Object-based attentional expectancies in virtual reality

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Promotionsordnung vom 10.05.2010

vorgelegt von

Michael Wiesing

aus Paderborn

März 2022

Abstract

Modern virtual reality (VR) technology has the promise to enable neuroscientists and psychologists to conduct ecologically valid experiments, while maintaining precise experimental control. However, in recent studies, game engines like Unreal Engine or Unity, are used for stimulus creation and data collection. Yet game engines do not provide the underlying architecture to measure the time of stimulus events and behavioral input with the accuracy or precision required by many experiments. Furthermore, it is currently not well understood, if VR and the underlying technology engages the same cognitive processes as a comparable real-world situation. Similarly, not much is known, if experimental findings obtained in a standard monitor-based experiment, are comparable to those obtained in VR by using a head-mounted display (HMD) or if the different stimulus devices also engage different cognitive processes.

The aim of my thesis was to investigate if modern HMDs affect the early processing of basic visual features differently than a standard computer monitor.

In the first project (chapter 1), I developed a new behavioral paradigm, to investigate how prediction errors of basic object features are processed. In a series of four experiments, the results consistently indicated that simultaneous prediction errors for unexpected colors and orientations are processed independently on an early level of processing, before object binding comes into play.

My second project (chapter 2) examined the accuracy and precision of stimulus timing and reaction time measurements, when using Unreal Engine 4 (UE4) in combination with a modern HMD system. My results demonstrate that stimulus durations can be defined and controlled with high precision and accuracy. However, reaction time measurements turned out to be highly imprecise and inaccurate, when using UE4's standard application programming interface (API). Instead, I proposed a new software-based approach to circumvent these limitations. Timings benchmarks confirmed that the method can measure reaction times with a precision and accuracy in the millisecond range.

In the third project (chapter 3), I directly compared the task performance in the paradigm developed in chapter 1 between the original experimental setup and a virtual reality simulation of this experiment. To establish two identical experimental setups, I recreated the entire physical environment in which the experiments took place within VR and blended the virtual replica over the physical lab. As a result, the virtual environment (VE) corresponded not only visually with the physical laboratory but also provided accurate sensory properties of other modalities, such as haptic or acoustic feedback. The results showed a comparable task performance in both the non-VR and the VR experiments, suggesting that modern HMDs do not affect early processing of basic visual features differently than a typical computer monitor.

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im (Juli 2022) angenommen (Beschluss des Promotionsausschusses vom 20.10.2010).

Acknowledgements

Foremost, I would like to thank my supervisors without whom this would not have been possible: PD Dr. Ralph Weidner, who supported and encouraged me all the way in this journey and supported me and gave me the freedom to follow my own ideas. And Prof. Dr. Simone Vossel for her support and here invaluable feedback and advice.

I also want to thank Prof. Dr. Fink for his advice and for giving me the opportunity to call the INM-3 my scientific home for such a long time. And Prof. Dr. Weiss-Blankenhorn, for giving me the opportunity to come as an intern to the INM-3 in 2009 and where I made my first experiences with the wonderful world of cognitive neuroscience.

Furthermore, I would like to thank all my former colleagues at the INM-3 for the cooperation and support, as well as fun times during coffee breaks and retreats. My special thanks go to Dr. Sabine Bertleff, Dr. Shivakumar Viswanathan, Dr. Rouhollah Abdollahi, Dr. Paola Mengotti, and Dr. Eva Nießen for both heated discussion and a lot of fun times and conversations.

Moreover, I am grateful for the support and help of my family and friends. I want to thank Hendrik Steinkönig and Patrick Wolff for all their support and for being the best friends you can wish for. I especially want to thank Adelheid Wiesing for unconditional love and support and for being the best and strongest mother in the world. Rolf Hoppe for your support and being an important part of our family. Like you have been there for my mother, we will now be with you in these hard times. I also want to thank my sister, Sabrina Jüde for all their support and love. And Thorsten Jüde, who is not only an awesome brother-in-law but also loving father of my wonderful niece and nephew, Talea and Taio.

Table of Contents

List of Abbreviations

API	application programming interface
BBTK	Black Box Toolkit
EEG	electroencephalography
FIT	Feature Integration Theory
FOV	field of view
FPS	frames per second
HMD	head-mounted display
RT	reaction time
UE4	Unreal Engine 4
VE	virtual environment
VR	virtual reality
LCD	liquid crystal displays
CRT	cathode-ray tube

Introduction

Research in cognitive neuroscience aims to understand human behavior and its underlying brain functions. A prerequisite to successfully relate behavior and brain activity to specific cognitive functions are experiments with a high internal validity. Internal validity describes to which extent a study can draw conclusions on cause and effect and how well alternative explanations can be ruled out. One critical factor for internal validity is high experimental control, i.e., minimizing the impact of variables other than the variable of interest on the experimental outcome. To achieve high experimental control, experiments in cognitive neuroscience are preferably conducted using simplified and minimalistic stimuli as compared to real-world scenarios. Minimalistic stimuli allow to precisely control and manipulate certain stimulus features. Behavioral responses are similarly often restricted and repetitive, as scientists aim to obtain behavioral measures, which are comparable between participants.

However, it has been criticized that these sterile laboratory experiments poorly relate to real-world phenomena, making it difficult to transfer experimental findings to real-life situations. Hence, many researchers have argued for higher *ecological validity* in behavioral experiments to promote the generalizability of experimental findings (Kingstone et al., 2008; Parsons, 2015). Ecological validity refers to the relation between real-world phenomena and the experimental context. For example, in 1976 Neisser criticized that “contemporary studies of cognitive processes usually use stimulus material that is abstract, discontinuous, and only marginally real. It is almost as if ecological **invalidity** were a deliberate feature of the experimental design” (Neisser, 1976, p. 34) (accentuation by the author). However, although Neisser’s criticism is almost half a century old, researchers remained hesitant to strive for increasing ecological validity in behavioral experiments. The reasons for this situation are manifold. For example, the concept of ecological validity lacks a clear and generally accepted definition and its overall usefulness has been questioned (Schmuckler, 2001). However, beyond theoretical considerations for or against ecological validity, one important factor that has kept researchers to increase ecological validity is experimental control. Both concepts have been considered as trade-offs (Loomis et al., 1999a), emphasizing tensions between researchers striving for the one and those striving for the other (Parsons, 2015).

