-CAVE". Demographic information and the answers to the questionnaires about familiarity of use with computers, video games, and virtual environments were collected once from each participant, at the starting of the first session. Training and Practice were repeated at every session.

Measures and Data Analysis

The data collected in the experimental phase were analysed in order to answer the following research questions:

1. Is an AB present in the tested population?
2. Is the lag-1 sparing effect present in the tested population?
3. Does the display condition makes a difference on the AB performance?

Questions 1. and 2. addressed the more general issue about whether the classical findings arising with the AB paradigm are replicated in the present study. Instead, question 3. is the key one of this study, addressing the differences in AB performance possibly induced by the VE, and between the auditory distraction, and the kind of distraction introduced by the VE.

The experimental design adopted for this study foresaw a single group of participants, tested under experimental conditions based on two three-level factors: "Display Condition" and "lag". Possible levels of the former factor were "Desktop", "Distraction" and "mini-CAVE", while "lag" had three possible values — 1, 3 or 7 — corresponding to the serial position of T2 relative to T1 position in the RSVP. A further factor was the block, i.e., whether the measures of reporting performance came from the first or second block of the session experimental phase. However, in no case a statistically significant main effect or interaction of the "block" factor was found on the selected outcome measures.

In order to tackle the three questions, the primary outcome measure calculated for each participant in each experimental condition was the proportion of accurate T2 reporting trials contingent on T1 correct reporting, a.k.a. conditional accuracy, calculated as the mean across each of the two blocks run in the experimental phase of every session. The proportion was expressed as a per centum and is indicated in the following as $P(T2|T1)$. Also the mean T1 accuracy across trials, $P(T1)$, has been calculated as the percentage of T1-correct trials (including T2-incorrect trials) to serve as an unimpaired accuracy benchmark. $P(T2|T1)$ scores T2 reporting accuracy in the attempt to ensure a common processing and memory load (based on T1 accomplished identification and unimpaired report). Although some previous studies used unconditional accuracy on T2 trials (see for example Potter et al., 2002), the majority of recent studies used the P(T2|T1) measure, and also elaborations about the lag-1 sparing effect were chiefly conducted on conditional T2 report accuracy (Martin and Shapiro, 2008; Hommel and Akyürek, 2003).

A General Linear Model (GLM) was used to implement Repeated Measures ANOVA, since each participant was scored for every level of the two experimental factors (Lag and Display Condition). Separate analyses were performed adopting P(T2|T1), and P(T1) as the response variable. A fully related 2-way ANOVA for "lag" and "display condition" was performed over the whole dataset. Then, in order to interpret a slightly significant interaction effect of "display condition"×"lag" revealed by the latter analysis, *post-hoc* tests based on Bonferroni and Tukey's methods were performed to analyse simple effects of the factors. For the "Distraction" session, the factor "block" was introduced in a two-factor repeated measures ANOVA to reveal a possible difference in reporting performance due to the explicit additional questions about the regularity of the concurrent auditory stimulus posed in one of the two experimental blocks of such a session.

All statistical analyses were carried out by means of two software packages, namely MINITAB® 16 Statistical Software (Minitab Inc., State College, PA, USA) and MATLAB® Statistics Toolbox™ (MATLAB R2011b, The MathWorks Inc., Natick, MA, USA).

5.3. Study 2 Results

A summary of the mean proportions of correct T2|T1, and T1 trials adopted as response variables, and calculated across participants and experimental blocks, is reported in Table 5.2.

The GLM implementing the 2-way Lag×Display Condition repeated measures ANOVA carried out on the whole dataset yielded the results summarised in Table 5.3.

Is an AB present in the tested population?

The factor "Lag" has a main effect that turns out to be statistically significant (p-value $< 0{,}001$) on both T1 and T2|T1 accuracies. On the other hand, a main effect of "Display Condition" is significant only for P(T2|T1) (p-value $= 0{,}015$), indicating that T1 accuracy is not affected by the concurrent auditory distraction, and by the 3-D presentation of the RSVPs in a stimuli-related context. The interaction "Lag x Display Condition" is not highly significant for any of the tested response variables, with only a weak evidence of possible significant interaction (p-value $= 0{,}068$) on P(T2|T1).

The graphical presentation of of P(T2|T1) and P(T1) as a function of lag, and across display conditions, is depicted in Figure 5.5. T2|T1 accuracy is different relative to

(a) Summary of means for P(T2|T1) across participants and experimental blocks.

P(T2|T1) – Cell means, Marginal means, and Grand mean				
	Display Condition			
Lag	"Desktop"	"Distraction"	"mini-CAVE"	Marginal means
1	93,2096 %	92,6382 %	90,8748 %	92,2409 %
3	92,2777 %	96,5887 %	88,7476 %	92,5380 %
7	97,5646 %	96,7739 %	93,0872 %	95,8086 %
Marginal means	94,351 %	95,334 %	90,903 %	$\mu_G = 93,529\,\%$

(b) Summary of means for P(T1) across participants and experimental blocks.

P(T1) – Cell means, Marginal means, and Grand mean				
	Display Condition			
Lag	"Desktop"	"Distraction"	"mini-CAVE"	Marginal means
1	80,3241 %	82,4074 %	79,8611 %	80,8642 %
3	96,2963 %	95,1389 %	97,4537 %	96,2963 %
7	96,7593 %	96,7593 %	96,0648 %	96,5278 %
Marginal means	91,13 %	91,44 %	91,13 %	$\mu_G = 91,2294\,\%$

Table 5.2.: Summary of the means of proportions of correct trials in the different experimental conditions (μ_G denotes the grand mean) calculated across participants and experimental blocks. Conditional accuracy of T2 report, P(T2|T1), was analysed for T2. P(T1), calculated by including every trial with a correct T1 report independently of T2 performance, was taken as a measure to index performance on T1. The values in these tables refer to means estimated from the whole experimental dataset, by collapsing the two experimental blocks in a single pool, and averaging across all participants in the experiment.

T1 accuracy at lag 3 and lag 1, while the two proportions are similar at lag 7. Specifically, T1 accuracy at lag 1 (80,86 %) is lower than T2|T1 accuracy at lag 1 (92,24 %), whereas T2|T1 at lag 3 is 92,54 % against a T1 accuracy basically steady at lag 3 and lag 7 (96,3 % and 96,53 % respectively). This indicates a rather high T2|T1 accuracy values, and a little AB on the overall collected dataset.

Is the lag-1 sparing effect present in the tested population?

No lag-1 sparing pattern is present on the overall dataset. Instead, there seems to be a T2 superiority at lag 1. In order to have a graphical reference to compare accuracies involving T2 report with T1 report rates, and because of the substantial independence of P(T1) on the display condition, P(T1) calculated as a mean accuracy across Display Conditions and against lag levels has been considered as the term of reference with the interaction plot of P(T2|T1) as reported in Figure 5.6a.

Looking at the interaction plot of P(T2|T1) for the different display conditions (Figure 5.6a), which is only weakly significant from a statistical point of view, the situation somehow changes since at lag 1 T2 conditional accuracy is higher than at lag 3 for the "Desktop", and "mini-CAVE" conditions, and for the latter the

	P(T2|T1)		P(T1)	
	F-ratio	p-value	F-ratio	p-value
Lag	$F_{2,22} = 6,3$	0,007	$F_{2,22} = 56,32$	< 0,001
Display Condition	$F_{2,22} = 5,11$	0,015	$F_{2,22} = 0,02$	0,984
Lag×Display Condition	$F_{4,152} = 2,23$	0,068	$F_{4,152} = 0,83$	0,509

Table 5.3.: Summary of statistical results from a Repeated Measures ANOVA on mean report accuracies using Lag and Display Condition as fixed factors. P(T2|T1), which is the usual conditional measure for indexing the AB magnitude, undergoes significant main effects of both Lag and Display Condition, and a weakly significant interaction effect between Lag and Display Condition. In contrast, only Lag has a significant main effect on P(T1), for which neither evidence about an effect of any Display Condition, nor an interaction of the experimental factors are found.

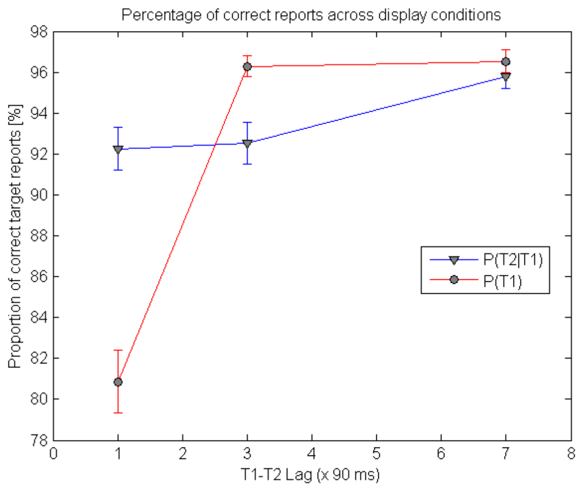
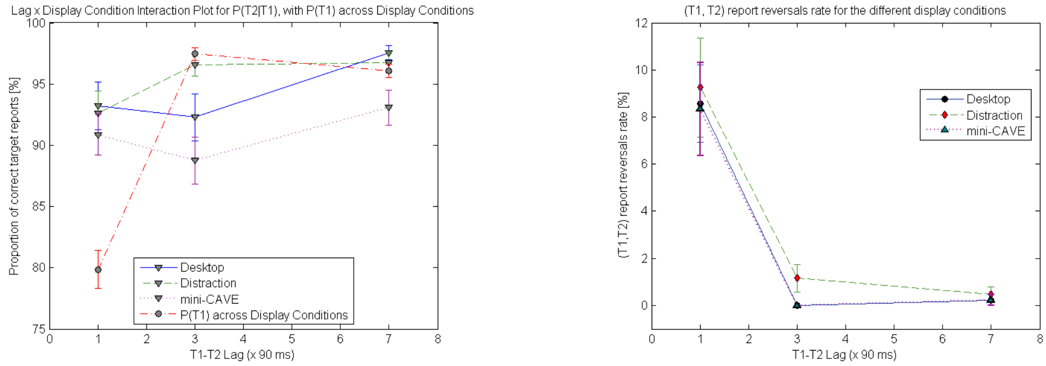


Figure 5.5.: Study 2 – Mean proportions of T2|T1 and T1 correct report versus T1–T2 Lag across display conditions. Error bars indicate standard errors of the mean.

rates are lower. A particular shape characterises the “Distraction” condition, with a T2 conditional accuracy at lag 3 similar to that at lag 7. Therefore, the U-shaped characteristic of T2 conditional accuracy versus T1–T2 lag is only partially found when considering the interaction effect of Lag and Display Condition (this is the case for “Desktop” and “mini-CAVE” conditions), but it is completely absent when considering the “Distraction” condition, or the whole dataset across display conditions.

The rate of T1 correct reports at lag 1 is quite lower than at lag 3 and lag 7 (the difference in means is larger than 10 % within every display condition), while the difference between the means of P(T1) at lag 3 and lag 7 is quite small (less than 2 %): this latter finding concords with the tendency of T1 report accuracy as a function of T1–T2 lag commonly reported in scientific literature about the AB (Dux and Marois, 2009; Martin and Shapiro, 2008; Dell’Acqua et al., 2007).



(a) Lag×Display Condition interaction plot of P(T2|T1), with P(T1) means across display conditions as a reference. Error bars indicate standard errors of the mean.

(b) Percentage of reversed reports of T1 and T2 in the different display conditions. The reversals frequencies for "Desktop" and "mini-CAVE" are substantially overlapped. Anyway, reversals at lag 1 are much more than at lag 3 and lag 7. Error bars indicate standard errors of the mean.

Figure 5.6.: Study 2 – Lag×Display Condition interaction plot, and percentage of reversed reports of T1 and T2 in the different display conditions.

The superiority of T2 at lag 1 may be related to a further effect emerged from this study: the higher order of reversals — i.e., reversed reports of T1 and T2 — associated to lag-1 trials (Figure 5.6b).

Does the display condition makes a difference on the AB performance?

The size of the main effect of Display Condition on P(T2|T1) (Figure 5.7) is more relevant when a concurrent auditory distraction is present, and when the RSVP paradigm is delivered stereoscopically in a virtual motorway context — the 3-D VE delivered in the mini-CAVE system. However, the sign of this effect is different in the two conditions, improving P(T2|T1) accuracy in the "Distraction" condition (as expected based on the beneficial effects documented in past experiments), while deteriorating it in the "mini-CAVE" condition. Specifically, the mean accuracy P(T2|T1) is maximum when there is an auditory distraction concurrent to the RSVP (95,334 %), and minimum when the dual-target task is to be accomplished in the mini-CAVE visualisation system (90,996 %). The size of the main effect for the "Desktop" display condition is smaller. Applying *post hoc* tests to the data for P(T2|T1), Bonferroni and Tukey's pairwise comparisons agree as they find a significant difference of the "mini-CAVE" condition from the others (adjusted p-value $< 0{,}004$). However, the difference between "Distraction" and "Desktop" mean conditional accuracies as measured by P(T2|T1) is not significant. As for the *post hoc* analysis on the Lag factor, the results show a significant difference of P(T2|T1) at lag 7 (adjusted p-value $< 0{,}007$), but without statistical evidence inferring a significant difference between the means of P(T2|T1) at lag 1 and lag 3. This confirms that no lag-1 sparing is present, except if we consider an interaction of the experimental factors.

