FACULTEIT PSYCHOLOGIE EN
PEDAGOGISCHE WETENSCHAPPEN

Collecting EEG data during Virtual Reality experience with a headset

Aranka van Dongen
0100105

Supervisor: Prof. Dr. Daniele Marinazzo
Mentor: Dr. Elena Patricia Nùñez Castellar

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of
Master in Clinical Psychology
Academic year: 2016 - 2017

UNIVERSITEIT
GENT

Acknowledgments

I would like to thank everyone who helped and stood by me in some way during this process. It has been a really challenging experience, where I sometimes lost hope and had to rely on perseverance and support from my family, friends and supervisors.

First, I thank professor Daniele Marinazzo for giving me the opportunity to start this project even though I had no experience. You were always very patient with me through all of my doubts and insecurities, and pushed me to be independent and being open to learn new things.

I also want to thank Dr. Elena Patricia Nùñez Castellar from standing by me through the first part of the thesis. Despite you having to leave the project, I learned a lot from you and you comforted me when it was difficult. It was stressful for me when you told me you needed to leave, yet it brought me in contact with some amazing people. Thank you Hannes for supporting me during the programming, and Klaas for explaining all the EEG ways to me and standing by me through all the steps. Alessandra, no testing day was ever too much for you. We spent a lot of time together doing the experiments and it was always a great time with you.

Natuurlijk wil ook mijn ouders bedanken. Bedankt, Mama en Maurice om me de kans te geven om te studeren en om altijd te blijven geloven dat ik zou afstuderen, ook al geloofde ik dat soms zelf niet. Ook bedankt om zoveel geduld te hebben tijdens alle stressmomenten en slechte buien. Ik heb erg veel gehad aan jullie steun.

Tijdens het schrijven van de thesis heb ik enorm veel steun ontvangen van mijn vrienden, waarvoor dank. Ik wil vooral Dagmar, Sofie en Roel bedanken voor alle tijd die ze hierin gestoken hebben.

Ten slotte wil ik graag mijn vriend, Alexander, bedanken. Niet alleen ben je er geweest wanneer het moeilijk ging en was geen moment te veel om me te helpen, doorheen mijn hele studie heb je meegeleefd. Dit was tijdens dit proces niet anders. Merci, Selsje.

Abstract

The original flow theory (Csikszentmihalyi, 1990) states that every activity can be flow-inducing, as long as there is a perfect balance between challenge and skill. Sherry (2004) applied the flow theory to the activity of gaming whereas Weber and colleagues (2009) attempted to objectify flow by defining the phenomenon in neurological terms. In this theory, flow is defined as “a discrete, energetically optimized, and gratifying experience resulting from the synchronization of attentional and reward networks under condition of balance between challenge and skill” (Weber et al., 2009, p.412). The current study combines all former approaches. Attention during flow is measured through both behavioural and neuronal paradigms, whilst subjects play an adapted version of a commercial game in three conditions (boredom, flow and frustration). Additionally, a comparison is made between attention during flow when playing on a PC and with a virtual reality HMD. An auditory oddball task revealed no significant differences in reaction times and error rates between conditions and devices. The neuronal correlates did not differ significantly between conditions as well, where the amplitude and latency of the P300 showed no significance difference between boredom, flow and frustration on a PC or in VR. The Flow Questionnaire (Sherry et al., 2006) showed a problem with the operationalization of the conditions. Plus, scores on the Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993) contained a large variability, where some subjects experienced all symptoms of simulator sickness and had problems finishing the game, and others reported no symptoms at all. Since the current study partially replicated Núñez Castellar et al. (2016), discrepancies between prior and current research are furtherly elaborated in the discussion such as the challenges and pitfalls accompanying the attempt of a novel research design.

Nederlandstalige samenvatting

De originele theorie over flow (Csikszentmihalyi, 1990) stelt dat elke activiteit flow-uitlokkend kan zijn, zolang er een perfecte balans plaatsvindt tussen de uitdaging en iemands competentie. Sherry (2004) past de theorie over flow toe op de activiteit van *gaming*, daar Weber en collega's (2004) trachten om flow te objectiveren door het fenomeen te definiëren in neurologische termen. In deze theorie is flow gedefinieerd als "een discrete, energetisch geoptimaliseerde, en bevredigende ervaring als resultaat van de synchronisatie van aandachts- en beloningsnetwerken met als voorwaarde een balans tussen uitdaging en iemands competentie" (Weber et al., 2009, p.412). De huidige studie combineert al deze invalshoeken. Aandacht gedurende flow wordt gemeten aan de hand van gedragsmatige en neuronale paradigma's, terwijl subjecten een aangepaste versie van een commercieel spel spelen in drie condities (saaiheid, flow, frustratie). Bovendien wordt er een vergelijking gemaakt tussen aandacht gedurende flow, wanneer men speelt op een PC en met een virtual reality bril. Een auditieve oddball taak toonde geen significante verschillen in reactietijd en foutpercentage tussen condities en apparaten. De neuronale correlaten verschilden eveneens niet significant tussen condities, daar de amplitude en latency van de P300 geen significante verschillen laten zien tussen saaiheid, flow en frustratie op een PC of in VR. De Flow Questionnaire (Sherry et al., 2006) onthulde een probleem in de operationalisatie van de condities. Overigens bevatten de scores op de Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993) een grote variabiliteit, waar sommige participanten alle symptomen van *simulator sickness* ervaarden en moeilijkheden kenden om het spel af te maken, en anderen geen enkel symptoom rapporteerden. Daar deze studie een gedeeltelijke replicatie is van het onderzoek van Núñez Castellar et al. (2016), worden discrepanties tussen eerder en huidig onderzoek verder uitgewerkt in de discussie, net als de uitdagingen en valkuilen die gepaard gaan met een vernieuwd onderzoeksdesign.

Content

Virtual reality (VR) has been around for more than sixty years, with its technology evolving rapidly. However, in the last decade, many companies discovered the possibilities of the head mounted display. For example, most recent version of the Oculus Rift (DK2) has been sold 119.000 times worldwide since its release (Statistica, 2015). Recent statistics of ESA (Entertainment Software Association) show 63% of the American households having at least one gamer. Out of those frequent gamers, 40% intends to purchase a head mounted display within the year. Of a total $23.5 billion spent on videogames in 2015, a budget of $2.1 billion was spent on accessories. These budgets grow every year, proving an increase in interest in the gaming industry and specifically in virtual reality (ESA, 2016). Despite the popularity of VR and its usage in various fields, little research has been done involving the experience itself.

An immersive experience implies the feeling of presence, a sense of really being in the virtual environment (Sanchez-Vives & Slater, 2005). With the peripheral vision covered, the focus of attention is narrowed to the stimuli on the screen (Csikszentmihalyi, 2014). A state of utter concentration is achieved. With attention fully invested in the task at hand, this scenario comprises the ultimate opportunity to feel flow, a state in which an individual operates at full capacity (Nakamura & Csikszentmihalyi, 2014). Next to the behavioural components of the experience, neural correlates are examined using EEG methods.

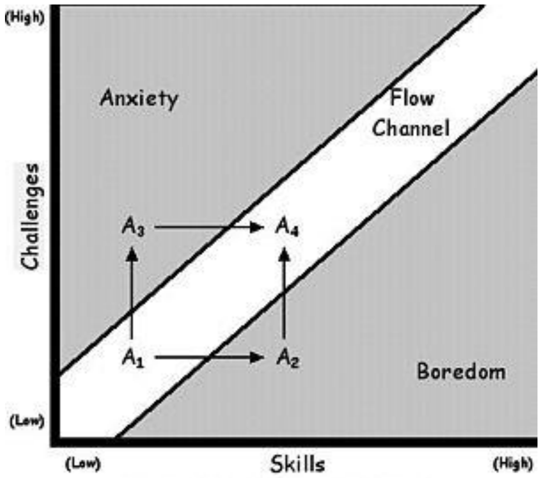
Flow and Synchronized Theory of Flow

Origin.

Concepts such as satisfaction, fun, and enjoyment are termed as the essence of play. In an activity characterized by play, an individual depends the least on external incentives, but is defined by an internal motivation to do so. The act of play itself is spontaneous and relatively free of needs (Csikszentmihalyi, 2014). During play, one acquires the feeling of freedom, creativity, and the sensation of being whole (Brown, 1959). Throughout literature, many purposes and applications have been set forward regarding play, forgetting the critical component in this story; play being enjoyable (Csikszentmihalyi, 2014). Being intrinsically motivated implies doing something because it is inherently interesting or enjoyable to the person (Ryan & Deci, 2000). In other words, the reward attached to this activity is a state of consciousness that is autotelic, meaning the activity is perceived simply as rewarding on itself. This particular activity is self-contained, without any expectations other than enriching oneself. Activities that once were exotelic, where attention is focused on external consequences, can evolve into autotelic experiences. These exotelic activities contain a certain need to achieve an external goal other than simply enjoying oneself (Csikszentmihalyi, 1990). Note that not only enjoyment determines whether an activity is autotelic, challenge is also a key factor. Enjoyment may not be inherent to challenge, the satisfactory feeling one experiences after overcoming difficulties is nevertheless essential (Steels, 2004).

Definition.

Flow is a state of effortless concentration and enjoyment (Csikszentmihalyi, 1997). It is subjective, and reported when one experiences complete involvement in the performed task. Various forms of play and enjoyment are part of the experience, although flow can also occur in situations that are not linked to play. The activity itself plays a small role in the possibility of experiencing flow (Csikszentmihalyi, 1997), whereas the perceived balance between one's skills and the challenge affects this experience in a greater way (Csikszentmihalyi, 2014). In this manner, actions that are not perceived as enjoyable by others can be so for individuals whose skills are in perfect balance with the executed task (Csikszentmihalyi, 1997). The principle of flow can be illustrated with a well-known diagram shown in figure 1, with "challenge" on the vertical axis, and "skills" on the horizontal one. When the challenge of an activity matches the skills that one masters, flow occurs, and the activity is perceived as enjoyable or satisfactory (A1). Once each next step of the task is controlled, the acquired skills exceed the required ones. If the same difficulty level of the task is pursued, one is bound to feel boredom (A2). Skills become more advanced and task difficulty stays the same, making the challenge too easy. On the other hand, if the individual wants to proceed to a more difficult part, the level of skill needs to increase proportionally with the amount of challenge (A4). If not, one can feel anxiety (A3). The level of skill stays the same or does not increase as much as task difficulty, making the challenge too hard (Csikszentmihalyi, 1990).



From Flow: The Psychology of Optimal Experience
by Mihaly Csikszentmihalyi (page 74)

Figure 1: Optimal experience of flow from Csikszentmihalyi

In his original theory, Csikszentmihalyi (1990) defines flow by referring to six key elements which also have been used as an explanation for play and enjoyment. Note that these principles only take place when the assumption of balancing challenge and skill is met. First key element is the merging of action and awareness. Normally, there is a sense of awareness of oneself and the environment. When experiencing flow, this duality becomes absent, and the activity almost feels automatic. There is only awareness for the performed actions (Csikszentmihalyi, 2014). Second, to achieve and maintain a sensation of flow, goals need to be clear and feedback immediate. These goals need to be feasible but not trivial, otherwise success does not lead to satisfaction. It does not matter how ambiguous or open-ended a goal is, setting them on itself ensures enjoyment. Setting a clear goal allows one to focus entirely on the activity and its objectives (Csikszentmihalyi, 1990). The third principle of flow comprises a transient focus. This temporary focus keeps us from wandering thoughts and enables a complete and utter concentration (Csikszentmihalyi, 1990). Irrelevant stimuli need to be kept out of attention in order to concentrate on one's actions (Nakamura & Csikszentmihalyi, 2014). A sensation of control is the fourth key element of flow. When asked, people who experience flow mention a feeling of not worrying of what may go wrong (Csikszentmihalyi , 1990). In reality this contains a contradiction, whereas it is the possibility of control one perceives instead of the actuality (Csikszentmihalyi, 2014). In a state of flow, perfection seems attainable (Csikszentmihalyi , 1990). Fifth, a loss of self-consciousness can occur during flow (Csikszentmihalyi, 2014). It is "the loss of the sense of a self separate from the world around it" (Csikszentmihalyi, 1990, p.63). An ego-less state arises, where one is in harmony with the environment. This loss of self-conscience does not mean one loses touch with their physical reality. Mostly, one even becomes more aware of the internal processes (Csikszentmihalyi, 2014). Finally, most proclaimed is a distortion of time. Actual progression of time doesn't seem to match perceived progression. An action taking only seconds can seem like minutes when utter concentration is demanded. The same goes for hours of enjoyment that feel like only minutes (Csikszentmihalyi, 1990).

Application.

Since the flow theory was proposed, literature has been flooded with numerous applications in all sorts of fields. Activities such as sports, playing games, sex, and music are commonly acknowledged as enjoyable and therefore have been linked to the concept of flow. However, more relevant for the present study is the application of the concept of flow during media enjoyment. In most cases, media is purposefully chosen in expectation to enjoy oneself (Sherry, 2004). Although all media usage can lead to a sensation of flow, playing videogames holds most of the requirements to actually achieve this state. In the work of Sherry (2004), components of gaming are linked to flow. Concrete goals are provided, either ultimate or through overmatching one self's former performance. To accomplish these goals, a difficulty level of preference can be chosen that equals one's competence. The amount of skill needed will increase with the complexity of the challenge. A sense of control is gained through being able to choose the amount of challenge opposed (Sherry, 2003). Like in flow, this control is perceived by the user, and does not have to correspond to reality (Csikszentmihalyi, 1990). By distributing clear feedback, information about progression in beating the challenge is provided. This information is plentiful and provided through multiple senses at once. Visual, tactile and aural impressions facilitate concentration and keep the gamer from being distracted (Sherry, 2003).

Sweetser and Wyeth (2005) propose "GameFlow", a detailed exposition of how the elements of flow manifest in computer games. They expand the original six components to eight and define each one in terms of gaming, with immersion and social interaction added. By comparing two commercial games, they manage to validate these adapted criteria of flow. In both real-time strategy games, these additional elements of flow could be derived. Extensively created universes and detailed chronicles demand utter concentration to complete tasks and are characterized by a feeling of immersion. Other than in the original theory of flow (Csikszentmihalyi, 1990), social interaction also plays a role in feeling part of the game and experiencing flow. In the in-game environment, different kinds of opponents and settings enhance the difficulty, and at the same time grant a sense of control to the player. Followed paths and strategies are also free to choose according to difficulty. While the user attempts to achieve goals presented by the introductions and followed narrative, skills are acquired with the aid of tutorials, cues and rewards. Constant feedback is granted by succeeding or failing missions, objectives and scores.

Jegers (2007) uses this model to map flow to more pervasive gaming. This kind of gaming expands the gaming experience into the real world, allowing more mobility. Information about the player's context is acquired and used to provide a gaming experience that changes according to the specific context of the user (Benford, Magerkurth, & Ljungstrand, 2005). In this study, three major characteristics are put forward regarding this specific genre of gaming, being: place-independent gameplay, the integration of physical and virtual worlds, and social interaction. Implementing these features requires some adaptations of the previously defined concepts. When the gamer can move freely throughout the physical environment, distractions and obstacles have to be taken into account. Therefore, the acquirement of undivided attention and utter concentration experienced in traditional gaming cannot be met in pervasive gaming. Hereby, shifting focus and integrating worlds is necessary. Attention needs to be divided between the physical and virtual environment, without losing immersion and the feeling of presence. Also the skill- and narrative driven challenges have to make room for a more socially based system, where a new dimension of social interaction creates a more flexible format. Personal goals become less important, although goals should stay clear and immediate.

Synchronized Theory of Flow.

Criteria regarding flow have been examined in the context of entertainment, but this supposedly unconscious state is never actually been objectively measured. The experience sampling method (ESM; Csikszentmihalyi & Larson, 2014) was put forward by the originator of flow theory, measuring flow in a natural environment. This research methodology utilizes notes of participant's experiences in real time. Despite its dynamic view, it was a rather impractical tool for examining flow because of the low likelihood of a flow experience during the actual registration (Weber, Tamborini, Wescott-Baker, & Kantor, 2009). Dietrich (2004) tried to adapt the concept of flow to ensure a better operationalization. In this work, a distinction is made between an explicit and an implicit system. The explicit system is rule governed, increases cognitive flexibility, and is accessible to consciousness. The implicit system on the contrary, is experience based, efficient, and is unconscious. A constant trade-off between these two systems results in the effortless information processing one experiences during flow.