A promising middle ground comes in form of Virtual Reality (VR). VR has the potential to bring worlds into the laboratory and it allows transferring experimental paradigms from the laboratory into naturalistic but highly controlled scenarios (Kothgassner & Felnhofer, 2020). Although VR’s benefits have been recognized for decades in neuroscientific research (Bohil et al., 2011; Loomis et al., 1999b), it has been highly underutilized in cognitive neuroscience and experimental psychology. Yet, advances in 3D rendering and virtual reality technology in the recent years, as well as the decreasing cost of associated equipment have led to an increasing interest in VR as a research tool.

VR places the participant in computer-generated and ideally multi-sensory three-dimensional environments, while at the same time shutting out sensory input from real-world stimuli as much as possible. In other words, the idea of VR is to immerse the participant into a simulated reality, which is experienced as if it were real. However, an effective VR simulation constitutes not only sophisticated devices for sensory stimulation, but instead requires the ability to closely monitor the participant’s behavioral responses and to translate these into virtual interactions in real-time. By default, typical state-of-the-art consumer VR systems come with motion tracking systems for head and hand movements. Some systems, such as the SteamVR tracking system, allow to increase the number of tracked object or body parts, by including additional tracking sensors. Also, eye-tracking, finger tracking or tracking of facial gestures are available on the consumer level. On the professional sector, VR systems combined with physiological sensors such as heart rate, galvanic skin-responses, or even integrated electroencephalograms (EEG) are available. Hence, a typical VR system is both, a device for stimulus delivery but also a sensitive and versatile measurement tool. Not surprisingly, with the release of low-cost consumer VR systems, numbers of research papers within neuroscience and experimental psychology utilizing VR, has been growing steadily in the recent years (Vasser & Aru, 2020).

However, although VR found its way into research labs, to date, there is a lack of standards and good practice guidelines regarding VR experiments, a situation which has been compared to the "Wild West" (Birckhead et al., 2019). For example, within the scientific literature the definitions of Virtual Reality range from "computer-generated world" to wall projector such as a Cave system (X. Pan & Hamilton, 2018b; Slater, 2018; Takac et al., 2021). Hence, the term needs some clarification. Within the context of this introduction as well as the chapters 2 and 3 of this dissertation, everything discussed about VR will primarily refer to head-mounted display (HMD) based VR systems. The main reasoning is to keep the introduction short and concise as well as the research questions of the studies reported in chapter 2 and 3 specifically focused on HMD-based VR systems. In chapter 2, I investigated the precision and accuracy of stimulus timing and time measurements in experiments based on HMDs and associated rendering processes. In chapter 3, I examined if the simple fact that an experiment takes place in VR, i.e., stimuli were presented via an HMD, changes the behavior of the participants when compared to a standard monitor-based setup.

Furthermore, the recent surge in VR studies is mainly driven by the release of consumer HMD systems, such as HTC Vive or Valve Index. While for example, CAVE systems come at a high cost and require a lot of space and technical expertise, a state-of-the art HMD system can be bought already for a few hundred Euro and is considerably easier to install and to use. Therefore, for the foreseeable future, HMDs will be the dominant form of VR systems that can be found in research labs.

Currently most research experiments have been created by using modern game engines, such as Unity or Unreal Engine. While game engines provide powerful tools for the creation of stimulus environments and interactions in VR, they do not contain certain key validated technical features that are critical for neuroscientific experiments. For example, game engines run in a so-called game-loop. The game-loop contains several processes such as physics simulations, input processing or drawing of objects, and typically iterates once with every display refresh, when using HMDs. Consequently, sampling rates for data collection, e.g., motion tracking data or button presses, are directly limited by the refresh rate of the HMD. Typical refresh rates of modern HMDs range between 80 Hz and 144 Hz, which limit the precision and accuracy with which for example stimulus timing can be controlled and with which time-sensitive measurements can be obtained within the game-loop.

Furthermore, although VR comes with the great promise towards more naturalistic and generalizable experimental designs, currently it is not well understood if and how the VR setting itself might bias behavioral responses differently than the real world. In other words, it is not known if virtual settings engage the same cognitive processes as the equivalent real-word situation (Kulik, 2018; Pan & Hamilton, 2018a; Vasser & Aru, 2020).

If the simple fact that an experiment takes place in VR, results in different behavioral responses than the equivalent real-world setting, the generalizability of research findings from VR studies would need to be questioned. However, the same is obviously true for non-VR experiments, e.g., experiments in which visual stimuli are presented on a computer screen. In both experiments participants enter an

alternate reality, which is different from reality we know from our everyday life (Pan & Hamilton, 2018a).
However, this implies that VR experiments might also engage different cognitive processing than standard non-VR experiments, questioning the transferability of experimental findings from VR to non-VR experiments and vice versa.
Consequently, an important issue to address is whether VR experiments evoke the same behavioral responses as non-VR experiments. Given the technological differences between a standard computer display and the rather complicated optics on an HMD, behavior differences could for example be driven by the stimulus presentation device itself. Furthermore, the current generation of HMDs still suffers from various technological limitations like lenses with a fixed focal length, resulting in side-effects like vergence-accommodation conflicts (Kramida, 2015). However, so far, studies directly comparing monitor-based and VR experiments are rare.

Here, my aim was to examine if reaction time costs induced by unexpected basic visual features differ between visual stimulus presentation with an HDM and a standard monitor. The central idea for this venture was to create two identical experimental setups, with ideally the only difference being that one experiment takes place in VR, while the other one is conducted with a standard non-VR experimental setup and to test if both experiments will yield the same results.

