International Journal of Human-Computer Studies 126 (2019) 64–80

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Measuring presence in video games: An investigation of the potential use of physiological measures as indicators of presence

Thomas Terkildsen*, Guido Makransky

Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 København, Denmark

ARTICLE INFO

Keywords:
Event-Related Potential (ERP)
Galvanic Skin Response (GSR)
Measuring presence
Games User Research (GUR)
User Experience (UX)

ABSTRACT

Presence has become an increasingly central component of Games User Research (GUR) as developments in technology continuously make modern video games more conducive to the sensation of 'being there' in virtual environments. The quality of games is now commonly evaluated based on how reliably they elicit presence; however, no standardized objective measure of presence currently exists. This study investigated two physiological measures, Galvanic Skin Response (GSR) and task-irrelevant Event-Related Potentials (ERPs), as potential objective indicators of presence in games. A total of 34 participants were divided into low or high presence groups based on their self-reported presence evoked from experiencing a horror game while task-irrelevant tones were being played. It was hypothesized that presence is associated with attentional resources being fully absorbed by the game, which would lead to less or insufficient perceptual resources available for processing the concurrent game-irrelevant oddball-task. This effect was expected to manifest as a measurable decrease in early ERP component amplitudes. It was also hypothesized that presence would make players react to emotion-eliciting events as if they were real, which would result in more GSR peaks throughout the game while not impacting event response magnitude. ERP components (N1, MMN and SW), GSR peaks/min and response magnitude were compared between the presence groups revealing significant differences in GSR peaks/min and early ERP components of N1 and MMN, but not in GSR response magnitude. The findings support the hypotheses and show that GSR peaks/min, N1 and MMN correlate with presence and have potential as presence indicators.

1. Introduction

Physiological assessment of user experience (UX) in video games through the use of biometrics is a young, but rapidly evolving field within Games User Research (GUR; Nacke, 2018; Charij and Oikonomou, 2015). Psychophysiological measurements have only been used for objective evaluation of subjective gaming experiences within GUR in the last decade (Mandryk and Nacke, 2016), primarily focusing on assessment of emotions and physiological arousal (Yannakakis et al., 2016; Nacke, 2018). The goal of GUR as a field is to create methods, techniques and tools to facilitate data collection and player experience assessment in order to optimize usability and UX in video games (Mandryk and Nacke, 2016). One of the most discussed gameplay experience factors in GUR literature is the phenomenon of presence (Nacke et al., 2010), which has seen a surge in applicability as highly immersive Virtual Reality (VR) technologies develop and become more accessible (Tamborini and Skalski, 2005; Belini et al., 2016), however no standardized objective measure of presence exists as of now (Mandryk and Nacke, 2016).

Presence is generally defined as an overall subjective sensation of "being there" in mediated content to the point where the virtuality of the virtual environment (VE) goes unnoticed and feels like the dominant reality (Barfield et al., 1995; Ijsselsteijn et al., 2000). Given the subjective nature of the phenomenon the most frequently used measurement tool has been retrospective self-report questionnaires (Ijsselsteijn et al., 2000). While these post-questionnaires are considered a reliable and empirically supported way of measuring presence they have the distinct limitation of not being able to provide continuous insight into the potentially varying experience of presence during gameplay (Insko, 2003). Think-aloud protocols have been suggested to combat this disadvantage, however requiring respondents to continuously report their feeling of presence is advocated to be detrimental to the experience of presence by necessitating them to, by definition, not be present in the VE while reporting (Wiederhold et al., 2003). Physiological measures on the other hand are advocated to be continuous by nature and do not require any obtrusive interaction from the respondents. Even though physiological indicators, such as heart rate, Galvanic Skin Response (GSR) and Event-Related Potentials (ERP) have

* Corresponding author.
E-mail address: tst@psy.ku.dk (T. Terkildsen).

https://doi.org/10.1016/j.ijhcs.2019.02.006
Received 7 June 2018; Received in revised form 21 January 2019; Accepted 10 February 2019
Available online 11 February 2019
1071-5819/

been suggested as objective operational measurements of presence, relatively little empirical research is readily available that specifically examines and shows how biometrics relate to the sense of presence, particularly within the context of GUR (Mandryk and Nacke, 2016; Ijsselsteijn et al., 2000; Sadowski and Stanney, 2002).

The purpose of the present study is to help rectify this by conducting an experiment, where all participants play a video game for 25 min while their physiological responses are measured and where their subjective experience of presence evoked by the game stimulus is retrospectively assessed through a self-report questionnaire. The experiment is conducted to further the investigating of GSR as a presence indicator and to further examine the potential of using a more novel Electroencephalography (EEG)-based approach for measuring presence in games through task-irrelevant ERP. These two physiological measures were specifically chosen for further inspection as they have respectively been empirically shown to be capable of reflecting physiological arousal (GSR) and attention allocation (task-irrelevant ERP), which we hypothesize to be essential concepts for operationalizing the psychophysiological measurement of presence. The overall aim of the present investigation is to bring us closer to a standardized objective measure of the experience of presence in video games. To meet this aim two specific overarching research questions will be explored:

Research Question 1: Can GSR be used as an objective indicator of presence?
Research Question 2: Can game-irrelevant ERPs be used as an objective indicator of presence?

These research questions are deemed relevant because a continuous and objective measure of presence would enable UX researchers to test and identify what ‘works’ in practice when it comes to eliciting and increasing presence in media. This would lead to better experiences that would undoubtedly attract more consumers by maximizing this highly desirable attribute of presence (Sadowski and Stanney, 2002).

1.1. Presence in video games

The experience of presence evoked during game play has become an increasingly central component of GUR as developments in technology and design trends make video games more inherently conducive to the sensation of presence by continuously increasing the sensory vividness of haptics, sounds, and visuals and by increasing interactivity (Tamborini and Bowman, 2010). Vividness as a term refers to technology's capability to construct a rich sensory environment that is high in breadth i.e. the number of sensory channels stimulated and high in depth, which is describing the degree of resolution within each sensory channel. Interactivity describes the user's ability to influence the VE (Steuer, 1992). These concepts relate to presence in the sense that humans phylogenetically have developed a strong automatic tendency to accept any incoming stimuli as being non-mediated, unless there is strong counterevidence, which will trigger a reappraisal process (Lee, 2004a). So the higher the level of vividness and interactivity in the VE the less likely it is to be automatically labeled as mediated.

Although many theories of presence exist (Schuemie et al., 2001; Järvinen, 2017), current conceptualizations of presence in academic literature have identified three commonly agreed-upon sub-dimensions: Social presence, self presence & spatial presence (Lee, 2004b; Tamborini and Bowman, 2010). Social presence is a psychological state in which the artificialness of virtual social agents goes unnoticed leading to players interacting with the artificial agents as if they were real. This should result in elicitation of automatic social responses, such as social smiles etc. (Lee, 2004b). Self presence refers to the perceptual state in which players experience their virtual avatar's self as if it was their actual self. Manifesting as a feeling of having game events happening to them and not just happening to an expendable virtual character (Tamborini and Skalski, 2005). The experience of self presence in VEs is generally linked with a sense of virtual embodiment i.e. the feeling of ownership and self-location within a virtual body (Biocca, 1997; Kilteni et al., 2012). Spatial presence (or physical presence as it is also regularly termed) describes the binary (on/off) sensation of feeling physically located in a VE, including feeling and interacting with a VE (and objects in it) as if it was real, resulting in mental capacities being bound by the mediated environment instead of reality (Slater, 2002; Wirth et al., 2007). According to Witmer and Singer (1998) the emergence of presence is dependent on the interplay of both technological and psychological contributing factors. A player needs to selectively attend and be willing to suspend disbelief, which means that they have to be ready to disregard technological shortcomings of the game and ignore inconsistencies, in order to feel involvement and immersion, but a game's technological qualities can understandably make it easier to do so (Christou, 2014). A game's capacity to focus the players’ attention is a central determinant of spatial presence (Fontaine, 1992). Witmer and Singer (1998) advocate that fully focused energy and attention is what leads to the two psychological states of involvement – a form of mental vigilance characterized as being cognitively engrossed and immersion – a feeling of being absorbed by the VE and isolated from other surrounding stimuli (Witmer and Singer, 1988). These two psychological states are considered as the essence of experiencing presence (Tamborini and Skalski, 2005). This suggests that a heightening of one's attention on game events and stimuli within a VE would increase involvement, which in turn would make the illusion of virtual presence more likely to occur (Sadowski and Stanney, 2002). Highlighting the interplay between external and internal contributing factors, Tamborini (2000) argues that increasing technological features such as interactivity and vividness also heightens the user's sense of involvement and immersion resulting in a sense of spatial presence (Tamborini, 2000).

Wirth et al. (2007) have proposed a model of the formation of spatial presence, which expands upon the descriptors of spatial presence presented above and divides the formation process into two major steps. The first step involves the user constructing a mental representation of the space portrayed by the mediated environment by processing multimodal spatial cues and by relating the perceived VE to relevant personal spatial memories and cognitions. This mental model is called the “Spatial Situation Model” (SSM) and it is regarded by Wirth et al. (2007) as a precondition for spatial presence to occur. The process of constructing a SSM requires attention allocation directed towards the mediated environment, which in most cases ensue due to an interplay of media factors that facilitate involuntary attention allocation and user factors, which motivate the individual to selectively direct attention resources towards the VE. Media factors can trigger automatic attention allocation through short-term orienting events, such as when surprising or interesting stimuli is presented. Attention allocation towards the medium should be consistent and should not be interrupted if the constructed SSM is to be maintained, which is why certain media factors, such as vividness and interactivity, are thought to elicit persistent automatic resource allocation. They are advocated to do so by providing a more realistic and congruent stream of information (spatial cues) that the user can continuously adjust and base their SSM on. If a user finds the VE interesting then they may voluntarily direct attention towards the VE, which will facilitate the construction of a SSM, even if there are no attention-catching media factors.

The second step of the model refers to the actual formation of spatial presence, which automatically emerges from the constructed SSM through confirmation of a “medium-as-PERF-hypothesis” (PERF = Primary Egocentric Reference Frame), which is a perceptual hypothesis that continuously investigates whether the user is experiencing the VE (and not the real world) as the PERF, i.e. the dominant reality in which they are located. The probability of the hypothesis being confirmed or rejected depends on the strength of the user's SSM and it depends on whether the VE remains consistent with our perceptual expectations. A strong SSM binds our cognitive resources to the

VE and suppresses processing of real world information, which challenges the user's inherent perception of the non-mediated reality as PERF. Incongruent information or interferences that would usually lead to the “medium-as-PERF-hypothesis” being rejected can automatically and unconsciously be disregarded depending on an individual's suspension of disbelief and trait absorption, which refers to an individual's innate ability to become intensely involved in the VE without much effort. Wirth et al. (2007) argue that with enough trait absorption and suspension of disbelief even a weak “medium-as-PERF-hypothesis” might be sufficient to evoke Spatial Presence. They also maintain that Spatial Presence may occur in cases with low involvement and low suspension of disbelief, if the media factors facilitated by the presence-inducing technology is convincing and attention-triggering enough.

Wirth et al. (2007)’s model is consistent with the presented conceptualization of presence as an attentional resource-dependent experience of “being there” in a mediated environment, which is facilitated by a combination of technological and psychological factors. In this study, the subjective experience of presence conceptualized in this section as the sum of its three subdomains (i.e. Social presence, self presence, and spatial presence) will be measured using a self-report questionnaire “The Multimodal Presence Scale”(MPS; Makransky et al., 2017). MPS has been shown to provide a short (15 items) yet valid subjective measure of presence, based on a definition of the term that is consistent with the multi-dimensional conceptualization provided in this section (Makransky et al., 2017). The participants’ subjective ratings on this retrospective presence scale will be used to divide them into a low and a high presence group. The GSR- and ERP-based measures examined in this study will be compared across these two presence groups and correlated with the subjective measure as a way of directly exploring their potential as operational measures of the experience of presence evoked by a video game.

In the next section on physiological measures, both of the objective measurement methods will be discussed and the rationale for using them will be explicated.

1.2. Physiological measures

1.2.1. Galvanic skin response

GSR, also commonly known as Skin Conductance (SC) or Electrodermal Activity (EDA), reflects the variation in electrical characteristics of the skin in response to autonomic arousal, which triggers the eccrine sweat glands resulting in a measurable increase in skin conductance (Nogueira et al., 2011). Eccrine sweat glands are of the highest density in palms of the hand and the soles of the feet and they are mostly involved in emotional responses as these glands are innervated by the sympathetic nervous system (Dawson et al., 2007).

A raw GSR signal is comprised of two primary components: (1) Skin Conductance Level (SCL) and (2) Skin Conductance Response (SCR). Skin conductance level refers to the tonic baseline that slowly varies (tens of seconds to minutes) within each respondent resulting in a gradually rising or declining SCL depending on factors such as hydration, skin dryness or autonomic regulation. Tonic level can be significantly different between respondents, which means that SCL is rarely informative on its own (Boucsein, 2012; iMotions, 2018a). SCR describes the phasic signal that rides on top of the tonic changes. Phasic SCR has been found to reflect the emotional arousal, which usually occur around 1–5 s after onset of an arousing event stimulus. The phasic component is only a small fraction of the overall GSR signal and it consists of short-lived changes that manifests significantly quicker than the tonic variations resulting in steep inclines, referred to as “GSR Peaks”, which are then followed by slower declines back to the tonic baseline (iMotions, 2018a; Benedek and Kaernbach, 2010).