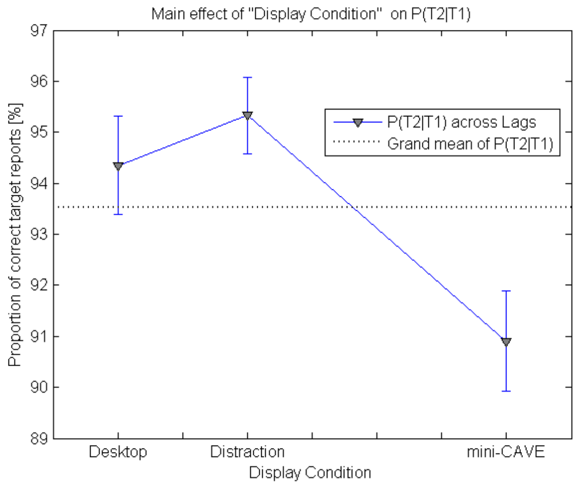


Figure 5.7.: Study 2 – Main effect of Display Condition on mean T2|T1 report accuracy. Error bars indicate standard errors of the means.

For the distraction session further aspects had to be investigated. On one side whether there was actually a beneficial effect of the auditory distraction on AB performance with respect to the other display conditions. On the other side, whether there is a different performance in terms of report accuracy under the "Distraction" condition when, beside the instruction to participants of listening to the concurrent sound for detecting its regularity, also the explicit question about sound regularity is made after the usual questions to report T1 and T2. Concerning the latter issue, an analysis conducted by adding a "block type" factor to a repeated measures ANOVA on Lag effect within the "Distraction" sessions revealed no significant difference (p-value = 0,339) between the block where participants were only instructed to pay attention to the concurrent sound distraction, and the block where, at the end of the trials, they were also explicitly asked about sound regularity. In this case the result of the repeated measures ANOVA is that there is a main effect of lag on P(T2|T1) (p-value = 0,023), but the significant difference is only between the accuracy value at lag 1 and those at lag 3 and lag 7. No significant difference was found between P(T2|T1) at lag 3 and at lag 7. However, as depicted in Figure 5.8 it seems in this session no AB has been arisen since all T2|T1 accuracy values are over 90 %, and T1 correct report proportion is very close to the conditional accuracy also at lag 3. It is like no impairment affected T2 conditional upon T1 correct reports at any lag, while at lag 1 order reversals impaired T1 reports as found in the other display conditions. This seems a further evidence of the beneficial effect of the concurrent auditory distraction on the 2-D RSVP dual-target task.

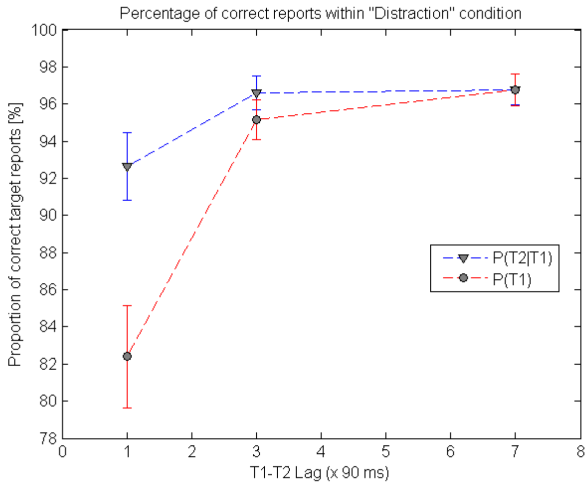


Figure 5.8.: Study 2 – Mean proportions of correct reports versus Lag levels within the "Distraction" display condition. Error bars represent standard errors of the mean.

5.4. Study 2 Discussion

The first objective of Study 2 was to investigate the impact on AB of using road signs in an RSVP, at a rate — one visual item every 90 ms — that with both simpler stimuli (Dux and Marois, 2009), and with pictures (Einhäuser et al., 2007), already proved to cause an AB.

The AB consists of a decrease in T2 report accuracy relative to T1 report accuracy when the two targets appear at a short temporal separation from one another. Such a fall then recovers for longer lags, that is when T2 appears after T1 identification is consolidated. The extent of the temporal window where T2 accuracy is impaired, and the depth of the impairment depend upon stimuli characteristics such as letters and digits versus natural pictures, or same-category versus different-category targets, on similarities between targets and distractors, and on possible cuing effects priming the second target. A special characteristic accompanying AB is lag-1 sparing, a robust effect that is often, although not always found, when T1 and T2 appear consecutively, or anyway within 100 ms–150 ms from one another. Visser et al. (1999) defined the lag-1 sparing based on the difference in T2 report accuracy between lag 1 and the lag when T2 report accuracy is minimal (typically at lag 3): when this difference exceeds 5 % then a lag-1 sparing is there. More liberal definitions of AB have been considered too, as reported in Martin and Shapiro (2008), as well as more conservative ones (Dell'Acqua et al., 2007). Lag-1 sparing usually occurs if some requirements are met (Visser et al., 1999): T1 and T2 must appear in the same spatial location, their SOA must fall within a temporal integration window, and they need not require an attentional set switch between one another. Most of the manifested AB effects, and especially lag-1 sparing, assume that T2 accuracy is calculated conditionally upon T1 correct report (T2|T1 accuracy).

Our experiment considered three possible values for T1–T2 SOA (lag 1, lag 3, and lag 7, i.e., 90 ms, 270 ms, and 630 ms) applied in three different display conditions (2-D display on a computer desktop, 2-D display on a computer desktop with a concurrent auditory distraction, and presentation in a 3-D VE meaningfully contextualising the RSVP of road signs). The research questions guiding the analyses of the collected data towards the first objective were those about the capability of the experiment to cause an AB, and in case to show a lag-1 sparing pattern.

In terms of reduced T2|T1 accuracy our experiment found a statistically significant decrease at lag 1 and lag 3 compared to lag 7 across display conditions. The magnitude of this AB was not as large as found in previous studies using pictures and line drawings, and showed no lag-1 sparing effect. The small magnitude of the AB could be related to the reduced number of possible targets to be identified, which might have facilitated the formation of short-cut strategies by the participants, and to the reporting modality, because participants had to choose the two targets from a 10-item list, and not to "freely" report them as it is common in other experiments addressing the AB.

The fact that no lag-1 sparing was observed is in line with the results of the study by Hommel and Akyürek (2003), which differentiated the interpretation of lag-1 sparing between the condition when the task is to report T1 and T2 order along with the targets identities, and when the task is to report targets identities regardless their order of appearance. They found that when also targets order had to be correctly reported for having a trial response counted as correct, the lag-1 sparing disappeared or was very less evident. This is the case for the results of our experiment, for which both correct identities, and correct order of T1 and T2 should be reported to accomplish a correct trial. Another related finding is shared with Hommel and Akyürek (2003) (and also with Potter et al., 2002 and Dell'Acqua et al., 2007): there is a relevant rate of order reversals at lag 1, which determine a T1 accuracy reduction (down to approximately 80 %). The effect is compatible with two-stage accounts (Chun and Potter, 1995; Potter et al., 2002), and with the the eSTST model (Wyble et al., 2009). The order reversals may in fact depend from an integration of T1 and T2 in the same attentional episode, yet at the expense of episodic information. Thus, T2 superiority at lag 1 likely depends on both a trade-off with T1 due to integration of the two in the same attentional episode, and on the fact that T2 either wins the competition with T1 for attentional resources, as predicted by the model of Potter et al. (2002), or it benefits of the transient attention response triggered by T1, as postulated in the two-stage account by Chun and Potter (1995). Also the model proposed in Evans and Treisman (2005) for explaining rapid processing of natural scenes may explain why lag-1 sparing was not observed: T1 and T2 features are both recognised in a first parallel stage, but the serial processing needed for identification determines the AB even at lag 1. The latter explanation is, however, rather independent of the higher rate of order reversals found at lag 1, and appears less likely also because no direct measure of accuracy at lag 1 was taken in the series of experiments by Evans and Treisman (2005).

We speculate that the general high T2|T1 report accuracy emerged from our data might also be related to the generally high frequency of use of computers made by the population participating in the present study, and on the experience as video-gamers of a relevant group of them. In fact, an attenuation in the AB had been previously shown in expert video-gamers compared to a control group naïve at using video games (Green et al., 2003). Actually, the participants in our experiment took a simple questionnaire during the preparation phase of the first experimental session they underwent. The questionnaire (see section A.4) included questions in order to self-rate on a 5-point scale the familiarity of the participant with three media: personal computers, video games, and virtual environments. The results are illustrated by the chart in Figure 5.9. Performing a GLM implementing a repeated measures ANOVA with the Participants' random factor nested within the "Familiarity with video games use" fixed factor based on the collected self-rates, it emerged a significant effect of the latter factor on the T2|T1 accuracy, but no significant interaction with the "Lag" factor was found. Specifically, the familiarity-factor had two

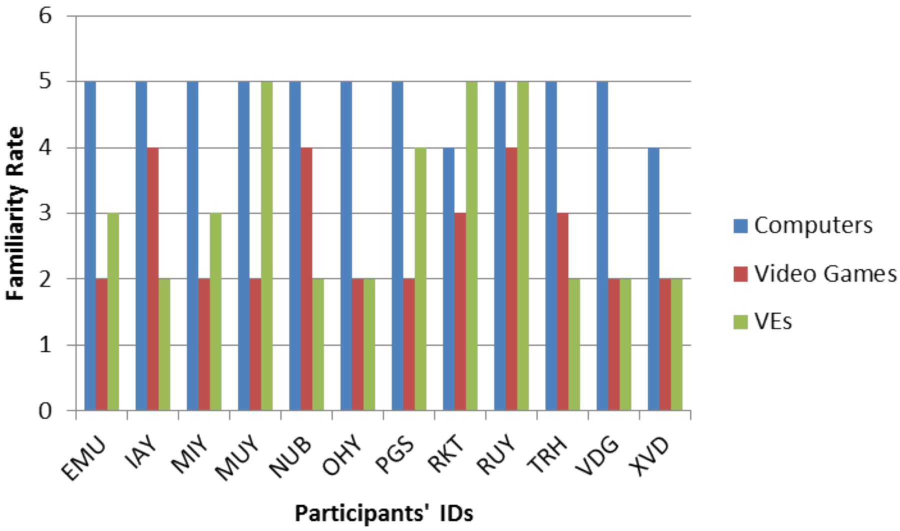


Figure 5.9.: Study 2 – Self-rating of participants' familiarity with computers, video games, and virtual environments. The rates were made on a 5-point scale ranging between 1–no experience, and 5–assiduous use. Questions are reported in section A.4.

levels, either "High" (self-rate≥3) or "Low", and its main effect on T2|T1 accuracy was statistically significant (p-value = 0,008), with the mean T2|T1 accuracy across lags and display conditions for participants having an "High" familiarity with video games larger than that of those having a "Low" familiarity (96,12 % vs 90,97 %).

As for the second objective of the present study, it chiefly pointed to the effect on AB of an immersive 3-D VE providing a meaningful context for the stimuli presented in the RSVP — the "mini-CAVE" display condition. The effect has been compared to the AB performance in a classical 2-D display condition, and when an auditory distraction is delivered concurrently to the 2-D display condition. The results to discuss in this case are mainly those related to the third guiding research question, that is whether the display condition makes a difference on the AB performance.

Actually, the AB was present in the “mini-CAVE” condition, and it showed the lowest T2 conditional accuracy, with a difference that was statistically significant from both the “Desktop” and the “Distraction” conditions. This might depend on the context-relatedness of the visual stimuli (road signs presented at a virtual motorway junction), on perceptual issues connected to the VE display (e.g., stereoscopy), or on both of these aspects. It is difficult to isolate the different contributions, but a putative role of the virtual scene could be expected to improve T2|T1 accuracy, as previously found with visual background manipulations by Arend et al. (2006). Furthermore, in the “Distraction” condition of the present experiment, a substantial immunity to the AB — even of little magnitude as found in the other display conditions — has been revealed by analysing the simple effect of lag. And this seems in line with the findings and hypothetical explanations of Olivers and Nieuwenhuis (2006) for beneficial effects of distraction at alleviating the AB magnitude. Following these considerations, the expectation would have been to observe a less profound dip of the T2|T1 curve with respect to those in the 2-D, less immersive conditions, but this is not the case. In addition to perceptual issues of the stereoscopic display and head tracking, a possible explanation is related to the *involvement* component that might accompany more immersive virtual experiences. Involvement and absorption are terms used in studies about what cognitive factors and personality traits contribute to determine the sense of presence in VEs, e.g., in the form of *immersive tendency* (Wallach et al., 2010; Sas and O’Hare, 2003; Takatalo et al., 2008). There is empirical evidence that experience of an immersive VE is affected by immersive tendency, a personality trait affecting the level of absorption in the VE. Absorption is said to require a total attention, fully open to experience, but also fully concentrated (Sas and O’Hare, 2003). In this framework, our participants’ absorption in the immersive “mini-CAVE” VE could correspond to either a more diffuse, or an heightened-focused attention allocation. In the former case a diffuse state compatible with the over-investment hypothesis of Olivers and Nieuwenhuis (2006) would have predicted a reduced AB magnitude, differently for what actually found in our “mini-CAVE” condition. The same absorption construct interpreted as an heightened focused attention, would instead correspond to an over-investment of resources, and predict what is actually found, i.e., a deteriorated AB performance compared to 2-D, less immersive conditions.