In chapter 1, I developed a new behavioral paradigm, which served for the direct comparison between VR and non-VR setup in chapter 3. In a series of four experiments, I investigated how prediction errors of basic visual stimulus features, such as color and orientation, are formed and on which levels of processing they arise. In the first experiment, participants saw two Gabor patches on a computer screen, of which one feature (color) was manipulated on one grating and another feature (orientation) on the other grating. Expectations were implicitly manipulated by presenting a specific color and orientation more frequently than another one. However, these features were rendered completely task irrelevant. Instead, participants had to differentiate whether the spatial frequencies of both gratings were identical or different.
The results showed that responses were slower when the color or the orientation was unexpected, without an indication for an interaction between the prediction errors. In a second experiment, I tested if there is a mutual influence of both types of predictions errors, when both features belong to the same object. In general, the paradigm was identical to Experiment 1 with the only difference that both colors and orientations were manipulated on each grating simultaneously. Again, the results clearly indicated prediction errors of both features affected the task performance independently, without any evidence for a mutual interaction. In Experiment 3, feature expectations were manipulated explicitly by the means of textural cues indicating the most likely color and orientation in the next trial. Additionally, we were concerned that the task irrelevance of the features resulted in participants not paying enough attention to them and thereby inhibit feature-binding, a process that requires attention according to influential theories like the Feature Integration Theory (FIT) (Treisman & Gelade, 1980). Hence, to increase the relevance of the color and orientation, after each block of 64 trials, participants had to rate the proportion of trials in which the cue correctly predicted upcoming features. Although the analysis of the cue ratings showed that participants paid attention to the cue as well as the features, the results showed again no evidence for an interaction of both prediction errors. In the last experiment, I evaluated if the missing interaction was the result of two peripheral presented objects, requiring participants to divide their attentional resources between the stimuli. In this experiment,

only one central grating was presented, and participants had to judge if the spatial frequency was high or low. As in the previous experiments, I was able to demonstrate the impact on the task performance for both types of prediction errors independently without any evidence for a mutual influence between them. Taken together, the results suggest that prediction errors for object features are formed and resolved on an early level of visual processing, when the features are still processed in parallel. Furthermore, the consistent results of all four experiments demonstrate that the observed effects are robust and replicable when tested with a standard setup. This was crucial for the experiments planned for the study reported in chapter 3, in which I tested if I could replicate the results when the same paradigm is conducted within VR. Without a robust effect, it would become unclear if behavioral differences between the experimental setups would be the result of the different hardware or just the result of a flakey behavioral effect.

As outlined above, most VR experiments in cognitive neuroscience and experimental psychology use game engines for the stimulus creation and data collection. However, game engines come with known limitations like limited sampling rates. For the experimental paradigm developed in chapter 1, a minimum requirement of the stimulus software is the capability to tightly control stimulus timing and to measure precise and accurate reaction times. Hence, in chapter 2 I determined the level of accuracy and precision for both stimulus timing and reaction time measurements when using the combination of Unreal Engine 4 (UE4), SteamVR and the HTC Vive VR system.

In a first experiment, the accuracy and precision of pre-defined stimulus durations were tested. Objective measurements were provided by means of the Black Box Toolkit (BBTK), a specialized device for the validation of several timing parameters and time measurements in behavioral experiments. A white square was presented for a pre-defined duration on the display. A photo-sensor, connected to the BBTK, was used to measure the duration of each stimulus with a sub-millisecond precision. While the stimulus durations turned out highly precise, the measured stimulus durations always exceeded the pre-defined duration a bit. In my tests, I defined stimulus durations in terms of displays refreshes or ticks, which indicates that the observed inaccuracies can be explained by the refresh rate of the HTC Vive. This was confirmed after calculating the exact frequency, with which new frames were presented. Across all my tests, including two different computers and two different HTC Vive HMDs, the number of frames per second (FPS) turned out to be exactly 89.53 FPS, indicating that the HTC Vive has a true refresh rate of 89.53 Hz instead of the officially stated 90 Hz. Overall, the results of the first experiment indicate that the VR setup can present stimuli with a high precision and, when taking the exact refresh rate into account, with a high accuracy.

In a second experiment, I evaluated the precision and accuracy of reaction time measurements of the VR setup and compared it to results I got with the same test procedure with a standard monitor setup and both Presentation and PsychoPy. While I observed precise and accurate RT measurement with both standard setups, reaction times measured with the VR setup were highly inaccurate and imprecise. In the following, I explain the reasons for the discrepancies between expected and observed reaction times based on limitations resulting from the architecture underlying game engines and VR rendering. Furthermore, I proposed a new software-based method that circumvents these limitations. Benchmarking results revealed that the method is capable to measure reaction time with an accuracy and precision, which is on par with both PsychoPy and Presentation.