1.2.1.1. GSR and arousal. GSR measurements have been linearly correlated with arousal (Leon et al., 2007; Chanel et al., 2006; Haag et al., 2004), which has led to it being considered and used as a valuable tool for assessing emotional activity (Mandryk et al., 2006; See Larsen et al., 2010 for a general overview of how psychophysiology, including GSR, has been used to identify emotional states previously).

The initial evidence that substantiated using GSR as an indicator of underlying emotions originated from findings by Ekman et al. (1983) and (Levenson et al. (1990)), who provided evidence for emotion-specific autonomic nervous system (ANS) patterns using heart rate, finger temperature, GSR and forearm flexor muscle tension as physiological measures. They found distinct arousal patterns across these measures for six basic emotions, namely sadness, fear, disgust, anger, surprise and happiness. This led to them suggesting that each discrete emotion is associated with an innate affect response package that serves to mobilize distinct central and autonomic nervous system resources in order to facilitate optimal coping with the challenging events that triggered the particular emotion affect program. The response package for fear involves eliciting adjustments that narrows attention towards the threatening stimulus while providing physiological support for the purposeful behavior of avoidance/flight through an overall increase in arousal and a redirection of blood flow from periphery towards muscles of locomotion (Levenson, 1999, 1992). GSR is a commonly used arousal and emotions measurement tool within the field of GUR (Nacke, 2018) and its various and distinct applications within the field is well-documented across numerous studies (Weber et al., 2009; Poels et al., 2012; Mandryk et al., 2006; Chalfoun and Dankoff, 2018).

1.2.1.2. GSR and presence. The rationale for using GSR as an indicator of presence is advocated to be rooted in its functionality as an emotion measurement tool. Presence was previously defined as the perceptual state in which the virtuality of the VE goes unnoticed inferring that the VE feels real to the player in that moment. According to Lazarus and Folkman (1987) emotional responses are triggered as a result of a two-step automatic appraisal process that unconsciously evaluates whether you feel as if you personally have something at stake and whether you believe that you can effectively cope with the environmental demand in your current state. If you do feel as if you have something at stake and you don't feel like you can cope successfully in your current state, then an appropriate affect response package is triggered for physiological support (Lazarus and Folkman, 1987). As a result, emotional activity will only be elicited if the triggering mediated stimulus is currently perceived to be real, i.e. sense of spatial presence, and of importance to the individual, i.e. sense of self presence.

While the experience of presence or lack thereof theoretically will determine whether an affect program should be evoked or not in a mediated environment, the intensity of the emotional experience primarily depends on the characteristics of the eliciting stimulus in the VE (Brehm, 1999; Sonnemans and Frijda, 1995), such as size and motion (Lee, 2004a). With fear for instance, the emotional intensity will, according to Ekman (2003a,b) and Brehm (1999), depend on how threatening to your well-being the triggering stimulus is automatically perceived to be and on the amount of physiological support deemed necessary to optimally cope with the threatening situation (Brehm, 1999; Ekman, 2003a; Lee, 2004a). The notion that presence does not directly influence arousal intensity is argued to be supported by studies done by Reeves et al. (1993, 1996). One of their studies showed that manipulation of visual fidelity, which is generally thought of as a key contributing media factor of presence (Tamborini and Bowman, 2010; McMahan et al., 2012), did not significantly impact physiological arousal (Reeves et al., 1993). It is suggested that humans have a large tolerance for variance in image fidelity given that the world is commonly observed through a lower fidelity peripheral field of vision, so fidelity is not a salient feature in the automatic stimulus appraisal process (Lee, 2004a). Whereas manipulation of media factors, such as screen size and motion (also associated with presence) that directly influence the perceived characteristics and salience of the stimulus does significantly affect physiological arousal levels (Reeves and Naas, 1996; Detenber and Reeves, 1996).

Overall, it is argued that if the measured physiological response to a mediated environmental demand is equivalent to what would be expected if the emotion-inducing stimulus was non-mediated, then this could serve as an indicator of presence (Slater et al., 2009). This theoretically limits GSR's applications for indicating presence to stress- or emotion-inducing VEs where the physiological response would be significant and apparent if the virtual situation was perceived to be non-mediated. A horror game was therefore selected as experimental stimulus in this study because of its intended and supported ability to elicit the arousal through emotional responses of fear throughout the game.

As mentioned, only a relatively limited amount of studies have explicitly investigated the relationship between arousal, as measured by GSR, and self-reported presence. These studies have generally demonstrated varying results regarding the potential of GSR as an indicator of presence (Meehan et al., 2002; Wiederhold et al., 2003; Drachen et al., 2010). While GSR as a raw physiological signal is objective by nature, the choice of which measurement methodology to use in order to extract the phasic signal from the SCL is deemed subjective. In GUR studies utilizing GSR, this extraction process is commonly performed based on either recordings of participants' GSR signal during a pre-stimulus baseline resting-period, which can then be subtracted from the average stimulus GSR signal or they subtract the GSR amplitude recorded at the beginning of the stimulus from the amplitude measured at the end of the stimulus (Nacke, 2018). It is advocated that the varying findings regarding the relationship between GSR and presence could result from different measurement techniques being utilized across studies and from potential problems with these observed extraction methods in distinct experimental contexts. Mandryk et al. (2006) detail how they have found that arousal resting rates recorded during a pre-stimulus baseline period frequently have been higher than the average game play GSR measures due to factors such as anticipation and nervousness, which artificially inflates the resting baseline and confounds results. It is advocated that this problem would be particularly probable to occur when participants were anticipating playing an unknown horror game. It is also advocated that, due to the relatively long game stimulus exposure time of 25 min in this study, any potential differences in overall averaged GSR signal, even if successfully baseline-corrected, would be obscured or flattened out, since SCRs, as previously mentioned, only make up a tiny portion of the overall signal and because spontaneous emotional responses are short-lived usually only lasting between 2/3 of a second and four seconds (Ekman, 2003b). For these reasons, a peak detection algorithm was chosen as extraction method in this study, since this procedure is done by applying a sliding window median to the raw signal. Using this procedure means that no pre-stimulus baseline period is necessary and that the length of the stimulus exposure time doesn't matter as the current measurement window is baseline-corrected against the window preceding it to extract the phasic signal. The total amount of peaks detected in the phasic signal is operationalized to be an overall measure of how many emotionally arousing events that participant has experienced during their entire gameplay session.

In the present study, the overall GSR measurements of peaks/min detected during the video game stimulus was used as an operational measure of presence given that it was theoretically advocated in the previous sections that the consistency of the presence sensation would determine how many game events would be perceived as non-mediated and thus evoke a peak-inducing emotional response. To corroborate that the GSR activity measured were due to emotional responses and to further explore if presence leads to differences in emotional arousal, two independent arousal self-report instruments were administered: (1) A questionnaire by Rottenberg et al. (2007), which assesses the overall experience of specific emotions, and (2) A self-report scale by Bradley and Lang (1994) that measures general level of felt emotional arousal. An event-specific GSR measure was also included in the study by examining participants' response magnitude to a jump scare game event across the two presence groups. A jump scare is a scare-technique often used in both horror games and films, where an abrupt and unexpected event is triggered causing a sudden visual change, often accompanied by a loud sound. Average GSR peak amplitude, i.e. the intensity of arousal peaks, was used as an overall measure to further examine the connection between presence and response magnitude. These measures were incorporated to explore the potential of response magnitude and overall arousal as possible indicators of presence elicited by a video game.

Based on the advocated mechanical and theoretical relation between GSR, arousal and presence, three hypotheses will be explored.

H_1: It is hypothesized that a high presence group will have significantly more arousing experiences as measured by both phasic GSR peaks per minute (H_{1A}) and self-reported emotional arousal (H_{1B}) than the low presence group. It is proposed that this is a likely result of the participants in the high presence group experiencing spatial presence and self presence for a larger portion of the play session.

H_2: It is also hypothesized that no significant differences will be found between the high and the low presence groups in GSR response magnitude, as explored by investigating both GSR response to a scare event and average peak amplitude throughout the entire gameplay session. The theoretical reasoning behind this hypothesis is that an arguably unexpected scare event would facilitate spatial presence in the moment as an event like this is advocated to involuntarily orient attentional resources to the game through strong media factors as they occur. Considering that the mediated stimuli is the same and is displayed in the same way for both groups throughout the game, no overall or event-specific differences in emotional intensity, as measured by GSR, is expected.

H_3: Given the expected relation between arousal and presence, the last hypothesis of this research question proposes that peaks/min as measured by GSR can be used to indicate presence and predict low or high presence group memberships.

1.2.2. Electroencephalogram and event-related potentials

An EEG measures the gross electrical activity of the brain generated by millions of neurons firing at the same time, which produces a large enough electrical potential that it is measurable along the scalp (Pinel, 2011; Breedlove and Watson, 2013). Unlike the continuous EEG signal, which represents spontaneous brain activity, ERPs are generated as a response to specific events and is usually averaged across many samples in order to obtain a reliable noise-free estimate of the relevant time-locked electrical potentials that are evoked by a selected sensory stimulus (Andreassi, 2007). Averaged sensory-evoked ERPs reflect different stages of stimulus-processing through a series of positive and negative voltage deflections, which are termed as peaks, waves or components, occurring at distinct latencies (time after stimulus onset; Luck, 2014).

1.2.2.1. ERPs and attention. ERP components have frequently been studied and supported as indicators for allocation of cognitive processing resources (Luck et al., 2000; Kok, 1997). The most valuable method for studying attention resource allocation is advocated to be the dual-task paradigm (Karatekin et al., 2004), which involves performing a primary task and a secondary task concurrently (Gosselin and Gagné, 2010). Dual-task paradigms are based on the underlying assumption that there is a limited capacity of cognitive resources available at a given time (Kahneman, 1973; Wickens, 2002). This is an assumption that has been investigated and empirically supported numerous times in literature on divided attention and selective attention (Matlin, 2009). The rationale behind the dual-task paradigm is that performance on the primary task takes priority and utilizes the required attentional resources while performance on the secondary task is dependent on the amount of left-over cognitive capacity. Studies by Lavie (1995) show that performance on a secondary task only decreases if the primary task requires such a high perceptual load that it exceeds processing capacity to the point where no or not enough resources are available for the secondary task. In low load primary conditions performance on the secondary task doesn't decrement. In task-irrelevant probing techniques, which are based on the same underlying rationale, the secondary task is extraneous to the primary task and does not require any response from the participants. An auditory oddball paradigm, where the subject is tasked with distinguishing between frequent and deviant tones (i.e. oddballs) that are continuously presented in random series (Strüber and Polich, 2002), is generally used as the secondary probing task and it is assumed that this extraneous discrimination task, will absorb attentional resources left-over from the primary task, if there are any to spare (Kok, 1997).

The application of task-irrelevant probes as indicators of cognitive resource allocation has been well-documented and established in studies examining mental workload during cognitive tasks performed in mediated or non-mediated environments (Finnigan et al., 2010; Allison and Polich, 2008; Kramer et al., 1995; Miller et al., 2011; Ullsperger et al., 2001). In these workload studies, a wide range of ERP components have been investigated, such as N1, N2, Mismatch Negativity (MMN), P1, P2, P3 and late negative Slow Waves (SW). The significant differences found in the components of N1, MMN and P3 across conditions in these studies were attributed to differences in attention allocation rather than differences in workload or task-difficulty. The P3-component however has been shown to not be pronounced when subjects are not required to consciously respond to the occurrence of oddball tones (Kramer et al., 1995; Pugnetti et al., 1996; Mager et al., 2000; Kober and Neuper, 2012).

N1 is a negative fronto-central component, which generally peaks in the latency range of 100–200 ms after stimulus onset (Luck, 2014; Finnigan et al., 2010). N1 is sensitive to attention (Luck, 2014; Näätänen et al., 2011) and has long been interpreted to reflect orientation of attention and perceptual resource allocation (Luck et al., 1990; Kok, 1997; Hillyard et al., 1973; Hackley et al., 1990). N1 as an early ERP component is similarly assumed to reflect effects of attention on cortical areas that encode elementary stimulus features in order to determine whether the perceived stimulus has the characteristics of a non-target or a target, which might require further cognitive response (Hillyard et al., 1978). Kok (1997) advocates that N1 activation will be strongest for target oddball stimuli, which is attended to and it will decrease with inattention.

The MMN component is particularly elicited in auditory oddball sequences when auditory irregularities, i.e. oddballs, are automatically detected (Näätänen et al., 2007). Activation of MMN will usually prompt conscious perception of the sequence violation due to its involuntary attention-shifting properties. However, MMN and the accompanying conscious perception won't be triggered if the change in ERP stimulus tone is not noticed due to for instance inattention or insufficient perceptual resources available for early stage stimulus discrimination (Näätänen et al., 2011). The auditory MMN component is a deviant-standard difference-waveform that usually occurs fronto-central between 100 and 250 ms after stimulus onset (Luck, 2014).

1.2.2.2. ERPs and presence. The relation between presence and attention has already been theoretically explicated and it is supported in studies investigating attention as a likely candidate measurement that correlates with self-reported presence in dual-task paradigms, where the attentional demands of competing virtual and real world experiences are measured (Darken et al., 1999; Hecht and Reiner, 2006). However, these studies have not measured attention through an objective physiological measure and both experimental tasks in these studies have required actions and responses from the participants, which is advocated to inhibit the experience of presence in a VE.