However, functional magnetic resonance imaging (fMRI) studies showed that attentional networks are inherent to flow (Raz & Buhle, 2006). Hence, Dietrich's approach is too simple to grasp this concept. As a response, Weber and colleagues (2009) suggest a theoretical model to the operationalization of flow based on a neurophysiological perspective (Weber, Sherry, & Mathiak, 2008). This model contains a detailed description of the underlying brain mechanisms. They propose gamma oscillations as a mechanism for awareness of information, memory, and attention. When in flow, concentration and attention are high, awareness is limited to the actions itself and subjects are focused on the goal. This requires a lot of mental effort. In this theory, flow is defined as "a discrete, energetically optimized, and gratifying experience resulting from the synchronization of attentional and reward networks under condition of balance between challenge and skill" (Weber et al., 2009, p.412).

Taken together, flow is defined as being "completely involved in something to the point of forgetting time, fatigue, and everything else but the activity itself" (Csikszentmihalyi, Abuhamdeh, & Nakamura, 2014, p. 230). Flow occurs when one's skills are in proportion with the opposed challenge. Hence, whether flow is achieved or not is independent from the activity itself. However, key elements leading to flow can also be applied to the activity of gaming, containing all of the needed ingredients for experiencing flow. Three elaborations of these key elements are made, adapted to three different theories regarding gaming. Despite various applications of flow, a proper operationalization still lacked. Through a neurophysiological perspective on flow, the synchronized theory of flow (Weber et al., 2009) has been proposed.

Virtual reality

Conceptualization and current role.

Virtual reality can be defined as an artificially created environment in which the presence of the user is simulated. It is computer-generated and in 3D, which gives the feeling of a complete and real existence (Desai, Desai, Ajmera, & Mehta, 2014). This implies a user's viewpoint and control of the simulated surroundings (Brooks, 1999). In a virtual environment, users have the possibility to interact with their simulated surroundings by various movements. When the participant's senses are stimulated continuously by inputs of the virtual environment, an egocentric frame of reference occurs. The user experiences himself in the virtual environment as a totality (Ellis, 1991). The intuitive nature of the experience allows the user to perform tasks as in the "real world", making virtual reality an optimal interface between human and computer. Note that not every virtual environment is immersive and completely engages the user. In exocentric systems, individuals interact with the simulated surroundings using an external screen (Usoh & Slater, 1995). In the present study however, virtual reality is defined as a programmed world in which the user experiences the environment in an immersive and an egocentric manner, feeling as if being really there.

Facilitating the immersive experience, head mounted displays (HMD) are the preference for submerging in virtual environments. They move along with the user's head and completely cover vision. When wearing a HMD, the individual is able to explore the environment only with eye-and head movements. Besides from scouting, one can interact with objects and even "humans" in the virtual environment (Sanchez-Vives & slater, 2005). Virtual reality found its way to everyday entertainment, but also to educational, medical and psychological settings (Kaufmann, Schmalstieg, & Wagner, 2000; Haluck & Krummel, 2000). Mathematicians and surgeons are trained using virtual reality, people suffering from Parkinson's disease learn postural control (Yen et al., 2011), and psychologists try to overcome social phobia's in virtual environments (Gebara, Barros-Neto, Gertsenschtein, & Lotufo-Neto, 2015). The concept of virtual reality is not necessary recent, although a new generation of HMDs emerged (Avila & Baily, 2014). Technology is becoming more user friendly and the market is flooded with high performance devices. As prototypes are being transformed into commercial models, virtual reality gradually becomes accessible for the general public. The path to commercial success lies in the entertainment business. Going to the movies is much more immersive than it used to be, 3D technologies even are available for home systems. However, computer games seem to have it all. With large open worlds that are simulated and even without virtual reality, a player can interact with the environment and feel absorbed (Boas, 2013).

The head mounted display.

In regard to the rising success of virtual reality, many devices were developed to compensate for the large demand. The HTC Vive is a lightweight goggle device, also known as HMD, and was composed for experiencing virtual environments (Stengel et al., 2015). The intensity of this experience is enhanced by head tracking. Through the combined usage of a gyroscope, an accelerometer, and a magnetometer, the orientation of the user's head can be established. The precision of the measured head motion parallels the motion of the image (Byagowi, Singhal, Lambeta, Albada, & Moussavi, 2014). In this manner, the virtual environment feels more real, which causes increased sensation of immersion. Virtual reality itself is no longer solely intended for gaming purposes. Non-game contexts assimilate game design elements for various applications, such as psychotherapy and different kinds of training. This phenomenon is called gamification (Deterding, Dixon, Khaled, & Nacke, 2011). The HTC Vive, in this regard, is i.a. used for virtual reality exposure therapies (Miloff et al, 2016), pain control therapy (Hoffman et al., 2014) and stress therapy (Mahalil, Rusli, Mohd Yusof, Mohd Yusof, & Zainudin, 2014).

Presence and immersion

Conceptualization.

Presence can be defined as the sense of "being there", and is understood as a response to immersion. This can occur when an experience is overwhelming, and evokes an illusion of an actual place produced by simulated sensory data. In a state of presence, the body behaves as if it is part of the illusion (Sanchez-Vives & Slater, 2005). Experiencing presence seems to require a certain degree of focus, and occurs when attention is intentionally shifted between a physical locale and a virtual environment. A specific amount of attentional resources is needed to allocate attention to the extent of feeling presence (Witmer & Singer, 1998). Though presence often happens during enjoyment, one does not have to be interested or even involved in the activity to feel presence (Slater, 2003). An important factor for experiencing presence, however, is the meaningfulness of the information in the virtual environment. Selective attention facilitates a focus on these essential stimuli whilst excluding irrelevant stimuli (Witmer & Singer). Well-known circumstances inducing presence are reading, watching a movie, and listening to music. Recently, gaming became prominent in entertainment media and created the perfect opportunity for experiencing presence. The player actually sees through the eyes of the device used at the moment. The awareness of the machine becoming the body takes over, and one feels and reacts like actually being there. This specific sensation is called telepresence (Sanchez-Vives & Slater, 2005). Presence seems to be influenced by synchronization between the user's actions and what happens on screen, auditory sensations, haptic feedback, and virtual body representation (Sanchez-Vives & Slater, 2005). To measure the intensity of this subjective experience in a virtual environment, the presence questionnaire (PQ) was composed (Witmer & Singer, 1998). It uses a seven-point measurement scale on which participants need to rate various characteristics of presence. The virtual environment is displayed through a flight helmet and a joystick was used for control.

Slater (2003) defines immersion as an objectively measured unit and states it takes place when various modalities equal reality. In this definition, immersion is conceptually separated from presence. Presence, in fact, is a subjective reaction to immersion. When virtual environments are similar to reality, the player gets caught up in the activity (Mcmahan, 2003). The relationship between these two concepts is perfectly grasped by the work of Sanchez-Vives and Slater (2005), who state that "the degree of immersion is an objective property of a system that can be measured independently of the human experience that it engenders" (p.333). In this way, "presence is the man response to the system" (p.333). Immersion can occur in any medium offering a certain narrative. The user feels completely besieged by the story and its elements, making their surroundings disappear completely (Murray, 1997).

In extent of the present study, the focus lies on video gaming, being the perfect medium for experiencing immersion and presence. Most games carry a certain storyline that needs to be played out, although games seemingly without a narrative are popular too. Although this chronical, diegetic way of playing is sufficient to evoke immersion, it's not essential. Deep play, a nearly obsessive way of engaging in an activity, can also be induced at a nondiegetic level (Mcmahan, 2003). This level contains gaining points, working out strategy, and boast progress to opponents, and can lead to immersion and presence as well. Parameters of immersion include various technical elements, such as field of view, quality of the displayed image, and the latency of the head tracking. The simulated sensory data in particular has to complement proprioception. The latter is most important in systems using a head mounted display, where every movement of the head needs to match the motion on screen (Sanchez-Vives & Slater, 2005)

Effect of VR on presence and immersion.

Creating virtual environments (VE) and being able to experience them through head mounted displays offers a new dimension to presence and immersion. Realistic sensory inputs are being produced by the used devices, and the sensation arises of being an extended object in one's simulated surroundings. This phenomenon is called virtualization and extends beyond the actual space the user is located in (Ellis, 1991). Factors regarding virtual reality that intensify the feeling of presence are head tracking with a lowered latency, depth perception, virtual body representation, and body engagement. Also sounds and haptic feedback amplify presence, although both are difficult to attain (Sanchez-Vives & Slater, 2005). Followed study also showed that presence during an fMRI was higher when the view of the user wasn't obstructed with a small white square. This unobstructed view was conceptualized as a high-tech virtual environment and lead to an even greater sense of presence when experienced with virtual reality through a HMD (Hoffman, Richards, Coda, Richards, & Sharar, 2003). Unexpectedly, virtual realism does not lead to more presence. Even stylized virtual environments that do not resemble reality can induce presence. Interaction however, does cause an increase of presence (Sanchez-Vives & Slater, 2005). "The sense of 'being there' in a virtual environment is grounded on the ability to 'do' there" (Sanchez-Vives & Slater, 2005, p.333). In a VE, both top-down and bottom-up processing occurs. When the VE contains salient stimuli and offers a recognizable and coherent experience, the feeling of presence increases (Nichols, Haldane, & Wilson, 2000).

The effectiveness of immersion is influenced by the excitement and comfort of the environment during the experience, sound, and quality of the images. A sense of control and desire to repeat the experience influenced the degree of which one feels immersion (Bangay & Preston, 1998). Parameters that enhance immersion in traditional contexts apply to virtual reality as well. The field of view is even more extended with a head mounted display, offering a 360° view. Seeing the VE from an egocentric point of view and experiencing a synchronization between the real and the virtual body influence immersion as well. This allows the user a certain degree of control (Sanchez-Vives & Slater, 2005).

Attention and its measures

Posner: three networks of attention.

The cognitive process where one selectively focusses on one fraction of the environment, ruling out distraction is one of the most defined concepts in psychology. Attention has been examined for hundreds of years, yet in this study only the framework of Posner is used. In a revision (Petersen & Posner, 2012) of the original work (Posner & Petersen, 1989), based on recent neuroimaging techniques, attentional networks are set out. These networks are believed to be anatomically separate from other processing systems and from each other, each with its own cognitive functions.

the alerting network.

When reacting to sudden and external cues, vigilance and task performance increases. This rapid change is considered phasic, whereupon a resting state is replaced with one of arousal. During this phase of increased arousal, reaction time decreases in response to the warning signal. External cues provide information about when or where a specific target will occur, in order to prepare the body for an adapted response. Norepinephrine (NE) is extracted during this short period of time, along with more activity in the locus coeruleus releasing this neurotransmitter (Petersen & Posner, 2012).

the executive network.

In contrast to previous mentioned network, the executive network ensures a top-down regulation. Arguments to support this statement are found in the revised work of Posner (Petersen & Posner, 2012). Activity in related brain areas is higher when looking at targets than when looking at non-targets. This Increased activity also occurred for conflict and error trials. To monitor attention, awareness is necessary. Monitoring awareness applies to targets, attentional conflicts, and errors. The Anterior Cingulate Cortex (ACC) shows higher activation while controlling these processes (Petersen & Posner, 2012).

Posner and Petersen (2012) propose two separate top-down control networks. To demonstrate, they differentiate between three sorts of signals related to control. The first kind of signals occur at the start of a task and are temporal. The second kind is distributed during the task to maintain control. The latter signals are observed when feedback is provided at the end of the task. The cingo-opercular system manages preservation of control during the task, and the fronto-parietal system is operative at the beginning of the task, where starting cues are observed.

the orienting network.

The orienting network is responsible for internally shifting attention to a certain external stimulus (Petersen & Posner, 2012). This shift is a method of selecting information, mostly using head and eye movements, although not every response is overt. Saccadic movement and internally shifting attention without moving the head are covert responses. While orienting towards a certain stimulus, a network of brain regions is active (Posner & Dehaene, 1994). Research states that the P300 component represents such orienting responses to novel stimuli (Barceló, Periáñez & Knight, 2002; Donchin et. al, 1984; Donchin, 1981). Such as an orienting response, the P300 is elicited by novel events that are improbable and relevant to the task at hand (Donchin, 1984). More recent attempts to support this theory were done by Barceló and colleagues (2002), in which the P300 is linked to executive control of cognitive set shifting. Current study makes further use of the relevance of the P300 in the measurements of attention.

EEG methods.

Electroencephalography (EEG) is an electrophysiological technique that measures event related potentials and records electrical activity of the brain with scalp electrodes. It is a method with excellent temporal resolution, but contains poor spatial information due to the difficulty of separating the current sources in the brain by the tissues of the head (Handy, 2005). Ionic currents, which cause field potentials, are measured along the cell-membranes. As there are excitatory and inhibitory synapses, activity can be perceived closer or further from the surface respectively. These potentials, which are extracellular, form an EEG when the time constant is one second or less. This implies a highly frequent variation in the amplitude of the signal (Niedermeyer & Da Silva, 2005). As these potentials are measured by electrodes, coherence between these points can be assessed. In this manner, one can examine the cortical areas during cognitive processes (Handy, 2005). The frequency component of an ongoing EEG is determined by the changes in activity of interactions between different neurons. These neural networks oscillate, which can be synchronized or desynchronized (Pfurtscheller & Da Silva, 1999). These oscillations can be interpreted as a certain manner of correspondence between cortical cells. When these oscillations obtain a certain frequency, they are called alpha-waves, and can be specifically measured with EEG methods (Palva & Palva, 2007). EEG is broadly used in cognitive psychology. Processes such as attention allocation and memory are assessed by extracting ERP's from the signals (Polich & Kok, 1995).

Event Related Potentials.

Event related potentials (ERPs) are electrophysiological reactions to external or internal stimuli. These reactions contain small changes of the electrical activity in the brain and are extracted from the EEG (Luck, 2005). When measuring ERPs, current brain activity is represented with no delay. This timing is an advantage when examining the difference between two conditions or the influence on sensory activity when investigating mental processes (Luck, 2005). When a neural process takes place in a particular brain region, a certain voltage is deviated from the ERP waveform. This deflection is defined as an ERP component (Luck, 2005). Early in an ERP, attentional components are reflected posteriorly (C1, N1,P1) whereas afterwards higher order cognitions (N2, P300) can be distinguished (Luck & Kappenman, 2012). A relevant component regarding the current study is the P300. This response is elicited when a certain stimulus is unpredictable or infrequent and peaks around 300ms poststimulus. The P300 is evoked when a stimulus is novel and improbable (Donchin, 1984). However, the amplitude of the P300 is lower when the task difficulty is increased, whereas the latency is elongated in demanding circumstances (Polich, 2007).

Current study

Recently, efforts (Núñez Castellar, Antons, Marinazzo &, van Looy, 2016) have been made to measure flow (Csikszentmihalyi, 1990) objectively. Through an auditory oddball paradigm, attention during a state of boredom, flow and frustration was measured whilst subjects played a game on a desktop. Reaction times and error rates functioned as measures for attention. Additionally, an EEG was conducted during all conditions and neuronal correlates were examined through complex methods.