In chapter 3, I report the final study of this dissertation, which tried to replicate the results obtained in Experiment 2 of chapter 1, in both a standard non-VR experimental setup as well as in VR, with the same group of participants.
Results from previous studies, which directly compared a VR and a non-VR experimental setup, indicate that HMDs might affect how participants allocate attentional resources (Li et al., 2020) and spatial processes (e.g., Anglin et al., 2017) as well as impairs motor-learning and increases cognitive load (Juliano et al., 2021). However, so far, no study has investigated on which level of processing these differences arise. Here, I aimed to determine if modern HMDs already affect early visual processing levels differently than a typical monitor setup.
Furthermore, I was especially interested in potential hardware-related differences. As mentioned above, the optical system of HMDs is rather complex, when compared to an ordinary computer screen. HMDs contain lenses with a fixed-focal length, resulting in incorrect depth cues and a conflict between vergence and accommodation (Kramida, 2015). Previous research indicates that vergence-accommodation conflicts impair visual performance (Hoffman et al., 2008) and might reduce attentional resources, to compensate for the incongruent depth cues and remain clear vision (Daniel & Kapoula, 2019). Similarly, the weight of HMDs and the limited FOV have been found to affect head and eye coordination (Pfeil et al., 2018).
However, the discrepant findings between VR and non-VR reported by some studies could be the result of inconsistencies of the stimulus presentation between the VR and non-VR conditions. For example, Li et al. (2020) found increased attention allocated toward stimuli presented in an HMD as compared to the corresponding monitor version. The stimulus material used in their study involved a small three-dimensional scene, which was presented in stereoscopic 3D in the VR condition, while the stimuli were presented monocular in the non-VR version. This leaves open the question if the increased allocation towards stimuli in VR originate from technological particularities inherent to HMDs or if the effects were merely driven by differences in how the stimuli were presented, i.e., 3D vs 2D.
In the study reported in chapter 3, I aimed to avoid inconsistencies of the stimulus material as well as the context in which it was presented, by achieving the highest possible correspondence between the VR and the non-VR experiment.
Hence, I aimed to design the experimental setup of the VR experiment as close as possible to the non-VR experiment. A virtual environment (VE) was developed that matched the visual appearance of the physical laboratory, in which both experiments took place. However, the virtual replicate corresponded with the physical laboratory also 1:1 with respect to scale and location. This was used to present the VE as an overlay blended on top of the real laboratory. Consequently, everything visible through the HMD did also exist physically, providing for example realistic and consistent haptic and tactile stimulation.
Having two identical experimental setups, a group of 16 participants was tested in each version in counterbalanced order. The results showed for both experiments the same reaction time pattern, which I already observed in chapter 1, indicating independent processing of prediction errors for unexpected colors and orientation. In particular, the statistical analysis did not provide any evidence for behavioral differences between the VR and the non-VR experiment. The findings indicate that early visual feature processing is not differently affected by the stimulus-presentation device.

Lastly, in chapter 4 I will discuss some of the main findings reported in a bit broader context and discuss some ideas in more detail, which did not find its way in the original papers.

Chapter 1 – Combined Expectancies

Wiesing, M., Fink, G. R., Weidner, R., & Vossel, S. (2020). Combined expectancies: the role of expectations for the coding of salient bottom-up signals. *Experimental brain research*, *238*(2), 381-393.

Author contributions

MW, S.V., and R.W. conceptualized and designed the research; MW collected the data; wrote software, analyzed and visualized the data; S.V. and R.W. supervised the research project; MW, S.V., G.R.F., and R.W. wrote the manuscript.

Experimental Brain Research (2020) 238:381–393
https://doi.org/10.1007/s00221-019-05710-z

RESEARCH ARTICLE

Combined expectancies: the role of expectations for the coding of salient bottom-up signals

Michael Wiesing[1] · Gereon R. Fink[1,2] · Ralph Weidner[1] · Simone Vossel[1,3]

Received: 17 June 2019 / Accepted: 12 December 2019 / Published online: 13 January 2020

Abstract

The visual system forms predictions about upcoming visual features based on previous visual experiences. Such predictions impact on current perception, so that expected stimuli can be detected faster and with higher accuracy. A key question is how these predictions are formed and on which levels of processing they arise. Particularly, predictions could be formed on early levels of processing, where visual features are represented separately, or might require higher levels of processing, with predictions formed based on full object representations that involve combinations of visual features. In four experiments, the present study investigated whether the visual system forms joint prediction errors or whether expectations about different visual features such as color and orientation are formed independently. The first experiment revealed that task-irrelevant and implicitly learned expectations were formed independently when the features were separately bound to different objects. In a second experiment, no evidence for a mutual influence of both types of task-irrelevant and implicitly formed feature expectations was observed, although both visual features were assigned to the same objects. A third experiment confirmed the findings of the previous experiments for explicitly rather than implicitly formed expectations. Finally, no evidence for a mutual influence of different feature expectations was observed when features were assigned to a single centrally presented object. Overall, the present results do not support the view that object feature binding generates joint feature-based expectancies of different object features. Rather, the results suggest that expectations for color and orientation are processed and resolved independently at the feature level.

Keywords Probabilistic context · Feature expectancies · Feature binding · Object binding · Prediction error

Introduction

Perception is not passive and not exclusively determined by the physical properties of sensory stimuli. Rather, it is affected by internal settings such as prior beliefs and probabilistic expectations about upcoming sensory events (Von Helmholtz 1867; Gregory 1997). Prior knowledge in form of experience-based expectancies about behaviorally relevant stimulus features alters how fast and accurately visual stimuli can be detected and perceived. Stimulus features or object properties that are consistent with prior expectations lead to a faster and more accurate stimulus detection, whereas performance is impaired when these features or properties are inconsistent and hence violate current expectations (Dombert et al. 2016b; Kuhns et al. 2017). Expectations not only facilitate stimulus detection (Stojanoski and Niemeier 2015), but also affect object recognition and enhance perceptual sensitivity (Wyart et al. 2012; Stein and Peelen 2015). Such expectations can be induced by either varying the frequency of occurrence of different features in an experiment or by cues that indicate certain features with a specific probability. Rarely occurring or invalidly cued features are assumed to elicit prediction error signals that slow down response times. Compelling evidence for the influence of expectations and prediction error signaling also comes from neuroimaging and electrophysiological studies, in which the perception of unexpected stimuli produces larger neuronal responses than expected ones (Mars et al. 2008; Summerfield and Egner 2009; Kok et al. 2012; Richter et al. 2017; Stefanics et al. 2018).

Communicated by Melvyn A. Goodale.

Ralph Weidner and Simone Vossel have contributed equally to this work and thus share the senior authorship.

✉ Michael Wiesing
mi.wiesing@fz-juelich.de

Extended author information available on the last page of the article

Behavioral studies suggest that expectations can be formed about single visual features such as color or orientation (Cheadle et al. 2015; Dombert et al. 2016a; Jabar et al. 2017). However, expectations may not only concern single features but also refer to fully integrated object representations. For instance, contextual probabilities can affect perceptual performance, such that object recognition is facilitated when an object is perceived within a scene that is typical for that particular object (e.g., a couch in the living room) compared to when an object is embedded in an untypical environment (e.g., a couch on the beach) (Bar 2004; Summerfield and Egner 2009; Zhao et al. 2013). A key question is how expectations from different processing levels relate to each other. The brain may form expectations simultaneously and independently on different processing levels and within different processing modules. One possibility is that expectations concerning one entity of visual information would be unaffected by expectations formed about other aspects of visual information.