In the present study, the irrelevant probing technique was used to objectively assess allocation of attentional resources between the concurrent primary video game task and the secondary auditory oddball task. The underlying principle behind using task-irrelevant oddball ERPs as a secondary probing task in a dual-task paradigm to operationally measure presence is that, if a player is present in the VE then, based on the previous conceptualization of presence, all their attentional resources would be allocated towards the game in that moment, which means that the task-irrelevant beeping tones, and changes thereof, would not be perceived or attended to. So in the high presence group, where the primary task is assumed to be high in perceptual load, an overall measurable decrease in both N1 amplitude, elicited by deviant oddball tones, and MMN difference-wave amplitude during gameplay, is expected when compared to a low presence group as a result of perceptual suppression of the unattended secondary oddball task. An inherent advantage of using ERPs as an indicator of presence compared to GSR is that its applications are not theoretically limited to emotion- and arousal-inducing VEs.

Few studies are available that explicitly have investigated the use of task-irrelevant ERPs as an indicator of presence. A study by Kober and Neuper (2012) examined the potential of using resource allocation towards task-irrelevant ERP-tones as an indicator of presence using a VR simulation of a city, where 40 respondents were primarily tasked with navigating this virtual city with the goal of finding the shortest route possible between target buildings. They investigated the ERP components of N1, MMN, and SW and only found significant differences in SW, divided into two latency windows SW1 (400–650 ms) and SW2 (650–900 ms), between a low and high presence group.

Kober and Neuper (2012) propose that the significantly higher SW amplitudes in the low presence group compared to the high presence group found in their study indicated that the low presence group was not as involved in the VE and consequently more attentional resources were allocated to the task-irrelevant tones, which they argue is reflected by the increased SW amplitudes.

A study by Burns and Fairclough (2015) referenced and utilized Kober and Neuper (2012)'s methodology to further investigate the role of SW in immersion by having respondents play a video game, "WipeOutHD Fury", with three different difficulty levels and on different display types. Burns and Fairclough (2015) found that immersion significantly scaled with task-difficulty i.e. the harder the difficulty, the higher level of subjectively reported immersion, which is a finding that has been documented in other studies as well (Ravaja et al., 2004). Their analyses showed that the SW component exhibited a significant decline when game difficulty increased from easy to hard and from easy to impossible. They advocate that SW amplitudes in response to a secondary auditory oddball task is sensitive to challenge-based immersion, which describes a feeling of immersion elicited by task-difficulty on the primary game task (Ermi and Mayra, 2005). Burns and Fairclough (2015) didn't find any SW effects due to display type, which they initially hypothesized would induce sensory immersion due to superior media factors. In the previously cited study by Allison and Polich (2008) they examined workload in video games using ERPs and they also found results indicating that the early slow wave (SW1) amplitude declined as game difficulty increased (Allison and Polich, 2008). Muluh (2011) reviewed ERP components employed in mental arithmetic processing (MAP) studies and found that slow wave activation has been associated with MAP and linked to general task demands (Muluh, 2011). These recent findings demonstrate that SW components are generally thought to reflect task-difficulty, and are not directly linked to attentional resource allocation as suggested by Kober and Neuper (2012).

This present study aims to further investigate the role of N1 and MMN in indicating presence in videos games and even though it's been advocated that the SW component is more directly related to task-difficulty (Rösler et al., 1997; Kok, 1997) than the experience of presence, it will be included in this present study as well to further a discussion on its relevance to indicating presence in games. To keep findings comparable with the other mentioned studies that examine SW in relation to presence in games, the SW component will be divided into SW1 and

Fig. 1. Screenshots from 'Don't Knock Twice' by Wales Interactive. Generally, all four scenes from the screenshots were experienced by the players during the 25 min of gameplay.

SW2 using the latency windows defined by Kober and Neuper (2012). In order to investigate the second and last research question of this article, which involves examining the potential of ERPs as a predictor of presence, five underlying hypotheses will be explored:

H_4: Since N1 activation is thought to reflect orientation of attention and perceptual resource allocation, it is hypothesized that N1 amplitudes elicited by deviant oddball tones extraneous to the game will be significantly decreased in the high presence group compared to the low presence group.

H_5: MMN amplitudes are hypothesized to be significantly less pronounced in the high presence group in comparison to the low presence group. This is expected given that the high presence group is thought to have less perceptual resources leftover for the secondary oddball discrimination task so they won't perceive the change in tone as much as the low presence group.

H_6: Given that SW-components are expected to be related to task-difficulty, it is hypothesized that when comparing the two presence groups, no significant differences will be found in SW1 amplitudes evoked by deviant oddball tones.

H_7: As such, it is also hypothesized that no significant differences will be found in SW2 amplitudes elicited by deviant oddball tones between the groups.

H_8: Lastly, based on the explicated link between attention and presence, it is hypothesized that ERP components can be used to measure self-reported presence and predict low or high presence group memberships.

2. Methods and materials

2.1. Participants and experimental design

A total of 46 healthy adults took part in this study. The participants were students from a large European University that were recruited from the university's social media groups and from flyers handed out on the university campus. All participants gave written informed consent. Out of these 46 initial participants, 12 of them had to be excluded from the study due to excessive EEG/GSR noise artifacts or other technical issues (such as equipment disconnects). The data analysis was therefore based on a sample of 34 participants (23 males and 11 females, age $M = 24.63$, $SD = 5.10$).

To address the aforementioned hypotheses, the 34 participants were administered a dual-task paradigm experimental design, where the primary task was playing a horror game for 25 min and the concurrent secondary irrelevant probing task was an auditory oddball paradigm stimulus.

For the analysis, the participants were divided into a low and high presence group based on a median split of their post-game self-reported presence questionnaire (MPS) ratings (Median = 2.57). This was based on Kober and Neuper (2012)'s methodology in order to facilitate a between-subject experimental design, where the described physiological measurements of interest and self-reported emotional experiences could be compared between the two presence groups. A total of 11 males and six females who had an average MPS score below 2.57 were placed in the low presence group ($n = 17$), whereas 12 males and five females with average ratings above the established median value were categorized as belonging in the high presence group ($n = 17$).

All participants were required to have proficiency in English, since all the materials used in this study were in English. It was also a prerequisite that the participants had normal (corrected-to-normal) vision and no diagnosed hearing problems.

2.2. Materials

2.2.1. Game stimulus

As game stimulus, a 2017 horror game by Wales Interactive called 'Don't Knock Twice' was used (Fig. 1). The game is described by its developers as a "highly-immersive first-person horror game based on a psychologically terrifying urban legend" ("Don't Knock Twice on Steam", 2018). A commercially available game was chosen to enhance the ecological validity of the study and to highlight the practical application of the proposed methodologies.

The actual gameplay involved the participants exploring a haunted house at their own pace searching for a game character by the name of Chloe while encountering multiple scripted jump scare events, such as a bursting window, lightning strike and a painting being flung off the wall. The game was advocated to be a suitable stimulus for experimental use seeing as it did not require or really benefit from any prior video game competency or knowledge as the participants could not lose or die in the game during the experience and the control scheme was easy to learn after a short tutorial. The general game mechanics were also relatively simple as they only involved opening doors, interacting with objects and moving around. Presentation of game sound was done through in-ear headphones. The game's sound levels were thoroughly tested relative to volume of the beeping ERP sounds to make sure that both the game sound and the task-irrelevant beeps played from speakers were clearly audible at all times.

2.2.2. Task-irrelevant ERP stimulus

During a training period and for the entirety of the gameplay session task-irrelevant beeping tones were played through a set of speakers that were situated two meters in front of the participant's position. Speakers were chosen as medium for the beeping tones instead of relaying them into the participants' in-ear headphones (where the game audio was being played) because the tones were supposed to feel external and separate from the gaming experience. The participants did not have to respond to the tones. The participants were told that the tones were extraneous to the game and that they were just there to make sure that the EEG measurements worked properly. The tones were played at a volume of 72 decibel (average) with an inter-stimulus interval of one second. They were randomized between a standard 1200 Hz tone (probability of 0.8) and a deviant 2000 Hz tone (probability of 0.2) with the condition that a deviant tone could not directly follow another deviant tone to avoid issues with oddball stimulus refractory periods. Each tone was presented for 100 ms. Methodological decisions regarding tone-frequency, tone-duration, probability and inter-stimulus interval were guided by the previously mentioned study by Kober and Neuper (2012).

2.2.3. Apparatus

The measurement equipment, questionnaires, training session and the game stimulus was run on a powerful high-end gaming PC (CPU: Intel® Core ™ i7-6850 K @ 3.60 GHz; RAM: 32 GB; GPU: NVIDIA Geforce GTX 1080) and displayed on a 27-inch BenQ ZOWIE XL2720 144 hz gaming monitor at a 1920 × 1080 resolution. A standard mouse and keyboard from Logitech were used by the participants as control devices. The task-irrelevant ERP stimulus application was executed from a separate laptop computer (CPU: Intel® Core ™ i7-4710HQ @ 2.50 GHz; RAM: 16 GB; GPU: NVIDIA Geforce GTX 860 M) during the experiment.

2.2.4. Pre- and post-session questionnaires

The pre-session questionnaire consisted of demographic questions (gender, age etc.) and questions about participants' familiarity with video games and the horror genre (five-point scale from very rarely use to very often use). The post-session questionnaire consisted of instruments measuring self-reported presence, general emotional arousal, emotional-specific intensity and one question regarding perceived awareness of the beeping tones. They are presented here in the order in which they appeared in the post-session questionnaire. MPS by Makransky et al. (2017) was used as instrument for measuring subjective presence ratings. The MPS consists of three previously mentioned underlying presence dimensions: Social, self and spatial presence. Each subdomain is probed by five specific items and they are all rated on a Likert scale from 1 ('strongly disagree') to 5 ('strongly agree'; Makransky et al., 2017). The arousal scale from the Self-Assessment Manikin (SAM) by Bradley and Lang (1994) was used to evaluate the participants' overall level of emotional arousal elicited by the game stimulus. SAM is a non-verbal pictorial assessment instrument that measures arousal on a nine-point scale with one being the lowest amount of arousal and nine being the highest (Bradley and Lang, 1994). For evaluating the participants' overall emotional experiences of specific emotions in response to the gameplay stimulus, a Post-Film questionnaire developed by Rottenberg et al. (2007) was utilized after being slightly adapted to fit the use of a video game as the eliciting stimulus instead of a film. This questionnaire retrospectively measures the overall felt emotional experience across twenty different emotional categories on a nine-point scale from 0 (none) to 8 (extreme). However, in this study only the emotion ratings for fear was used in the analysis as it is advocated to be the most relevant feeling when examining the affective response to a horror game stimulus. The post-session questionnaire also contained a question probing the respondents' thoughts about their overall perceived awareness of the beeping stimulus. This question was scored on ten-point scale with 1 denoting "not at all" and 10 being "completely".

2.3. Procedure

When the participants arrived at the lab, they were first asked to read and sign a consent form. The participants were told that the study investigated the physiology of horror games. When the consent form was signed the participants were fitted with the psychophysiological measurement equipment. The data recording quality of the equipment was then tested before asking the participants to fill out the pre-session questionnaire. After having filled out the pre-session questionnaire the participants were then asked to sit and calmly look at a white cross on the monitor in front of them, while the task-irrelevant ERP beeping stimulus was played for three minutes. This served as a 'training session' for the ERP portion of the study, where the participants could be unconsciously familiarized with the beeping tones as recommended by Kramer et al. (1995). After this, the participants were introduced to the game controls and once they felt comfortable with them, the game and the ERP stimulus was started and the participants were asked to play for twenty-five minutes. Once the twenty-five minutes had gone, the participants were lastly asked to immediately fill out the post-game questionnaires. The experimenter left the room each time after instructions were provided for the following segment, so that the participants were alone in the room while experimental tasks were performed. Each participant was compensated with a gift card valued at 100 Danish crowns (approximately 16 USD) upon completion.

2.4. Data collection and analysis

The iMotions Biometric Research Platform (iMotions, 2018b) was used to setup the experimental design and to collect the data from the physiological measurement equipment. SurveyMonkey was used to collect responses from both the pre- and post-session questionnaires; however, it was integrated into the iMotions platform for a seamless transition between the game-stimulus and the post-game questionnaires. IBM SPSS Statistics version 25 was used for statistical analysis.

2.4.1. GSR data collection and analysis

Raw GSR measurements were performed using a Shimmer3 GSR hardware module that records electrodermal activity at a sampling rate of 128 Hz through a pair of bipolar Ag/AgCl electrodes. To avoid movement interference or artifacts from pressing buttons on the keyboard during gameplay the electrodes were placed on the left central wrist with an inter-electrode distance of three centimeters (measured from center to center) instead of the more common finger-based sensor location. The reason why the central wrist was specifically chosen as the most optimal recording site in this experiment was due to a study done by van Dooren et al. (2012). Their findings showed that the central wrist was the location on the arm that correlated the highest with the traditionally preferred sensor placement on the distal phalanges of the fingers (Scerbo et al., 1992).