Our preference goes to the considerations that in this experiment the “mini-CAVE” condition reduces T2|T1 accuracy due to perceptual issues related to stereoscopy, although this requires to be demonstrated with further experiments. In fact, although immersive, the interaction of the participant with the VE seemed not so engaging to trigger a major contribution of subjective immersive tendencies on the AB task accomplishment. Anyway, an AB is present when the RSVP is presented in the VE, and the effect of the virtual scene is not similar to that of a moving 2-D background, or of a concurrent auditory distraction. On the contrary, the auditory distraction consisting of a sound engine being either idle or accelerated, resulted in a suppression of the AB, which was actually not very profound in the other display conditions.

6. Conclusions

The present dissertation stemmed from a research work aimed at investigating the use of virtual environments in cognitive sciences, but at the interface with medicine. The stated objectives were essentially two. Firstly, to outline a framework for situating applications of VEs to cognitive sciences, by starting from how VEs are applied for the assessment and therapy of neurological and mental disorders, and from what are the grounds for this kind of applications. Secondly, to provide an empirical case study illustrative of the potential of VEs for studying specific aspects of cognitive processes, not necessarily in the medical domain. The two objectives have been pursued by individuating specific sets of research questions, and undertaking them with two different approaches, one directed towards the first objective, and another towards the second. The conclusions drawn are summarised in the following sections.

6.1. A Framework for Virtual Environments Applications to Cognitive Sciences

In order to delineate a framework situating the applications of VEs to cognitive sciences, the approach has been to consider medical applications, and to formulate three guiding questions:

- What are the grounds underpinning the applications of VEs for the assessment and therapy of neurological and mental disorders?
- How are VEs applied for the assessment and therapy of neurological and mental disorders?
- Is it possible to extend the framework hosting the applications for the assessment and therapy of neurological and mental disorders to all the applications of VEs to cognitive sciences?

These questions are entangled, and their answers emerged in parallel in the course of the work. VEs have been applied to diverse medical domains, namely for educational purposes, for surgery and diagnostics, for medical rehabilitation (including neurorehabilitation, and general rehabilitation), and for treating mental disorders. The applications of VEs to assess neuropsychological disorders, and help cognitive function(s) rehabilitation fall under the neurorehabilitation heading (or the more broad category of virtual rehabilitation), whereas mental disorders cover applications of VEs to psychotherapy and psychiatry. The applications to these two sub-categories, i.e., neurorehabilitation and mental disorders, share the emphasis on the interactive opportunities offered by VEs, in contrast to simulation opportunities which are more central for the remaining categories of VE applications to medicine. This consideration has led to the definition of interaction-centred VEs as a master category for applications of VEs to cognitive sciences from the medical point of view (Riva, 2006).

The criterion underlying the distinction between interaction-centred applications and simulation-centred applications has the advantage to be less dependent on the application domain, and to highlight the focus of *how* and *why* VEs are used in a certain context. In the case of the interaction-centred applications considered in this thesis the *why* has been summarised in a number of prerogatives making VEs potentially valuable for applications to cognitive assessment and rehabilitation, and to psychotherapy and psychiatry. Concerning the *how*, sub-categories for such interaction-centred applications have been proposed, and supported with examples drawn from the scientific literature. Finally, the separate categorisations have been unified in a single framework, including also non-medical applications, although these have been only briefly mentioned.

Altogether, the main conclusions related to the directive research questions may be restated as follows.

> *What are the grounds underpinning the applications of VEs for the assessment and therapy of neurological and mental disorders?*

The prerogatives have been much discussed and analysed for applications to neuropsychological assessment and cognitive rehabilitation, but they are substantially founding also applications to motor rehabilitation, and for the study of neuroscience, and cognitive sciences in general. Such prerogatives are:

- Environmental enrichment
- Improved ecological validity
- Systematic and controlled delivery of 3-D multimodal stimuli
- Immediate multimodal feedback
- Cueing stimuli to guide successful performance
- Complete digital performance capture
- Safety testing and training environments

> *How are VEs applied for the assessment and therapy of neurological and mental disorders?*

The categorisations adopted to describe and discuss how VEs have been applied are reported in Figure 2.3 and Figure 2.4.

As for neuropsychological assessment (NA) and cognitive rehabilitation (CR), VEs have been proposed to address impairments of attention, memory, spatial abilities, executive functions, and language. Generally, all the applications yielded positive results when evaluated, demonstrating that the potential of VEs for NA and CR is concrete. A large number of VEs applications are oriented to assessment, because the generalisation of psychometric assessment outcomes from the laboratory setting to daily life is a crucial issue of NA. Actually, studies on psychometric tests in VE usually correlated with standard tests, and, when the point has been addressed, also with functioning in the real-world. This is a confirmation that environmental enrichment and improved ecological validity are likely the most important prerogatives of VEs for applications to NA. Also CR applications yielded positive results, and relied on environmental enrichment, but also much on the other prerogatives. However, some limitations are still affecting these applications: first, a lack of large-scale studies comparing the effect of VEs with those of standard assessment procedures and rehabilitation protocols, and a lack of standardisation of VR-based protocols; second, the requirement of special competences for handling the VR equipment, both during the intervention itself, and for preparing the VR-based tasks. The first limitation may be probably overcome with the time, and by counteracting the second limitation. This can be achieved by means of two concurrent actions: to promote the establishment of multidisciplinary teams involving clinicians, engineers, and computer scientists, that cooperate since the design of the VE, through its development and validation, and during its operating life; and to take into consideration since the design phase the target population of the VE, in terms of patients, and also the needs of the therapists, enabling them to dynamically operate on the objects and behaviours of a VE at a high abstraction level, and easily adapt the task to the patient's specific needs and real-time performance.

Considering the specific areas of application in NA and CR, language disorders seem to have been the less attractive for VEs applications, likely because contexts as executive dysfunctions, and memory impairments more evidently and directly took advantage from the impact of ecologically valid environments. An area apparently less undertaken is that of attention, probably because it is well covered by non-VR-based assessment tests, and can be addressed as part of VEs applications addressing more complex impairments, as those descending from executive dysfunctions.

The applications in psychotherapy seem to be somewhat more advanced, at least for the application of Virtual Reality Exposure Therapy for the treatment of anxiety disorders. This is due to the longer history of this kind of applications, with a greater number of controlled studies carried out to compare them with standard psychotherapies. Moreover, the psychological connotation of the applications for mental disorders is relevant also for a more general and basic research about how VR can affect body perception and representation, the sense of self, and the sense of agency. Another aspect standing out in applications for mental disorders is the resort to virtual characters. This helps to ground and spread applications promoting social interaction, and social inclusion, which are now primary goals for assisting significant portions of the ageing population, and the increasing number of youngsters on the autism spectrum disorder.

Lastly, a general remark emerging from the discussions of the different VEs applications, independently of whether they address NA and CR, or mental disorders, is that they are not intended and suitable to replace the current tools. Rather, they deploy themselves as valuable adjuncts which solve limitations of current standard interventions, and offer additional capabilities not conceivable with mainstream paradigms.

> *Is it possible to extend the framework hosting the applications for the assessment and therapy of neurological and mental disorders to all the applications of VEs to cognitive sciences?*

The applications of VEs to cognitive neuroscience, and not directed to medical conditions focus on spatial cognition and navigation, and, more recently, on the ability of VR to induce body-transfer illusions. This scope of application will probably expand because VR systems are becoming compatible with neuroimaging techniques, and because of the advances in brain-computer interfaces. Indeed, the latter may be used to establish a direct connection between the nervous system and a VE. This may have several applications, and a major one is again in rehabilitation and training, for instance for patients who have to learn the control of prostheses, or for patients as quadriplegics, who may control the real environment through actions effected on virtual objects by using their nervous signals.

At any event, the applications of VEs to medical conditions associated to cognitive impairments are more numerous than the applications to non-medical conditions, and, more important, the applications to NA, CR, and mental disorder essentially encompass all the prerogatives of VEs useful for cognitive science and beyond. Consequently, a single framework has been proposed, as in Figure 2.5.

An observation regarding how a typology of VEs applications may be transversal to the proposed frameworks is prompted by serious games, i.e., 3-D video games, often including typical VR technologies as those for motion capture and stereoscopic vision. This sort of video games is spreading as a means to implement rehabilitation exercises as game-like tasks. The aim is to increase patient's motivation, engagement, and attention in the rehabilitation process. Both sensorimotor training and cognitive rehabilitation may be addressed with serious games applications, even in the same environment. And also psychotherapy, as for impulse-related disorders, has been a target for the application of serious games. Therefore, serious games are a good example of how VEs applications for cognitive science, in this case in the medical field, may generalise to several domains as identified in the proposed frameworks.

6.2. The Empirical Studies of Temporal Attention

Attention, among the different cognitive domains included in the proposed framework, plays a crucial role. Indeed, it affects multiple perceptual and cognitive operations, and is a fundamental requirement for effectively dealing with the information coming from the environment, and use them to consistently behave in view of specific goals. Furthermore, the time course of attention, in terms of how it is allocated at the level of the human information processing flow, is an issue which has attracted a comparatively little interest in association to VEs. These were the prime motives for the two empirical studies undertaken and presented in chapter 4 and chapter 5.

The research questions taken on about attention were indeed split in two experimental studies. Study 1 did not involve VEs, and investigated whether the distribution of attentional resources as indexed by the AB performance could be changed following a 7-week Tibetan Yoga training undertook by habitual meditators. The conclusions of the study were a confirmation of the fact that a relatively short training involving meditation, and specifically a mindfulness meditation style as in *Tsa Lung* Tibetan Yoga is capable of improving the AB performance, reflecting a more efficient allocation of attentional resources over time. The different amounts of meditation practised by participants in the 4 weeks prior to the training start did not statistically affect their baseline performance in the AB task. Possible confounding factors in this experiment were the lack of a control group, and the fact that not all the participants practised the same exact amount of meditation and physical exercise during training. In fact, Tibetan Yoga is different with respect to pure mental training in that a mild physical exercise associated with visual imagery is a fundamental part of it, along with mindfulness meditation. The different amounts of practice by the participants enabled to test whether the contributions to AB alleviation of the mindfulness meditation component, and of the physical exercise component were significantly different. However, it was not possible with a general regression to find a statistically significant contribution of either of the two terms, average amount of hours of physical exercise per week, and average amount of hours of mindfulness meditation per week during the 7-week training, on the mean T2|T1 report accuracy.

A future elaboration of Study 1 could be to reconsider the experimental design for distinguishing the contribution of the body and mind components of the training, using a larger number of participants, and a control group.

Study 2 introduced a VE in the experiment, and attempted to answer two questions: whether rapid visual stimuli which are not as simple as letters or numbers, and which may be meaningfully contextualised in a VE, elicit an AB with similar characteristics as that arisen with abstract and simple visual stimuli; and whether there is an effect of the VE on the AB magnitude, similar to that of an auditory distraction. Road signs were used as visual stimuli for assessing the AB magnitude, and they were presented on a 3-D virtual motorway junction to have them in a meaningful context.

The AB magnitude measured as the difference between T2|T1 mean report accuracy at lag 3 and lag 7 was very little, and no lag-1 sparing was found. The magnitude of the AB may be relevantly affected by the composition of the group of participants, who were all assiduous computer users, with a number of them being frequent video game players. The latter aspect has been already demonstrated as a factor improving the AB performance (Green et al., 2003). The absence of a lag-1 sparing effect might depend on the fact that also the order of reporting T1 and T2 had to be right for counting the trial in the correct reports. Besides, the number of possible targets, although belonging to two different categories, was small (5 animals and 5 vehicles), and reporting was provided by selecting the responses from a comprehensive multiple-options list.

The effect of the VE was not beneficial, i.e., it did not alleviate the AB, as is the case for an auditory distraction (a finding at least partially replicated in the present experiment, albeit the AB "baseline level" to be benefited was already a high performance). Specifically, the AB magnitude in the immersive 3-D VE, the "mini-CAVE" condition, was lower than both the 2-D "Desktop" condition, and the 2-D "Distraction" condition. No specific measures of immersive tendency of the participants were made, and this difference in personality traits might actually affect the performance in the VE. Nonetheless, given the limited interaction with the immersive VE required in the present experiment, it is more likely that perceptual issues were the reasons for a stronger AB in the "mini-CAVE" system. This aspect should be further studied in an experiment that uses a participant group without biases related to previous experience with video games and VEs, including a measure of immersive tendency of the participants, taking into account more in details the perceptual properties of the VE display system, and trying to devise a task for the AB that implies more interaction with the VE. This latter point is important, because it would be of significant help to elicit a sense of presence in the VE, and the consequent effects at the psychological and sensorimotor levels in the participants.