Evidence for this assumption comes from a recent fMRI study by Stefanics et al. (2019). The authors tested whether the same physical stimulus produced distinct feature-specific prediction errors for the color and emotional expression of faces. Their results suggest that violations of different feature expectations are processed in different brain regions and do not interact when the features are unattended and task-irrelevant.

Alternatively, expectations from various processing levels may be combined to form a joint expectation when they are perceptually related, for instance, when they refer to the same object. Recent evidence in favor of the latter assumption comes from a study by Jiang, Summerfield and Egner (2016). The authors performed a behavioral experiment in which colored moving dots were presented. The dots could be either red or green and move upwards or downwards. An auditory cue indicated the upcoming color and movement direction with a validity of 75%, and participants were instructed to attend to either the color or motion of the stimuli presented. In particular, participants were asked to identify the color or the motion direction of the dots. Three competing hypotheses were tested. First, expectations about color and motion operate independently, and the violation of one feature expectation does not affect the other ("independence model"). Second, a prediction error for one feature spreads to the other due to an expectation at the object level ("reconciliation model"). Third, the conflict between expected and unexpected features will result in the perception that both features do not belong to the same object and produce segregated representations for each feature ("segregation model"). The results suggested that expectancy-related reaction time benefits did affect not only the attended but also the unattended feature, thereby supporting the assumptions of the reconciliation model that prediction errors referring to the same object are combined. Yet, at least in principle, prediction errors emerging within different visual dimensions may interact irrespective of whether or not they are bound to the same object. Testing this hypothesis requires an experimental variation assigning prediction errors from different dimensions to different objects. Accordingly, feature expectations were manipulated on the same or different objects in the current study. To prevent potential confounding effects originating from response-consistent perceptual and motor expectations, feature expectations induced in the present experiment were defined as task-irrelevant and were, hence, not related to any particular motor response.

We hypothesized that for combined object-level expectancies (i.e., for expectancies about features on the same object), the simultaneous violation of two feature expectations would result in an interaction of both prediction errors, which would reflect a mutual influence of both prediction error signals in that the joint prediction error signal is less or more than the sum of its parts. In contrast, the prediction errors for both features should be independent (i.e., additive) and not interact when the features are distributed to separate objects.

Experiment 1

Experiment 1 was conducted to determine whether it is possible to manipulate feature expectations independently on two task-irrelevant dimensions when the features are separated on different objects. We presented two sinusoidal grating stimuli, where one feature (color) was manipulated on one grating and the other feature (orientation) on the other grating.

Expectations were manipulated by presenting specific feature configurations more frequently than others, assuming that the biased probabilities of the features of the target objects would be learned implicitly. This setup resulted in four experimental conditions: (1) color and orientation expected, (2) color expected and orientation unexpected, (3) color unexpected and orientation expected, and (4) color and orientation unexpected.

The participants' task was focused on yet a further stimulus feature and they were asked to report whether the spatial frequencies of both target gratings were identical or different. Thereby, the participants' task required them to keep track of and to respond to this third dimension (e.g., frequency), so that it was orthogonal to the expectations related to color and orientation.

We hypothesized that violations of feature expectations for color and orientation will affect behavior (response times) independently (i.e., additively), and will, hence, not interact even when both feature expectations are violated simultaneously.

Materials and methods

Participants

Sixteen participants (five women, mean age: 29.2 years, age range: 19—46, one left-handed) took part in Experiment 1. All participants had a normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Normal color vision in all participants was assessed by pseudo-isochromatic color plates (Velhagen and Broschmann 2003). Before the experiment, written informed consent was obtained following the Declaration of Helsinki. The study was approved by the ethics committee of the German Society of Psychology, and participants were remunerated for their time.

Apparatus

Stimuli were presented on a 22-in. Samsung SyncMaster monitor (spatial resolution 1680 X 1050; refresh rate 120 Hz) at a distance of 60 cm. A chin and forehead rest preserved the distance. The presentation of stimuli and response recording were controlled using PsychoPy psychology software for Python (Peirce 2007, 2008). Participants were provided with button response pads (NAtA Technologies) for each hand and responded by pressing the corresponding button on the button response pad with the left and right index fingers.

Stimuli and task

The visual stimuli consisted of two horizontally arranged target stimuli (see Fig. 1). A central black plus sign (0.57° × 0.57°) was placed in between serving as a fixation point. All stimuli were presented on a gray background. Participants were instructed to fixate the cross throughout the experiment.

The target stimuli were grating stimuli which consisted of a 4°×4° sine wave grating windowed by a two-dimensional Gaussian envelope with a standard deviation of 0.66° with two possible spatial frequencies (low frequency: 1.5 cycles per degree and high frequency: 2.5 cycles per degree). All combinations of frequencies were presented randomly and with an equal probability (e.g., 50% same and 50% different).

Furthermore, one grating (color target) was always colored (either red/green or blue/yellow) and was oriented vertically (0°). The other grating (orientation target) was in grayscale, but could have two different orientations (45°, 90°). The side on which the color and the orientation targets were presented was held constant during the experiment (e.g., color was always left and orientation always right), but was counterbalanced across participants.