A jump scare event was examined in order to investigate the event-related GSR response magnitude between the presence groups. The event happened by the end of the designated play time, where the player is walking down a barely lit hallway, when a window abruptly

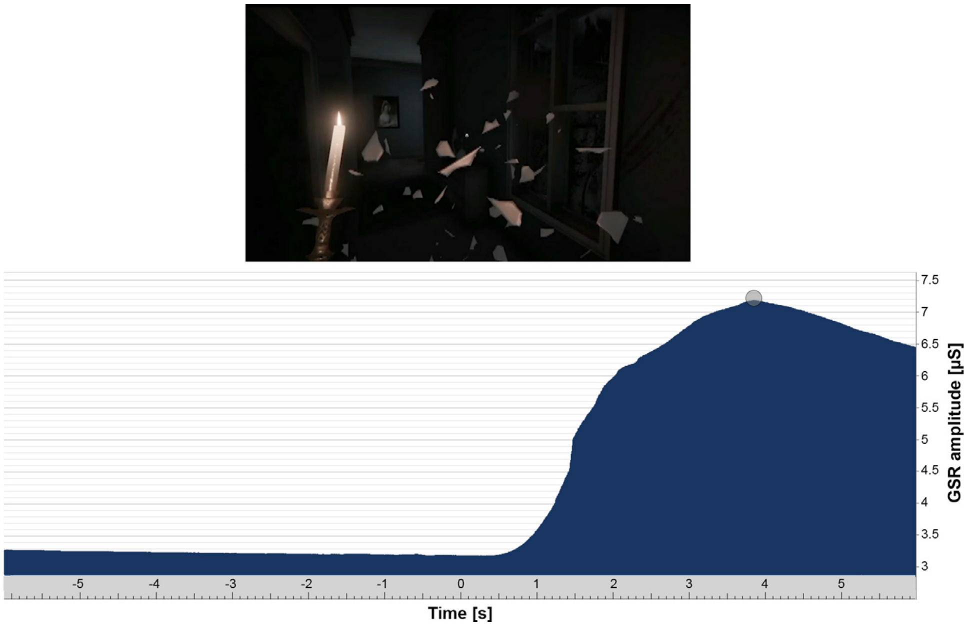


Fig. 2. Visual example of an event-related GSR peak following the selected scare event (onset: 0 s). The baseline window ranges from -6 s to 0 s and the response window ranges from 0 s to 6 s. The peak amplitude is marked with a grey circle.

bursts next to them sending shards of broken glass flying towards their virtual avatar. The event was manually segmented in the participants' data recordings as a 12 s window with six seconds before and six seconds after stimulus onset. The event-related response magnitude is calculated and baseline-corrected by subtracting a baseline value, which is an average of the GSR signal from the six seconds prior to stimulus onset, from the largest GSR value found in the six second window following stimulus onset (Fig. 2).

Some participants didn't make it to the event during the allotted play time, which means that the analysis of this event is done with a sample of 29 participants (Low presence group = 16, High presence group = 13). GSR peaks were calculated by running a peak detection algorithm that applies a sliding window median filter [−4 s; +4 s] to the raw GSR signal in order to extract the faster emotional related changes in phasic data from the underlying tonic signal. A low-pass filter of 5 Hz is then applied to the extracted phasic data to remove line noise. Peaks are then detected in the phasic data by identifying onsets (> 0.01 µS) and offsets (< 0 µS) as a way to define peak regions. The maximum amplitude within a region is categorized as a peak if it is larger than the set peak amplitude threshold, which in this case was set at 0.005 µS above the GSR value at onset. A signal jump threshold of 0.01 µS was also used as an artifact rejection method within the peak detection algorithm. The purpose of this threshold was to reject false positive peaks due to sudden jumps in the data, such as those that could arise from movement artifacts etc., if they were not picked up by the previous filters. After conducting the peak detection algorithm the amount of peaks for all 34 participants were divided by the exact time that each participant spent playing the game in order to produce a comparable GSR peaks/min-measure. The amplitudes of each GSR peak found with the described peak detection algorithm was averaged for each presence group as a measure of overall peak intensity throughout the game.

2.4.2. EEG data collection and analysis

The EEG data was collected and digitalized at 256 Hz using Advanced Brain Monitoring (ABM) X-10, wireless 9-channel EEG system with a linked mastoid reference. Nine Ag/AgCl electrodes were spread across the scalp, located at Fz, F3, F4, Cz, C3, C4, POz, P3 and P4, in accordance with the international 10–20 electrode placement system. Conductivity gel from Synapse was applied to the electrodes and electrode impedances were kept below 10 kΩ. To ensure accurate collection of the time-locked and phase-locked EEG data (ERP), the onset of the auditory beeping stimuli was marked and synchronized using an ABM External Sync Unit in combination with a Cedrus StimTracker.

The data pre-processing and analysis was performed in EEGlab and ERPlab while being blind to what presence group each participant would belong to. The first pre-processing step was to apply a basic FIR bandpass filter from 0.01 Hz to 30 Hz to the continuous EEG data. After this, a visual inspection of the continuous data was conducted in order to reject extreme noise before running an ICA-based artifact correction procedure to remove components containing ocular artifacts, such as blinks and horizontal eye movement. The epochs were then extracted from the continuous data to form frequent (standard tone) and infrequent (deviant tone) events ranging from – 100 to 900 ms after stimulus onset. The 100 ms prior to stimulus onset was used as a baseline for the following ERP signal. Each participants had approximately 1500 epochs (300 deviant ERPs and 1200 standard ERPs) based on the set game exposure time and probability of standard and deviant tones. Epoch extraction facilitated the use of epoch-based automatic artifact detection approaches in ERPlab that aim to reject noisy epochs due to for instance muscle activation. The first approach applied was a maximum voltage threshold of ± 120 µV. Second step was to administer a 200 ms Moving Window Peak-to-Peak threshold with the rejection criteria of ± 50 µV (using window steps of 100 ms). These two artifact detection measures led to an averaged epoch rejection percentage of 16.22%.

All the previously mentioned a priori theoretically- and empirically-guided latency windows of interest were adjusted based on visual inspection of the grand averaged signals from across all participants in order to identify the exact time windows where the peaks and periods of interest occur within this experimental setup. Measures of mean area amplitude were therefore derived from averaging the amplitude of the ERP signal for the following latency windows: N1 = 120–200 ms, MMN = 150–200 ms, SW1 = 400–650 ms, SW2 = 650–900 ms. MMN is a difference wave that is analyzed by subtracting the deviant ERP amplitude from the standard ERP amplitude. Overall, no hemispheric differences were found during preliminary assessment of the laterally placed electrodes, so only midline electrodes Fz, Cz and POz were considered in the further analyses.

Table 1
Results of Independent-samples *t*-tests comparing presence scores between the two presence groups.

	Low presence		High presence					
	M	*SD*	*M*	*SD*	*t*	*Df*	*Sig. (2-tailed)*	*d*
Presence (Total)	1.74	0.42	3.16	0.41	−9.94	32	.000**	3.42
Self-Presence	1.35	0.41	2.51	0.59	−6.60	32	.000**	2.28
Spatial presence	2.14	0.72	3.62	0.60	−6.54	32	.000**	2.23
Social presence	1.72	0.66	3.34	0.87	−6.13	32	.000**	2.10

*$p < .05$.
** $p < .001$.

3. Results

The difference in self-reported presence scores between the low (participants scoring below 2.57 on the MPS) and high (participants scoring above 2.57 on the MPS) presence group was examined using independent samples *t*-tests. The high presence group scored significantly higher in overall presence score and in all three subdomains in comparison to the low presence group (Table 1). The differences all lead to very sizable effect sizes, which indicates that the two presence groups are very different in overall sense of presence, which is deemed advantageous for the following investigations of research question 1 and 2.

Two independent-samples *t*-test were conducted to investigate if there were between-group differences in familiarity with video games and the horror genre in general as reported on the pre-session questionnaire, before investigating the two research questions. No significant difference between the low ($M = 3.71$, $SD = 1.21$) and high ($M = 3.53$, $SD = 1.07$) presence groups were found in familiarity with the video games ($t(32) = 0.45$, $p = .655$, $d = 0.16$) and no significant difference was found between the low ($M = 2.16$, $SD = 0.57$) and high ($M = 2.09$, $SD = 0.37$) presence groups in familiarity with the horror genre ($t(32) = 0.43$, $p = .671$, $d = 0.15$). This indicates that the participants in both groups did not differ significantly in games and horror familiarity prior to this experiment's horror game experience.

3.1. Research question 1: can GSR be used as an objective indicator of presence?

In these following subsections research question 1 will be investigated through individual examination of the associated hypotheses.

3.1.1. Hypothesis 1: significantly more arousing experiences in the high presence group than the low presence group

This hypothesis was explored through two sub-hypotheses. Hypothesis 1A examines the between-group differences in the objective measure of GSR peaks/min, which reflects the amount of arousing events experienced during the play session. Hypothesis 1B investigates the difference in participants' self-reported experience of overall emotional arousal and experience of fear throughout the game between the two presence groups.

3.1.1.1. Hypothesis 1A: significantly more GSR peaks/min in the high presence group when compared to the low presence group. An independent-samples *t*-test comparing the amount of GSR peaks/min during gameplay in the low and high presence group was conducted. The results show that the groups differ significantly with the participants in the high presence group ($M = 7.42$, $SD = 4.63$) having more GSR peaks/min than participants in the low presence group ($M = 4.51$, $SD = 2.46$); $t(32) = -2.29$, $p = .029$, $d = 0.79$.

3.1.1.2. Hypothesis 1B: significantly more self-reported overall emotional arousal experienced during gameplay in the high presence group when compared to the low presence group. Two independent-samples *t*-tests were performed to compare the overall self-reported emotional arousal and the overall experience of fear between the two presence groups. The findings show that the participants in the high presence group ($M = 5.41$, $SD = 2.27$) reported feeling significantly more overall fear throughout the game stimulus than did the respondents in the low presence group ($M = 3.47$, $SD = 2.37$); $t(32) = -2.44$, $p = .020$, $d = 0.84$. The high group ($M = 7.06$, $SD = 1.56$) also reported feeling more overall arousal as measured by SAM than did the low group ($M = 4.76$, $SD = 2.05$); $t(32) = -3.68$, $p = .001$, $d = 1.26$.

3.1.2. Hypothesis 2: no significant differences in GSR response magnitude between presence groups

As expected, no significant differences were found in event-specific GSR response magnitude between the low ($M = 1.26$, $SD = 1.45$) and high ($M = 1.37$, $SD = 1.42$) presence group; $t(27) = -0.21$, $p = .838$, $d = 0.08$. This suggests that a jump scare event elicit approximately the same arousal response across low and high sensations of presence. The examination of average amplitude of peaks detected during the gameplay session between low ($M = 0.12$, $SD = 0.10$) and the high ($M = 0.13$, $SD = 0.09$) presence group showed no significant differences; $t(32) = -0.32$, $p = .752$, $d = 0.11$. This reflects that the overall experienced intensity of arousal-inducing game events didn't differ between presence groups.

3.1.3. Hypothesis 3: the GSR-based peaks/min-measure can predict the experience of presence in video games

A significant correlation was found between the participants' MPS total presence score and GSR peaks/min, $r = 0.41$, $p = .016$. Significant correlations were also found between GSR peaks/min and presence scores for two of the three sub-dimensions of the MPS: Self-Presence, $r = 0.41$, $p = .017$ and spatial presence, $r = 0.38$, $p = .027$. No significant correlation was found between the sub-dimension of social presence and GSR peaks/min, $r = 0.30$, $p = .090$. A standard linear regression was performed in order to investigate the relationship between GSR peaks/min and presence. This method was conducted to determine how well GSR peaks/min can account for the variance in overall presence score. A significant regression equation was found when predicting presence based on GSR peaks/min ($F(1,32) = 6.49$, $p = .016$), with an R^2 of 0.17 indicating that GSR can explain 17% of the variance in overall presence score. As a way to further explore GSR's potential for indicating or predicting presence, a linear discriminant analysis (LDA) was conducted as a way to test whether GSR could reliably be used to divide participants into their respective presence groups. In other words, this classification technique examined the degree to which the produced discriminant function could predictively assign individual participants into either the low or high presence group based solely on GSR peaks/min as predictor variable ($F(1,32) = 5.25$, $p = .029$; Wilks' $\lambda = 0.86$; Canonical Correlation = 0.38; $x^2(1) = 4.78$, $p = .029$). The significant discriminant function based on GSR peaks/min allowed for accurate classification of 70.6% (Low presence = 76.5%, High presence = 64.7%; Cross-validated = 70.6%) of participants in agreement with their original self-report-based presence group membership. Classification results are shown in Table 2.

Table 2
Classification results[a, c] for GSR-based LDA.

		Predicted group membership		
		Low presence	*High presence*	*Total*
Original count	Low presence	13	4	17
	High presence	6	11	17
(%)	Low presence	76.5	23.5	100
	High presence	35.3	64.7	100
Cross-validated count[b]	Low presence	13	4	17
	High presence	6	11	17
(%)	Low presence	76.5	23.5	100
	High presence	35.3	64.7	100

[a] 70.6% of original grouped cases correctly classified.
[b] Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
[c] 70.6% of cross-validated grouped cases correctly classified.