Acknowledgements

Firstly, I thank Professor Massimo Bergamasco, who gave me the opportunity to undertake the Ph.D. course, and supervised my research as an insightful and liberal advisor.

I also express my appreciation to Prof. Francesco Fornai, and Prof. Amelio Dolfi, who followed my Ph.D. student career at the University of Pisa.

The research presented in this dissertation has been mostly carried out at the Perceptual Robotics Laboratory (PERCRO) of Scuola Superiore Sant'Anna, where I also live my everyday working life, and to which, as both an institution and a group of people I wish to show sincere acknowledgements.

I express a particular gratitude to Dr. Marcello Carrozzino, and Dr. Emanuele Ruffaldi, who helped me for the experiments presented in this dissertation and beyond. Marcello introduced me to the e*X*treme *V*irtual *R*eality environment, and I learned fundamental information from his lectures on Virtual Environments. He especially contributed to Study 2 set up, and to tune the experiment for the mini-CAVE equipment. Emanuele provided me with many suggestions for experimental design and analysis, and with sometimes cryptic, but always efficacious computer code. Our lunch meetings are generally very productive, and originated ideas that fertilised this dissertation too. I believe he embodies the researcher spirit, and this was very instructive to me.

I am especially thankful to Dr. Elisabetta Sani, a friend and colleague, who always spoke words of support throughout all the years of my Ph.D. course, and in the last months also shouldered much work to let me concentrate on this dissertation.

Study 1 presented in the dissertation was carried out at the Lama Tzong Khapa Institute in Pomaia. I thank again Prof. Bergamasco, who introduced me to the Institute, and to the studies on Tibetan Yoga, which were going ahead there. Hence, I am grateful to Joan Nicell for the hospitality of the Institute, and to Dr. Rafael Ferrer-Estrems, who conducted the Tibetan Yoga training, and discussed with me the interpretation of Study 1 experimental results.

I acknowledge the willing and active disposition of all the participants to the experiments presented herein, both the colleagues at the PERCRO laboratory, and the people from the Lama Tzong Khapa Institute. Most of them contributed useful remarks in the experiment debriefing.

At last, thanks to my family, who unrelentingly listen to all my doubts and wishes, and encourage me to move on with dignity.

A. Appendixes

A.1. Tibetan Yoga (*Tsa Lung*) Exercises Instructions

The instructions for *Tsa Lung* exercises included in the 7-week training programme whose effects were tested in Study 1 mentioned the following main points (adapted from a guide prepared by Dr. Rafael Ferrer Estrems, who trained the participants at the Institute Lama Tzong Khapa in Pomaia, Italy).

Brief Guide on Tsa Lung Exercises

- 5 *Tsa Lung* exercises. Each performed 5 times.
- They can be used both as a preliminary to other meditation practice, and as a primary practice for enhancing the experience of open awareness in day-to-day life.
- Take a comfortable posture. The most important is to have the spine straight.
- Gently hold the breathing for the whole duration of each exercise.
- 2-phase inhalation: (1) inhale, then hold the breath; (2) re-inhale to expand further the chest; then, while holding, perform the exercise.
- Exercise 1: attention on the throat. Imagine your breath held at the throat level, and concentrate on that point. Head turns with no neck movements are requested.
- Exercise 2: attention on the chest. Lift and rotation of the arms. Hand closure/opening. Chest rotation with no shoulders movement.
- Exercise 3: attention to a point between the perineum and abdomen. Perineum tension. Abdomen rotation and expansion. Imagine the air encapsulated between perineum and diaphragm.
- Exercise 4: attention to no specific point. Arms and hands movements to vigorously rubbing down the head, face, shoulders, arms, chest, back, and legs. Stretch an arm as if holding a bow, and imagine you are shooting an arrow in the direction of your gaze.
- Exercise 5: attention to the perineum. Movements of the legs crossed at the ankles levels, and rotation of the pelvis.

A.2. Instructions for Participants in Study 1

Below is reported the text that was used by the experimenter in Study 1 to provide instructions to the participants before they took part in the experiment session (both prior to and after the 7-week Tibetan Yoga training). Details might be rephrased by the experimenter upon request from the participant.

Instructions for the Attentional Blink Experiment

These are the instructions for the subjects participating in the "attentional blink" experiment.

The objective of the experiment is to assess how an individual manage her/his own attentional resources over time, when confronted with visual targets embedded in a stream of visual distractors.

The Participant is asked to comfortably sit in front of a computer screen, at a viewing distance of approximately 60 cm, and to look at the centre of the screen, where different messages and symbols will appear throughout the whole experiment.

The whole experiment takes less than 30 minutes and is composed of multiple trials, organized in 3 blocks.

- Each trial starts with a fixation cross-hair located at the centre of the screen, followed by a stream of symbols rapidly appearing at the centre of the screen one after another.
- The fixation cross-hair aim is to indicate the point of the screen where the Subject has to look, and it disappears as soon as a stream starts.
- The stream is composed of alphabetic letters and two different single-digit numbers.
- These two numbers are the targets the Subject has to look for and remember through each trial.
- The two numbers are separated by two or more letters.
- The letters are not important, the Subject does not have to remember them.
- In each trial, the first digit is called "the FIRST number", while the second digit is called "the SECOND number".

Thus, at the end of each trial, two questions are posed to the Participant:

1. **"What was the FIRST number?"**

2. **“What was the SECOND number?”**

- When question 1) appears on the screen, the Participant should type on the keypad of the computer keyboard in front of her/him the first number she/he saw during the trial.
  - If he did not see any number, the Participant should type 0.
  - Instead, if the Participant saw a number, but is not sure about which number it was the recommendation is to guess that number and type it.
- When question 2) appears on the screen, the Participant should type on the keypad of the computer keyboard in front of her/him the second number she/he saw during the trial.
  - If he did not see it, the Participant should type 0.
  - Instead, if the Participant saw a number, but is not sure about which number it was the recommendation is to guess that number and type it.

There is no time constraint when responding to the two questions at the end of each trial.

As soon as the Participant has typed the second number (i.e., as she/he has answered to question 2), a new trial begins with a fixation cross-hair at the screen centre.

Overall, the Participant will complete 152 trials, subdivided into **3 blocks**:

- **Block No.1** is a **“Practice Block”**, and is intended for the Participant to become familiar with the experimental set-up and its requirements. The Practice Block comprises **8 consecutive trials**: Participant’s responses to such trials will not be included in the successive analysis process. Only during the Practice Block after the Participant’s response to each question the number actually appeared during the trial (i.e., the correct answer) will be presented to the Participant as a feedback. Such a feedback will not be present during the successive experimental blocks.
- The remaining **2 blocks** are called **“Experimental Blocks”** and form the actual experiment. Each of the Experimental Blocks consists of **72 trials**, and the Participant’s responses to such trials will be collected and analysed for getting information on the attention processing skills of the Participant.
- Between two consecutive experimental blocks, the Participant should take a two-minute rest. The experimenter will press a key to let the successive block to start.

At the beginning, before starting the Practice Block, the Participant will have to provide the experimenter with some information (ID, age and problems with vision if any). The experimenter will collect such data that will remain anonymous in possible future publications.

You may refer to the sketch below for a graphic summary of the experiment procedure (Figure A.1).

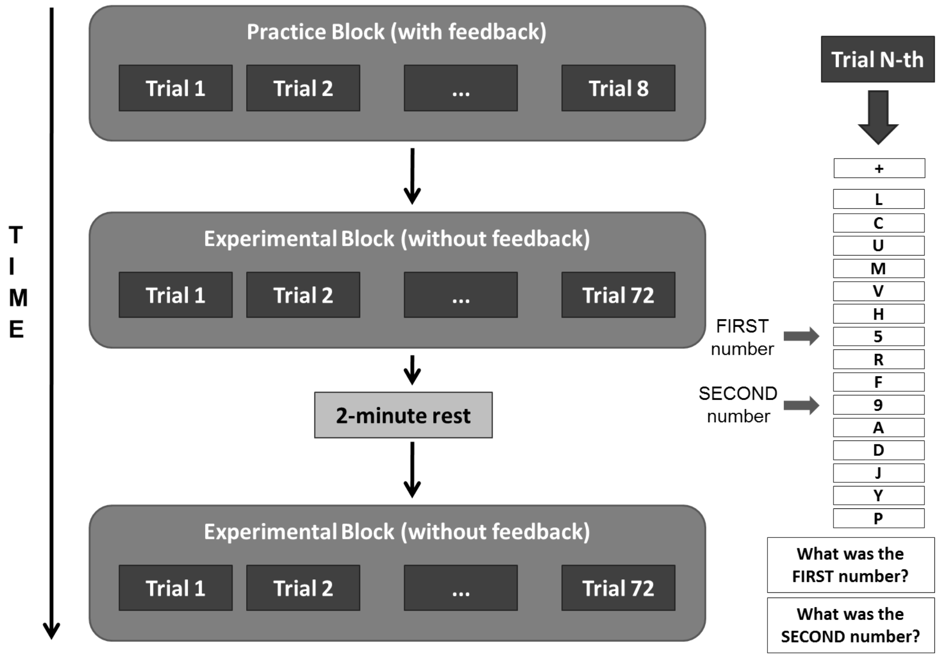


Figure A.1.: Sketch of Study 1 experimental procedure included in instructions for participants. The original caption wrote: "Sketch of the overall experimental procedure and of a single trial (the depicted sequence is for illustrative purpose only)".

A.3. Instructions for Participants in Study 2

Below is reported the text that participants in Study 2 were asked to read before taking part in the experiment. Details about the specific session were provided in addition by the experimenter upon request from the participant. An equivalent version of these instructions written in Italian was used for the Italian-speaking participants.

Instructions for the Attentional Blink Experimental Sessions

3 experimental sessions are foreseen for each participant. Each session should be undertaken in a different day. The 3 sessions are referred to as: "Desktop", Distraction", and "mini-CAVE". The first 2 sessions are run by watching to a computer screen and using a keyboard to provide responses, while in the "mini-cave" session the visualization is provided by an immersive 3D virtual environment, and the responses are still delivered through a keyboard. In every session, the participant is asked to watch a sequence of visual stimuli appearing all at the same spatial location on the visual display. The participant has to look for some special stimuli, that we refer to as targets, and that s-/he is asked to remember and later report in the correct order as they appeared in the just finished sequence. All the visual elements, included targets, are road signs. The targets are road signs depicting animals and vehicles, while the other road signs only show arrows. The sequences differ in the type of targets they include, and in the temporal separation between their onsets on the visual display during each sequence; in addition, in the "Distraction" session, the sequences also differ about whether a concurrent sound is either regular, or deviant during the sequences visualization. The sound will be that of an idle car engine, as waiting in front of a traffic light. All the required responses are delivered by the participant by means of a computer keyboard. The participant is instructed to provide unspeeded responses, and to guess the response anyway when not sure of it. No collection of the participant's response time is performed. The "mini-CAVE" session is substantially identical to the "Desktop" session, except for the use in the former of an immersive virtual environment representing a motorway for the presentation of the visual stimuli. During one of the blocks of trials in the "Distraction" session the participant is additionally asked to recognize if a sound playing concurrently with the visual sequence is regular or not, based on the definition of regular and deviant sounds learned in a training phase. Each session consists of 3 phases:

- A training phase, during which the participant first observes and learns the names of the target stimuli s-/he will have to look for in the sequences of the subsequent phases, and then takes a test to check if s-/he learned those target names.
- A practice phase, during which the participant practices a few sample trials, similar to those that will constitute the actual experiment in the next experimental phase.
- An experimental phase, where participants responses to 2 blocks of trials are collected. These 2 blocks are separated by a rest interval lasting 2 minutes.

Messages to warn participants about the phase starting, and to ask participants' input will always appear on the screen at appropriate moments.

A.4. Questionnaire for Participants in Study 2

Below is reported the questionnaire administered to the participants in the Study 2 experiment prior to starting their first session.

Attentional Blink in Virtual Environments

- Session info:
- Date/Data:
- Participant ID / ID Partecipante:
- Sex/Sesso: ☐ F ☐ M
- Age/Età:
- Vision/Vista: ☐ Lenses or spectacles/Lenti o occhiali da vista?

1. Rate the intensity of your previous experience at using the computer (from 1 – No experience, to 5 – Assiduous use) / Quantifica la tua esperienza pregressa nell'uso del computer (da 1 – Nessuna esperienza a 5 – Assiduo utilizzo)

   1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

2. Rate the intensity of your previous experience with video games (from 1 – No experience, to 5 – Assiduous use) / Quantifica la tua esperienza pregressa nell'uso di video games (da 1 – Nessuna esperienza a 5 – Assiduo utilizzo)

   1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

3. Rate the intensity of your previous experience with virtual environments (from 1 – No experience, to 5 – Assiduous use) / Quantifica la tua esperienza pregressa come utente di ambienti virtuali (da 1 – Nessuna esperienza a 5 – Assiduo utilizzo)

   1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Bibliography

Abramowitz, J., S. Taylor, and D. McKay (2009). Obsessive-compulsive disorder. *The Lancet 374*(9688), 491–499.