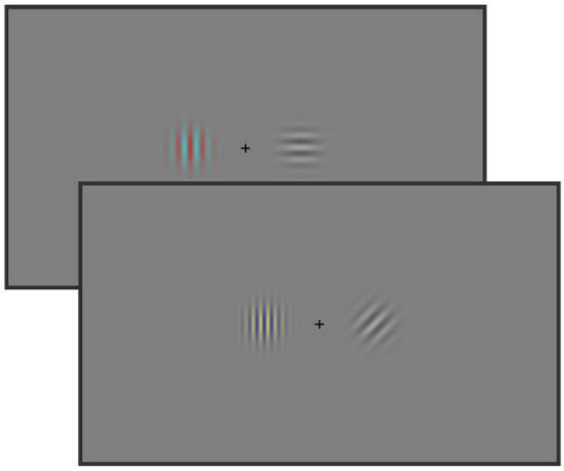

Fig. 1 Stimulus examples of Experiment 1. The participants were asked to indicate by button presses whether the two gratings had the same or different spatial frequency. The probabilities of occurrence of the colors of one grating and the orientations of the other grating were manipulated to induce feature expectations

For the color target, one color combination (e.g., blue/yellow) was defined as the "expected color" and the other combination as "unexpected color". Likewise, one orientation (e.g., 45°) was defined as the "expected orientation" and the other orientation as "unexpected orientation". For both, color and orientation, the expected feature was presented on 87.5% of the trials.

A 2×2 factorial design with the factors Color Prediction Error (high, low) (*ColPE*) and Orientation Prediction Error (high, low) (*OriPE*) resulted in four experimental conditions: *ColPE_low/OriPE_low* (color expected and orientation expected), *ColPE_high/OriPE_low* (color unexpected and orientation expected), *ColPE_low/OriPE_high* (color expected and orientation unexpected), and *ColPE_high/OriPE_high* (both color and orientation unexpected).

The experiment consisted of 14 blocks, each comprising 64 trials, resulting in 896 trials. The experiment comprised 700 *ColPE_low/OriPE_low* trials (78.125%), 84 *ColPE_high/OriPE_low* trials (9.375%), 84 *ColPE_low/OriPE_high* trials (9.375%), and 28 *ColPE_high/OriPE_high* trials (3.125%).

Each trial started with the presentation of the two target stimuli until a response was given. An inter-trial interval, which randomly varied between 500 and 1000 ms, separated the trials.

Each block was followed by a break that could be terminated via button press.

The participants' task was to indicate whether both target stimuli were identical or different concerning spatial frequency by pressing the corresponding response button with the left or right index fingers. The task was independent of the expectation manipulations of color and orientation, to avoid any confounding effects of response preparation to the features. Participants were asked to respond as fast and accurately as possible. An erroneous response produced the message "Fehler" (i.e., the German word for "error") on the screen for 750 ms.

To familiarize the participants with the task, participants performed a training session of 128 trials before they started with the experiment. During the training, all trials were *ColPE_low/OriPE_low* trials. This was intended to let the participants form expectations about the most likely color and orientation of the target stimuli.

All participants were informed that the color and the orientation could change during the main experiment. Furthermore, they were told that color or orientation changes are irrelevant to their task.

Analysis

The free statistical software R (R Foundation for Statistical Computing, Vienna, Austria; https://www.r-project.org) was used for behavioral data analysis. For each participant, mean RTs and error rates were calculated. Error trials, and trials following errors and trials with RTs differing more than two standard deviations from the mean were excluded from RT analysis.

Repeated-measures ANOVAs for the RTs and error rates were conducted with the within-subject factors *ColPE* (high, low) and *OriPE* (high, low). The reported mean values for expected and unexpected color and orientation were calculated by collapsing all trials with the specific feature being expected or unexpected (e.g., the mean values for the expected color reflect the mean of all *ColPE_low/OriPE_low* and *ColPE_low/OriPE_high* trials).

Results

The overall amount of incorrect responses was very low with on average 2.19% (±0.34 SEM) errors. The ANOVA of the error rates yielded a significant main effect of *OriPE* ($F(1,15) = 5.025$, $p < 0.05$, $\eta_p^2 = 0.186$) with lower error rates for expected orientations (2.04%) compared to unexpected orientations (3.29%). Neither the main effect *ColPE*, with 2.02% errors for expected colors versus 3.40% errors for unexpected colors, was significant ($F(1,15) = 3.437$, $p = 0.0835$, $\eta_p^2 = 0.251$), nor was the interaction between *ColPE* and *OriPE* ($F(1,15) = 1.453$, $p = 0.247$, $\eta_p^2 = 0.088$).

The ANOVA of the mean RTs revealed a significant main effect for *ColPE* ($F(1,15) = 13.31$, $p < 0.05$, $\eta_p^2 = 0.470$) with 662 ms for expected colors versus 682 ms for unexpected colors, reflecting RT costs for the unexpected colors. Moreover, we observed a significant main effect for *OriPE* ($F(1,15) = 5.735$, $p < 0.05$, $\eta_p^2 = 0.277$), with 660 ms for expected orientations versus 692 ms for unexpected orientations, reflecting a cost for unexpected orientation. Thus, both unexpected colors and unexpected orientations resulted in significantly higher RTs compared to the expected features, although both features were task-irrelevant and orthogonal to the actual task. The interaction of *ColPE* × *OriPE* was not significant ($F(1,15) = 0.401$, $p = 0.536$, $\eta_p^2 = 0.026$), providing no evidence that prediction errors for both features influenced each other. The mean RTs and error rates are shown in Fig. 2.

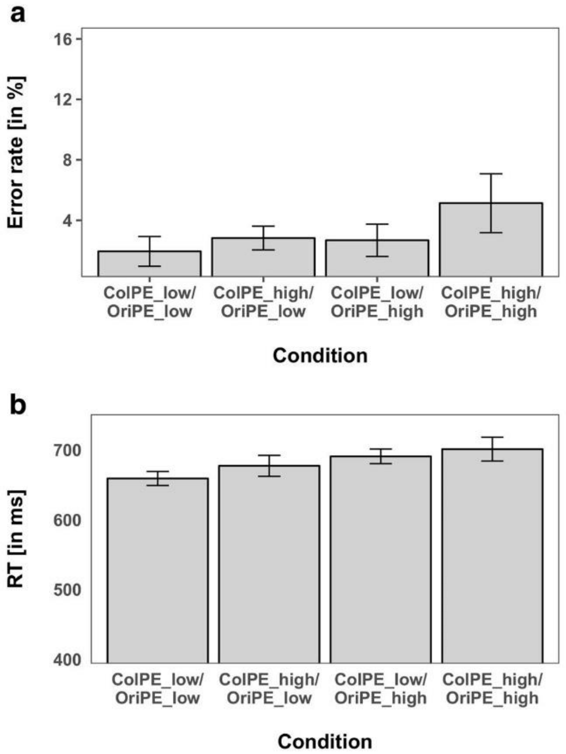

Fig. 2 Performance measures of each combination of color and orientation manipulations of Experiment 1. **a** Error rates. **b** Reaction times. Error bars reflect the 95% confidence intervals