3.2. Research question 2: can Game-irrelevant ERPs be used as an objective indicator of presence?

Research question 2 was firstly investigated by conducting separate independent-samples *t*-tests comparing the mean peak area amplitudes of N1, MMN, SW1 and SW2, measured from midline electrodes Fz, Cz and POz, between the low and high presence group. This was done to investigate hypotheses four through seven. The results of these *t*-tests are briefly summarized in the next sections and illustrated in Fig. 3, which depicts the groups' average N1, SW1 and SW2 amplitudes elicited by deviant tones, and in Fig. 4, where the difference wave MMN is illustrated along with the standard and deviant tone ERPs for both groups. In both figures, the previously discussed latency windows of interest are highlighted. The mean amplitudes of the selected components were computed for each participant and for each midline electrode. They were then separated and averaged depending on presence group membership. The last hypothesis of this research question, H8, builds upon the results of the investigations of the other hypotheses underlying this research question in order to investigate the proposal that task-irrelevant ERPs can predict the self-reported experience of presence and predict the assigned presence group memberships in this study.

An independent-samples *t*-test revealed that participants in the low presence group ($M = 3.53$, $SD = 1.33$) reported being significantly more aware of the beeping sounds overall than the high presence group did ($M = 2.53$, $SD = 1.28$); $t(32) = 2.24$, $p = .033$, $d = 0.77$.

3.2.1. Hypothesis 4: N1 mean amplitudes will be significantly decreased in the high presence group compared to the low presence group

Significant differences in mean N1 peak amplitude elicited by deviant tones were found across the three selected midline electrodes when comparing the high and low presence groups (See Table 3). While the N1 peak amplitude was largest over electrode position Cz, followed by Fz and lastly POz (Fig. 3), the effect size was most pronounced at POz and second largest over Cz and smallest over Fz. The negative N1 peak amplitudes were overall significantly decreased over all electrodes for the high presence group when compared to the low presence group.

3.2.2. Hypothesis 5: MMN mean amplitudes will be significantly decreased in the high presence group compared to the low presence group

Significant differences and large effect sizes between the high and low presence groups were also found in mean MMN peak amplitude for the examined electrodes (See Table 3). The MMN findings follow the same pattern as the N1 results with the MMN peak amplitude being most pronounced over Cz, then over Fz and lastly over POz (Fig. 4). The largest effect size was found over POz and second most prominent effect over Cz and smallest over Fz. Generally, the high presence group had significantly smaller negative MMN amplitudes across all midline electrodes than the low presence group did.

3.2.3. Hypothesis 6: no significant difference in SW1 between presence groups

No significant differences were found when comparing the SW1 amplitudes elicited by deviant tones between the two presence groups (See Table 3).

3.2.4. Hypothesis 7: no significant difference in SW2 between presence groups

The high and low presence groups showed no significant differences in mean SW2 amplitude elicited by deviant tones over any of the electrode sites (See Table 3).

3.2.5. Hypothesis 8: game-irrelevant ERP components can be used to predict the experience of presence in video games

Significant correlations were found between the participants' total MPS presence scores and the early ERP components of N1 and MMN across all electrodes. This was also the case for the three presence-subdomains, except the relationship between N1 Fz and social presence, which did not reach significance ($p = .073$). SW1 and spatial presence had a positive correlation across all electrodes, but it didn't quite reach significance over POz ($p = .055$). No other significant correlations were found. The correlation matrix can be viewed in Table 4.

A standard multiple regression was conducted to examine how well the ERP variables could explain the variance in overall reported presence. To limit issues with multicollinearity only the examined ERP variables for POz were used (N1, MMN, SW1, and SW2), since this was the electrode that generally showed the most pronounced effects across the variables and correlated most highly with presence overall. This led to a significant regression equation ($F(4, 29) = 6.84, p = .001$), with an R^2 of 0.49 suggesting that ERP variables can be used to explain 49% of the variance in self-reported presence ratings. The predictive capabilities of game-irrelevant deviant ERPs were further analyzed by running a LDA with the ERP variables that significantly differed between the two presence groups (N1 and MMN) as predictor variables, but for the same reasons as mentioned above with the multiple regression analysis only POz was included. This revealed a significant discriminant function (Wilks' $\lambda = 0.65$; Canonical Correlation = 0.59; $x^2(2) = 13.40$, $p = .001$) that successfully predicted presence group membership for 76.5% of participants (Low presence = 70.6%, High presence = 82.4%; Cross-validated = 73.5%). Classification results are presented in Table 5.

4. Discussion

The present study has investigated the potential use of GSR (research question 1) and ERPs (research question 2) as physiological measures of presence in a horror video game. The participants in this study were split into a low or high presence group depending on their ratings on the administered subjective presence questionnaire. This produced two presence groups that were significantly different with very prominent effect sizes in overall presence ($d = 3.42$) and in the subdomains of self presence ($d = 2.28$), spatial presence ($d = 2.23$), and social presence ($d = 2.10$). No prior differences in familiarity of the horror genre or video games in general were found between the groups. In the following sections the results from the exploration of each research question and the methodological, empirical and practical implications hereof are discussed.

4.1. GSR as an objective indicator of presence

To explore research question 1, the three underlying hypotheses were examined by comparing the relevant GSR metrics between these two presence groups.

Firstly, it was hypothesized that the high presence group would

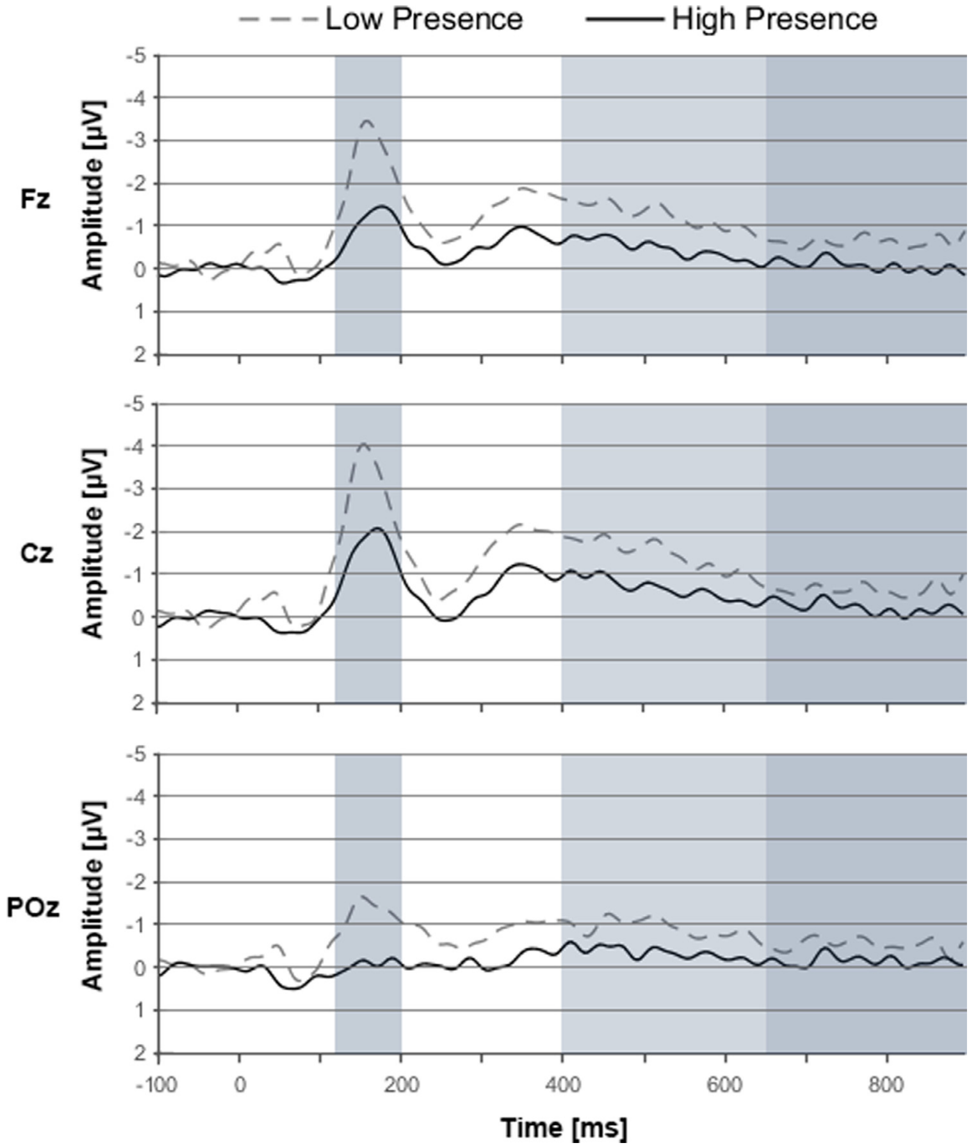


Fig. 3. The figure illustrates the grand average event-related potentials elicited by deviant tones from electrodes Fz, Cz and POz for both presence groups. The marked areas indicate different latency windows of interest; N1 = 120–200 ms, SW1 = 400–650 ms, SW2 = 650–900 ms measured after deviant stimulus onset.

have significantly more arousing experiences as measured by phasic GSR peaks/min and self-reported emotional arousal than the low presence group. The results show that the high presence group had significant more GSR peaks/min than the low presence group, which suggests that there is a connection between the subjective experience of presence and GSR peaks/min. To underpin and understand the basis of this significant and approximately large sized difference in GSR peaks/min ($d = 0.79$) across presence groups, the participants' self-reported overall arousal and emotional experience of fear was examined. The investigation showed that the high presence group rated significantly higher overall levels of fear ($d = 0.84$) and arousal ($d = 1.26$) than the low presence group did. This corroborates the differences found by the objective GSR peaks/min measure and together it indicates that the high presence group experienced significantly more emotional arousing events throughout the gameplay session than the low presence group did, which directly supports the first hypothesis of this research question and it underpins the general theoretical relationship between GSR, arousal and presence, which were explored in the earlier sections of this paper.

It was also hypothesized that no significant differences would be found between the presence groups in GSR response magnitude. This hypothesis was first investigated by comparing the averaged and baseline-corrected physiological arousal levels elicited in response to a jump scare event between the groups. Secondly, the hypothesis was examined by averaging the peak amplitude for all detected peaks in both groups. Both these exploratory steps showed no significant differences between the groups, which supports the hypothesis that arousal intensity is stimulus-dependent and not intensified by the experience of presence. This finding has methodological implications for future empirical work on GSR measurements of presence as it indicates that it can be problematic to use distinct experimental conditions for manipulating the evoked sense of presence, such as comparing a desktop (lower presence) and a VR version (higher presence) of the same VE stimulus, if these different presence conditions would influence the characteristics of the emotion-inducing event stimuli in the game. For instance with the scare event examined in this study, where a window bursts sending shards of glass flying at the player, these shards of glass would arguably appear larger when presented on a VR Head-mounted display that is closer to the eyes and has a wider field of view than on a regular desktop monitor. Lee (2004a) advocates that object size is one of the most primitive cues humans use to phylogenetically base the amplitude of their reaction on and the bigger the object the

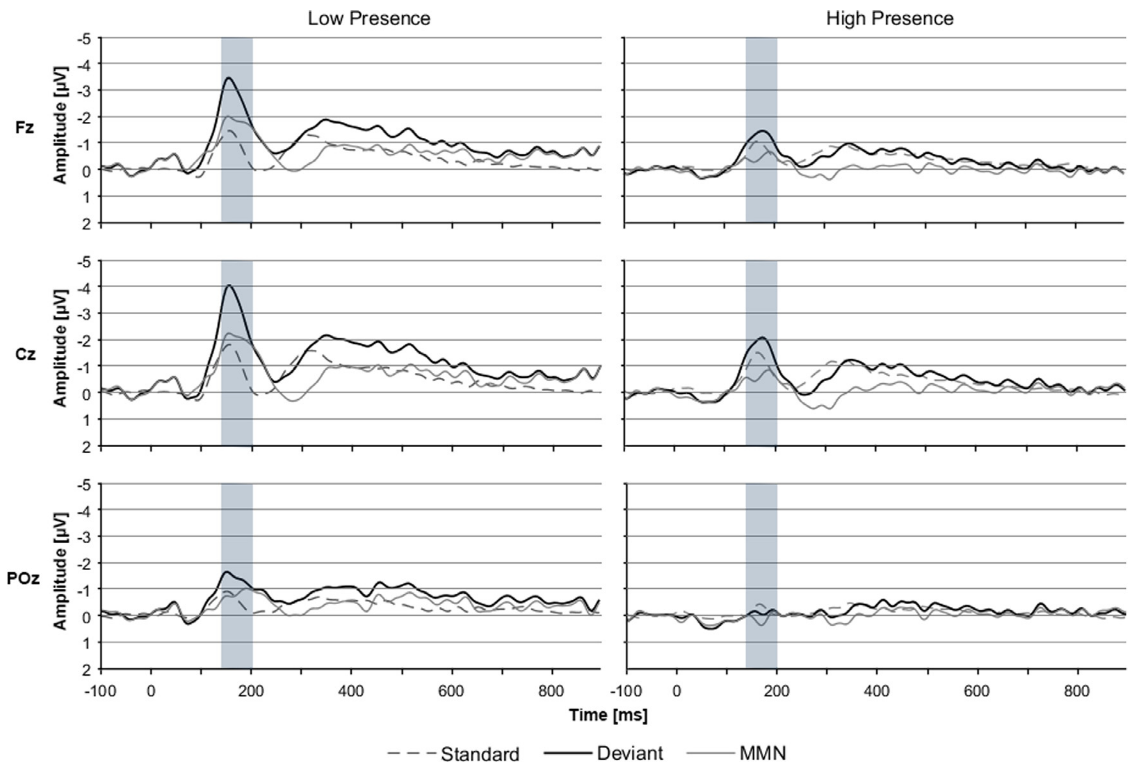


Fig. 4. Grand average ERPs for each group and each midline electrode. The figure displays the response to the standard and deviant tone and the MMN, which was analyzed by subtracting the deviant ERP response from the standard ERP amplitude values. The area highlighted illustrates the latency window of interest for MMN (150–200 ms after stimulus onset).