Adamovich, S., G. Fluet, E. Tunik, and A. Merians (2009). Sensorimotor training in virtual reality: a review. *NeuroRehabilitation 25*(1), 29–44.

Alcorn, A., H. Pain, G. Rajendran, T. Smith, O. Lemon, K. Porayska-Pomsta, M. Foster, K. Avramides, C. Frauenberger, and S. Bernardini (2011). Social communication between virtual characters and children with autism. In G. Biswas, S. Bull, J. Kay, and A. Mitrovic (Eds.), *Artificial Intelligence in Education*, Volume 6738 of *Lecture Notes in Computer Science*, pp. 7–14. Springer Berlin Heidelberg.

Alonso, M., F. Vexo, and D. Thalmann (2008). *Stepping into virtual reality.* Springer-Verlag New York Inc.

Alvarez, G. and P. Cavanagh (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science 15*(2), 106–111.

Anderson, P., E. Zimand, L. Hodges, and B. Rothbaum (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety 22*(3), 156–158.

Arend, I., S. Johnston, and K. Shapiro (2006). Task-irrelevant visual motion and flicker attenuate the attentional blink. *Psychonomic Bulletin & Review 13*(4), 600–607.

Arnell, K., K. Killman, and D. Fijavz (2007). Blinded by emotion: target misses follow attention capture by arousing distractors in RSVP. *Emotion 7*(3), 465.

Aschoff, A., A. Ernst, H. Brambs, and M. Juchems (2008). CT colonography: an update. *European radiology 18*(3), 429–437.

Aston-Jones, G., J. Rajkowski, and J. Cohen (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological psychiatry 46*(9), 1309–1320.

Awh, E., B. Barton, and E. Vogel (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological science 18*(7), 622.

Awh, E., J. Serences, P. Laurey, H. Dhaliwal, T. van der Jagt, and P. Dassonville (2004). Evidence against a central bottleneck during the attentional blink: Multiple channels for configural and featural processing. *Cognitive psychology 48*(1), 95–126.

Awh, E., E. Vogel, and S. Oh (2006). Interactions between attention and working memory. *Neuroscience 139*(1), 201–208.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences 4*(11), 417–423.

Baddeley, A. (2010). Working memory. *Current Biology 20*(4), R136 – R140.

Baddeley, A., G. Hitch, et al. (1974). Working memory. In *The psychology of learning and motivation*, Volume 8, pp. 47–89. Academic Press New York.

Baer, R. A., E. Walsh, and E. L. B. Lykins (2009). Assessment of mindfulness. In F. Didonna (Ed.), *Clinical Handbook of Mindfulness*, pp. 153–168. Springer New York.

Baheux, K., M. Yoshizawa, and Y. Yoshida (2007). Simulating hemispatial neglect with virtual reality. *Journal of NeuroEngineering and Rehabilitation 4*(1), 27.

Barnard, P., S. Scott, J. Taylor, J. May, and W. Knightley (2004). Paying attention to meaning. *Psychological Science 15*(3), 179–186.

Battista, G., C. Sassi, R. Schiavina, A. Franceschelli, E. Baglivo, G. Martorana, and R. Canini (2009). Computerized tomography virtual endoscopy in evaluation of upper urinary tract tumors: initial experience. *Abdominal imaging 34*(1), 107–112.

Bellani, M., L. Fornasari, L. Chittaro, and P. Brambilla (2011). Virtual reality in autism: state of the art. *Epidemiology and Psychiatric Sciences 20*(3), 235–238.

Bergamasco, M. (2012). Virtual environments and augmented reality for skill training. See Bergamasco et al. (2012).

Bergamasco, M., C. Avizzano, G. Di Pietro, F. Barbagli, and A. Frisoli (2001). The museum of pure form: system architecture. In *Robot and Human Interactive Communication, 2001. Proceedings. 10th IEEE International Workshop on*, pp. 112–117. IEEE.

Bergamasco, M., B. G. Bardy, and D. Gopher (Eds.) (2012). *Skills training in multimodal virtual environments.* CRC Press, Taylor & Francis Group, Boca Raton (FL, USA).

Bergamasco, M., A. Frisoli, and C. Avizzano (2007). Exoskeletons as man-machine interface systems for teleoperation and interaction in virtual environments. In M. Ferre, M. Buss, R. Aracil, C. Melchiorri, and C. Balaguer (Eds.), *Advances in Telerobotics*, Volume 31 of *Springer Tracts in Advanced Robotics*, pp. 61–76. Springer Berlin Heidelberg.

Biocca, F. (1992). Communication within virtual reality: Creating a space for research. *Journal of Communication 42*(4), 5–22.

Biocca, F. (1997). The cyborg's dilemma: Embodiment in virtual environments. In *Cognitive Technology, 1997.'Humanizing the Information Age'. Proceedings., Second International Conference on*, pp. 12–26. IEEE.

Biocca, F. and B. Delaney (1995). Immersive virtual reality technology. In F. Biocca and M. Levy (Eds.), *Communication in the age of virtual reality*, pp. 57–124. Lawrence Erlbaum.

Bishop, S., M. Lau, S. Shapiro, L. Carlson, N. Anderson, J. Carmody, Z. Segal, S. Abbey, M. Speca, D. Velting, et al. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology Science and Practice 11*(3), 230–241.

Bohil, C., B. Alicea, and F. Biocca (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience 12*(12), 752–762.

Botden, S. and J. Jakimowicz (2009). What is going on in augmented reality simulation in laparoscopic surgery? *Surgical endoscopy 23*(8), 1693–1700.

Bowman, H. and B. Wyble (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review; Psychological Review 114*(1), 38.

Brainard, D. (1997). The psychophysics toolbox. *Spatial vision 10*(4), 433–436.

Brefczynski-Lewis, J. A., A. Lutz, H. S. Schaefer, D. B. Levinson, and R. J. Davidson (2007). Neural correlates of attentional expertise in long-term meditation practitioners. *Proceedings of the National Academy of Sciences 104*(27), 11483–11488.

Broadbent, D. and M. Broadbent (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Attention, Perception, & Psychophysics 42*(2), 105–113.

Brooks, B., F. Rose, J. Potter, E. Attree, S. Jayawardena, and A. Morling (2002). Assessing stroke patients' ability to remember to perform actions in the future using virtual reality. In *Proceedings of the 4th International Conference on Disability, Virtual Reality and Associated Technologies*, pp. 239–245.

Burdea, G. (2003). Virtual rehabilitation–benefits and challenges. *Methods of information in medicine 42*(5), 519–523.

Caglio, M., L. Latini-Corazzini, F. D'agata, F. Cauda, K. Sacco, S. Monteverdi, M. Zettin, S. Duca, and G. Geminiani (2009). Video game play changes spatial and verbal memory: rehabilitation of a single case with traumatic brain injury. *Cognitive processing 10*, 195–197.

Cambridge Dictionaries Online (2012). "attention". In *Cambridge Dictionaries Online*. Cambridge University Press.

Carrozzino, M. and M. Bergamasco (2010). Beyond virtual museums: Experiencing immersive virtual reality in real museums. *Journal of Cultural Heritage 11*(4), 452–458.

Carrozzino, M., F. Tecchia, S. Bacinelli, C. Cappelletti, and M. Bergamasco (2005). Lowering the development time of multimodal interactive application: the real-life experience of the xvr project. In *Proceedings of the 2005 ACM SIGCHI In-*

ternational Conference on Advances in computer entertainment technology*, pp. 270–273. ACM.

Carter, O., D. Presti, C. Callistemon, Y. Ungerer, G. Liu, and J. Pettigrew (2005). Meditation alters perceptual rivalry in Tibetan Buddhist monks. *Current biology 15*(11), R412–R413.

Cheng, Y. and J. Ye (2010). Exploring the social competence of students with autism spectrum conditions in a collaborative virtual learning environment–the pilot study. *Computers & Education 54*(4), 1068–1077.

Chiesa, A., R. Calati, and A. Serretti (2011). Does mindfulness training improve cognitive abilities? a systematic review of neuropsychological findings. *Clinical Psychology Review 31*(3), 449–464.

Childs, M. (2010). A conceptual framework for mediated environments. *Educational Research 52*(2), 197–213.

Chun, M., J. Golomb, and N. Turk-Browne (2010). A Taxonomy of External and Internal Attention. *Psychology 62*(1), 2011.

Chun, M. and M. Potter (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance 21*(1), 109.

Coates, G. (1992). Program from invisible site-a virtual show, a multimedia performance work presented by george coates performance works. *San Francisco, CA*.

Coelho, C., J. Tichon, T. Hine, G. Wallis, and G. Riva (2006). Media presence and inner presence: The sense of presence in virtual reality technologies. *From Communication to Presence: Cognition, Emotions and Culture towards the Ultimate Communicative Experience*, 25–45.

Cohen, L., C. Warneke, R. Fouladi, M. Rodriguez, and A. Chaoul-Reich (2004). Psychological adjustment and sleep quality in a randomized trial of the effects of a tibetan yoga intervention in patients with lymphoma. *Cancer 100*(10), 2253–2260.

Corbetta, M. and G. Shulman (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience 3*(3), 201–215.

Cruz-Neira, C., D. Sandin, and T. DeFanti (1993). Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pp. 135–142. ACM.

Cushman, L., K. Stein, and C. Duffy (2008). Detecting navigational deficits in cognitive aging and alzheimer disease using virtual reality. *Neurology 71*(12), 888–895.

de Jong, P., E. Koster, R. Van Wees, and S. Martens (2009). Emotional facial expressions and the attentional blink: Attenuated blink for angry and happy faces irrespective of social anxiety. *Cognition and Emotion 23*(8), 1640–1652.

de Jong, P., E. Koster, R. van Wees, and S. Martens (2010). Angry facial expressions hamper subsequent target identification. *Emotion 10*(5), 727.

de Vignemont, F. (2010). Body schema and body image–pros and cons. *Neuropsychologia 48*(3), 669 – 680.

Dehaene, S., C. Sergent, and J. Changeux (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. In *Proceedings of the National Academy of Sciences*, Volume 100, pp. 8520. National Acad Sciences.

Dell'Acqua, R., P. Jolicœur, P. Sessa, and M. Turatto (2006). Attentional blink and selection in the tactile domain. *European Journal of Cognitive Psychology 18*(4), 537–559.

Dell'Acqua, R., J. Pierre, A. Pascali, and P. Pluchino (2007). Short-term consolidation of individual identities leads to lag-1 sparing. *Journal of Experimental Psychology: Human Perception and Performance 33*(3), 593.

Desimone, R. and J. Duncan (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience 18*(1), 193–222.

Di Lollo, V., J. Kawahara, S. Shahab Ghorashi, and J. Enns (2005). The attentional blink: Resource depletion or temporary loss of control? *Psychological Research 69*(3), 191–200.

Difede, J., J. Cukor, N. Jayasinghe, I. Patt, S. Jedel, L. Spielman, C. Giosan, H. Hoffman, et al. (2007). Virtual reality exposure therapy for the treatment of posttraumatic stress disorder following september 11, 2001. *Journal of Clinical Psychiatry 68*(11), 1639–1647.

Duncan, J., S. Martens, and R. Ward (1997, June). Restricted attentional capacity within but not between sensory modalities. *Nature 387*, 808–810.

Dux, P. and I. Harris (2007). Viewpoint costs occur during consolidation: Evidence from the attentional blink. *Cognition 104*(1), 47–58.

Dux, P. and R. Marois (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics 71*(8), 1683.

Einhäuser, W., C. Koch, and S. Makeig (2007). The duration of the attentional blink in natural scenes depends on stimulus category. *Vision research 47*(5), 597–607.

Elkind, J., E. Rubin, S. Rosenthal, B. Skoff, and P. Prather (2001). A simulated reality scenario compared with the computerized wisconsin card sorting test: An analysis of preliminary results. *Cyberpsychology & Behavior 4*(4), 489–496.

Evans, K. and A. Treisman (2005). Perception of objects in natural scenes: is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance 31*(6), 1476.

Faraone, S., J. Sergeant, C. Gillberg, and J. Biederman (2003). The worldwide prevalence of adhd: is it an american condition? *World Psychiatry 2*(2), 104.

Ferlazzo, F., S. Lucido, F. Di Nocera, S. Fagioli, and S. Sdoia (2007). Switching between goals mediates the attentional blink effect. *Experimental psychology 54*(2), 89.

Fernández-Aranda, F., S. Jiménez-Murcia, J. Santamaría, K. Gunnard, A. Soto, E. Kalapanidas, R. Bults, C. Davarakis, T. Ganchev, R. Granero, et al. (2012). Video games as a complementary therapy tool in mental disorders: Playmancer, a european multicentre study. *Journal of Mental Health 21*(4), 364–374.