Discussion

Experiment 1 investigated whether two simultaneous feature expectations can be manipulated independently. We hypothesized that feature expectation would be processed independently when the features are separated on different objects. Consistent with the hypothesis, the experiment provides evidence that separated feature expectations are processed independently.

When color and orientation features were expected due to their higher probability of occurrence, participants responded faster than in trials with one unexpected feature. Response times increased further in trials with two unexpected features compared to one unexpected. However, the RTs for two unexpected features increased additively rather than interactively, indicating independent effects for both features. These effects were observed despite color and orientation being irrelevant for the task of the subjects.

To investigate whether multiple simultaneous feature expectations of the same object result in a combined object-level expectation, we conducted a second experiment.

Experiment 2

Experiment 2 was designed to determine whether concurrent color and orientation expectancies combine interactively when both features are bound to the same object. In Experiment 1, the color and orientation features were distributed to separate objects, which, consequently, did not result in an interaction of both types of prediction errors. Experiment 2 followed the procedure of Experiment 1, except that in Experiment 2, both color and orientation expectations were manipulated on both gratings simultaneously. We hypothesized that simultaneous violations of both feature expectations for color and orientation within the same object would result in an interaction, indicating combined object-level expectancies.

Materials and methods

Participants

Sixteen participants (eight women, mean age: 29.69 years, age range: 20–45, two left-handed) took part in Experiment 2. Seven of them already participated in Experiment 1. All participants had a normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Normal color vision in all participants was assessed by pseudo isochromatic color plates (Velhagen and Broschmann 2003). Before the experiment, written informed consent was obtained following the Declaration of Helsinki. The study was approved by the ethics committee of the German Society of Psychology, and participants were remunerated for their time.

Stimuli, design, and procedure

In general, procedures were similar to Experiment 1. Again, the participants' task was to indicate whether the spatial frequencies of two gratings stimuli were identical or different. In contrast to Experiment 1, however, both color and orientation expectation were manipulated on the two gratings simultaneously (see Fig. 3). The probabilities for the expected and unexpected features were identical to those in Experiment 1.

Like in Experiment 1, participants performed a training session of 128 trials with 100% expected features before they started with the experiment.

Analysis

In general, the analysis of Experiment 2 was identical to Experiment 1. In an additional analysis, we compared data across Experiment 1 and Experiment 2, to test whether simultaneous prediction errors of color and orientation on the same objects resulted in a combined expectancy effect. Therefore, for both experiments, the mutual interaction of prediction errors in the different dimensions was estimated by calculating an interaction score (ISC) that contrasts

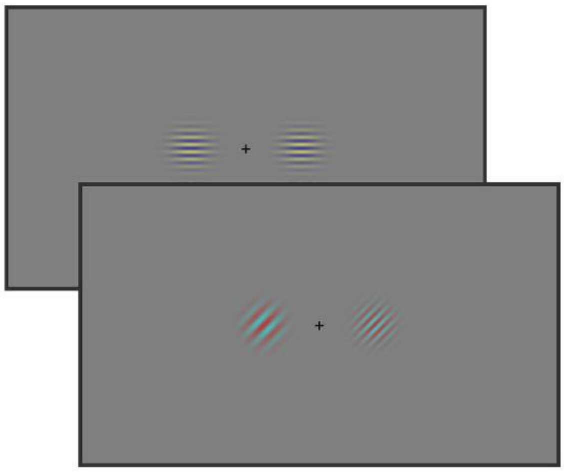

Fig. 3 Stimulus examples of Experiment 2. As in Experiment 1, participants were asked to respond to the spatial frequency of the two gratings, which could be the same or different. The probabilities of occurrence of both color and orientation of the identical gratings were manipulated to induce feature expectations

the effects of simultaneous high prediction errors in both dimensions with high prediction errors in only one dimension. Trials with only a prediction error in one dimension (*ColPE_low/OriPE_high and ColPE_high/OriPE_low*) were subtracted from the sum of trials with consistent prediction errors in both dimensions *(ColPE_high/OriPE_high and ColPE_low/OriPE_low)*. In particular, the calculation can be expressed by the interaction score ISC = *(ColPE_high/OriPE_high + ColPE_low/OriPE_low)—(ColPE_high/OriPE_low + ColPE_low/OriPE_high)*. Please note that in both terms, the same number of high and low prediction errors are involved, the only difference being that in one term prediction errors occurred simultaneously. In case there is absolutely no interaction (i.e., when the effects of a prediction error in one dimension is independent of a prediction error in another visual dimension), the ISC would be expected to be 0. Otherwise, if prediction errors in both dimensions jointly generate a higher prediction error, the ISC value will be positive. ISC values smaller than zero would be indicative of reduced prediction errors when both belong to the same object.

This calculation was conducted for RTs as well as for error rates. The resulting scores from both experiments were then compared using a *t* test for partially depending samples (Derrick et al. 2017).

Results

Similar to Experiment 1, the mean error rate was very low with an average of 3.28% (±0.38 SEM).

The ANOVA of the error rates yielded a significant main effect for *ColPE* ($F(1,15) = 9.785$, $p < 0.05$, $\eta_p^2 = 0.395$) with lower error rates for expected colors (2.97%) compared to unexpected colors (5.47%) and a significant main effect for *OriPE* ($F(1,15) = 11.65$, $p < 0.05$, $\eta_p^2 = 0.437$) with lower error rates for expected orientations (2.95%) compared to unexpected orientations (5.58%). The interaction was not significant ($F(1,15) = 0.904$, $p = 0.357$, $\eta_p^2 = 0.057$).