Table 3
Results of Independent Samples *t*-tests of N1, MMN, SW1 and SW2 across the three midline electrodes comparing low and high presence groups.

	Low presence		High presence					
	M	*SD*	*M*	*SD*	*t*	*df*	*Sig. (2-tailed)*	*d*
N1 Fz	−2.49	1.78	−1.09	1.05	−2.79	25.91	.010*	0.96
N1 Cz	−3.03	1.73	−1.42	1.00	−3.31	25.67	.003*	1.14
N1 POz	−1.30	1.14	0.01	0.71	−4.02	32	.000**	1.38
MMN Fz	−1.74	1.60	−0.48	1.06	−2.71	32	.011*	0.93
MMN Cz	−2.05	1.68	−0.56	1.18	−2.99	32	.005*	1.03
MMN POz	−0.93	0.96	0.24	1.03	−3.45	32	.002*	1.18
SW1 Fz	−1.14	1.26	−0.50	1.02	−1.64	32	.111	0.56
SW1 Cz	−1.35	1.35	−0.71	1.07	−1.53	32	.135	0.53
SW1 POz	−0.86	1.24	−0.35	0.82	−1.42	32	.167	0.49
SW2 Fz	−0.61	1.31	−0.05	1.12	−1.34	32	.188	0.46
SW2 Cz	−0.61	1.28	−0.23	1.15	−0.91	32	.372	0.31
SW2 POz	−0.50	1.24	−0.15	1.08	−0.88	32	.383	0.30

* $p < .05$.
** $p < .001$.

Table 4
Pearson Correlation matrix showing correlations between ERP-variables and presence, including subdomains.

	Total presence	Self presence	Spatial presence	Social presence
N1 Fz	.462**	.388*	.503**	.311
N1 Cz	.502**	.399*	.469**	.422*
N1 POz	.609**	.499**	.574**	.499**
MMN Fz	.513**	.434*	.539**	.360*
MMN Cz	.563**	.449*	.537**	.464**
MMN POz	.625**	.524**	.508**	.576**
SW 1Fz	.271	.309	.368*	.062
SW1 Cz	.245	.214	.354*	.082
SW1 POz	.266	.264	.332	.114
SW2 Fz	.220	.314	.271	.031
SW2 Cz	.145	.134	.225	.029
SW2 POz	.212	.186	.244	.125

* $p < .05$.
** $p < .001$.

more likely it is that it will become a challenge to, or an opportunity for, survival. He argues that this innate size-judgment module is still automatically applied when seeing virtual objects on a screen, which results in larger screens yielding more arousal even if the content is the same (Lee, 2004a). This suggests that a VR version would result in the perception that the larger stimulus was more threatening and therefore would induce more intense measurable arousal in response. It is advocated, that this change or increase in event-specific response arousal between such media conditions could easily be misinterpreted as a function of the experience of presence, whereas the findings of this study indicate that self-reported presence does not directly influence response magnitude to game events when the media format and the characteristics of the event stimuli remain consistent across presence groups.

Together the support for the first two hypotheses empirically reinforces this paper's theoretical conceptualization of presence as a moment-to-moment binary experience that determines whether a current mediated event is perceived as real or not, while not influencing the response intensity of an event once it has been appraised as non-mediated. The high presence group is consequently corroborated, based on the GSR measures, to have experienced presence more reliably and frequently throughout the game as indicated by their larger amount of GSR peaks/min resulting from having treated more game events as being real than the low presence group, where the sensation of presence is suggested to have been less consistently evoked. This underpins the practical importance of being able to objectively measure presence as the design goal of entertainment media products, such as horror games or movies, is often to evoke emotional responses and in this study presence is supported to be a central determinant of whether mediated

Table 5
Classification results[a, c] for ERP-based LDA.

		Predicted group membership		
		Low presence	*High presence*	*Total*
Original count	Low presence	12	5	17
	High presence	3	14	17
(%)	Low presence	70.6	29.4	100
	High presence	17.6	82.4	100
Cross-validated count[b]	Low presence	11	6	17
	High presence	3	14	17
(%)	Low presence	64.7	35.3	100
	High presence	17.6	82.4	100

[a] 76.5% of original grouped cases correctly classified.
[b] Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
[c] 73.5% of cross-validated grouped cases correctly classified.

events throughout the virtual experience elicit the intended emotionally arousing response. As mentioned the high presence group experienced more peak-inducing events than the low presence group during the allotted play time, which indicate that there have been some events that were perceived as non-mediated in the high presence group and mediated in the low group. It is possible that those events, which inconsistently elicited an arousal response between the groups, might have been less attention-orienting and less unexpected in nature and that these types of events would be better indicators of presence if they can be reliably identified. Future work should examine different types of game events and events of varying intensities to explore the possibility that a certain type of less salient events might be better suited as a presence indicator than the intense and attention-orienting event investigated in this study. However as mentioned, this study also examined the average amplitude of peaks and found no difference between the groups, so the findings suggest that event-based response intensity in general is not a promising indicator of the overall experience of presence.

The third hypothesis utilized the significant differences in GSR found between the two presence groups and deals with the main part of this research question regarding GSR's predictive capabilities when it comes to the experience of presence. It was hypothesized that GSR could be used to indicate presence and predict participants' presence group memberships. A significant positive correlation ($r = 0.41$) was found between GSR peaks/min and presence indicating a linear relationship of moderate strength between the objective GSR measure and the subjective experience of presence. GSR peaks/min had a moderate positive relationship between the two presence subdomains of spatial presence and self presence, however no significant correlation was found between GSR peaks and social presence. This is consistent with what was theoretically expected based on Lazarus and Folkman (1987)'s automatic emotion appraisal process that was previously described, where both the sensation of self presence and spatial presence were inferred to be necessary prerequisites in order for emotional activity to be elicited, whereas social presence was not required. This finding is also in line with previous empirical research, which demonstrates, using the rubber hand illusion, that significant GSR peaks in response to threats towards the perceived self-location only occur if a sense of self presence and embodiment is experienced (Armel and Ramachandran, 2003; Ehrson et al., 2007; Yuan and Steed, 2010). A significant linear regression was also found as part of the investigation of this hypothesis. It showed that GSR peaks alone could significantly explain 17% of the overall variance in self-reported presence. These measures supported the hypothesis that GSR had significant predictive capabilities in regards to presence, but as a test of the practical applications of this, a LDA was conducted to explore how reliably GSR could predict the participants' self-report-based presence group memberships in this study. The discriminant function developed based on GSR peaks/min allowed for accurate classification 70.6% of the time, which is higher than chance. These findings combined are advocated to generally support the hypothesis by underpinning that there is a relationship between GSR peaks/min and presence, which shows that GSR peaks/min has capabilities for predicting and indicating overall presence in a horror video game.

The findings of this study suggest that the GSR measurement of overall peaks/min can be used as an appropriate presence indicator in very emotionally arousing VEs, such as horror games. The findings also suggest that response magnitude to an attention-orienting game event and overall average peak amplitude cannot be supported as potential indicators of presence. It is possible that unexpected events such as jump scares are not good indicators of presence because they inherently demand attention allocation to the VE, which is argued to momentarily facilitate spatial presence. It is also possible, given that a previous study has found that GSR generally increases with presentation of an unexpected stimulus and decreases with habituation (Yaremko et al., 1970), that intense emotional events are not good indicators of presence because they generally increase arousal, which might obscure any differences caused by presence. The results of this study has also suggested that the GSR-based measure of peaks/min is significantly linked to the experience of self and spatial presence, but not social presence, which is an insight that has not been empirically demonstrated. Generally, the results derived from the exploration of research question 1 has practical implications for UX researchers since it underpins that the GSR measurement of peaks/min can serve as an indicator of presence, and thereby overall quality, of media products where the designed VEs aim to evoke emotion-arousing responses through facilitation of spatial and self presence. The results also generally support the assumption that presence in arousal-inducing VE can be operationalized as the amount of emotionally arousing game events experienced as non-mediated, which is advocated and supported to be measurable by the proposed GSR peaks/min metric.

4.2. Game-irrelevant ERPs as an objective indicator of presence

To investigate the second research question of this study, distinct ERP components (N1, MMN, SW1 and SW2) elicited by task-irrelevant tones during gameplay were explored as possible indicators of presence.

It was hypothesized that N1 amplitudes elicited by deviant tones would significantly decrease in the high presence group compared to the low presence group. This hypothesis was supported as significant and large to very large differences were observed in N1 mean peak amplitudes across all midline electrodes (Fz, Cz and POz) when comparing the two presence groups with the high presence group having the lowest negative N1 peak amplitudes overall. The high presence group also had significantly lower MMN difference-waveform amplitudes across all electrodes than the low presence group did leading to a large effect size difference, which supports the second hypothesis of this research question. The N1 component has as previously mentioned been demonstrated to be sensitive to attention with amplitudes increasing with attention and decreasing with inattention to the oddball tones, so these findings suggest that significantly more perceptual resources were allocated to the task-irrelevant ERP stimuli in the low presence group during the experiment. The high presence group subjectively reported being significantly less aware of the beeping tones than the low presence group. This corroborates the theory that the high presence group's attentional resources were more frequently fully-absorbed by the primary gaming task to the point where no or not enough perceptual resources were leftover to consistently process and perceive the ERP tones or changes hereof in the secondary task. This indicates that automatic perceptual processing of real world information, as probed by the extraneous beeping tones, were generally suppressed more in the high presence group, due to a lack of available resources, than in the low presence group. This effect is illustrated in Fig. 4, where it is visually observable that the high presence group's grand averaged oddball and standard tones across electrodes are more similar in N1 amplitudes than the low presence group, where the averaged amplitudes elicited by oddball and the standard tones are markedly different. This is numerically reflected by the mean amplitudes of MMN as this standard-deviant-difference-wave component is elicited by automatic detection of oddballs in the auditory sequence and it shows that the low presence group more reliably detected the difference between deviant and standard tones than the high presence group did. In accordance with Lavie (1995), this is suggested to be a result of the low presence group more frequently having sufficient attentional resources leftover from the primary gaming task that the performance on the secondary task, i.e. early stage oddball-stimulus discrimination, did not significantly decrement. These findings are inconsistent with the results of Kober and Neuper (2012)'s study where no differences were found between the low and high presence groups in N1 and MMN. The fact that no difference in these early processing components were found indicate that both groups in their study detected the change between standard and oddball tones similarly suggesting that the groups didn't significantly differ in attentional resources allocated to the simulation despite having rated significantly different presence scores.

The present findings that N1 and MMN amplitudes were significantly decreased in the high presence group compared to the low presence group is in line with the theoretical operationalization of presence as an experience that necessitates full attentional resource allocation to the mediated content before a strong SSM can be formed, which is argued by Wirth et al. (2007) to be a prerequisite for spatial presence to occur. However, the task-irrelevant probe technique used in this study has the limitation that it does not directly reflect the increase in the allocation of attentional resources to the game task, but rather it reflects the decrease in the allocation of perceptual resources to the task-irrelevant beeping probes. It is therefore necessary to assume that the total amount of attentional resources available at a given time is constant and that whatever perceptual resources are not allocated to the secondary ERP task is instead being absorbed by the primary game task. The results of this study support this assumption as the early ERP components, associated with early stimuli processing, elicited by the task-irrelevant tones differed as a function of the participants' self-reported presence experience evoked by the game. This indicates that a measurement of the degree of attention to a secondary irrelevant probing task is a valid practical operationalization of the subjective experience of presence in a primary video game task

The amplitudes of N1 and MMN were generally most pronounced over Fz and Cz for both groups, which supports that these components are fronto-centrally innervated. However, the differences between the two presence groups were largest over POz for both N1 (d = 1.38) and MMN (d = 1.18), which was unexpected. A potential explanation for this is advocated to be that EEG signal strength continuously decrease as it travels further from the source (Pinel, 2011), so only the significantly larger component amplitudes in the low presence group is still measurable at POz, whereas with the high presence group hardly any ERP component activation is observable at this posterior electrode site. This finding is argued to have methodological implications for future empirical studies that aim to measure presence through early ERPs, since it suggests that even though the ERP components of interest are fronto-centrally innervated the posterior electrode site has shown the greatest capabilities for indicating presence.