With the commercial success of virtual reality, prior study was replicated with a head mounted display and different stimulus material. Apart from furtherly investigating said method for measuring flow, this study is prone to propose new insights in virtual reality and gaming studies. First, a comparison is made between gaming on a desktop and gaming in virtual reality. The difference in immersion and presence between both devices open doors for new hypotheses. Second, a distinction is made between flow, and states of boredom and frustration so as in the study of Núñez Castellar et al. (2016). However, an additional component of examining flow is provided by the virtual reality condition. Not only can flow be compared to other states of being, flow while gaming on a desktop can be directly compared to gaming in virtual reality. Hereby, implications regarding immersion can be made. Furtherly, this research combines the use of a virtual reality headset and EEG measurements, being relatively new in research. This opposed a challenge, due to movement restrictions of the EEG and combining delicate scalp measurements with large and rather heavy goggles. To deal with this challenge, original stimulus material was constructed in which the HMD functions as a monitor controlled with a mouse and keyboard. In this case, minimal head movement is necessary to play the game, still providing the possibility to emerge oneself in the virtual environment. Since an important goal of this study is to present stimulus material that is enjoyable and resembles commercial gaming, a game needed to be created fulfilling both criteria. Result is a shooter resembling popular games, yet largely controllable to fit experimental research. In the neuronal part of the study, attention is measured by examining the P300 in every condition and each with device. Based on this design, following hypotheses are made:

(1) Participants commit more errors in the oddball task during the flow condition
(2) Reaction times during the oddball task will be higher in the flow condition
(3) Participants commit more errors in the oddball task during flow condition in virtual reality in comparison to a non-virtual reality condition
(4) Reaction times during the oddball task will be higher in the flow condition in virtual reality in comparison to a non-virtual reality condition
(5) The amplitude of the P300 will be smaller in the flow condition
(6) The latency of the P300 will be longer in the flow condition
(7) The amplitude of the P300 will be smaller during flow in virtual reality in comparison to a non-virtual reality condition
(8) The latency of the P300 will be longer during flow in virtual reality in comparison to a non-virtual reality condition

Method

Sample

The total sample included 18 participants, recruited through online sampling. This community sample (*N*=18) consisted of 15 men (83,33%; *M* = 23 years old; *SD* = 2,1) and 3 women (16,67%; *M* = 25,67 years old; *SD* = 2,52). Throughout the entire sample, the mean age was 23,44 years old, ranging from 19 to 28 years old. Most of the participants (83,33%) had a Belgian nationality and were born in Belgium. Since this sample was mainly drawn from a student population, the majority (94,44%) of the participants was highly educated or still attending university. Few of them (16,67%) acquired or will acquire their PhD. Merely 5,56% completed solely a lower education. Due to circumstances furtherly explained in the paragraph ‘procedure’, the majority of the subjects had gaming experience. 61,11% classified themselves as casual gamers, 33,33% identified as experts and 5,55% had never gamed before. 44,44% had already played the commercial version of presented game. All included subjects participated on a voluntary basis and signed an informed consent.

Instruments

Equipment.

Regarding hardware, current design required an extensively equipped lab. For the primary task, an Alienware gaming PC , a 46 inch Phillips television screen and a HTC Vive HMD was used. The system model of the gaming PC was Alienware Area-51 R2, with a i7-5820K processor and a RAM memory of 16384 MB. The graphical card was a NVIDIA Geforce GTX 1070 with a total memory of 16222 MB. These specifications fit the requirements for usage of the HTC Vive. This HMD offers a resolution of 2160 x 1200 (with 1080 x 1200 per eye), global lighting and AMOLED-displays of 90Hz. This task was performed with a keyboard and mouse, and was auditory supported by a DELL A215 MultiMedia speaker of 3 Watt in function of the in-game sounds such as gunshots. A DELL desktop and Trust sound system with a total RMS output of 15 Watt and peak power of 30 Watt provided the secondary task, whereas a Cedrusbox (RB-830) was used to respond. Due to spatial issues, a USB extension cord connected the response box with associated desktop computer. Latter mentioned computer was connected to an ASUS laptop comprising the EEG software by an optical receiver. For the EEG itself, the BioSemi ActiveTwo measurement system with a 64 channel layout was used. Both pin-type and flat-type electrodes were used, supplemented by skin conductance electrodes employed on the forearm. The set-up was completed by an Optional Passive Channel Headbox to employ the flat-type electrodes for ocular measurement and monitor the heartrate.

The map of the primary task was constructed with CS:GO-SDK Hammer World Editor. The goal was to achieve constructing a game as close to commercial gaming as possible. After all, this software allows to standardize, control and manipulate every element of the game. The game itself was programmed in Notepad++ in the object-oriented programming language, Squirrel. This is an open source software and the only language compatible with Hammer World Editor. In order to run the game, CS:GO was opened through Steam, an online gaming platform developed by Valve. To convert CS:GO to Virtual Reality, VorpX was purchased, a 3D-driver for virtual reality headsets with full head tracking support. Notepad++ was also used to adapt the code of the oddball task. The original code was programmed in Tscope and had to be manually converted to Tscope5. Finally, ActiView software was installed to record the EEG signals.

Questionnaires.

The **Flow Questionnaire** (FQ) by Sherry and colleagues (Sherry et al., 2006) was utilised to measure engagement during the game. This instrument consists of 12 items stated on a seven-point Likert scale (going from 1 = '*Strongly Disagree*' to 7 = '*Strongly Agree*') and quantifies seven dimensions. An example item of this questionnaire is: *"I was caught up in the game".* Dimensions "Skill", "Difficulty", "Anxiety" and "Boredom" directly measure flow based on the basic premises of the original flow theory (Csikszentmihalyi, 1990). "Temporal distortion", "Concentration" and "Loss of self-consciousness" on the other hand, are important key elements of flow which are stated earlier (Csikszentmihalyi, 1990). The reliability of this instrument in a community sample was calculated in the study of Sherry (Sherry et al., 2006) by using Cronbach's Alpha. For the dimensions "Difficulty" (Cronbach's α = .83) and "Skill" (Cronbach's α = .88), the internal constancy is marked as 'good' as well as for the main facet "engagement" (Cronbach's α = .85). The internal consistency of the dimension "Boredom" (Cronbach's α = .92) was marked as excellent. Although most participant were Dutch speaking, the questionnaire was taken in English. Since no translation was available, translating to Dutch would have disadvantaged the few participants with a different ethnicity. The researches were available for questions at any given moment.

Every subject receives the **Simulator Sickness Questionnaire** (SSQ; Kennedy, Lane, Berbaum & Lilienthal, 1993) directly after finishing the three conditions in virtual reality. This symptom checklist includes 16 symptoms of simulator sickness scored on four levels of severity (*'None', 'Slight', 'Moderate' and 'Severe'*), coded by respectively 0, 1, 2, and 3. This self-report instrument contains three subscales: The Nausea subscale, with example item: "*Salivation Increasing"*, an Oculomotor subscale, with example item: "*Eye strain"* and " Disorientation, with example item: "*Dizziness with eyes open".* No Cronbach's Alpha values of internal consistency were reported in the original article. However, Moss and Muth (Moss & Muth, 2011) state the reliability and validity of the instrument by reporting results of a series of factor analyses. Reliability is demonstrated by calculating the split-half correlation (r = 0.78), because test-retest reliability could suffer under adaptation effects. No Dutch translation is found and the questionnaire is presented in English for reasons stated above.

Stimulus material.

Primary task.

As a primary task, subjects were instructed to play a custom made game. This game is a first person shooter derived from the commercial success Counter Strike Global Offensive (CS:GO). For the stimulus material, a shooter seemed fitting as it is straightforward to play, with clear goals and immediate feedback, being elements that induce flow (Csikszentmihalyi, 1990; Sherry, 2004). The player starts in a practice room with three targets on the wall. An automatic gun is the weapon of choice to provide a player of a scope to facilitate aiming and granting the feeling of control. It takes an effort to control the munition fired, which is needed due to a limited amount of bullets. This perceived control is also an important element of flow (Csikszentmihalyi , 1990). The subject can start the game by triggering the slide door with the inscription "*Start".* In-game targets are cardboard cut-outs of enemies, since artificial intelligence is problematical to control. However, research shows that realism in games doesn't necessarily leads to more immersion (Sanchez-Vives & Slater, 2005). Each target needs to be shot twice, to counter random spraying and accidental striking. To keep inducing flow throughout the game, players can only proceed to the next room when every target in a room is hit. This to avoid a speed-accuracy trade-off since the instruction is given to play as accurately and fast as possible. Taking this trade-off into account, players are rewarded for accuracy by skipping levels, and are able to achieve a much higher level when playing fast.

Immediate feedback is provided in three different manners: The target goes down when hit, players are encouraged by the script *"Good job"* on every door that opens and the subject can keep track of his progress and munition on scoreboards above every door in the game.

Three conditions, being boredom, flow and frustration were constructed by manipulating two different features of the game. Players have to repeat the game for 8 minutes straight in every condition. Since progress is solely feasible in flow, players need to repeat one level continuously in boredom and frustration, probably inducing weariness. First manipulation in which the conditions were constructed is the moving speed of the targets, adapted during gameplay. In the boredom condition, targets are stationary, creating a rather easy environment. When starting in the flow condition, targets move at a low speed from left to right an back, placed on a rail. During flow, players can progress when playing accurately through 12 levels. An alternating scheme was conducted, to create a salient and linear acceleration over levels, still keeping the change moderate (Table 1). In the frustration condition, targets move at the highest speed presented in the game.

Targets	*Speed units*	*Bullets per target*
[1,1,1,1]	[50,50,50,50]	10
[1,2,1,2]	[50,100,50,100]	9
[2,2,2,2]	[100,100,100,100]	9
[2,3,2,3]	[100,150,100,150]	8
[3,3,3,3]	[150,150,150,150]	8
[3,4,3,4]	[150,200,150,200]	7
[4,4,4,4]	[200,200,200,200]	7
[4,5,4,5]	[200,250,200,250,200]	6
[5,5,5,5]	[250,250,250,250]	5
[5,6,5,6]	[250,300,250,300]	3
[6,6,6,6]	[300,300,300,300]	3
[7,7,7,7]	[350,350,350,350]	2

Table 1: Acceleration scheme and bullets per targets in flow condition

Second manipulation concerned the ammunition granted in each condition. When in boredom, players start with 12 bullets per target, which appeared to be enough and to spare. In flow, foregoing table (Table 1) depicts used scheme for distributing ammunition throughout the levels in this condition. When the player is out of ammunition, one has to repeat the currently played level. Subjects are granted only two bullets per target in frustration, giving them no ammunition to spare. Prior decisions were made based on the premise that both novices and experts were able to experience the intent of the conditions. Through multiple testing on different types of players, boredom seemed tedious for every player, whereas frustration could not be “won” by experts and even provoked expressions of anger. Flow allows subjects to play and progress based on their skill level. When participants were 65% or 85% accurate they could skip respectively one and two levels. Subjects were informed on the possibility of skipping levels in advance.

The idea for the map of the game was based on different custom made maps for said commercial game. It consists of eight rooms in total (Figure 2), in which five are part of the actual game and contain targets. Participants can practice and get used to virtual reality in the first space while given instructions. After players finish one run, they enter a final room in which the scoreboard is presented and a portal that brings them back to the first room where they can start again until the eight minutes are over. When time runs out, players are being teleported to a small room disconnected from the map with the inscription “*Out of Time”.* After a few seconds, players can enter the starting room to progress to the next condition.

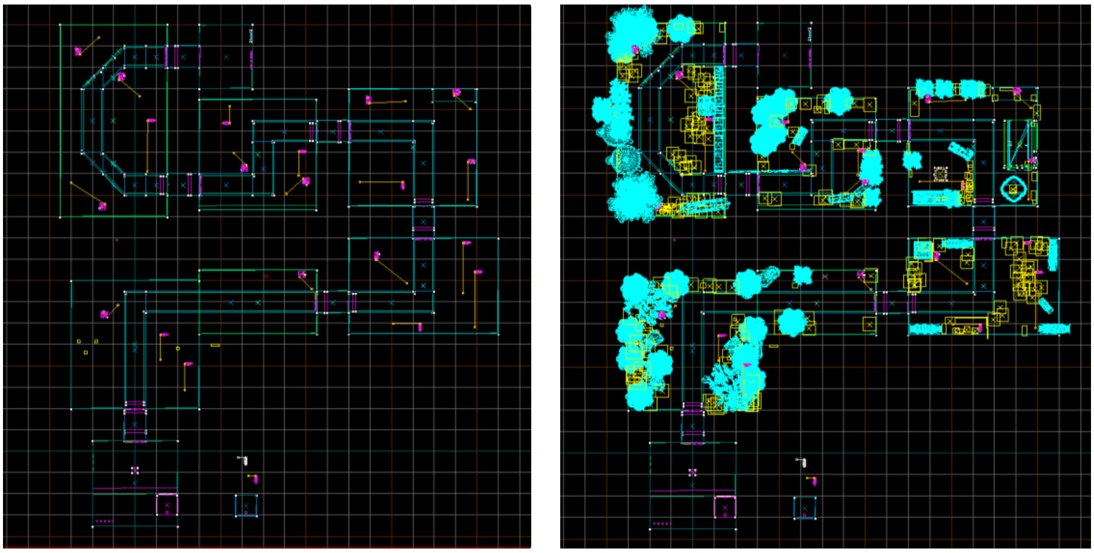


Figure 2: Basic floorplan of the game with target placement (left) and floorplan with decoration (right)

All rooms are connected in series in function of the simplicity. Airlocks are created between rooms to keep players from clearing a room when still in previous spaces. Participants are solely allowed to go forward and are blocked by the sliding doors if they return to previous rooms. Although players can move relatively freely in one chamber, walking lanes with raised edges to avoid random movement. These edges furtherly keep players from lining up with targets in order to hit them more easily.

To apply this game to current research design a few measures had been taken. First, conditions were programmed separately so the order can be adjusted, whereas conditions are counterbalanced between participants. These adaptations need to be made in the associated code before the game starts. Second, all of the in-game controls were unbended since players were only allowed to walk, shoot and open a scope. The conventional keys (W,Q,S,D) were changed to the arrow keys, to facilitate employment of the keys in combination with the response pad. Third, music volume was turned down to avoid interference with the auditory task. However, in-game sounds like footsteps and gunshots were enabled to increase immersion and were kept stable over sessions. Finally, to run the game in virtual reality, VorpX settings needed to be adapted. Default settings allowed the 3D diagonals (X,Y,Z) to move along with the player's movements, inducing motion sickness. Chosen setting was 'Z-normal', where the coordinate for depth was fixed.

Secondary task.

The experimental task consisted of a novelty oddball paradigm as conducted in the study of Núñez Castellar et al. (2016) and originally in Debener and collegues (Debener, Makeig, Delorme, & Engel, 2005). The auditory stimuli were two sinusoids (350 *Hz* and 650 *Hz*) with a duration of 339*ms* and 96 unique novelty sounds (Fabiani, Kazmerski, Cycowicz, & Friedman, 1996) with a mean duration of 338*ms* (ranging from 161*ms* to 402*ms*). Each novelty sound was presented at least once per subject and had no restricted order. In total 960 sounds were presented, being 960 trials. These trials were divided among three conditions (Boredom, Flow and Frustration) of 320 trials and 8 minutes each, which comprises 10 blocks of 32 trials. Every block holds 80% standard tones, 10% novelty sounds, and 10% oddball sounds. To avoid confounders, the low (350 *Hz*) and high (650 *Hz*) sinusoids were counterbalanced across participants alternating as frequent or rare (oddball) sounds. Waiting times between sounds differs randomly among 960, 1060, 1160, 1260 and 1360*ms*. Participants were instructed to react as fast and accurate as possible through a response box below their keyboard. The instructor emphasized the importance of performing well on both tasks.

Procedure

Design.

In this study, a blind randomized within subject design was chosen for a variety of reasons. First, this design holds a comparison between two gaming devices. Gaming on a desktop is treated as a control condition whilst gaming with a virtual reality headset operates as the experimental manipulation. In case of a within subject design, variability within subjects can be observed more precisely (Conaway, 1999). Second, not only comparison between gaming on two devices can be made, fluctuations of one participant's skill are controlled for (Bakeman, 2005). Since the stimulus material consists of three conditions on two different devices, each condition is played twice. Thus, in total participants perform six conditions of eight minutes each. To counter unwanted effects of boredom and repetitiveness, two versions (Figure 3) of the same map were created. Following elements were kept as equal as possible: Target placement, direction in which targets move, the amount of obstructions to see and hit targets, and finally the type of decoration in certain places in the game.

Figure 3: Gameplay depicted in two versions of the map

Sampling.

Due to time restrictions, participants were recruited through online convenience sampling, mostly in the social networks of the researchers. Information about the current study and an invitation to participate was posted in a closed Facebook group of 656 students staying in a particular Ghent University living unit. A total of 18 participants was collected through this method. As many men as women participated, and both people with and without gaming experience signed up. However, small pilot studies indicated that people without gaming experience became extremely nauseated, forcing the researchers to diverge from the original sample method. In this way an online snowball sample happened, otherwise known as a referral sample, where a chain reaction is made through different people knowing one another and the sample is extended in every step (Berg, 1988). For example: someone in the private circle of the researched was approached to participate, whereas a contact of the first person also participated. Because of the small sample size, the number of waves in this snowball sample is limited.

Experiment.