Fernandez-Duque, D. and M. Johnson (2002). Cause and effect theories of attention: The role of conceptual metaphors. *Review of general psychology 6*(2), 153.

Ferrer-García, M. and J. Gutiérrez-Maldonado (2012). The use of virtual reality in the study, assessment, and treatment of body image in eating disorders and nonclinical samples: a review of the literature. *Body Image 9*(1), 1–11.

Foa, E. and M. Kozak (1986). Emotional processing of fear: exposure to corrective information. *Psychological bulletin 99*(1), 20.

Franchi, J. (1994). Virtual reality: An overview. *TechTrends 39*, 23–26.

Freeman, D. (2008). Studying and treating schizophrenia using virtual reality: a new paradigm. *Schizophrenia bulletin 34*(4), 605–610.

Freeman, J., S. Avons, R. Meddis, D. Pearson, and W. IJsselsteijn (2000). Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators & Virtual Environments 9*(2), 149–164.

Freeman, J., S. Avons, D. Pearson, and W. IJsselsteijn (1999). Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators & Virtual Environments 8*(1), 1–13.

Frisoli, A., F. Salsedo, M. Bergamasco, B. Rossi, and M. Carboncini (2009). A force-feedback exoskeleton for upper-limb rehabilitation in virtual reality. *Applied Bionics and Biomechanics 6*(2), 115–126.

Gallagher, A., E. Ritter, H. Champion, G. Higgins, M. Fried, G. Moses, C. Smith, and R. Satava (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of surgery 241*(2), 364.

Garcia-Palacios, A., H. Hoffman, A. Carlin, T. Furness III, and C. Botella (2002). Virtual reality in the treatment of spider phobia: a controlled study. *Behaviour Research and Therapy 40*(9), 983–993.

Geiszt, Z., M. Kamson, C. Lányi, and J. Stark (2006). Do we need high-scale flexibility in virtual therapies? *International Journal on Disability and Human Development, Special issue on disability, virtual reality and associated technologies, 5 3*, 251–256.

Gibson, J. (1986). *The ecological approach to visual perception*. Lawrence Erlbaum.

Gosselin, F., F. Ferlay, S. Bouchigny, C. Mégard, and F. Taha (2010). Design of a multimodal vr platform for the training of surgery skills. In A. Kappers, J. van Erp, W. Bergmann Tiest, and F. van der Helm (Eds.), *Haptics: Generating and Perceiving Tangible Sensations*, Volume 6192 of *Lecture Notes in Computer Science*, pp. 109–116. Springer Berlin / Heidelberg.

Green, C., D. Bavelier, et al. (2003). Action video game modifies visual selective attention. *Nature 423*(6939), 534–537.

Greenbaum, P. (1992). The lawnmower man. *Film and video 9*(3), 58–62.

Gregg, L. and N. Tarrier (2007). Virtual reality in mental health: a review of the literature. *Social Psychiatry and Psychiatric Epidemiology 42*(5), 343–354.

Gutierrez, T., J. Rodriguez, Y. Velaz, S. Casado, A. Suescun, and E. Saánchez (2010). Ima-vr: a multimodal virtual training system for skills transfer in industrial maintenance and assembly tasks. In *RO-MAN, 2010 IEEE*, pp. 428–433. IEEE.

Haroush, K., L. Y. Deouell, and S. Hochstein (2011). Hearing while blinking: Multisensory attentional blink revisited. *The Journal of Neuroscience 31*(3), 922–927.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments 1*(2), 262–271.

Henderson, A., N. Korner-Bitensky, and M. Levin (2007). Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery. *Topics in stroke rehabilitation 14*(2), 52–61.

Hitch, A. and G. J. Baddeley (2010). Working memory. *Scholarpedia 5*(2), 3015.

Hoffman, H., D. Patterson, J. Magula, G. Carrougher, K. Zeltzer, S. Dagadakis, and S. Sharar (2004). Water-friendly virtual reality pain control during wound care. *Journal of Clinical psychology 60*(2), 189–195.

Hoffman, H., D. Patterson, E. Seibel, M. Soltani, L. Jewett-Leahy, and S. Sharar (2008). Virtual reality pain control during burn wound debridement in the hydrotank. *The Clinical journal of pain 24*(4), 299–304.

Hölzel, B., S. Lazar, T. Gard, Z. Schuman-Olivier, D. Vago, and U. Ott (2011). How does mindfulness meditation work? proposing mechanisms of action from a conceptual and neural perspective. *Perspectives on Psychological Science 6*(6), 537–559.

Hommel, B. and E. Akyürek (2003). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *The Quarterly Journal of Experimental Psychology Section A 58*(8), 1415–1433.

Ijsselsteijn, W., H. De Ridder, J. Freeman, and S. Avons (2000). Presence: Concept, determinants and measurement. In *SPIE proceedings series*, pp. 520–529. Society of Photo-Optical Instrumentation Engineers.

IJsselsteijn, W. and M. Reiner (2004). On the importance of reliable real-time sensorimotor dependencies for establishing telepresence. In *Presence, 7th Annual International Workshop on*, pp. 298.

IJsselsteijn, W. and G. Riva (2003). Being there: The experience of presence in mediated environments. In *Being there: Concepts, effects and measurements of user presence in synthetic environments*, Volume 5, pp. 3–16. IOS Press/Ohmsha Amsterdam/Washington, DC/Tokyo.

James, W. (1890). *The Principles of Psychology.* New York: Holt,Rinehart & Winston.

Jha, A., J. Krompinger, and M. Baime (2007). Mindfulness training modifies subsystems of attention. *Cognitive, Affective, & Behavioral Neuroscience 7*(2), 109–119.

Johansson, B. (2000). Brain plasticity and stroke rehabilitation: the Willis lecture. *Stroke 31*(1), 223–230.

Johnson, A. and R. Proctor (2004). *Attention: Theory and practice.* Sage Publications, Inc.

Johnstone, B. and H. Stonnington (2009). *Rehabilitation of neuropsychological disorders: a practical guide for rehabilitation professionals.* Psychology Press.

Jolicoeur, P. (1998). Modulation of the attentional blink by on-line response selection: Evidence from speeded and unspeeded task 1 decisions. *Memory & Cognition 26*(5), 1014–1032.

Jolicoeur, P. (1999). Restricted attentional capacity between sensory modalities. *Psychonomic Bulletin & Review 6*(1), 87–92.

Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical psychology: Science and practice 10*(2), 144–156.

Kahneman, D. (1973). *Attention and effort.* Prentice Hall.

Karhuketo, T., P. Dastidar, P. Ryymin, E. Laasonen, and H. Puhakka (2002). Virtual endoscopy imaging of the middle ear cavity and ossicles. *European archives of oto-rhino-laryngology 259*(2), 77–83.

Katz, N., H. Ring, Y. Naveh, R. Kizony, U. Feintuch, and P. Weiss (2005). Interactive virtual environment training for safe street crossing of right hemisphere stroke patients with unilateral spatial neglect. *Disability & Rehabilitation 27*(20), 1235–1244.

Kelly, K., A. Heilbrun, and B. Stacks (1989). Virtual reality; an interview with jaron lanier. *Whole Earth Review 64*(2), 108–120.

Kim, K., C. Kim, K. Cha, J. Park, K. Han, Y. Kim, J. Kim, I. Kim, and S. Kim (2008). Anxiety provocation and measurement using virtual reality in patients with obsessive-compulsive disorder. *CyberPsychology & Behavior 11*(6), 637–641.

Kim, K., J. Kim, J. Kim, D. Park, H. Jang, J. Ku, C. Kim, I. Kim, and S. Kim (2006). Characteristics of social perception assessed in schizophrenia using virtual reality. *CyberPsychology & Behavior 10*(2), 215–219.

Kim, K., S. Kim, K. Cha, J. Park, M. Rosenthal, J. Kim, K. Han, I. Kim, and C. Kim (2010). Development of a computer-based behavioral assessment of checking behavior in obsessive-compulsive disorder. *Comprehensive psychiatry 51*(1), 86–93.

Kleiner, M., D. Brainard, D. Pelli, A. Ingling, R. Murray, and C. Broussard (2007). What's new in psychtoolbox-3. *Perception 36*(ECVP Abstract Supplement), 14.

Kober, S. E. and C. Neuper (2012). Using auditory event-related eeg potentials to assess presence in virtual reality. *International Journal of Human-Computer Studies 70*(9), 577 – 587.

Kockro, R. and P. Hwang (2009). Virtual temporal bone: an interactive 3-dimensional learning aid for cranial base surgery. *Neurosurgery 64*(5), ons216–ons230.

Konietschke, R., A. Tobergte, C. Preusche, P. Tripicchio, E. Ruffaldi, S. Webel, and U. Bockholt (2010). A multimodal training platform for minimally invasive robotic surgery. In *RO-MAN, 2010 IEEE*, pp. 422–427. IEEE.

Krijn, M., P. Emmelkamp, R. Olafsson, M. Bouwman, L. van Gerwen, P. Spinhoven, M. Schuemie, and C. van der Mast (2007). Fear of flying treatment methods: Virtual reality exposure vs. cognitive behavioral therapy. *Aviation, space, and environmental medicine 78*(2), 121–128.

Krummel, T. (1998). Surgical simulation and virtual reality: the coming revolution. *Annals of surgery 228*(5), 635.

Lapointe-Goupil, R., C. Fortin, B. Brisson, and S. Tremblay (2011). Concurrency benefits in the attentional blink: Attentional flexibility and shifts of decision criteria. *Attention, Perception, & Psychophysics*, 1–15.

Lauscher, J., J. Ritz, A. Stroux, H. Buhr, and J. Gröne (2010). A new surgical trainer (bopt) improves skill transfer for anastomotic techniques in gastrointestinal surgery into the operating room: A prospective randomized trial. *World journal of surgery 34*(9), 2017–2025.

Lavie, N., A. Hirst, J. de Fockert, and E. Viding (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General 133*(3), 339.

Lee, K. (2004). Presence, explicated. *Communication Theory 14*(1), 27–50.

Lengenfelder, J., M. Schultheis, T. Al-Shihabi, R. Mourant, and J. DeLuca (2002). Divided attention and driving: a pilot study using virtual reality technology. *The Journal of Head Trauma Rehabilitation 17*(1), 26.

Lenggenhager, B., T. Tadi, T. Metzinger, and O. Blanke (2007). Video ergo sum: manipulating bodily self-consciousness. *Science 317*(5841), 1096–1099.

Levin, D. and D. Simons (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review 4*(4), 501–506.

Livingstone, S. and R. Skelton (2007). Virtual environment navigation tasks and the assessment of cognitive deficits in individuals with brain injury. *Behavioural brain research 185*(1), 21–31.

Lombard, M. and T. Ditton (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication 3*(2).

Loomis, J. (1992). Distal attribution and presence. *Presence: Teleoperators & Virtual Environments 1*(1), 113–119.

Loomis, J., J. Blascovich, and A. Beall (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods Instruments and Computers 31*(4), 557–564.

Louis, N., P. Desgranges, H. Kobeiter, M. Kirsch, and J. Becquemin (2009). Virtual angioscopy and 3-dimensional navigation findings of the aortic arch after vascular surgery. *Circulation 119*(7), 1052–1055.

Lu, J., J. Chen, H. Çakmak, H. Maass, U. Kuhnapfel, and G. Bretthauer (2009). A knee arthroscopy simulator for partial meniscectomy training. In *Asian Control Conference, 2009. ASCC 2009. 7th*, pp. 763–767. IEEE.

Luck, S., E. Vogel, et al. (1997). The capacity of visual working memory for features and conjunctions. *Nature 390*(6657), 279–280.

Luck, S., E. Vogel, K. Shapiro, et al. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature 383*(6601), 616–618.

Lutz, A., H. Slagter, J. Dunne, and R. Davidson (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences 12*(4), 163–169.

Mabrey, J., K. Reinig, and W. Cannon (2010). Virtual reality in orthopaedics: Is it a reality? *Clinical Orthopaedics and Related Research® 468*(10), 2586–2591.

MacLean, K., E. Ferrer, S. Aichele, D. Bridwell, A. Zanesco, T. Jacobs, B. King, E. Rosenberg, B. Sahdra, P. Shaver, et al. (2010). Intensive Meditation Training Improves Perceptual Discrimination and Sustained Attention. *Psychological Science 21*(6), 829.

Malloy, K. and L. Milling (2010). The effectiveness of virtual reality distraction for pain reduction: a systematic review. *Clinical Psychology Review 30*(8), 1011–1018.

Marescaux, J., J. Leroy, M. Gagner, F. Rubino, D. Mutter, M. Vix, S. Butner, and M. Smith (2001). Transatlantic robot-assisted telesurgery. *Nature 413*(6854), 379–380.

Marois, R., D. Yi, M. Chun, et al. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron 41*(3), 465–472.

Martens, S. and A. Johnson (2005). Timing attention: Cuing target onset interval attenuates the attentional blink. *Memory & cognition 33*(2), 234–240.

Martens, S. and A. Johnson (2009). Working memory capacity, intelligence, and the magnitude of the attentional blink revisited. *Experimental Brain Research 192*(1), 43–52.