It was hypothesized that neither the SW1 nor the SW2 component would be significantly different between the two presence groups. As anticipated, no significant differences in SW were found, which suggests that the results of this study cannot directly support the previously mentioned findings of Kober and Neuper (2012). While no significant correlation was found between SW components and overall presence score, a weak relationship was observed between SW1 and the subdomain of spatial presence across multiple electrodes. Burns and Fairclough (2015) found a link between challenge-based immersion and SW in their study, so it is possible that perceived difficulty of the present gaming task, although not explicitly manipulated or measured in this study, has had a degree of influence on the participants' experience of spatial presence. With a game that is more goal-oriented and skill-dependent this effect might be more pronounced, but as the game stimulus in this study was partly chosen due to it not requiring any prior video game competency it is possible that the effects of challenge-based immersion wasn't essential enough to the experience that it would manifest as a significant difference in overall presence ratings. It is possible that the challenge of the goal-oriented navigation task (involving elements of mental rotation and memory search to complete) in the virtual simulation used by Kober and Neuper (2012) was sufficiently essential to the experience to the point where it significantly influenced the overall presence ratings. This could mean that the participants in their study were divided into low or high presence groups based on challenge-based immersion resulting from individual differences in how difficult the participants found the primary navigation task and not significantly based on interpersonal differences in user factors of attention, as indicated by the lack of significant differences in N1 and MMN between the presence groups in their study. This could potentially explain why Kober and Neuper (2012) found a difference in SW in their study, similarly to when their methodology was replicated by Burns and Fairclough (2015). However future work should continue to explore the relevance of SW components for measuring presence in video games, as previous and present findings are inconsistent in this regard.

Lastly, it was hypothesized that the measured ERP components could be used to significantly predict presence in this study. Both the N1 and MMN component correlated significantly with overall presence score across all electrodes. The strength of the correlation was most prominent over POz, where the correlation coefficient went above the established conventions for interpreting a strong linear relationship. N1 and MMN also correlated significantly with all three presence subdomains. This could suggest that these ERP components are sensitive to all three subdomains of presence, whereas the GSR-based measure was significantly correlated with spatial and self presence. A standard multiple regression analysis revealed that the measured ERP variables could significantly explain 49% of the overall variance in self-reported presence ratings, which is argued to be a very respectable percentage considering that it is an objective measure which is being used to predict variance in a subjectively rated experience. As with the exploration of the predictive capabilities of the GSR measurement, a LDA was also conducted to investigate how accurately the ERP components that significantly differed between presence groups can be used to classify each participant's presence group membership. This discriminant function could correctly assign participants into their original self-report-based presence groups 76.5% of the time, which is higher than chance and slightly more precise than what was achieved with the GSR-based discriminant function.

Multiple independent studies have currently found a significant difference in distinct ERP components when comparing groups with different self-reported levels of presence in virtual environments. This present study is the first to our knowledge to find a strong significant correlation between the subjective experience of presence and the early ERP components of N1 and MMN, including demonstrating significant links between ERP components and all three presence subdomains. We hope that these findings will provide researchers with new methodological tools and insights that will further the exploration and understanding of the phenomenon of presence in VEs and how to objectively measure it.

5. Future directions and limitations

The findings of this study show the potential of using ERP components and GSR measures as predictors of presence and its underlying components in video games. However, results across studies are inconsistent regarding which ERP components and GSR metrics show the most promise as presence indicators, so further work is deemed necessary in order to explore this. In ecological settings, UX researchers in GUR envision a plug-and-record system that can easily be setup to provide physiological assessments of the player's gameplay experiences of presence, which can then be used to promptly identify what parts of the game reliably engrosses the player and which parts fail to demand the players attention consistently (Nacke, 2018). Before this can become a reality, a standardized objective measure of presence must be proposed and be thoroughly and repeatedly validated. This study has taken steps towards this by proposing and testing two independent physiological measures of presence using a horror game as experimental stimulus. While, in general, both the ERP- and the GSR-based approach seem to be promising methods for measuring presence as indicated by the results of this study, more work is needed to further validate and replicate the methods and findings across different games and different game genres before the sensitivity of the measurements can be fully understood and be generalized to a broader GUR context.

A limitation of this study's methodology is that presence groups were derived from a median split of the participants' subjective presence ratings after they had all been exposed to the same game stimulus and not derived by manipulation of media factors, such as graphical fidelity or interactivity. It therefore follows that the results of this study is a reflection of changes in user factors of presence, such as motivation, willing suspension of disbelief, trait absorption and selective attention, and not media factors The findings of this study is argued to support that the proposed ERP-measurement technique is sensitive to attention and thereby presence as attention is what the overall ERP-based operationalization of presence is centered around. Given that both media and user factors are theoretically advocated to impact presence through direction of attention, it suggests that this measure is sensitive to presence independently of what factors evoked the experience. However, future studies should explore and validate the measurement techniques' sensitivity to presence facilitated through conditions where the strength of media factors, such as vividness and interactivity, are manipulated.

This study has supported that the participants overall subjective presence experience correlated with averaged physiological measures based on data recordings of the entire gameplay session. In order to someday realize the goal of being able to objectively assess the presence-sensation induced by individual parts of a game, future studies should investigate the potential of using the probing methods of this study more continuously as time-series measurements as a way to investigate whether time-varying qualities of presence are observable. Methodologically though, both of the investigated measurement techniques have certain potential limitations as continuous measures. While the ERP-measurement approach theoretically probes the participants' presence levels continuously around every five seconds (depending on the probability of an oddball tone being played), the reliability of ERP experiments in general is contingent on averaging across a large number of sample trials in order to get a noise-free signal reflecting the cognitive process of interest, which arguably complicates its applications as a time-series measure. Future research should investigate whether this method is reliable with small segments of the total playtime, which might facilitate a more continuous insight into the presence experience. The GSR measure is conceptualized as probing presence every time a participant would be feeling physiologically aroused if the mediated VE was real, such as when a stress-inducing game event occurs. However, this study has demonstrated the limitation that intense unexpected stimulus-events are not a good presence indicator as there were no significant difference in arousal response magnitude between the presence groups, which suggest that the jump scare event was similarly perceived as either mediated or non-mediated across the groups. Based on the promising relationship between presence and GSR peaks per minute shown in this study, it would be interesting to investigate further whether using the amount of non-specific and event-related GSR peaks in per minute-segments could reliably provide practical insights into the moment-to-moment experience of presence in arousing VEs.

The current study has assumed and intentionally treated presence as a somewhat binary construct (bifurcated into non-telepresent or telepresent), where the 'scalability' comes from the percentage of time a player experiences a sense of presence during the game. This was done based on the theoretical conceptualization of presence presented in this article. However, no consensus has been reached as of yet about whether presence should be operationalized as either dimensional or binary or possibly something in between (Slater and Steed, 2000). Future studies could potentially benefit from using the proposed objective measures of this paper to investigate this further.

In terms of accessibility and practical application, the proposed physiological measures are considered to be unobtrusive to the experience of presence during gaming, but currently they require rather expensive equipment and significant know-how and expertise for both setup and analysis. Future work should attempt to improve usability by automating the analyses and constructing 'presence classifiers' that can be used in GUR practice. The GSR measurement approach is a lot closer to reaching the goal of convenience and accessibility than the ERP-approach. However, as mentioned the applicability of GSR is argued to be limited to stress- or emotion-inducing stimulus, whereas this limitation is not expected for the attention-based ERP-approach. Generally, the study and use of physiological measures of presence in GUR is still regarded as being in its infancy, so future work is deemed necessary following this study before we might see these methods being implemented in practice.

6. Conclusion

Presence is a relatively new concept within GUR that hasn't received a lot of empirical attention until recently. Understanding presence and how to objectively measure it is considered essential as modern games are now commonly evaluated based on how reliably they evoke the sensation of presence in their players (Meehan et al., 2002). This study found that dividing participants into a low and a high presence groups based on their self-reported experience of presence while playing a horror video game lead to measurable significant between-group differences across two independent and objective physiological measures, namely GSR and ERP. The physiological variables that differed between presence groups (GSR peaks/min, N1 and MMN) correlated significantly with presence and they were demonstrated to have potential as objective indicators of overall presence. The results of this study is therefore argued to provide empirical support for the use of physiological operational measures of presence in GUR and is thought to contribute to the discussion on the topic by increasing the relatively limited amount of GUR literature currently available, which explicitly explores the link between either of these two physiological measures and the subjective experience of presence.

Conflict of interest statement

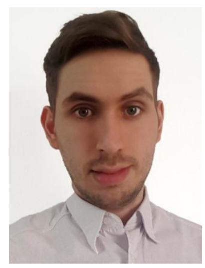
The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

Allison, B.Z., Polich, J., 2008. Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. Biol. Psychol. 77, 277–283.
Andreassi, J., 2007. Psychophysiology: Human Behavior and Physiological Response.