Preparatory, different forms were drafted for every participant, such as different questionnaires, instruction forms and an informed consent in both English and Dutch. A custom document was used for observation by the researcher, and contained the number and coded identification of the participant, the order of devices, conditions and versions, highest ranking and other remarks. The order of conditions was counterbalanced across participants, depending on subject number, whereas device and version were additionally counterbalanced to assure a different order for each person. In this manner, 18 of the 48 possible orders were randomly assigned. Other important preparations concerned the experimental set-up. First, the participant number was adjusted in the code of the oddball task, where the order of conditions and the nature of the sinusoid was displayed. For this particular step, it was important to correctly connect the response box and check the transferring of the triggers to the laptop on which the EEG was recorded. To equip the primary task, a few important steps are necessary. First, the order of the conditions prescribed by the oddball task had to be altered in the script. Second, the proper game had to be selected in Steam and the console was opened once access to the game-menu was granted. In the console screen, the command was given to start the right version of the game. Through the command: "*cl_draw_only_deathnotices 1;*", the user interface was cleared apart from the crosshair. When in the virtual reality condition, VorpX was configured and started, just as Steam VR. A room set up was administered to calibrate the virtual reality goggles in the used space.

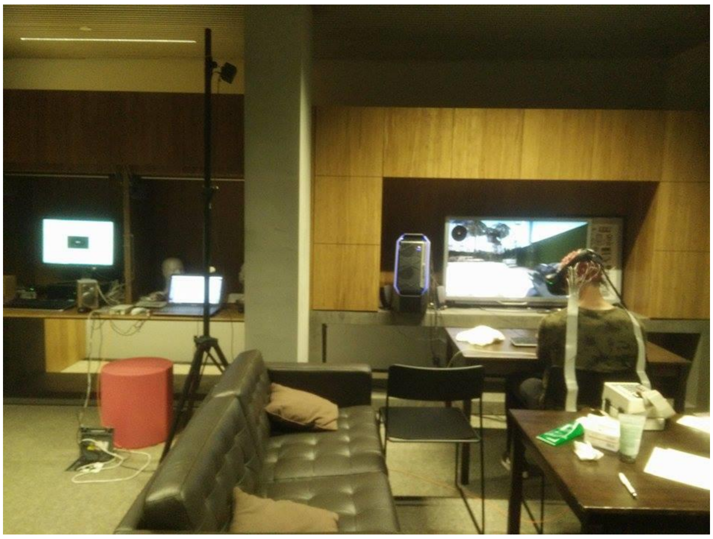


Figure 4:Photo of the experimental set-up with the participants playing the game in the virtual reality condition. The computer on the left provides the oddball task, whereas the laptop in the middle is used to record the EEG and the gaming computer with flat screen on the right provides the game. The subject is seated behind a table to reach the keyboard, response pas and mouse. The table behind the subject carries the EEG recording device and all necessary supplies.

Subjects were tested individually and seated down at approximately 1 meter of the screen behind a table. During the assemblage of the EEG cap and external electrodes, participants were asked to fill out the informed consent. In the first step of the actual experiment participants needed to focus on a fixation cross while alternating between no blinking and being relaxed for one minute, for six minutes in total. The cross remained on the screen during the whole run and every minute new instructions were given ("*Try not to blink"/"Be relaxed*"), resulting in three minutes of not blinking and three minutes of relaxing. After recording the baseline, most ocular electrodes were removed to assure comfort wearing the HMD. After, standardized instructions were read (Dutch or English) to the subjects while giving them a chance to practice navigation and shooting. The importance of performing well on both tasks was emphasized.

In between all six conditions, subjects filled out the FQ with the constant possibility of asking questions about presented items. After performing three conditions in virtual reality, the SSQ was administered. Due to various reactions on virtual reality such as nausea, headaches and eyestrain, time in-between conditions differed between subjects and within subjects. During the experiment, observations such as talking, emotional expressions, and performance in-game were noted.

Data-analysis

Behavioural data.

For the analysis of this study, SPSS (SPSS Statistics for Windows version 24.0) and R (R 3.4.1 for Windows) were utilized. To assess significance, a threshold of $p < .05$ will be used. However, because of the small sample size ($N = 18$), results with a p-value lower than $p < .1$ will be discussed and reported as being 'marginally significant'.

First, all variables used in the analysis were tested for normality to meet the assumptions of *t*-tests, repeated measures ANOVA, and principal component analysis. According to the Kolmogorov-Smirinov test of normality, mean reaction times in all conditions were normally distributed. However, apart from the error rates in PC boredom, the data for the error rates showed significance on the Kolmogorov-Smirnov test, indicating the data is not normal. Since this procedure is rather conservative (Crutcher, 1975), cut-off values of [-1.96;1.96] were used to examine skewness and kurtosis (Ghasemi & Zahediasl, 2012), which were violated in every condition apart from VR flow. Mean reaction times in VR boredom (*Skewness* = 2.85, *SE* = .54; *Kurtosis* = 9.53, *SE* = 1.04) , VR flow (*Skewness* =1.34, SE = .54; *Kurtosis* = 1.49, *SE* = 1.04), VR frustration (*Skewness* = 4.12, *SE* = .54; *Kurtosis* = 17.27, *SE* = 1.04), PC flow (*Skewness* = 6.56, *SE* = .54; *Kurtosis* = 6.56, *SE* = 1.04) and PC frustration (*Skewness* = 4.13, *SE* = .54; *Kurtosis* = 17.32, *SE* = 1.04) were positively skewed. Since the data comprised zero-values, a logarithmic transformation with an additive constant was administered, such as a square root and exponential transformation. Yet no methods could transform the error rates to fit the assumption of normality. Therefore, non-parametric tests will be conducted in analysing the error rates.

In order to carry out a principal component analysis on the FQ, responses were added over all conditions per item, resulting in one large dataset comprising information from every condition. When examining significance on the Kolmogorov-Smirnov test, every item seems normally distributed except for item seven and nine. However, values of skewness for both item seven (*Skewness* = .46, *SE* = .54; *Kurtosis* = -1.45, *SE* = 1.04) and item nine (*Skewness* = 1.39, *SE* = .54; *Kurtosis* = 1.39, *SE* = 1.04) stay within prior cut offs, and thus will be considered normal. After conducting a principal component analysis, the same tests were carried out once more to check if newly found components were normally distributed. Violations were found in the subscale difficult in every condition and the flow subscale in VR boredom and PC flow, yet a logarithmic transformation was administered on all subscales. After transformation, significance on the Kolmogorov-Smirnov test was found for some subscales (Flow in VR boredom, difficult in VR flow, Difficult in PC boredom and Flow in PC flow), yet values for skewness and kurtosis also stayed within cut off scores.

Finally, scores on the subscales of the SSQ were not normally distributed as well. After a square root transformation, the subscale Nausea (*Skewness* = .99, *SD* = .54; *Kurtosis* = .90, *SD* = 1.04) still showed significance, yet with values of skewness and kurtosis lied within the cut off interval.

All means, standard deviations and ranges were calculated for both mean reaction time and error rate. A *t*-test for independent samples was conducted for reaction times grouped by sex. In this manner, differences between male and female subjects can be discovered. Means, standard deviations, values for the *t*-test and Cohen's *d* were computed. Grouping was done with gaming experience as well, however, none of the outcomes were significant. A repeated measures ANOVA was conducted solely for reaction time of which partial Eta squared, value for the F-test and the p-value were calculated. Since within-subjects test showed no significance, no post-hoc tests were conducted. Because the error rates were not normally distributed, even after multiple transformations, non-parametric tests were conducted. A Willcoxon signed-rank test is considered a non-parametric counterpart for the *t*-test whereas the repeated measures ANOVA can be replaced by a Friedman test (Zimmerman & Zumbo, 1993). Note that for every repeated measures ANOVA in this study, Mauchly's test of sphericity was consulted. In case of any violations, the Greenhouse-Geisser correction was used (Abdi, 2010).

No scoring for the original dimensions of the FQ was found. To match the conditions in this study, a principal component analysis was conducted through the R package FactomineR. Three main components were extracted, equivalent to the three conditions used in the experimental design (Boredom, flow, frustration). Results were organized by factor loadings and labelled dimensions, whereas communalities were computed afterwards. In order to calculate the internal consistency, all items must be positively formulated to become a total score for flow. Items loading negatively on the first component, being flow, were reversed. Based on this organization of the items, a correlation matrix was computed. All correlations were corrected for the pairwise comparisons problem by administering a Benjamini Hochberg procedure (Benjamini & Hochberg, 1995). Further analysis of the FQ was conducted with the transformed data of priory extracted components. First, the internal consistency the FQ and the new scoring method in every condition was calculated, along with means, standard deviations and range. Regarding the interpretation of Cronbach's alpha, guidelines from George and Mallerey (2003) were used: α >.90 is excellent, α >.80 is good, α >.70 is acceptable, α >.60 is questionable, α >.50 is poor and α <.50 is unacceptable. To detect differences in reported flow per condition, for men and women separately, an independent *t*-test was administered. Finally, a repeated measures ANOVA examined main and interaction effects of the reported scores on the subscales with both devices and in all conditions. Post hoc tests were conducted using a Bonferroni correction, providing pairwise comparisons for both main and interaction effects. A violin plot was constructed to display the median, variation and dispersion of the scores on the FQ in every condition.

Next, means, standard deviations and range were calculated for the SSQ, alongside with the internal consistency. An independent *t*-test provided differences in simulator sickness between male and female participants. Note that the transformed data solely was used to administer a *t*-test.

Neuronal correlates.

The EEG was recorded with a sampling rate of 2048 *Hz* and pre-processed with a bandpass filter of 0.01 – 30 *Hz*. To analyse the raw datasets, Matlab version 9_1 with the toolboxes EEGlab and ERPlab was used. The quality of the filtered EEG sets was manually examined and large episodes of noise were deleted. All datasets were eventually adopted in current study. P2 showed excessive high frequency noise in most datasets and was therefore deleted and interpolated in all sets. Furtherly, PoZ and Cz were deleted and interpolated in some sets. Referencing was done with the electrode FPz whereas results were equal to the initial referencing of AFz and average referencing. In a next step, epochs were extracted with a time window of -200 to 1000 *ms*, and epoch rejection was automatically applied marking epochs containing activity below a threshold of -250 *Hz* and above an upper threshold of 250 *Hz*. Between 2% and 15% of the trials were deleted across the sample, independent of condition. ERPs were extracted time locked to the stimulus onset (oddball, standard, novelty), and a grand average over all conditions was computed. Waveforms showing the different electrodes and topographies were based on the grand average, whereas latency and amplitude were calculated per subject and condition. Further, a graphical representation of the ERPs evoked by an oddball sound was provided, measured only in Pz. Grand averages were calculated for every condition separately and depicted in colour, complemented by the ERPs of all subjects separately, showing dispersion. In this manner, a more elaborated way of depicting data other than a grand average across participants is provided (Rousselet, Foxe, & Bolam, 2016). All plots and statistics were based on the middle-lines electrodes Pz, Fz and Cz (Polich, 2007; Allison & Polich, 2008).

The values for amplitude and latency were later used in a repeated measures ANOVA, to discover an effect of device, condition and electrode on both latency (*ms*) and amplitude (*µV*). During the computing of latency, an error occurred in the files of two participants. As a result, these calculations are based on sixteen (88,89%) participants instead of the total sample. Normality tests were conducted for both amplitude and latency, to meet the assumptions of a repeated measures ANOVA. When examining the Kolgormorov-Smirnov test for amplitude, VR boredom in Fz (*Skewness* = 2.29, *SD* = .54; *Kurtosis* = 6.93, *SD* = 1.04), VR flow in Fz (*Skewness* = 2.32, *SD* = .54; *Kurtosis* =6.29, *SD* = 1.04)and PC flow in Cz (*Skewness* = 1.72, *SD* = .54; *Kurtosis* = 3.94, *SD* = 1.04) showed significance. These conditions also showed skewness and kurtosis values outside of the cut-off. When applying the same procedure to latency, only the condition PC flow in Cz *(Skewness* = -1.40, *SD* = .56; *Kurtosis* = .86, *SD* = 1.09) shows significance, still with values for skewness and kurtosis within the cut-off. Despite several transformations, the amplitude of the three conditions was not normally distributed. However, since all conclusions of the amplitude regarding the P300 will be made based on the measurement in Pz, a repeated measures ANOVA was conducted anyhow.

Results

Reaction times and error rates

Descriptive statistics.

In table 2, the means, standard deviations and range of the reaction times and error rate in every condition are depicted. Reaction time is shown in seconds, whereas error rate is displayed in percentages.

Table 2

Means, standard deviations and range for reaction time(s) and error rate(%)

	Reaction Time				Error Rate			
Condition	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*
VR Boredom	.746	.088	.617	.948	2,00%	2.27%	0.33%	10%
VR Flow	.745	.100	.577	.969	2,27%	2.08%	0.00%	7.6%
VR Frustration	.738	.081	.591	.882	5.07%	12.15%	0.30%	53%
PC Boredom	.725	.079	.604	.887	2.05%	1.95%	0.00%	6.67%
PC Flow	.741	.100	.592	.979	2.25%	2.53%	0.00%	10.56%
PC Frustration	.759	.115	.590	.988	5.07%	12.98%	0.00%	56.67%

Note. VR = Virtual Reality with a HTC Vive HMD; PC = Personal computer with flat screen

An independent *t*-test discerns differences in reaction times between sexes (Table 3). A significant effect of reaction time was found in the flow condition for both VR and PC between men (*n* = 15) and women (*n* = 3). In both cases, women showed higher reaction times than men.

Table 3

Means, standard deviations and independent t-tests of the reaction times for the male and female subjects

	Women		Men			
Condition	*M*	*SD*	*M*	*SD*	*t*	*d*
VR Boredom	.819	.114	.731	.0782	-1.66	.90
VR Flow	.856	.100	.722	.0824	-2.50*	1.46
VR Frustration	.773	.083	.731	.0822	-.80	.51
PC Boredom	.743	.132	.721	.0711	-.43	.21
PC Flow	.888	.080	.711	.0765	-3.62*	2.25
PC Frustration	.847	.122	.742	.109	-1.51	.91

Note. VR = Virtual Reality with a HTC Vive HMD; PC = Personal computer with flat screen; **p <.05*

With the error rates not being normally distributed, a non-parametric Willcoxon test was administered. Table 4 shows that the error rates differed significantly between male and female participants in the conditions VR boredom and PC frustration, where males committed fewer mistakes than women. A marginally significant difference in the same trend was found in the conditions VR flow and VR frustration.

Table 4

Mean rank, sum of ranks, Willcoxon values of error rate for male and female subjects

	Women		Men		
Condition	*M*	*Sum*	*M*	*Sum*	*W*
VR Boredom	15.67	47.00	8.27	124.00	4.00*
VR Flow	14.67	44.00	8.47	127.00	7.00[a]
VR Frustration	14.67	44.00	8.47	127.00	7.00[a]
PC Boredom	13.33	40.00	8.73	131.00	11.00
PC Flow	13.17	39.50	8.77	131.50	11.50
PC Frustration	17.00	51.00	8.00	120.00	.00**

Note. VR = Virtual Reality with a HTC Vive HMD; PC = Personal computer with flat screen;

Repeated measures ANOVA.

First, a two-way repeated measures ANOVA was conducted to compare the effect of device and condition on the mean reaction times measured in the conditions boredom, flow and frustration with both PC and VR. There was no significant main effect of device (η = .00, $F(1, 17)$ = .02, p =.905) nor was there a main effect of condition (*Eta* = .03, $F(16, 2)$ = .514, p = .603). Finally, no interaction effect was found between device and condition regarding reaction times (η = .08, $F(2, 16)$= 1.51, p =.239).

Friedman test.

A non-parametric Friedman test of differences among repeated measures was conducted for the error rate and rendered a Chi-square (5, N = 18) value of 3.84, which is non-significant. To detect possible differences between error rates in the conditions, a post-hoc test was administered for every feasible pair. A Wilcoxon Signed-Ranks Test indicated no significant differences in error rates between conditions and devices.

Flow questionnaire

Principal component analysis.

To obtain a scoring method for the FQ fitting the research design, a principal component analysis was conducted. Eventually, three components were found. The following figure offers a graphical rendition of the three extracted components and mutual proportions of the factor loadings. The component coding for boredom consists out of three items, each addressing difficulty or anxiety. Four items regarding engagement and presence constructed the factor coding for flow, whereas boredom consisted out of five items about tediousness and repetition.