Martens, S., M. Kandula, and J. Duncan (2010, 12). Restricted attentional capacity within but not between sensory modalities: An individual differences approach. *PLoS ONE 5*(12), e15280.

Martens, S., J. Munneke, H. Smid, and A. Johnson (2006). Quick minds don't blink: Electrophysiological correlates of individual differences in attentional selection. *Journal of Cognitive Neuroscience 18*(9), 1423–1438.

Martens, S. and N. Valchev (2009). Individual differences in the attentional blink. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie) 56*(1), 18–26.

Martens, S. and B. Wyble (2010). The attentional blink: Past, present, and future of a blind spot in perceptual awareness. *Neuroscience & Biobehavioral Reviews 34*(6), 947–957.

Martin, E. and K. Shapiro (2008). Does failure to mask t1 cause lag-1 sparing in the attentional blink? *Attention, Perception, & Psychophysics 70*(3), 562–570.

Maye, A. and A. Engel (2011). A discrete computational model of sensorimotor contingencies for object perception and control of behavior. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pp. 3810–3815. IEEE.

McDowd, J. (2007). An overview of attention: behavior and brain. *Journal of Neurologic Physical Therapy 31*(3), 98.

Meehan, M., B. Insko, M. Whitton, and F. Brooks Jr (2002). Physiological measures of presence in stressful virtual environments. In *ACM Transactions on Graphics (TOG)*, Volume 21, pp. 645–652. ACM.

Meehan, M., S. Razzaque, B. Insko, M. Whitton, and F. Brooks (2005). Review of four studies on the use of physiological reaction as a measure of presence in stressfulvirtual environments. *Applied psychophysiology and biofeedback 30*(3), 239–258.

Mégard, C., F. Gosselin, S. Bouchigny, F. Ferlay, and F. Taha (2009). User-centered design of a maxillo-facial surgery training platform. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pp. 265–266. ACM.

Merriam-Webster Dictionaries Online (2012). "attention". In *Merriam-Webster.com.* Encyclopaedia Britannica.

Minsky, M. (1980). Telepresence. *Omni 2*(9), 45–52.

Mitchell, P., S. Parsons, and A. Leonard (2007). Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders. *Journal of autism and developmental disorders 37*(3), 589–600.

Moore, D., Y. Cheng, P. McGrath, and N. Powell (2005). Collaborative virtual environment technology for people with autism. *Focus on Autism and Other Developmental Disabilities 20*(4), 231–243.

Morganti, F., A. Gaggioli, L. Strambi, M. Rusconi, and G. Riva (2007). A virtual reality extended neuropsychological assessment for topographical disorientation: a feasibility study. *Journal of NeuroEngineering and Rehabilitation 4*, 26.

Morganti, F., S. Marrakchi, P. Urban, G. Iannoccari, and G. Riva (2009). A virtual reality based tool for the assessment of "survey to route" spatial organization ability in elderly population: preliminary data. *Cognitive processing 10*(2), 257–259.

Morris, R., M. Kotitsa, J. Bramham, B. Brooks, and F. Rose (2002). Virtual reality investigation of strategy formation, rule breaking and prospective memory in patients with focal prefrontal neurosurgical lesions. In *Proceedings of the 4th International Conference on Disability, Virtual Reality and Associated Technology. Vresprem, Hungary: University of Reading (United Kingdom)*, pp. 101–108.

Most, S., M. Chun, D. Widders, and D. Zald (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review 12*(4), 654–661.

Murray, C., S. Pettifer, T. Howard, E. Patchick, F. Caillette, and J. Murray (2010). Virtual solutions to phantom problems: Using immersive virtual reality to treat phantom limb pain. *Amputation, Prosthesis Use, and Phantom Limb Pain*, 175–196.

Navon, D. and D. Gopher (1979). On the economy of the human-processing system. *Psychological review 86*(3), 214.

Nieuwenhuis, S., M. Gilzenrat, B. Holmes, and J. Cohen (2005). The role of the locus coeruleus in mediating the attentional blink: a neurocomputational theory. *Journal of Experimental Psychology: General 134*(3), 291.

Nieuwenstein, M. (2006). Top-down controlled, delayed selection in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance 32*(4), 973.

Nieuwenstein, M., M. Chun, R. Van Der Lubbe, and I. Hooge (2005). Delayed attentional engagement in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance 31*(6), 1463.

Nieuwenstein, M. and M. Potter (2006). Temporal limits of selection and memory encoding a comparison of whole versus partial report in rapid serial visual presentation. *Psychological Science 17*(6), 471–475.

Nieuwenstein, M., M. Potter, and J. Theeuwes (2009). Unmasking the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance 35*(1), 159.

Nolin, P., C. Martin, S. Bouchard, et al. (2009). Assessment of inhibition deficits with the virtual classroom in children with traumatic brain injury: a pilot-study. *Studies in Health Technology and Informatics 144*, 240–242.

Olivers, C. and M. Meeter (2008). A boost and bounce theory of temporal attention. *Psychological Review 115*(4), 836.

Olivers, C. and S. Nieuwenhuis (2005). The beneficial effect of concurrent task-irrelevant mental activity on temporal attention. *Psychological Science 16*(4), 265–269.

Olivers, C. and S. Nieuwenhuis (2006). The beneficial effects of additional task load, positive affect, and instruction on the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance 32*(2), 364.

Olivers, C., S. van der Stigchel, and J. Hulleman (2007). Spreading the sparing: Against a limited-capacity account of the attentional blink. *Psychological Research 71*(2), 126–139.

Olson, A., B. Eadie, C. Ernst, and B. Christie (2006). Environmental enrichment and voluntary exercise massively increase neurogenesis in the adult hippocampus via dissociable pathways. *Hippocampus 16*(3), 250–260.

O'Regan, J. and A. Noë (2001). What it is like to see: A sensorimotor theory of perceptual experience. *Synthese 129*(1), 79–103.

O'Regan, J., A. Noë, et al. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences 24*(5), 939–972.

Oxford Dictionaries Online (2012, April). "attention". In *Oxford Dictionaries.* Oxford University Press.

Pan, Z., A. D. Cheok, H. Yang, J. Zhu, and J. Shi (2006). Virtual reality and mixed reality for virtual learning environments. *Computers & Graphics 30*(1), 20–28.

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

Parsons, T., T. Bowerly, J. Buckwalter, and A. Rizzo (2007). A controlled clinical comparison of attention performance in children with adhd in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology 13*(4), 363–381.

Parsons, T. and A. Rizzo (2008). Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test. *CyberPsychology & Behavior 11*(1), 17–25.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance 10*(3), 358.

Pashler, H. (1998). Attentional limitations in dual-task performance. In *Attention.* Psychology Press/Erlbaum (UK) Taylor & Francis.

Pashler, H., J. Johnston, and E. Ruthruff (2001). Attention and performance. *Annual review of psychology 52*(1), 629–651.

Pednekar, A. and I. Kakadiaris (2000). Applications of virtual reality in surgery. In *Proceedings of the Indian Conference on Computer Vision, Graphics, and Image Processing (ICVGIP 2000)*, pp. 215–223.

Pelli, D. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision 10*(4), 437–442.

Perez-Marcos, D., M. Slater, and M. Sanchez-Vives (2009). Inducing a virtual hand ownership illusion through a brain-computer interface. *Neuroreport 20*(6), 589–594.

Plumert, J., J. Kearney, J. Cremer, and K. Recker (2005). Distance perception in real and virtual environments. *ACM Transactions on Applied Perception (TAP) 2*(3), 216–233.

Posner, M. and S. Petersen (1990). The attention system of the human brain. *Annual review of neuroscience 13*(1), 25–42.

Posner, M. and M. Rothbart (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology 58*, 1–23.

Potter, M., M. Chun, B. Banks, and M. Muckenhoupt (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and Cognition 24*(4), 979.

Potter, M., A. Staub, and D. O'Connor (2002). The time course of competition for attention: Attention is initially labile. *Journal of Experimental Psychology: Human Perception and Performance 28*(5), 1149.

Potter, M., B. Wyble, R. Pandav, and J. Olejarczyk (2010). Picture detection in rapid serial visual presentation: Features or identity? *Journal of Experimental Psychology: Human Perception and Performance 36*(6), 1486.

Proctor, R. and T. Van Zandt (2008). *Human factors in simple and complex systems.* CRC Press.

Pugnetti, L., L. Mendozzi, B. Brooks, E. Attree, E. Barbieri, D. Alpini, A. Motta, and D. Rose (1998). Active versus passive exploration of virtual environments modulates spatial memory in ms patients: a yoked control study. *The Italian Journal of Neurological Sciences 19*, 424–430.

Raffone, A. and N. Srinivasan (2010). The exploration of meditation in the neuroscience of attention and consciousness. *Cognitive processing 11*(1), 1–7.

Ramachandran, V. and D. Rogers-Ramachandran (1996). Synaesthesia in phantom limbs induced with mirrors. *Proceedings of the Royal Society of London. Series B: Biological Sciences 263*(1369), 377–386.

Rand, D., P. L. (Tamar) Weiss, and N. Katz (2009). Training multitasking in a virtual supermarket: A novel intervention after stroke. *The American Journal of Occupational Therapy 63*(5), 535–542.

Raspelli, S., , L. Carelli, F. Morganti, G. Albani, R. Pignatti, R. Mauro, B. Poletti, B. Corra, V. Silani, and G. Riva (2009). A neuro vr-based virtual version of the multiple errands test for the assessment of executive functions: a possible approach. *Journal of CyberTherapy and Rehabilitation 2*(4), 299–314.

Raymond, J., K. Shapiro, and K. Arnell (1992). Temporary suppression of visual processing in an rsvp task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance 18*(3), 849.

Rego, P., P. Moreira, and L. Reis (2010). Serious games for rehabilitation: A survey and a classification towards a taxonomy. In *Information Systems and Technologies (CISTI), 2010 5th Iberian Conference on*, pp. 1–6. IEEE.

Repetto, C., A. Gorini, C. Vigna, D. Algeri, F. Pallavicini, and G. Riva (2009). The use of biofeedback in clinical virtual reality: the intrepid project. *Journal of Visualized Experiments: JoVE* (33).

Riva, G. (2003). Applications of virtual environments in medicine. *Methods of information in medicine 42*(5), 524–534.

Riva, G. (2006). *Virtual Reality.* John Wiley & Sons, Inc.

Riva, G., J. Waterworth, E. Waterworth, and F. Mantovani (2011). From intention to action: The role of presence. *New ideas in psychology 29*(1), 24–37.

Rizzo, A., T. Bowerly, J. Buckwalter, D. Klimchuk, R. Mitura, and T. Parsons (2006). A virtual reality scenario for all seasons: the virtual classroom. *CNS spectrums 11*(1), 35–44.

Rizzo, A., T. Bowerly, J. Buckwalter, M. Schultheis, R. Matheis, C. Shahabi, U. Neumann, L. Kim, and M. Sharifzadeh (2002). Virtual environments for the assessment of attention and memory processes: the virtual classroom and office. In *Proceedings of the 4th International Conference on Disability, Virtual Reality and Associated Technologies. Reading, UK: University of Reading*, pp. 3–12.

Rizzo, A., J. Buckwalter, C. van der Zaag, U. Neumann, M. Thiebaux, C. Chua, A. van Rooyen, L. Humphrey, and P. Larson (2000). Virtual environment applications in clinical neuropsychology. In *Virtual Reality, 2000. Proceedings. IEEE*, pp. 63 –70.

Rizzo, A., J. Buckwalter, C. van der Zaag, U. Neumann, M. Thiebaux, C. Chua, A. Van Rooyen, L. Humphrey, and P. Larson (2002). Virtual environment applications in clinical neuropsychology. *The Handbook of Virtual Environments*, 1027–1064.

Rizzo, A., G. Reger, G. Gahm, J. Difede, and B. Rothbaum (2009). Virtual reality exposure therapy for combat-related ptsd. *Post-Traumatic Stress Disorder*, 375–399.

Rizzo, A., M. Schultheis, K. Kerns, and C. Mateer (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological rehabilitation 14*(1-2), 207–239.

Robertson, I. H. and S. M. Fitzpatrick (2008, September). The future of cognitive neurorehabilitation. In D. T. Stuss, G. Winocur, and I. H. Robertson (Eds.), *Cognitive Neurorehabilitation Evidence and Application 2nd Edition* (2nd ed.). Cambridge Univ Press.

Robillard, G., S. Bouchard, S. Dumoulin, T. Guitard, and É. Klinger (2010). Using virtual humans to alleviate social anxiety: Preliminary report from a comparative outcome study. *Annual Review of Cybertherapy and Telemedicine 2010: Advanced Technologies in Behavioral, Social and Neurosciences 154*, 57.

Rose, F., E. Attree, B. Brooks, and D. Johnson (1998). Virtual environments in brain damage rehabilitation: a rationale from basic neuroscience. *Studies in Health Technology and Informatics 58*, 233–242.