Lawrence Erlbaum.
Armel, K.C., Ramachandran, V.S., 2003. Projecting sensations to external objects: evidence from skin conductance response. Proc. R. Soc. B 270 (1523), 1499–1506.
Barfield, W., Zeltzer, D., Sheridan, T.B., Slater, M., 1995. Presence and performance within virtual environments. In: Barfield, W., Furness, T.A. (Eds.), Virtual Environments and Advanced Interface Design. Oxford University Press, Oxford, pp. 473–541.
Belini, H., Chen, W., Sugiyama, M., Shin, M., Alam, S., Takayama, D., 2016. Virtual & Augmented Reality: Understanding the Race for the Next Computing Platform. Goldman Sachs reportRetrieved 28 February 2018 from. http://www.goldmansachs.com/our-thinking/pages/technology-driving-innovation-folder/virtual-and-augmented-reality/report.pdf.
Benedek, M., Kaernbach, C., 2010. Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology 47 (4), 647–658.
Biocca, F., 1997. The Cyborg's Dilemma: embodiment in virtual environments. In: Presented at the 2nd International Conference on Cognitive Technology, 1997.
Boucsein, W., 2012. Electrodermal Activity, third ed. Springer Science & Business Media.
Bradley, M.M., Lang, P.J., 1994. Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatry 25 (1), 49–59.
Breedlove, S.M., Watson, N.V., 2013. Biological Psychology: An Introduction to Behavioral, Cognitive, and Clinical Neuroscience, seventh ed. Sinauer Associates, Inc.
Brehm, J.W., 1999. The intensity of emotion. Personal. Soc. Psychol. Rev. 3 (1), 2–22 Lawrence Erlbaum Associates, Inc.
Burns, C.G, Fairclough, S.H., 2015. Use of auditory event-related potentials to measure immersion during a computer game. Int. J. Hum.-Comput. Stud. 73, 107–144.
Chalfoun, P., Dankoff, J., 2018. Developing actionable biometric insight for production teams. Games User Research. Oxford University Press, pp. 281–299.
Chanel, G., Kronegg, J., Grandjean, D., Pun, T., 2006. Emotion Assessment: AArousal Evaluation Using EEG's and Peripheral Physiological Signals. pp. 530–537 Int. Workshop on Multimedia Content Representation.
Charij, S., Oikonomou, A., 2015. Using biometric measurement in real-time as a sympathetic system in computer games. Int. J. Game-Based Learn. 3 (3), 21–42.
Christou, G., 2014. The interplay between immersion and appeal in video games. Comput. Hum. Behav. 32, 92–100.
Darken, R.P., Bernatovich, D., Lawson, J., Peterson, B., 1999. Quantitative measures of presence in virtual environments: the roles of attention and spatial comprehension. Cyber-Psychol. Behav. 2 (4), 337–347.
Dawson, M.E., Schell, A.M., Filion, D.L., 2007. The electrodermal system. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (Eds.), Handbook of Psychophysiology. Cambridge University Press, New York, NY, US, pp. 200–223.
Detenber, B., Reeves, B., 1996. A bio-informational theory of emotion: motion and image size effects on viewers. J. Commun. 46 (3), 66–84.
Don't Knock Twice on Steam, 2018. (n.d.). Retrieved 25 February 2018. http://store.steampowered.com/app/328590/Dont_Knock_Twice/.
Drachen, A., Yannakakis, G., Nacke, L.E., Pedersen, A.L., 2010. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In: Proceeding Sandbox '10 Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games, pp. 49–54.
Ehrsson, H.H., Wiech, K., Weiskopf, N., Dolan, R.J., Passingham, R.E., 2007. Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. Proc. Natl. Acad. Sci. USA 104 (23), 9828–9833.
Ekman, P., 2003a. Emotions Revealed. Times Books.
Ekman, P., 2003b. Darwin, deception, and facial expression. N. Y. Acad. Sci. 1000, 205–221.
Ekman, P., Levenson, R.W., Friesen, W.V., 1983. Autonomic Nervous System activity distinguishes among emotions. Science 221, 1208–1210.
Ermi, L., Mayra, F., 2005. Fundamental Components of the Gameplay Experience: Analysing Immersion. In: de Castell, S., Jenson, J. (Eds.), DiGRA. DiGRA.
Finnigan, S., O'Connel, R., Robertson, I.H., 2010. ERP measures indicate both attention and working memory encoding decrements in aging. Psychophysiology 48 (4), 601–611.
Fontaine, G., 1992. The experience of a sense of presence in intercultural and international encounters. Presence 1 (4), 482–490.
Gosselin, P.A., Gagné, J., 2010. Use of a dual-task paradigm to measure listening effort. Can. J. Speech-Lang. Pathol. Audiol. 34 (1), 42–51.
Haag, A., Goronzy, S., Schaich, P., Williams, J., 2004. Emotion recognition using biosensors: first steps towards an automatic system. Affect. Dialogue Syst. 36–48.
Hackley, S., Woldoroff, M., Hillyard, S., 1990. Cross-modal selective attention effects on retinal, myogenic, brainstem and cerebral evoked potentials. Psychophysiology 27, 195–208.
Hecht, D., Reiner, M., 2006. Multimodal virtual environments: response times, attention, and presence. Presence Teleoperators Virtual Environ. 15 (5), 515–523.
Hillyard, S.A., Picton, T.W., Regan, D.M., 1978. Sensation, perception and attention: analysis using ERPs. In: Callaway, E., Tueting, P., Koslow, S. (Eds.), Event-Related Potentials in Man. Academic Press, New York, pp. 223–322.
Hillyard, S., Hink, R., Schwent, V., Picton, T., 1973. Electrical signs of selective attention in the human brain. Science 182, l77–180.
Ijsselsteijn, W.A., de Ridder, H., Freeman, J., Avons, S.E., 2000. Presence: concept, determinants and measurement. In: Proceedings of the SPIE, Human Vision and Electronic Imaging V, pp. 3959–3976.
iMotions A/S, 2018. GSR pocket guide. Retrieved from. https://imotions.com/guides/ 28 January 2018.
iMotions A/S, 2018. Imotions Biometric Research Platform 7.0. iMotions A/S, Copenhagen, Denmark.
Insko, B.E., 2003. Measuring presence: subjective, behavioral and physiological methods. In: Riva, G., Davide, F., IJsselsteijn, W.A. (Eds.), Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments. Ios Press, Amsterdam, pp. 109–118.
Järvinen, A, 2017. Designing for presence in VR. Retrieved January 15 2019, from Virtual Reality Pop. https://virtualrealitypop.com/designing-for-presence-in-vr-part-3-analysis-design-case-studies-5eedeb7b3a01.
Kahneman, D., 1973. Attention and Effort. Prentice-Hall, New Jersey.
Karatekin, C., Couperus, J.W., Marcus, D.J., 2004. Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. Psychophysiology 41, 175–185.
Kilteni, K., Groten, R., Slater, M., 2012. The sense of embodiment in virtual reality. Presence 21, 373–387.
Kober, S.E., Neuper, C., 2012. Using auditory event-related EEG potentials to assess presence in virtual reality. Int. J. Hum.-Comput. Stud. 70, 577–587.
Kok, A., 1997. Event-related-potential (ERP) reflections of mental resources: a review and synthesis. Biol. Psychol. 45, 19–56.
Kramer, A.F., Trejo, L.J., Humphrey, D., 1995. Assessment of mental workload with task-irrelevant auditory probes. Biol. Psychol. 40, 83–100.
Larsen, J.T., Berntson, G.G., Poehlmann, K.M., Ito, T.A., Cacioppo, J.T., 2010. The Psychophysiology of Emotion. Handbook of Emotions. The Guilford Press.
Lavie, N., 1995. Perceptual load as a necessary condition for selective attention. J. Exp. Psychol. 21 (3), 451–468.
Lazarus, R.S., Folkman, S., 1987. Transactional theory and research on emotions and coping. Eur. J. Personal. 1 (3), 141–169.
Lee, K.M., 2004a. Why presence occurs: evolutionary psychology. media equation, and presence. Presence 13 (4), 494–505.
Lee, K.M., 2004b. Presence, explicated. Commun. Theory 14 (1), 27–50.
Leon, E., Clarke, G., Callaghan, V., Sepulveda, F., 2007. A user-independent real-time emotion recognition system for software agents in domestic environments. Eng. Appl. Artif. Intell. 20 (3), 337–345.
Levenson, R.W., 1999. The intrapersonal functions of emotions. Cognit. Emot. 13, 481–504.
Levenson, R.W., 1992. Autonomic nervous system differences among emotions. Psychol. Sci. 3, 23–27.
Levenson, R.W., Ekman, P., Friesen, W.V., 1990. Voluntary facial action generates emotion-specific autonomic nervous system activity. Psychophysiology 27 (4), 363–384.
Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique, second ed. The MIT Press.
Luck, S.J., Heinze, H.J., Mangun, G.R., Hillyard, S.A., 1990. Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. Electroencephalogr. Clin. Neurophysiol. 75, 528–542.
Luck, S.J., Woodman, G.F., Vogel, E.K., 2000. Event-Related potential studies of Attention. Trends Cognit. Sci. 4, 432–440.
Mager, R., Bullinger, A.H., Roessler, A., Mueller-Spahn, F., Stoermer, R., 2000. Monitoring brain activity during use of stereoscopic virtual environments. CyberPsychol. Behav. 3 (3), 407–413.
Makransky, G., Lilleholt, L., Aaby, A., 2017. Development and validation of the Multimodal Presence Scale for virtual reality environments: a confirmatory factor analysis and item response theory approach. Comput. Hum. Behav. 72, 276–285.
Mandryk, R.L, Inkpen, K.M, Calvert, T.W., 2006. Using psychophysiological techniques to measure user experience with entertainment technologies. Behav. Inf. Technol. 25 (2), 141–158.
Mandryk, R.L., Nacke, L.E., 2016. Biometrics in gaming and entertainment technologies. Biometrics in a Data Driven World: Trends, Technologies, and Challenges. CRC Press, Boca Raton, FL, pp. 191–224.
Matlin, M.W., 2009. Cognitive Psychology. John Wiley & Sons (Asia) Pte Ltd.
McMahan, R.P., Bowman, D.A., Zielinski, D.J., Brady, R.B., 2012. Evaluating display fidelity and interaction fidelity in a virtual reality game. IEEE Trans. Vis. Comput. Graph. 18 (4), 626–633.
Meehan, M., Insko, B., Whitton, M., Brooks, F.P., 2002. Physiological measures of presence in stressful virtual environments. In: Proceeding SIGGRAPH '02 Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, pp. 645–652.
Miller, M.W., Rietschel, J.C., McDonald, C.G., Hatfield, B.D., 2011. A novel approach to the physiological measurement of mental workload. Int. J. Psychophysiol. 80, 75–78.
Muluh, E.T., 2011. A review of Event-Related Potential (ERP) components employed in mental arithmetic processing studies. J. Neurol. Neurophysiol. S6.
Näätänen, R., Kujala, T., Winkler, I., 2011. Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. Psychophysiology 48, 4–22.
Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118, 2544–2590.
Nacke, L.E., 2018. Introduction to biometric measures for games user research. Games User Research. Oxford University Press, pp. 281–299.
Nacke, L.E., Grimshaw, M.N., Lindley, C.A., 2010. More than a feeling: measurement of sonic user experience and psychophysiology in a first-person shooter game. Interact. Comput. 22, 336–343.
Nogueira, P.A., Rodrigues, R., Oliveira, E., Nacke, L.E., 2011. A Regression-Based Method For Lightweight Emotional State Detection in Interactive Environments. Springer-Verlag, Berlin Heidelberg.
Pinel, J.P.J., 2011. Biopsychology, eight ed. Pearson Education, Inc.
Poels, K., van den Hoogen, W., Ijsselsteijn, W., de Kort, Y., 2012. Pleasure to play, arousal to stay: the effect of player emotions on digital game preferences and playing time. Cyberpsychol. Behav. Soc. Netw. 15 (I), 1–6.
Pugnetti, L., Mendozzi, L., Barberi, E., Rose, F.D., Attree, E.A., 1996. Nervous system

correlates of virtual reality experience. In: Proceedings of the 1st European Conference on Disability, Virtual Reality & Assoc. Tech., Maidenhead. UK. pp. 239–246.
Ravaja, N., Salminen, M., Holopainen, J., Saari, T., Laarni, J., Järvinen, A., 2004. Emotional response patterns and sense of presence during video games: potential criterion variables for game design. In: Proceeding NordiCHI '04 Proceedings of the third Nordic Conference on Human-Computer Interaction, pp. 339–347.
Reeves, B., Nass, C., 1996. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, New York.
Reeves, B., Detenber, B., Steuer, J., 1993. New televisions: the effects of big pictures and big sound on viewer responses to the screen. In: Paper Presented to the Information Systems Division of the International Communication Association. Washington, DC.
Rösler, F., Heil, M., Röder, B., 1997. Slow negative brain potentials as reflections of specific modular resources of cognition. Biol. Psychol. 45, 109–141.
Rottenberg, J., Ray, R.D., Gross, J.J., 2007. Emotion elicitation using films. In: Coan, J.A., Allen, J.J.B. (Eds.), The Handbook of Emotion Elicitation and Assessment. Oxford University Press, London, pp. 9–28.
Sadowski, W., Stanney, K., 2002. Presence in virtual environments. In: Stanney, K.M. (Ed.), Human Factors and Ergonomics. Handbook of Virtual Environments: Design, Implementation, and Applications. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 791–806.
Scerbo, A.S., Freedman, L.W., Raine, A., Dawson, M.E., Venables, P.H., 1992. A major effect of recording site on measurement of electrodermal activity. Psychophysiology, 29, 241–246.
Schuemie, M.J., Straaten, P.V., Krijn, M., Mast, C.V., 2001. Research on presence in virtual reality: a survey. Cyberpsychol. Behav. 4 (2), 183–201.
Slater, M., 2002. Presence and the sixth sense. Presence 11 (4), 435–439.
Slater, M., Steed, A., 2000. A virtual presence counter. Presence 9 (5), 413–434. https://doi.org/10.1162/105474600566925.
Slater, M., Lotto, B., Arnold, M.M., Sanchez-Vives, M.V., 2009. How we experience immersive virtual environments: the concept of presence and its measurement. Anu. Psicol. 40 (2), 193–210.
Sonnemans, J., Frijda, N.H., 1995. The determinants of subjective emotional intensity. Cognit. Emot. 9 (5), 483–506.
Steuer, J., 1992. Defining virtual reality: dimensions determining telepresence. J. Commun. 42 (4), 73 .93.
Strüber, D., Polich, J., 2002. P300 and slow wave from oddball and single-stimulus visual tasks: inter-stimulus effects. Int. J. Psychophysiol. 45, 187–196.
Tamborini, R., Bowman, N.D., 2010. Presence in video games. In: Bracken, C., Skalski, P. (Eds.), Immersed in Media: Telepresence in Everyday Life. Routledge, New York, pp. 5–8.
Tamborini, R., Skalski, P., 2005. The role of presence in the experience of electronic games. In: Communication and Technology Divisionn of the International Communcation Association 55th Annual Conference, pp. 1–34.
Tamborini, R., 2000. The experience of telepresence in violent games. In: Paper presented at the annual conference of the National Communication Assosiation. Seattle, WA.
Ullsperger, P., Freude, G., Erdmann, U., 2001. Auditory probe sensitivity to mental workload changes—an event-related potential study. Int. J. Psychophysiol. 40, 201–209.
van Dooren, M., de Vries, J.J., Janssen, J.H., 2012. Emotional Sweating across the body: comparing 16 different skin conductance measurement location. Physiol. Behav. 106, 298–304.
Weber, R., Behr, K., Tamborini, R., Ritterfeld, U., Mathiak, K., 2009. What do we really know about first-person-shooter games? An event-related, high-resolution content analysis. J. Comput.-Mediat. Commun. 14 (4), 1016–1037.
Wickens, C.D., 2002. Multiple resources and performance prediction. Theor. Issues Ergon. Sci. 3 (2), 159–177.
Wiederhold, B., Kim, S.I., Jang, D.P., Kaneda, M., Cabral, I., Wiederhold, M.D., et al., 2003. An investigation into physiological responses in virtual environments: an objective measurement of presence. In: Galimberti, C., Riva, G. (Eds.), Towards CyberPsychology: Mind, Cognitions and Society in the Internet Age. IOS press, Amsterdam, pp. 175–182.
Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F.R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., Jäncke, P., 2007. A process model of the formation of spatial presence experiences. Media Psychol. 9 (3), 493–525.
Witmer, B., Singer, M., 1998. Measuring presence in virtual environments: a presence questionnaire. Presence, 7 (3), 225–240.
Yannakakis, G.N., Martinez, H.P., Garbarino, M., 2016. Psychophysiology in games. In Emotions in Games: Theory and Praxis. Springer, pp. 119–139.
Yaremko, R.M., Blair, M.W., Leckhart, B.T., 1970. The orienting reflex to changes in a conceptual stimulus dimension. Psychon. Sci. 5, 317–318.
Yuan, Y., Steed, A., 2010. Is the rubber hand illusion induced by immersive virtual reality? In: Paper presented at the 2010 IEEE Virtual Reality Conference (VR).



Thomas Terkildsen. Thomas Terkildsen, Msc. Psych, is a Research Assistant at the Department of Psychology at University of Copenhagen, Denmark, where he is currently managing the Virtual Learning Lab, which is a lab that primarily focuses on research relating to investigating process measure of learning in Virtual Reality. Terkildsen's personal research interests include Affective Computing, Physiological Measures, Assessment and Elicitation of Emotions, Games User Research, User Experience, Presence, Virtual Reality, Learning and Horror Games.

Guido Makransky. Dr. Makransky is an Associate Professor and leader of the Virtual Learning Lab, at the Department of Psychology, at the University of Copenhagen. His research topics include learning within immersive environments including virtual and augmented reality, multimedia learning, instructional design, motivation, self-efficacy, process measures of learning, presence, training, psychometrics, and psychological and educational measurement. His current research focuses on investigating the mechanisms underlying learning in immersive learning environments (including but not limited to virtual and augmented reality), by combining laboratory research with experiments in classrooms, using theories from cognitive psychology and modern psychometric and psychophysiological measures.