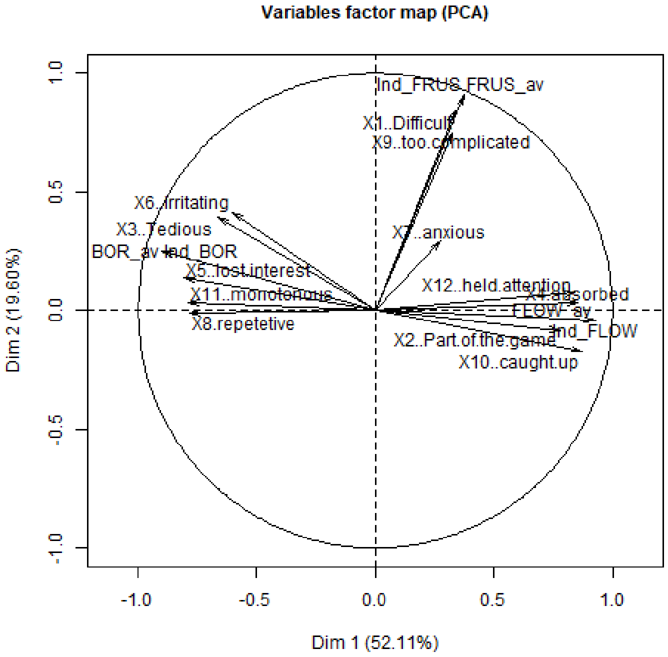


Figure 5. Graph of Principal Component Analysis in R for the FQ

Table 5 shows exact factor loadings and communalities. Items are ordered according to factor loadings, and communalities portrayed per item. Item numbers are included and comprise the same numbering as displayed in the correlation matrix in Table 6. Through confirmatory factor analysis, three components were constructed. However, the third component seems obsolete, since two groups of items show high factor loadings on factor one, in an opposite manner. This can also be observed in Figure 5, where two factors load highly on each extrema of the axis. Due to substantially high loadings, both groups of items are included as factor and used as subscales. The group of items correlating positively with component one will be the subscale for measuring flow (*Eigenvalue* = 4.84), whereas the ones correlating negatively with component one will be utilized for measuring the subscale easy (*Eigenvalue* = 3.22), functioning as an exact opposite of flow. Finally, the group of items correlating highly with component two functions as a measurement the subscale difficult (*Eigenvalue* = 9.95). The explained variance for these three factors totals 80.67% meaning that 80.67% of the total variance is explained by this model.

Table 5

Factor loadings, communalities and eigenvalues for each item of each component of the FQ

Denomination of factors	Components	Factor 1	Factor 2	Factor 3	Communality
Flow	10. I was caught up in the game	**.87**	-.17	.22	**.84**
	4. I felt absorbed by this game	**.85**	.04	.40	**.89**
	12. This game held my attention	**.84**	.07	.25	**.88**
	2. I felt like I was part of the game	**.78**	-.08	.45	**.82**
Difficult	1. I found this game difficult to play	.33	**.84**	-.1	**.83**
	9. This game was too complicated for me to play	.33	**.75**	-.13	**.69**
	7. I felt very anxious playing this game	.27	**.29**	.34	**.28**
Easy	5. I lost interest in this game pretty quickly	**-.81**	.14	.32	**.81**
	11. The game was very monotonous	**-.79**	.03	.33	**.77**
	8. This game felt very repetitive	**-.78**	-.01	.41	**.78**
	3. The game was very tedious to play	**-.67**	.39	.15	**.62**
	6. This game was very irritating to play	**-.60**	.42	.22	**.58**

Note. Items are numbered according to appearance in the FQ and are ordered according to factor loadings.

Table 6 shows the inter-item correlations between all the items of the FQ, corrected for the pairwise comparison problem. Almost all items of the subscale flow show significant correlations with every other item of said scale, solely item 2 shows no significant correlation whatsoever. Item 4 and 12 correlate exclusively to other items of the flow subscale, however item 10 also correlates significantly with an item of the difficult subscale. Item 3 of the boredom subscale correlates with every other item for the same scale, however, shows significant correlations with items of the flow scale. Every item of the boredom scale seems to correlate significantly to other items of the same scale, yet correlates significantly to random items from other scales as well. Finally, Item 1 of difficult does not correlate with the other items of the same scale, whereas Item 7 and 9 do not correlate which each other, nor with any other items.

Table 6

Inter-item Pearson correlations and p-values of all items of the FQ

Scale	1	2	3^b	4	5^b	6^b	7	8^b	9	10	11^b	12
Item 1												
Item 2	.88*											
P-value	<.0001											
Item 3^b	-.03	.22										
P-value	.917	.373										
Item 4	-.31	.64	.70*									
P-value	.218	.004	.001									
Item 5^b	.25	.37	.88*	.62								
P-value	.315	.133	<.0001	.006								
Item 6^b	.01	.20	.91*	.58	.86*							
P-value	.978	.433	<.0001	.013	<.0001							
Item 7	-.07	.15	.32	.48	.09	.39						
P-value	.798	.566	.194	.043	.736	.112						
Item 8^b	.28	.26	.70*	.50	.81*	.65*	.14					
P-value	.268	.303	.001	.033	<.0001	.003	.589					
Item 9	.26	.27	.06	.13	.02	.11	.42	.16				
P-value	.302	.271	.830	.608	.944	.675	.084	.526				
Item 10	.22	.47	.82*	.91*	.75*	.65	.36	.64	.02			
P-value	.382	.050	<.0001	<.0001	<.0001	.004	.149	.004	.924			
Item 11^b	.25	.22	.71*	.52	.81*	.69*	.21	.95*	.14	.67*		
P-value	.321	.386	.001	.026	<.0001	.002	.408	<.0001	.593	.002		
Item 12	.36	.62	.55	.85*	.54	.47	.44	.40	.28	.81*	.43	
P-value	.148	.007	.018	<.0001	.021	.049	.065	.100	.263	<.0001	.078	

Note. Easy = item 3, item 5, item 6, item 8, item 11; Flow = item 2, item 4, item 10, item 12; Difficult = item 1, item 7, item 9

**p* < critical value calculated with the Benjamini-Hochberg correction

^{b} Reverse item

Descriptive statistics.

Means, standard deviations, range and Cronbach's alpha for the FQ can be found in Table 7. The internal consistency of the subscales is calculated in every condition separately. The subscales easy and flow have an excellent internal consistency in the conditions VR flow, PC flow and PC frustration, whereas in condition PC boredom solely the subscale easy is excellent. Further, both subscales have a good internal consistency in VR boredom, whereas the subscale flow shows a good internal consistency in VR frustration and an acceptable value for Cronbach's alpha in PC boredom. The internal consistency for the subscale difficult is questionable in VR boredom and VR frustration, poor in PC frustration, and unacceptable in the conditions VR flow, PC boredom and PC flow.

Table 7

Means, standard deviations, Cronbach's Alpha and range for the Flow Questionnaire and its subscales in all of the experimental conditions

Condition		*M*	*SD*	*α*	*Min*	*Max*
VR Boredom		39.44	4.73		30.00	46.00
	Easy	21.89	7.23	.89	11.00	35.00
	Flow	13.17	5.29	.89	4.00	22.00
	Difficult	4.39	1.97	.66	3.00	11.00
VR Flow		37.33	7.28		25.00	54.00
	Easy	18.67	8.53	.94	8.00	35.00
	Flow	16.83	6.73	.94	4.00	27.00
	Difficult	6.17	2.28	.39	3.00	11.00
VR frustration		45.72	8.64		31.00	63.00
	Easy	21.44	7.08	.93	9.00	34.00
	Flow	14.78	7.07	.80	4.00	25.00
	Difficult	9.50	4.24	.66	3.00	20.00
PC Boredom		38.83	5.00		26.00	46.00
	Easy	23.28	6.03	.93	13.00	35.00
	Flow	11.11	5.12	.70	4.00	21.00
	Difficult	4.44	1.72	.01	3.00	8.00
PC Flow		39.22	5.00		28.00	45.00
	Easy	20.72	8.11	.93	10.00	35.00
	Flow	13.33	5.71	.92	400	20.00
	Difficult	5.17	1.65	.38	3.00	8.00
PC Frustration		44.50	6.96		31.00	62.00
	Easy	21.22	8.66	.92	7.00	35.00
	Flow	14.22	6.33	.92	5.00	23.00
	Difficult	9.06	3.32	.51	4.00	15.00

Note. VR = Virtual Reality with a HTC Vive HMD; PC = Personal computer with flat screen;

An independent *t*-test was conducted to examine differences between male and female participants in the scores for the FQ. Note that all parametric tests regarding the FQ were administered with the logarithmically transformed data. As shown in Table 8, no significant differences were found.

Table 8

Means, standard deviations and independent t-tests for the Flow Questionnaire and its subscales in all experimental conditions

		Women		Men			
Condition		*M*	*SD*	*M*	*SD*	*t*	*d*
VR Boredom		1.59	.06	1.59	.05	.04	.03
	Easy	1.33	.15	1.27	.14	.76	.43
	Flow	1.20	.12	1.05	.22	-1.10	.74
	Difficult	.55	.13	.63	.16	.76	.51
VR Flow		1.59	.17	1.56	.07	-.66	.43
	Easy	1.26	.33	1.22	.18	-.26	.00
	Flow	1.22	.14	1.18	.23	-.27	.03
	Difficult	.84	.06	.74	.17	-.93	.53
VR frustration		1.67	.16	1.65	.07	-.42	.31
	Easy	1.21	.24	1.33	.14	1.24	.57
	Flow	1.23	.08	1.08	.29	-.91	.53
	Difficult	1.11	.24	.90	.19	-1.69	1.04
PC Boredom		1.58	.06	1.59	.06	.26	.19
	Easy	1.37	.08	1.35	.12	-.25	.08
	Flow	1.02	.16	.99	.24	-.21	.03
	Difficult	.52	.07	.64	.16	1.26	1.00
PC Flow		1.58	.11	1.59	.05	.36	.09
	Easy	1.27	.27	1.29	.15	.134	.07
	Flow	1.09	.22	1.07	.23	-.12	.00
	Difficult	.59	.11	.71	.14	1.35	1.01
PC Frustration		1.64	.15	1.64	.05	-.06	.16
	Easy	1.25	.29	1.30	.19	.35	.06
	Flow	1.17	.13	1.09	.25	-.56	.23
	Difficult	.97	.20	.92	.16	-.46	.29

Note. VR = Virtual Reality with a HTC Vive HMD; PC = Personal computer with flat screen; *p <.05

Non-parametric descriptive analysis provides a validation of the three conditions in the primary task. This manipulation check shows if subjects in fact report more flow in the flow condition etcetera. Violin plots (Figure 6) display the median, variation and density of scores on the FQ in every condition, for PC and VR separately. In the boredom condition played in VR, subjects scored higher on the easy subscale (*M* = 4.38, *SD* = .34) than on the flow subscale (*M* = 3.29, *SD* = .31) and the difficult scale (*M* = 1.46, *SD* = .16), showing participants in fact experience more boredom in the boredom condition. Same trend can be observed in the PC condition, where subjects reported more boredom (*M* = 4.76, *SD* = .28) than flow (*M* = 2.78, *SD* = .30) and frustration (*M* = 1.48, *SD* = .14). In the flow condition, a discrepancy can be perceived between PC and VR. Whereas in VR participants actually report experiencing more flow (*M* = 4.21, *SD* = .40) than boredom (*M* = 3.73, *SD* = .40) and frustration (*M* = 2.06, *SD* = .18). However, when participants played the game on PC, self-reports showed more boredom (*M* = 4.16, *SD* = .38) than actual flow (*M* = 3.33, *SD* = .34). Subjects were also less frustrated during flow in the PC condition (*M* = 1.72, *SD* = 0.13). Moreover, participants reported overall more flow in the VR condition than on a PC. Finally, in frustration, identical tendencies in PC and VR can be observed. Subjects score highest on the subscale easy in both VR (*M* = 4,29, *SD* = .33) and PC (*M* = 4,24, *SD* = .41), whereas subjects also reported more flow in VR (*M* = 3.70, *SD* = .42) and PC (*M* = 3.55, *SD* = 0.37) than frustration. Although participants do report more frustration in the frustration condition than in any other, on both VR (*M* = 3.17, *SD* = .33) and PC (*M* = 3.02, *SD* = .26).

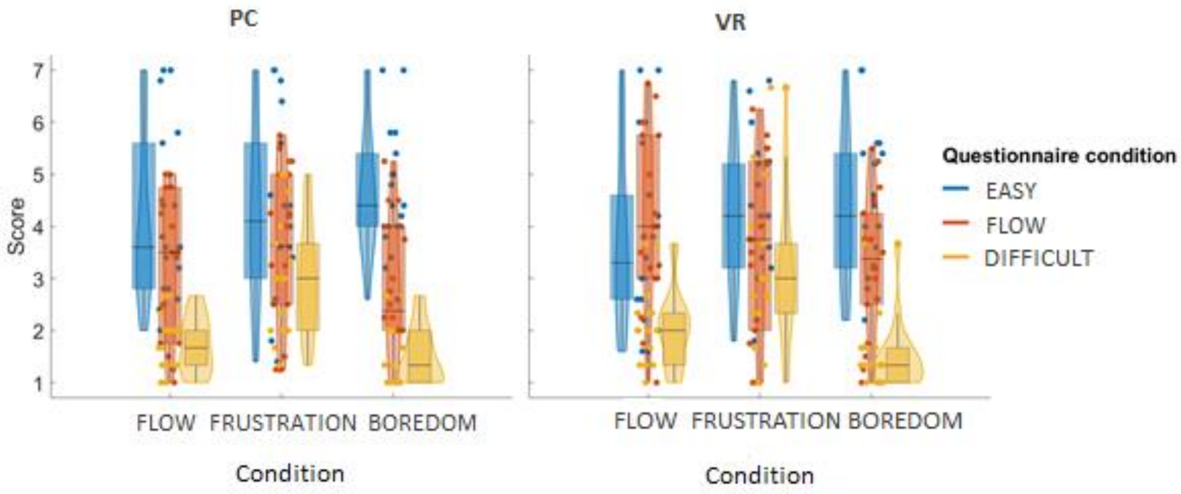


Figure 6: Violin plots displaying medians, variation and dispersion for every condition, with devices depicted separately.

Repeated measures ANOVA.

A factorial repeated measures ANOVA (2 x 3 x 3) was conducted to compare the main effects of device, condition and subscale, and the interaction effect on the scores of the questionnaires. *Device* included two levels (VR, PC), whereas *condition* (boredom, flow, frustration) and *subscale* (easy, flow, difficult) included both three levels. A post-hoc

Bonferroni procedure was conducted and showed significance in differences between the levels of every factor in the ANOVA, being device, condition and subscale of the FQ.

The tests of within-subject effects showed a marginally significant main effect of device ($\eta = .18$, $F(1, 17) = 3.73$, $p = .070$). Scores on the FQ are higher ($p_{\text{bonferroni}} = .070$) in the VR condition ($M = 1.06$, $SD = .02$) than in the PC condition ($M = 1.04$, $SD = .02$). The main effect of condition on scores of the FQ main effects of condition are significant ($\eta = .51$, $F(2, 16) = 17.64$, $p < .001$), whereas scores on the FQ are significantly lower in boredom ($M = .996$, $SD = .02$) than in flow ($M = 1{,}04$, $SD = .02$, $p_{\text{bonferroni}} = .039$) and frustration ($M = 1{,}10$, $SD = .03$, $p_{\text{bonferroni}} = <.001$). Scores of the FQ in flow are significantly lower ($p_{\text{bonferroni}} = .003$) than in frustration. The main effect of subscale was found significant ($\eta = .85$, $F(2, 16) = 96.27$, $p < .001$). Scores on the subscale difficult. Significant differences were also found between Easy ($M = 1.26$, $SD = 0.02$) and flow ($M = 1{,}12$, $SD = 0.4$, $p_{\text{bonferroni}} = 0.028$), easy and difficult ($M = .758$, $SD = 0.02$, $p_{\text{bonferroni}} = <.001$) and flow and difficult ($p_{\text{bonferroni}} = <.001$).