Rose, F., B. Brooks, and A. Rizzo (2005). Virtual reality in brain damage rehabilitation: review. *CyberPsychology & Behavior 8*(3), 241–262.

Rossi, J., D. Verma, G. Fujii, R. Lakhanpal, S. Wu, M. Humayun, and E. De Juan JR (2004). Virtual vitreoretinal surgical simulator as a training tool. *Retina 24*(2), 231.

Rothbaum, B., P. Anderson, E. Zimand, L. Hodges, D. Lang, and J. Wilson (2006). Virtual reality exposure therapy and standard (in vivo) exposure therapy in the treatment of fear of flying. *Behavior Therapy 37*(1), 80–90.

Ruffaldi, E., A. Filippeschi, C. Avizzano, B. Bardy, D. Gopher, and M. Bergamasco (2011). Feedback, affordances, and accelerators for training sports in virtual environments. *Presence: Teleoperators & Virtual Environments 20*(1), 33–46.

Ruffaldi, E., P. Tripicchio, C. Avizzano, and M. Bergamasco (2011). Haptic rendering of juggling with encountered type interfaces. *Presence: Teleoperators & Virtual Environments 20*(5), 480–501.

Sanchez-Vives, M. and M. Slater (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience 6*(4), 332–339.

Sarter, M., W. Gehring, and R. Kozak (2006). More attention must be paid: the neurobiology of attentional effort. *Brain Research Reviews 51*(2), 145–160.

Sas, C. and G. O'Hare (2003). Presence equation: An investigation into cognitive factors underlying presence. *Presence: Teleoperators & Virtual Environments 12*(5), 523–537.

Schuemie, M., P. Van Der Straaten, M. Krijn, and C. Van Der Mast (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior 4*(2), 183–201.

Schultheis, M. and A. Rizzo (2001). The application of virtual reality technology in rehabilitation. *Rehabilitation Psychology 46*(3), 296.

Sergent, C., S. Baillet, and S. Dehaene (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature neuroscience 8*(10), 1391–1400.

Seth, A., J. Vance, and J. Oliver (2011). Virtual reality for assembly methods prototyping: a review. *Virtual reality 15*(1), 5–20.

Shallice, T. and P. Burgess (1991). Deficits in strategy application following frontal lobe damage in man. *Brain 114*(2), 727–741.

Shapiro, K., J. Raymond, and K. Arnell (1994). Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. *Journal of Experimental psychology: Human perception and performance 20*(2), 357.

Sheridan, T. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators & virtual environments 1*(1), 120–126.

Sheridan, T. (1995). Teleoperation, telerobotics and telepresence: A progress report. *Control Engineering Practice 3*(2), 205–214.

Shipp, S. (2004). The brain circuitry of attention. *Trends in cognitive sciences 8*(5), 223–230.

Slagter, H. A., A. Lutz, L. L. Greischar, A. D. Francis, S. Nieuwenhuis, J. M. Davis, and R. J. Davidson (2007, 05). Mental training affects distribution of limited brain resources. *PLoS Biol 5*(6), e138.

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

Slater, M., P. Khanna, J. Mortensen, and I. Yu (2009). Visual realism enhances realistic response in an immersive virtual environment. *Computer Graphics and Applications, IEEE 29*(3), 76–84.

Slater, M., B. Lotto, M. Arnold, and M. Sanchez-Vives (2009). How we experience immersive virtual environments: the concept of presence and its measurement. *Anuario de psicología 40*(2), 193–210.

Slater, M., D.-P. Pertaub, and A. Steed (1999). Public speaking in virtual reality: facing an audience of avatars. *Computer Graphics and Applications, IEEE 19*(2), 6–9.

Slater, M. and A. Steed (2000). A virtual presence counter. *Presence: Teleoperators & Virtual Environments 9*(5), 413–434.

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

Slater, M. and S. Wilbur (1997). A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence Teleoperators and Virtual Environments 6*(6), 603–616.

Soto-Faraco, S. and C. Spence (2002). Modality-specific auditory and visual temporal processing deficits. *The Quarterly Journal of Experimental Psychology: Section A 55*(1), 23–40.

Spitzer, V. and M. Ackerman (2008). The visible human® at the university of colorado 15 years later. *Virtual Reality 12*(4), 191–200.

Standen, P. and D. Brown (2005). Virtual reality in the rehabilitation of people with intellectual disabilities: review. *Cyberpsychology & behavior 8*(3), 272–282.

Stein, T., M. Peelen, J. Funk, and K. Seidl (2010). The fearful-face advantage is modulated by task demands: Evidence from the attentional blink. *Emotion 10*(1), 136.

Steuer, J., F. Biocca, and M. Levy (1995). Defining virtual reality: Dimensions determining telepresence. In *Communication in the age of virtual reality*, pp. 33–56. Lawrence Erlbaum.

Stirk, J. A. and N. Foreman (2005). Assessment of visual-spatial deficits in patients with parkinson's disease and closed head injuries using virtual environments. *CyberPsychology & Behavior 8*(5), 431–440.

Stone, R. (1992). Virtual reality and telepresence. *Robotica 10*(05), 461–467.

Sveistrup, H. (2004). Motor rehabilitation using virtual reality. *Journal of NeuroEngineering and Rehabilitation 1*(10), 10.

Taatgen, N., I. Juvina, M. Schipper, J. Borst, and S. Martens (2009). Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive Psychology 59*(1), 1–29.

Takatalo, J., G. Nyman, and L. Laaksonen (2008). Components of human experience in virtual environments. *Computers in Human Behavior 24*(1), 1–15.

Tang, Y.-Y., Y. Ma, J. Wang, Y. Fan, S. Feng, Q. Lu, Q. Yu, D. Sui, M. K. Rothbart, M. Fan, and M. I. Posner (2007). Short-term meditation training improves attention and self-regulation. In *Proceedings of the National Academy of Sciences*, Volume 104, pp. 17152–17156.

Tecchia, F., M. Carrozzino, S. Bacinelli, F. Rossi, D. Vercelli, G. Marino, P. Gasparello, and M. Bergamasco (2010). A flexible framework for wide-spectrum vr development. *Presence: Teleoperators & Virtual Environments 19*(4), 302–312.

Tendick, F., M. Downes, T. Goktekin, M. Cavusoglu, D. Feygin, X. Wu, R. Eyal, M. Hegarty, and L. Way (2000). A virtual environment testbed for training laparoscopic surgical skills. *Presence: Teleoperators & Virtual Environments 9*(3), 236–255.

Thomas, B., M. Strother, E. Donnelly, and J. Worrell (2009). Ct virtual endoscopy in the evaluation of large airway disease: review. *American journal of roentgenology 192*(3 Supplement), S20–S30.

Thomas, G., L. Johnson, S. Dow, and C. Stanford (2001). The design and testing of a force feedback dental simulator. *Computer methods and programs in biomedicine 64*(1), 53–64.

Thomas, R., N. William John, and J. Delieu (2010). Augmented reality for anatomical education. *Journal of Visual Communication in Medicine 33*(1), 6–15.

Torralbo, A. and D. Beck (2008). Perceptual-load-induced selection as a result of local competitive interactions in visual cortex. *Psychological science 19*(10), 1045–1050.

Tremblay, S., F. Vachon, and D. Jones (2005). Attentional and perceptual sources of the auditory attentional blink. *Attention, Perception, & Psychophysics 67*(2), 195–208.

Tripicchio, P., A. Filippeschi, E. Ruffaldi, F. Tecchia, C. Avizzano, and M. Bergamasco (2010). A measuring tool for accurate haptic modeling in industrial maintenance training. *Haptics: Generating and Perceiving Tangible Sensations*, 377–384.

Tsirlin, I., E. Dupierrix, S. Chokron, S. Coquillart, and T. Ohlmann (2009). Uses of virtual reality for diagnosis, rehabilitation and study of unilateral spatial neglect: review and analysis. *CyberPsychology & Behavior 12*(2), 175–181.

van Leeuwen, S., N. G. Ḿuller, and L. Melloni (2009). Age effects on attentional blink performance in meditation. *Consciousness and Cognition 18*(3), 593 – 599.

Vanni, F., C. Conversano, A. Del Debbio, P. Landi, M. Carlini, C. Fanciullacci, M. Bergamasco, A. Di Fiorino, and L. Dell'Osso (In Press). A survey on virtual environment applications to fear of public speaking. *European Review for Medical and Pharmacological Sciences*.

Vining, D. (1996). Virtual endoscopy: is it reality? *Radiology 200*(1), 30–31.

Visser, T., W. Bischof, and V. Di Lollo (1999). Attentional switching in spatial and nonspatial domains: Evidence from the attentional blink. *Psychological Bulletin 125*(4), 458.

Visser, T., S. Zuvic, W. Bischof, and V. Di Lollo (1999). The attentional blink with targets in different spatial locations. *Psychonomic Bulletin & Review 6*(3), 432–436.

Wallach, H., M. Safir, and R. Samana (2010). Personality variables and presence. *Virtual Reality 14*(1), 3–13.

Ward, R., J. Duncan, and K. Shapiro (1996). The slow time-course of visual attention. *Cognitive Psychology 30*(1), 79–109.

Weiss, P., K. Kizony, U. Feintuch, and N. Katz (2005). Virtual reality in neurorehabilitation. *Textbook of Neural Repair and Neurorehabilitation 2*, 182–197.

Wender, R., H. Hoffman, H. Hunner, E. Seibel, D. Patterson, and S. Sharar (2009). Interactivity influences the magnitude of virtual reality analgesia. *Journal of cyber therapy and rehabilitation 2*(1), 27.

Wiederhold, B., D. Jang, M. Kaneda, I. Cabral, Y. Lurie, T. May, I. Kim, M. Wiederhold, and S. Kim (2001). An investigation into physiological responses in virtual environments: an objective measurement of presence. In *Towards CyberPsychology: Mind, Cognitions and Society in the Internet Age*, Volume 2. Ios Pr Inc.

Witmer, B. and M. Singer (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators & Virtual Environments 7*(3), 225–240.

Wyble, B., H. Bowman, and M. Nieuwenstein (2009). The attentional blink provides episodic distinctiveness: Sparing at a cost. *Journal of Experimental Psychology: Human Perception and Performance 35*(3), 787.

Wyble, B., H. Bowman, and M. Potter (2009). Categorically defined targets trigger spatiotemporal visual attention. *Journal of Experimental Psychology: Human Perception and Performance 35*(2), 324.

Zahorik, P. and R. Jenison (1998). Presence as being-in-the-world. *Presence: Teleoperators & Virtual Environments 7*(1), 78–89.

Zhang, L., B. Abreu, B. Masel, R. Scheibel, C. Christiansen, N. Huddleston, and K. Ottenbacher (2001). Virtual reality in the assessment of selected cognitive function after brain injury. *American journal of physical medicine & rehabilitation 80*(8), 597–604.

Zhao, S. (2003). Toward a taxonomy of copresence. *Presence: Teleoperators & Virtual Environments 12*(5), 445–455.

Zimmerman, T., J. Lanier, C. Blanchard, S. Bryson, and Y. Harvill (1987). A hand gesture interface device. In *ACM SIGCHI Bulletin*, Volume 18, pp. 189–192. ACM.

Nomenclature

3-D	Three-Dimensional
AB	Attentional Blink
ADHD	Attention-Deficit/Hyperactivity Disorder
ANT	Attention Network Test
AR	Augmented Reality
CAD	Computer-Aided Design
CAVE	CAVE Automatic Virtual Environment
CBT	Cognitive Behavioural Therapy
CR	Cognitive Rehabilitation
CT	Computed Tomography
EEG	Electroencephalogram
ERP	Event-Related Potentials
eSTST	Episodic Simultaneous Type/Serial Token
FA	Focused Attention meditation
fMRI	Functional Magnetic Resonance Imaging
GLM	General Linear Model
HMD	Head-Mounted Display
hrs	hours
ISI	Inter-Stimulus Interval
LC	Locus coeruleus
MBSR	Mindfulness-Based Stress Reduction

MIS	Minimally Invasive Surgery
MR	Magnetic Resonance
ms	milliseconds
NA	Neuropsychological Assessment
OCD	Obsessive-Compulsive Disorder
OM	Open Monitoring meditation
PET	Positron Emission Tomography
PI	Place Illusion
PRCS	Personal Report of Confidence as a Speaker
PRP	Psychological Refractory Period
Psi	Plausibility Illusion
PTSD	Post-Traumatic Stress Disorder
RSVP	Rapid Serial Visual Presentation
SAD	Social Anxiety Disorder
SC	Sensorimotor Contingency
SOA	Stimulus Onset Asynchrony
STM	Short-Term Memory
T1	First target stimulus appearing in a stream of non-target stimuli
T2	Second target stimulus appearing in a stream of non-target stimuli
TBI	Traumatic Brain Injury
TLC	Temporary Loss of Control
TOA	Target Onset Asynchrony
VE	Virtual Environment
VR	Virtual Reality
VRET	Virtual Reality Exposure Therapy

VSTM	Visual Short-Term Memory
WCST	Wisconsin Card Sorting Test
WM	Working Memory
μ_G	Grand mean of a sample population