There was no significant interaction effect between device and condition ($\eta = .059$, $F(2, 16) = 1{,}06$, $p = .356$). However, a significant interaction effect was found between device and subscale ($\eta = .363$, $F(2,16) = 9.68$, $p <.001$). The scores on the subscale Flow differ significantly ($p_{\text{bonferroni}} = .040$) in the VR condition from the PC condition, with self-reported flow begin higher in VR. The scores on subscales easy ($p_{\text{bonferroni}} = .478$) and Difficult ($p_{\text{bonferroni}} = .574$) were not significantly different in VR than in PC. There was also a significant interaction effect between condition and subscale ($\eta = .469$, $F(4, 14) = 14{,}99$, $p <.001$). In the boredom condition, differences between easy and flow were significant ($p_{\text{bonferroni}} = .002$), such as the difference between flow and difficult ($p_{\text{bonferroni}} <.001$) and easy and eifficult ($p_{\text{bonferroni}} <.001$). The same significance was found in the flow condition between easy and difficult, and flow and difficult ($p_{\text{bonferroni}} <.001$), however, the difference between the subscale flow and boredom was not found significant ($p_{\text{bonferroni}} = .433$). The same tendency was found in the frustration condition, were easy and difficult, and flow and difficult were significantly different from each other ($p_{\text{bonferroni}} <.001$), and easy and flow were not ($p_{\text{bonferroni}} = 1.00$). Finally, the three-way interaction between al factors was not significant ($p_{\text{bonferroni}} = .216$).

Simulator Sickness Questionnaire

Following table displays the means, standard deviations, range, and Cronbach's Alpha for the Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum & Lilienthal, 1993). The subscale nausea shows an excellent internal consistency, whereas the dimension oculomotor is good internal consistent and disorientation is labelled as acceptable. The total internal consistency of the SSQ is excellent.

Table 9

Means, standard deviations, Cronbach's Alpha, and range for the Simulator Sickness Questionnaire

Scale		*M*	*SD*	*α*	*Min*	*Max*
SSQ		39.06	43.13	.93	.00	172.04
	Nausea	38.69	51.68	.94	.00	190.80
	Oculomotor	32.00	30.05	.82	.00	113.70
	disorientation	39.44	48.04	.79	.00	194.88

Note. SSQ = Simulator Sickness Questionnaire

A *t*-test examines differences between female and male participants. Since this is a parametric test, only the scores on the SSQ after a square root transformation were used. As shown in table 10, there is a difference between women (*n*= 3) and men (*n*= 15) in simulator sickness. Female participants scored significantly higher than men on the subscales oculomotor and disorientation, yet so significant difference was found between men and women in the nausea subscale. Despite, an overall significant difference between male and female participants can be found, since *t*-scores for the total SSQ were significant. These results show that women report a higher severity of the simulator sickness symptoms than men.

Table 10

Means, standard deviations and independent t-tests of the Simulator Sickness questionnaire for the female and male subjects

		Women		Men			
Scale		*M*	*SD*	*M*	*SD*	*T*	*d*
SSQ		9.29	3.54	4.71	2.48	-2.75**	1.26
	Nausea	8.55	5.36	4.41	3.02	-1.92	.92
	Oculomotor	2.81	.21	2.36	.27	-2.60**	1.51
	Disorientation	9.59	3.90	4.04	3.37	-2.55**	1.23

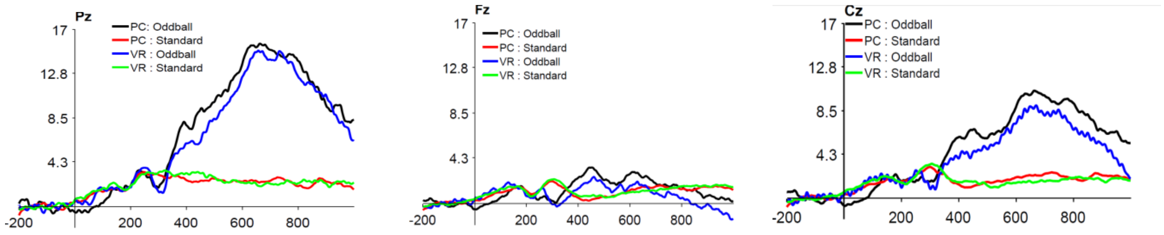
Note. SSQ = Simulator Sickness Questionnaire; **p* <.05, ***p* <.01

EEG

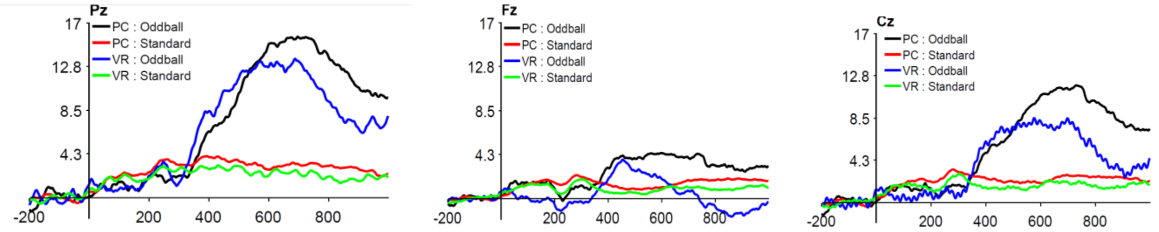
Descriptive analysis

Figure 7 shows the P300 elicited by the oddball tone based on the grand average, in all six conditions and measured in various cerebral sites. Said sites differ greatly from one another in amplitude, whereas Pz shows a much higher amplitude than Cz, and Fz does not seem to record any p300 at all. Hence, the P300 is measured mostly in parietal and central sites, yet less in frontal electrodes. First, a large effect of the oddball sound can be observed, showing a manipulation check. Neither of the recording sites measured an observable P300 to the standard tone, whereas oddball tones provoke clear waveforms in which the component can be distinguished. However, frontal measurements do not show a difference in response between standard and oddball tones. Peak amplitudes do not differ drastically across conditions, yet in Flow, a discrepancy between the peak amplitude of PC oddball (15,68 μV) and VR oddball (13,526 μV) is observed, depicted by respectively the black and the blue line . Hence, the oddball tone elicits a longer and stronger response in the PC condition in flow. This effect can be distinguished in both parietal and central sites.

Boredom



Flow



Frustration

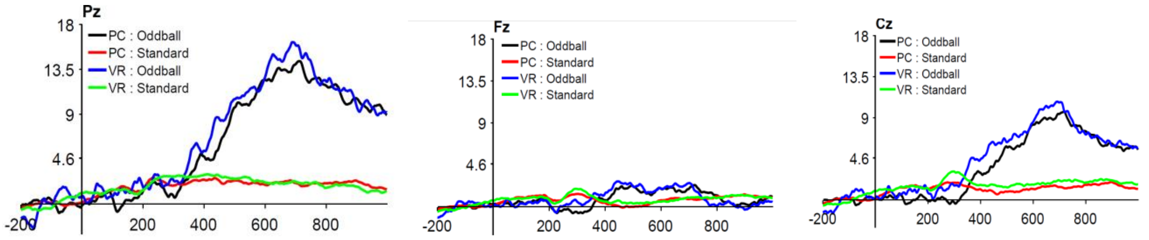
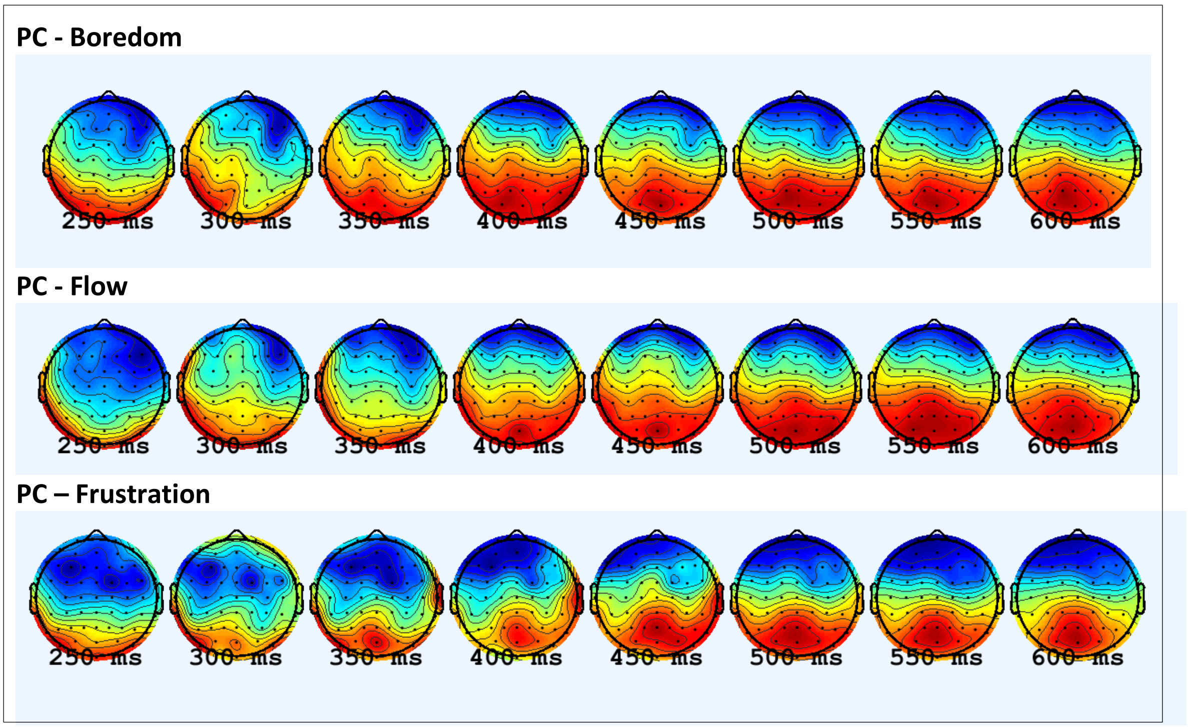


Figure 7: Waveforms of the P300 in a time window of 0-800 ms, measured in frontal (Fz), central (Cz) and parietal (Pz) sites for both oddball and standard trials, calculated from the grand average. With Microvolt on the y-axis and millisecond on the x-axis. The red line stands for a response to the standard tone in the PC condition, whereas green represents a standard tone in VR. Black shows the response to an oddball sound in PC and blue shows an oddball sound in VR.

Figure 8 depicts a topographical view over time of the activity provoked by an oddball sound in the PC condition. Most cerebral activity is parietally and centrally located, and almost no activity is observed in the frontal regions, confirming findings in figure 7. Visually, it seems that the amplitude and latency do not differ as much between conditions, however, intensity and latency of the response seems much higher in the Frustration condition. Oppositely, an earlier effect of the oddball sound and a higher intensity of response is shown in the Boredom condition.

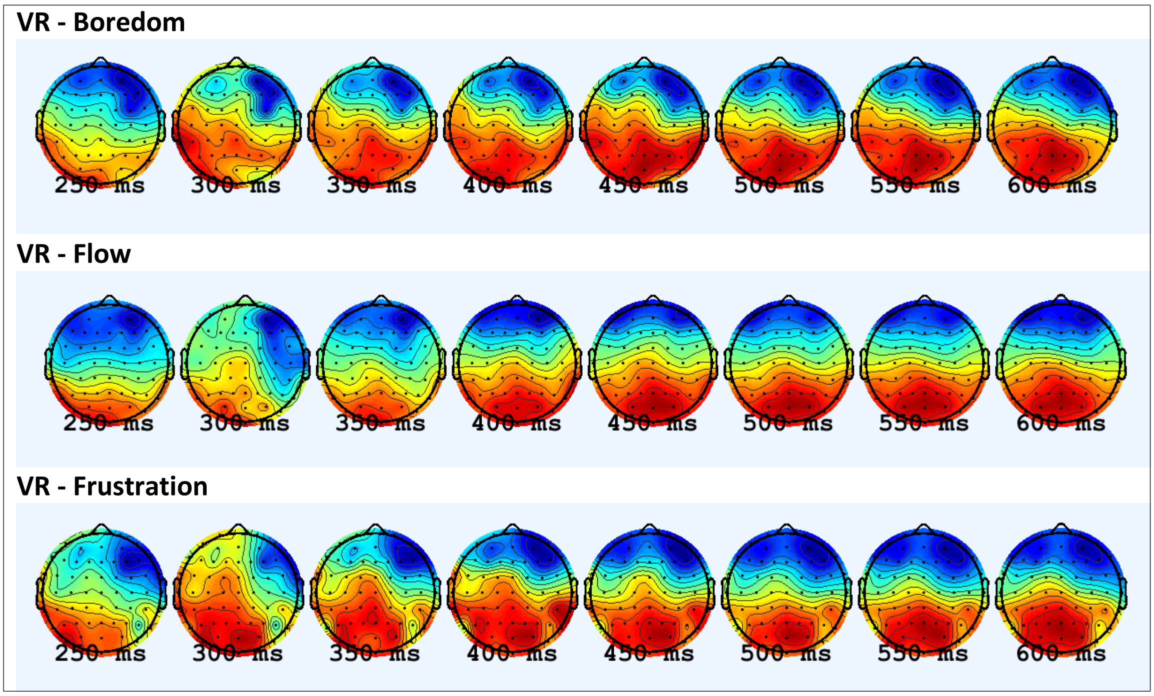

PC - Boredom

PC - Flow

PC – Frustration

Figure 8: topographies of instantaneous amplitude over time during the PC condition for Boredom, Flow and Frustration , calculated from the grand average

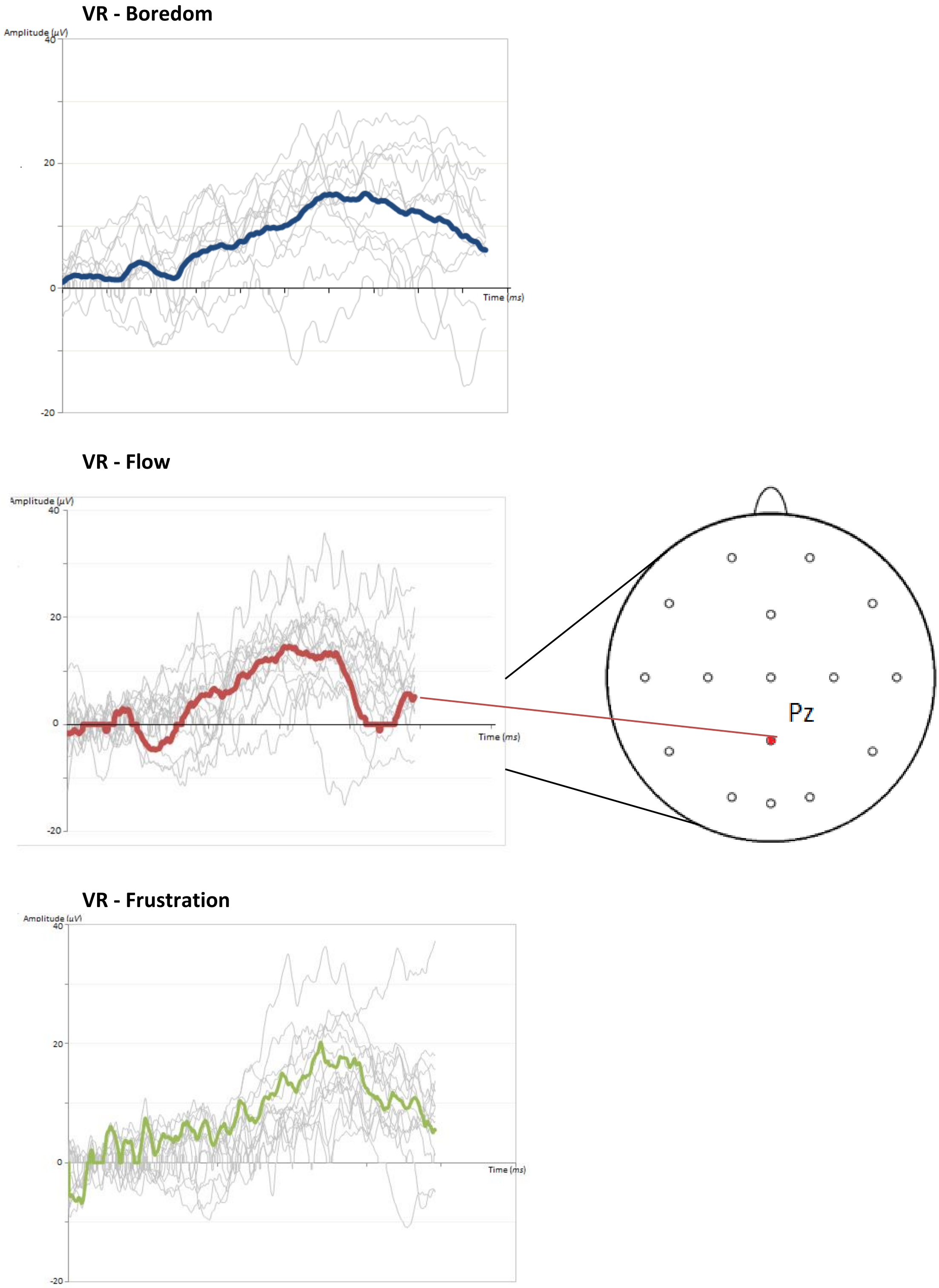
The intensity of the oddball effect seems slightly lower in VR (Figure 9), in every condition. The delayed response in the Frustration condition in PC cannot be observed in VR, with much more cerebral activity from 250 to 400 ms.


VR - Boredom

VR - Flow

VR - Frustration

Figure 9: topographies of instantaneous amplitude over time during the VR condition for Boredom, Flow and Frustration, calculated from the grand average


VR - Boredom

VR - Flow

VR - Frustration

Figure 10: ERP waveforms of every subject depicted per condition in virtual reality, with the grand average calculated for each condition separately in colour. All results shown are measured in Pz. Amplitude in µV on the vertical axis and time in milliseconds on the horizontal one.

Repeated measures ANOVA.

A repeated measures ANOVA (2 x 3 x 3) was conducted for both measurements of the ERPs, being latency and amplitude. In this manner effect of *device* (PC, VR), *condition* (Flow, Boredom, Frustration) and *electrode* (Pz, Fz, Cz) on both latency and amplitude can be examined.

Amplitude.

The main effect of device on amplitude was non-significant ($\eta = .01$, $F(1, 17) = .16$, $p = 699$), such as the main effect of condition ($\eta = .02$, $F(2, 16) = .30$, $p = .745$). The main effect of electrode on amplitude however, was significant ($\eta = .77$, $F(2, 16) = 55{,}36$, $p < .001$). Significant differences are shown in the pairwise comparisons for the electrodes. The amplitude of Pz ($M = 6{,}19$, $SD = .71$) is significantly ($p_{\text{bonferroni}} < .001$) larger than the amplitude of Fz ($M = 1{,}14$, $SD = .47$) and Cz ($M = 3.93$, $SD = .61$). The same tendency is found in the difference between Fz and Cz ($p_{\text{bonferroni}} < .001$). Further, the interaction effect of device and condition was not significant ($\eta = .00$, $F(2, 16) = .02$, $p = .976$), with the same trend for the interaction effect of device and electrode ($\eta = .07$, $F(2, 16) = 1.21$, $p = .299$) and the interaction effect of condition and electrode ($\eta = .04$, $F(4, 14) = 79$, $p = .537$). Finally, the three-way interaction of all three included factors was non-significant ($\eta = .121$, $F(4, 14) = 2.35$, $p = .103$).

Latency

Secondly, effects of device, condition and electrode on the latency of the P300 are examined. The main effect of device is marginally significant ($\eta = .182$, $F(1, 15) = 3{,}34$, $p = .088$). Pairwise comparisons showed a higher latency in PC ($M = 225{,}01$, $SD = 30.40$) than in VR ($M = 163{,}41$, $SD = 94{,}75$). The main effect of condition was non-significant ($\eta = .049$, $F(2, 14) = .77$, $p = .474$), whereas the main effect of electrode is significant ($\eta = .53$, $F(2, 14) = 17{,}011$, $p < .001$). The Pz ($M = 265{,}77$, $SD = 23{,}18$) electrode shows a significantly ($p_{\text{bonferroni}} = .001$) higher latency than Fz ($M = 89{,}37$, $SD = 39{,}34$, , $p_{\text{bonferroni}} < .001$) and Cz ($M = 227{,}50$, $SD = 31{,}90$, $p_{\text{bonferroni}} < .001$). The difference between Pz and Cz is not significant ($p_{\text{bonferroni}} = .552$). Both the interactions effects of device and condition ($\eta = .01$, $F(2, 14) = .21$, $p = .809$) and device and electrode ($\eta = .09$, $F(2, 14) = 1{,}27$, $p = .295$) are not significant. Yet, significant differences were found in the interaction effect of condition and electrode ($\eta = .21$, $F(4, 14) = 3{,}99$, $p = .006$). In Boredom, the latency of Pz ($M = 356{,}64$, $SD = 9.25$) is significantly higher than the latency of Fz ($M = 37{,}46$, $SD = 53{,}24$, $p_{\text{bonferroni}} < .001$), and Cz ($M = 217{,}53$, $SD = 31{,}33$, $p_{\text{bonferroni}} = .004$). In

Flow, the latency of Fz (*M* = 80,89, *SD* = 49,12) is significantly lower ($p_{\text{bonferroni}}$ = .027) than the latency of Cz (*M* = 228,76, *SD* = 61,28). In frustration, same tendency shows, with a significant difference ($p_{\text{bonferroni}}$ = .024) between Fz (*M* = 149,75, *SD* = 46,61) and Cz (*M* = 236,21, *SD* = 45,56). Finally, the three way interaction is marginally significant (η = .157, *F*(4, 14) = .16, *p* = .066).

Discussion

The aim of the current study consisted of two large aspects. First, the attempts of Núñez Castellar et al. (2016) in finding an objective measure for flow were replicated in order to confirm found method and pursue it further. Whilst subjects played a game that was consecutively boring, flow inducing and frustrating, attention was assessed through an auditory oddball task. Additionally, EEG was utilized to define neuronal correlates during all three conditions. Secondly, prior research was extended by replicating said study with virtual reality goggles. In this manner, possible discrepancies in immersion, flow and presence can be detected. Also, a difference in attention for the actual environment can be assessed through both behavioural and neuronal techniques. Attention was measured by computing the mean reaction time and error rate per participants. Both omissions and false positives were included in the calculation of the error rates. To measure the neuronal aspect of attention, the P300 was extracted and examined. In order to check used manipulation, being the conditions of the primary task, the FQ was administered, scored and analysed for every condition separately. To assess subject responses to Virtual Reality, a Simulator Sickness Questionnaires was administered as well.

Findings and propositions for further research

The behavioural aspect of current study consisted of measuring attention through an oddball paradigm. Four hypotheses were made regarding reaction times and committed errors: (1) Participants commit more errors during the oddball task in the flow condition, (2) Reaction times during the oddball task will be higher in the flow condition, (3) Participants commit more errors in the oddball task during flow condition in virtual reality in comparison to a non-virtual reality condition and (4) Reaction times during the oddball task will be higher in the flow condition in virtual reality in comparison to a non-virtual reality condition. As stated above, none of the hypotheses could be confirmed. (Núñez Castellar et al., 2016) expected higher reaction times and more committed errors in the flow condition as well. However, Núñez Castellar et al. (2016) do confirm said hypotheses. Despite partially replicating this study, several differences can be addressed between current and prior research. First, Núñez Castellar et al. (2016) used an existing game, offering a rather linear progress in difficulty and continuously ongoing gameplay. Hereby, subjects were caught up in the actual game during every oddball sound. Current study made use of a custom shooter game, designed to resemble commercial gaming. Subjects do not perform in the game continuously, since they are required to walk through all chambers. In this manner not every oddball sound was accompanied by actual game play, resulting in an impure measurement of flow. Second, samples differ from each other in gender diversity and objectified gaming experience. In prior research (Núñez Castellar et al., 2016), an objectified measure of gaming experience based on hours played per week was administered complementing self-reports. In current research, solely self-reports based on the subjects own estimate were used to classify subjects as expert, casual or novice gamers. In the study of Núñez Castellar et al. (2016), a more heterogeneous sample was taken, whereas in current research, the majority of the sample consisted of men who are gamers. Both these issues result in a less representative sample, possibly causing variability among results.

The experiment conducted in current study was accompanied by EEG recordings, extracting the P300 component to examine possible differences in attention. Four hypotheses were proposed: (1) The amplitude of the P300 will be smaller in the flow condition, (2) The latency of the P300 will be longer in the flow condition, (3) The amplitude of the P300 will be smaller during flow in virtual reality in comparison to a non-virtual reality condition, and (4) The latency of the P300 will be longer during flow in virtual reality in comparison to a non-virtual reality condition. Both hypothesis regarding amplitude were not confirmed. In Allison & Polich (2008), a similar research design was employed. With the primary stimulus material being a commercial game and hypothesis regarding amplitudes of the P300 measured in the same sites, findings are expected to be similar. However, current study conducts an oddball task to measure attention whereas Allison and Polich (2008) measure workload through a single-stimulus paradigm. Questions arise on the influence of the difference in difficulty on the variability found in the current study, since a single-stimulus paradigm is much simpler (Polich & Margala, 1997). However, performance on an oddball task seems equal to the performance on a single-stimulus paradigm. Moreover, the amplitude of the P300 is significantly larger in an oddball paradigm, measured in the same cerebral sites. (Polich & Margala, 1997) Studies show no disadvantage in using an oddball task, yet, the complex primary task should be taken into account.

Generally, identical issues as stated earlier regarding the primary stimulus can be addressed in not finding expected results. The premise of the commercial game in prior research was to shoot targets whilst avoiding be hit. In this manner, flow was constantly induced, whereas in the game in current study, participants had dull moments while walking through the map. Yet, artificial intelligence firing back causes additional randomness in the experimental design, which we tried to avoid in current study. Further, Allison & Polich (2008) added a condition where participants just watched the game being played as a control trial. By adding this condition, the additional difficulty of combining gaming and responding to the oddball sounds is eliminated, hypothetically showing less variability in the experimental conditions.

The hypotheses regarding latency were not only disconfirmed, but opposite results were found. Data-analysis showed a marginally significant difference in latency between devices, with the latency being longer in the PC condition. This effect shows that more attentional resources were allocated in the PC condition in comparison to the VR condition, whereas the VR condition was initially expected to call upon attention resources (Polich, 2007) since there is more immersion in VR (Mcmahan, 2003).

Partially replicating this study, the FQ was chosen based on the research of Núñez Castellar et al. (2016). However, a scoring system fitting the research design lacked. Therefore, a principal component analysis was conducted. Graphically, three components were found corresponding with the conditions in the design. Two components seem to complement each other, by each loading positively and negatively on the same factor. The components for the subscales Flow and Boredom appear to be exact opposites of each other, explaining the lack of divergent validity of both components. Both components contain convergent validity, correlating high reciprocally between all items of the same subscale. However, each item also correlates positively with almost every item of the other subscale, showing a lack of divergent validity. Still, with flow being a hypothetical concept, convergent validity is most important in order to define the components properly (Jackson & Marsh, 1996). These two subscales seem to each measure two complementary segments of the same component. Yet, in this case, these two components are expected to also complement each other in the self-reportings on the FQ in every condition. Nevertheless, Flow and Boredom seem to trend in the same way rather than interrelating, whereas both components are highly represented in all conditions and all devices. Subjects seem to report boredom and flow simultaneously. When examining the questionnaire on item-level, some items seem compatible, e.g. "*I felt like I was part of the game*" and "*this game felt repetitive*", whereas other items are exact opposites "*I lost interest in this game quickly*" and "*This game held my attention*". In this manner, participants can report flow whilst experiencing matters addressed by the Easy subscale and vice versa.

In order to measure flow, other options are available, such as the FQ based on Csikszentmihalyi and Figurski (1982) addressed in Moneta (2012). Said method is specifically developed to fit the key elements of flow, and uses an unambiguous and well conceptualized definition of flow. Yet, it comprises elaborate questions regarding flow and provides qualitative data, not fitting current design. Furthermore, a measure of average flow and the activity in which one experiences flow the most is assessed, rather than an actual state of flow induced by the balance of skill and challenge. Nevertheless, Moneta (2012) aims at assessing flow in two separate forms, being deep and shallow flow. Said addition could be interesting in conducting feature research, since observation during the experiments vouches for subjects experiencing flow in different manners. This discrepancy could hypothetically explain the entanglement of the subscales Flow and Easy. Other measures for Flow were proposed by Rheinberg, Vollmeyer and Engeser (2003), introducing the Flow Short Scale. Ten items address all key elements of flow (Engeser & Rheinberg, 2008), yet to components can be extracted, being automatic processing and absorption (Rheinberg, Vollmeyer & Engeser, 2003). This questionnaire was developed to measure flow in every activity, and thus would be suited for current research. However, no extinction can be made between boredom, flow and frustration since solely a total score of flow can be computed.

Limitations and strengths

A prior addressed limitation of this study is the finally used sample, with the majority being men with gaming experience. Initially, a more representative sample was intended, but a reassessment of used participants was necessary since some subjects felt nauseated after playing the game in virtual reality. Mostly participants without gaming experience showed symptoms of simulator sickness, such as extensively sweating, nausea, dizziness and increased salivation. With no-gamers dropping out of the experiment, a sample of gamers remained. Yet, subjects who self-reported themselves as gamers also suffered from simulator sickness, nevertheless still being able to finish the experiment. Resulting in a large variability of scores on the SSQ, varying from having no symptoms at all to experiencing every symptom of simulator sickness. No research has been done about the link between simulator sickness and game experience, however, several hypotheses were formed during the actual experiment of which some can be supported by former research. First, no-gamers use larger and uncontrolled movements since they are not trained in correlating vision with haptic feedback (Stanney, Kennedy, Drexler & Harm, 1999). This phenomenon could be a cause for this discrepancy between gamers and non-gamers. Secondly, subjects receive feedback from the VR system suggesting movement whilst the vestibular system does not respond in this manner (Cobb, Nichols, Ramsey & Wilson, 1999). Latter does not necessarily address gaming experience, but can be an important factor in the variability of subjects experiencing simulator sickness in this sample. Other factors in experiencing simulator sickness is the resolution and the weight of the device (Cobb, et al., 1999). During the experiment, subjects were complaining about the weight and the resolution of the device.

A second important limitation is the stimulus material for the primary task. Despite its value for external validity and attempting new techniques for research, the internal validity of this study decreases. First, as previously addressed, not every oddball sound was accompanied by actual gameplay, degrading the measurement of attention. Second, since the majority of the participants called themselves gamers, the frustration condition seemed rather flow inducing or boring than actual frustrating. This can be observed in the means per condition of the subscales of the FQ, with boredom and flow mostly reported in the frustration condition. These findings are supported by observations during the experiment, where subjects were determined to improve their former performance and were extremely focused on the game. In this manner, the frustration condition was more flow inducing than the actual flow condition, however solely for a selection of participants. Hence, a large variability in experiencing the different conditions is observed resulting in an indistinct manipulation.

Thirdly, due to a repeated measures design, subjects played the same stimulus material twice. Not only can learning effects appear, subjects could experience a condition in a different way playing it twice. To counter this issue, two versions of the same game were constructed. However, since the map and target placement remains identical, these effects can still appear. In order to assess this possible effect, one can analyse the FQ within conditions and ordered by which device subjects played first. If this effect exists in current study, scores on the subscales Easy and Difficult should be higher on the second played device, whereas scores for flow would be lower. When examining learning effects, solely 33.33% achieved a higher level the second time, whereas 27,78% scored lower the second time and 27,78% performed equally in both conditions. Statistical analysis was not performed in this matter, however, both possible assessments should be kept in mind in conducting new research.

An important strength of this study lies in a combination of various measures of flow, where an oddball paradigm was strengthened with an EEG recording and self-report questionnaires. Hereby, a strong foundation can be made in examining assessment techniques for the measurement of flow. Plus, the combination of EEG and Virtual Reality is rather new in experimental research. In facing all challenges that this combination implies, pitfalls and sore spots can be exposed in order to strengthen new studies aspiring such design. Moreover, collected datasets offer many opportunities in further examining neuronal correlates measured by EEG during a Virtual Reality experience.

Apart from several limitations caused by the primary task, a strength does lie in the construct of completely new stimulus material. The used software opens doors for gaming research, since every aspect of the constructed game is controllable. An attempt was made in creating an experimental manipulation resembling the reality of commercial gaming, and offers the opportunity to improve this technique. In choosing this material, a window was opened to reality intending to walk a different path than prior experimental research in gaming.

Implications

Facing several challenges that came with the combination of EEG and Virtual Reality, implications for following research can be made. The fact that EEG recordings require a pure signal without too many artefacts makes it difficult to combine with Virtual Reality. Subjects should be able to move their head freely and make use of all the opportunities VR offers. However, primary stimulus material was constructed in such way that head movement was not necessary to play the game, using the HMD as a monitor. Yet, because of this decision, simulator sickness seemed worse. A balance needs to be found when applying this design in experimental research. Despite issues with simulator sickness, the datasets collected in this study were of good quality. The expected practical issues of combining an EEG cap, ocular electrodes and a large headset were highly manageable, resulting in a solid dataset. This implies that combining these techniques actually can be a measurement for flow, and eventually other applications.

Conclusion

The need for objectified measures to examine flow shows in several studies. Many different techniques have been proposed, however, the search for enhanced and improved measuring continues. This study attempts to contribute to this search and examines flow in virtual reality with EEG recordings and an oddball paradigm to measure attention. Additionally, new stimulus material is proposed in studying flow, and through a repeated measures design with two factors, important comparisons can be made. However, the majority of the results showed insignificant. The reaction times and error rates of the oddball task did not differ between devices or conditions, nor did the amplitude of the P300. The used questionnaire to asses flow exposed issues in the definition and operationalization of the conditions, probably explaining the lack of significant results. However by attempting this novel research design, many new insights were offered and pitfalls were uncovered, opening doors for following research.

References

Abdi, H. (2010). The greenhouse-geisser correction. Encyclopedia of research design. *SAGE Publications*, 544-548.

Allison, B. Z., & Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological psychology*, 77(3), 277-283.

Avila, L., & Bailey, M. (2014). Virtual reality for the masses. *IEEE computer graphics and applications*, (5), 103-104. doi: 10.1109/mcg.2014.103

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, 37(3), 379-384. doi: 10.3758/BF0319270

Bangay, S., & Preston, L. (1998). An investigation into factors influencing immersion in interactive virtual reality environments. *Studies in health technology and informatics*, 43-51. doi: 10.3233/978-1-60750-902-8-43

Barceló, F., Periáñez, J. A., & Knight, R. T. (2002). Think differently: A brain orienting response to task novelty. *NeuroReport*, 13(15), 1887-1892. doi: 10.1097/00001756-200210280-00011

Benford, S., Magerkurth, C., & Ljungstrand, P. (2005). Bridging the physical and digital in pervasive gaming. *Communications of the ACM*, *48*(3), 54-57. doi: 10.1145/1047671.1047704

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Berg, S. (1988). Snowball sampling—I. *Encyclopedia of statistical sciences.*

Boas, Y. A. G. V. (2013). Overview of virtual reality technologies. *In Interactive Multimedia Conference.* Southampton.

Bressler, S. L., & Ding, M. (2006). Event-Related Potentials. *Wiley encyclopedia of biomedical engineering*. doi: 10.1002/9780471740360.ebs0455

Brooks Jr, F. P. (1999). What's real about virtual reality?. *Computer Graphics and Applications*, IEEE, 19(6), 16-27. doi: 10.1109/38.799723

Brown, N. O. (2012). *Life against death: The psychoanalytical meaning of history*. Wesleyan University Press.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences*, 4(6), 215-222. doi: 10.1016/s1364-6613(00)01483-2

Byagowi, A., Singhal, S., Lambeta, M., Aldaba, C., & Moussavi, Z. (2014). Design of a naturalistic navigational virtual reality using oculus rift. *Journal of Medical Devices*, *8*(3), 030946. doi: 10.1115/1.4027114

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research*, 22(3), 309-321. doi: 10.1016/j.cogbrainres.2004.09.006

Donchin, E. (1981). Surprise!…surprise?. *Psychophysiology,* 18(5), 493-513. doi: 10.1111/j.1469-8986.1981.tb01815.x

Donchin, E., Heffley, E., Hillyard, S. A., Loveless, N., Maltzman, I., Öhman, A., ... & Siddle, D. (1984). Cognition and Event-Related Potentials II. The Orienting Reflex and P300. *Annals of the New York Academy of Sciences*, 425(1), 39-57. doi: 10.1111/j.1749-6632.1984.tb23522.x

Desai, P. R., Desai, P. N., Ajmera, K. D., & Mehta, K. (2014). A review paper on oculus rift- A virtual reality headset. *arXiv preprint arXiv:1408.1173*. doi: 10.14445/22315381/ijett-v13p237

Cobb, S. V., Nichols, S., Ramsey, A., & Wilson, J. R. (1999). Virtual reality-induced symptoms and effects (VRISE). *Presence: teleoperators and virtual environments*, 8(2), 169-186. doi: 10.1162/105474699566152

Conaway, M. (1999). Repeated measures design. *Author. Retrieved from.*

Crutcher, H. L. (1975). A note on the possible misuse of the Kolmogorov-Smirnov test. *Journal of Applied Meteorology*, 14(8), 1600-1603. doi: 10.1175/1520-0450(1975)014<1600:ANOTPM>2.0.CO;2

Csikszentmihalyi, M. (2014). Attention and the holistic approach to behavior. In *Flow and the Foundations of Positive Psychology* (pp. 1-20). Springer Netherlands. doi: 10.1007/978-94-017-9088-8_1

Csikszentmihalyi, M. (1997). Finding flow. doi: 10.1017/cbo9780511621956.001

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* Harperperennial

Csikszentmihalyi, M. (2014). Play and Intrinsic Rewards. In *Flow and the Foundations of Positive Psychology* (pp. 135-153). Springer Netherlands. doi: 10.1007/978-94-017-9088-8_10

Csikszentmihalyi, M., & Figurski, T. J. (1982). Self-awareness and aversive experience in everyday life. *Journal of personality*, 50(1), 15-19.

Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In *Flow and the Foundations of Positive Psychology* (pp. 239-263). Springer Netherlands. doi: 10.1007/978-94-017-9088-8_16

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the Foundations of Positive Psychology* (pp. 35-54). Springer Netherlands. doi: 10.1007/978-94-017-9088-8_3

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9-15). ACM. doi: 10.1145/2181037.2181040

Dietrich, A. (2004). Neurocognitive mechanisms underlying the experience of flow. *Consciousness and Cognition*, 13(4), 746-761. doi: 10.1016/j.concog.2004.07.002

Ellis, S. R. (1991). Nature and origins of virtual environments: A bibliographical essay. *Computing Systems in Engineering*, 2(4), 321-347. doi: 10.1016/b978-0-08-051574-8.50096-0

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, 32(3), 158-172. doi: 10.1007/s11031-008-9102-4

ESA (2016). “Entertainment Software Association (ESA) - Essential Facts About the Computer and Video Game Industry.”

Fabiani, M., Kazmerski, V. A., Cycowicz, Y. M., & Friedman, D. (1996). Naming norms for brief environmental sounds: effects of age and dementia. *Psychophysiology*, 33(4), 462-475. doi: 10.1111/j.1469-8986.1996.tb01072.x

Gebara, C. M., Barros-Neto, T. P. D., Gertsenchtein, L., & Lotufo-Neto, F. (2015). Virtual reality exposure using three-dimensional images for the treatment of social phobia. *Revista Brasileira de Psiquiatra*, (AHEAD), 0-0. doi: 10.1590/1516-4446-2014-1560

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon.

Ghasemi, A., & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *Int J Endocrinol Metab*, 10(2), 486-489. doi: 10.5812/ijem.3505

Haluck, R. S., & Krummel, T.M. (2000). Computers and virtual reality for surgical education in the 21st century. *Archives of surgery,* 135(7), 786-792. doi: 10.1001/archsurg.135.7.786

Handy, T. C. (2005). *Event-related potentials: A methods handbook*. MIT press.

Hoffman, H. G., Meyer III, W. J., Ramirez, M., Roberts, L., Seibel, E. J., Atzori, B., ... & Patterson, D. R. (2014). Feasibility of articulated arm mounted Oculus Rift Virtual Reality goggles for adjunctive pain control during occupational therapy in pediatric burn patients. *Cyberpsychology, Behavior, and Social Networking*, *17*(6), 397-401. doi: 10.1089/cyber.2014.0058

Hoffman, H. G., Richards, T., Coda, B., Richards, A., & Sharar, S. R. (2003). The illusion of presence in immersive virtual reality during an fMRI brain scan. *CyberPsychology & Behavior*, *6*(2), 127-131. doi: 10.1089/cyber.2014.0058

Holdstock, J. S., & Rugg, M. D. (1995). The effect of attention on the P300 deflection elicited by novel sounds. *Journal of Psychophysiology*, 9(1), 18-31.

Jackson, S. A., & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of sport and exercise psychology*, 18(1), 17-35. doi: 10.1123/jsep.18.1.17

Jegers, K. (2007). Pervasive game flow: understanding player enjoyment in pervasive gaming. *Computers in Entertainment* (CIE), 5(1), 9. doi: 10.1145/1236224.1236238

Lin, C. T., Chuang, S. W., Chen, Y. C., Ko, L. W., Liang, S. F., & Jung, T. P. (2007, August). EEG effects of motion sickness induced in a dynamic virtual reality environment. In Engineering in Medicine and Biology Society, 2007. EMBS 2007. *29th Annual International Conference of the IEEE* (pp. 3872-3875). IEEE. doi: 10.1109/IEMBS.2007.4353178

Johnson, B. (2014). How the Oculus Rift Works. Retrieved November, 22, 2014.

Kaufmann, H., Schmalstieg, D., & Wagner, M. (2000). Construct3D: a virtual reality application for mathematics and geometry education. *Education and information technologies*, 5(4), 263-276. doi: 10.1023/A:1012049406877

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The international journal of aviation psychology, 3(3), 203-220. doi: 10.1207/s15327108ijap0303_3

Luck, S. J. (2005). An Introduction to the Event-Related Potential Technique (Cognitive Neuroscience).

Luck, S. J., & Kappenman, E. S. (2012). ERP components and selective attention. *The Oxford handbook of event-related potential components*, 295-328. doi: 10.1093/oxfordhb/9780195374148.013.0144

Mahalil, I., Rusli, M. E., Mohd Yusof, A., Mohd Yusof, M. Z., & Zainudin, A. R. R. (2014, August). Virtual reality-based technique for stress therapy. In *Engineering Technology and Technopreneuship (ICE2T), 2014 4th International Conference on* (pp. 295-300). IEEE. doi: 10.1109/ice2t.2014.7006265

McMahan, A. (2003). Immersion, engagement and presence. *The video game theory reader*, 67, 86.

Miloff, A., Lindner, P., Hamilton, W., Reuterskiöld, L., Andersson, G., & Carlbring, P. (2016). Single-session gamified virtual reality exposure therapy for spider phobia vs. traditional exposure therapy: study protocol for a randomized controlled non-inferiority trial. *Trials*, *17*(1), 1. doi: 10.1186/s13063-016-1171-1

Moneta, G. B. (2012). On the measurement and conceptualization of flow. *In Advances in flow research* (pp. 23-50). Springer New York. doi: 10.1007/978-1-4614-2359-1_2

Moss, J. D., & Muth, E. R. (2011). Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(3), 308-319. doi: 10.1177/0018720811405196

Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In *Flow and the Foundations of Positive Psychology* (pp. 239-263). Springer Netherlands. doi: 10.1007/978-94-017-9088-8_16

Nichols, S., Haldane, C., & Wilson, J. R. (2000). Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies*, *52*(3), 471-491. doi: 10.1006/ijhc.1999.0343

Niedermeyer, E., & da Silva, F. L. (Eds.). (2005). Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins.

Nuñez Castellar EP, Antons J, Marinazzo D, van Looy J. (2016) Being in the zone: Using behavioral and EEG recordings for the indirect assessment of flow. PeerJ Preprints 4:e2482v1 https://doi.org/10.7287/peerj.preprints.2482v1 Palva, S., & Palva, J. M. (2007). New vistas for α-frequency band oscillations. *Trends in neurosciences*, *30*(4), 150-158. doi: 10.1016/j.tins.2007.02.001

Page, B. (1999). Hamlet on The Holodeck: The Future of Narrative In Cyberspace. *MFS Modern Fiction Studies*, 45(2), 553-556. doi: 10.5860/choice.35-1342

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual review of neuroscience*, 35, 73. doi: 10.1146/annurev-neuro-062111-150525

Pfurtscheller, G., & Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*, *110*(11), 1842-1857. doi: 10.1016/s1388-2457(99)00141-8

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, 118(10), 2128-2148.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological psychology*, 41(2), 103-146. doi: 10.1016/0301-0511(95)05130-9

Polich, J., & Margala, C. (1997). P300 and probability: comparison of oddball and single-stimulus paradigms. *International Journal of Psychophysiology*, 25(2), 169-176.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in neurosciences*, 17(2), 75-79. doi: 10.1016/0166-2236(94)90078-7

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, 13(1), 25-42. doi: 10.1146/annurev.neuro.13.1.25

Raz, A., & Buhle, J. (2006). Typologies of attentional networks. *Nature Reviews Neuroscience*, 7(5), 367-379. doi: 10.1038/nrn1903

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Die erfassung des flow-erlebens. na.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1), 54-67. doi: 10.1006/ceps.1999.1020

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4), 332-339. doi: 10.1038/nrn1651

Sherry, J. L. (2004). Flow and media enjoyment. *Communication Theory*, 14(4), 328-347. doi: 10.1093/ct/14.4.328

Sherry, J. L., Rosaen, S., Bowman, N., & Huh, S. (2006). Cognitive skill predicts video game ability. *In annual meeting of the International Communication Association*, Dresden, Germany.

Slater, M. (2003). A note on presence terminology. *Presence connect*, 3(3), 1-5. doi: 10.1371/journal.pone.0067651

Stanney, K. M., Kennedy, R. S., Drexler, J. M., & Harm, D. L. (1999). Motion sickness and proprioceptive aftereffects following virtual environment exposure. *Applied ergonomics*, 30(1), 27-38. doi: 10.1016/S0003-6870(98)00039-8

Statistica (2015). “Number of DK1 & DK2 Oculus Rift development kits shipped worldwide up to June 2015, by region”

Steels, L. (2004). The autotelic principle. *In Embodied artificial intelligence* (pp. 231-242). Springer Berlin Heidelberg. doi: 10.1007/978-3-540-27833-7_17

Stengel, M., Grogorick, S., Eisemann, M., Eisemann, E., & Magnor, M. A. (2015, October). An affordable solution for binocular eye tracking and calibration in head-mounted displays. *In Proceedings of the 23rd ACM international conference on Multimedia* (pp. 15-24). ACM. doi: 10.1145/2733373.2806265

Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)*, 3(3), 3-3. doi: 10.1145/1077246.1077253

Usoh, M., & Slater, M. (1995). *An exploration of immersive virtual environments*. Endeavour, 19(1), 34-38. doi: 10.1016/0160-9327(95)98892-j

Walsh, B. J., Buonocore, M. H., Carter, C. S., & Mangun, G. R. (2011). Integrating conflict detection and attentional control mechanisms. *Journal of Cognitive Neuroscience*, 23(9), 2211-2221. doi: 10.1162/jocn.2010.21595

Weber, R., Sherry, J., & Mathiak, K. (2008). The neurophysiological perspective in mass communication research. *Biological dimensions of communication: Perspectives, methods, and research*, 41-71.

Weber, R., Tamborini, R., Westcott-Baker, A., & Kantor, B. (2009). Theorizing flow and media enjoyment as cognitive synchronization of attentional and reward networks. *Communication Theory, 19*(4), 397-422. doi: 10.1111/j.1468-2885.2009.01352.x

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, *7*(3), 225-240. doi: 10.1162/105474698565686

Yen, C. Y., Lin, K. H., Hu, M. H., Wu, R. M., Lu, T. W., & Lin, C. H. (2011). Effect of virtual reality-augmented balance training on sensory organization and attentional demand for postural control in people with Parkinson disease: a randomized controlled trial. *Physical therapy,* 91(6), 862-874. doi: 10.2522/ptj.20100050

Zimmerman, D. W., & Zumbo, B. D. (1993). Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. *The Journal of Experimental Education*, 62(1), 75-86. doi: 10.1080/00220973.1993.9943832