Again, the ANOVA of the mean RTs revealed a significant main effect for *ColPE* ($F(1,15) = 8.035$, $p < 0.05$, $\eta_p^2 = 0.349$), with 569 ms for expected colors versus 591 ms for unexpected colors, and a significant main effect for *OriPE* ($F(1,15) = 5.778$, $p < 0.05$, $\eta_p^2 = 0.278$), with 569 ms for expected orientations versus 593 ms for unexpected orientations, reflecting the cost for unexpected features. Thus, both unexpected color and unexpected orientation stimuli resulted in significantly higher RTs compared to the more frequently presented standard targets. The interaction of $ColPE \times OriPE$ was not significant ($F(1,15) = 0.329$, $p = 0.575$, $\eta_p^2 = 0.021$), providing no evidence for a mutual influence of prediction errors for both features. Both feature prediction errors were hence again processed independently, even when they were part of the same object. The mean RTs and error rates are shown in Fig. 4.

A joint analysis of the combined data of Experiment 1 and Experiment 2 revealed no significant difference concerning the interaction score ISC of error rates between experiments ($t(19.5) = -0.12$, $p = 0.909$), with an ISC of 1.57 in Experiment 1 and 1.79 in Experiment 2. Similarly, the comparison of the ISC related to RT costs revealed no significant difference between experiments ($t(19.5) = -0.15$, $p = 0.879$), with an ISC of -8 in Experiment 1 and -5 in Experiment 2.

Discussion

Experiment 2 was designed to determine whether concurrent color and orientation expectancies are combined interactively when both features are bound to the same object. We hypothesized that violations of both feature expectations simultaneously would result in mutual influence and hence in an interaction.

Consistent with Experiment 1, the RT costs associated with one feature being unexpected were also evident in Experiment 2 (as evidenced by the significant main effects

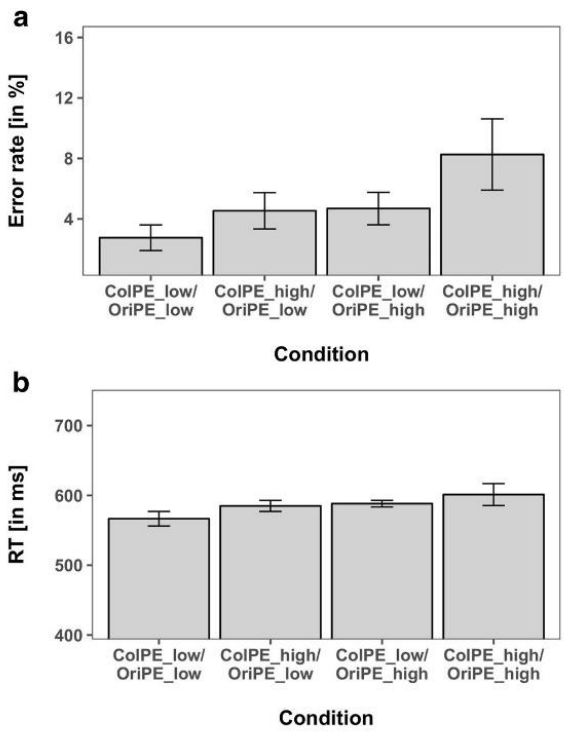

Fig. 4 Performance measures of the combination of color and orientation manipulations of Experiment 2. **a** Error rates. **b** Reaction times. Error bars reflect the 95% confidence intervals

for *ColPE* and *OriPE*). Contrary to our hypothesis, the results of Experiment 2 did not show an interaction between *ColPE* and *OriPE*, providing no evidence for a mutual influence of both types of prediction errors. This finding was confirmed by a direct comparison of the interaction pattern, as reflected in ISC determined in Experiment 1 and Experiment 2. In Experiment 1, the feature expectations for color and orientation were separated onto two different objects while the same feature expectations were combined and manipulated on the same objects simultaneously in Experiment 2. We hypothesized that combined object-level expectancies for color and orientation should be reflected in different ISCs related to the RT costs and error rates between experiments. However, this comparison revealed no significant difference and, again, did not provide any evidence for combined object-level expectancy effects.

These findings do not support the idea of combined feature expectancies on an object level, at least when expectancies are formed implicitly and when they relate to visual dimensions that are currently irrelevant for an ongoing task and are hence unattended. Attention has previously been suggested to play an essential role in feature binding. According to Treisman's Feature Integration Theory (FIT), basic visual features are first processed independently on a preattentive level before attentional binding combines them into a single-object representation (Treisman and Gelade 1980). Following this argumentation, one may expect that without attentional binding, not only feature representations per se, but also related expectancies are coded separately. Conversely, when these features are attended and are hence integrated to whole object representations via attentional binding, then expectancies may also be formed by combined feature information. To investigate whether the lack of interaction was indeed due to the lack of attentive processing of the implicitly learned target features, a third experiment was conducted.

Experiment 3

The results of Experiment 2 indicated that multiple feature expectations for task-irrelevant features influence behavior additively rather than interactively, even when the features belong to the same object. This result seems to be in line with the findings of Stefanics et al. (2019) who did not find an interaction of prediction errors of different features, when the features were unattended. Thus the absence of an interaction may be accounted for by a lack of attention assigned to the different features and hence by a lack of object binding in the current experiments. Therefore, Experiment 3 was designed to test whether the absence of an interaction between feature expectations was due to a lack of object binding. To promote object feature binding, we increased the subjects' need to attend to the gratings' color and orientation explicitly. Instead of manipulating feature probabilities of upcoming targets by their frequency of occurrence, we induced expectations of the upcoming feature configuration explicitly on a trial-by-trial basis using verbal cues.

Furthermore, we added a secondary task to the experiment, where participants had to estimate the percentage of cue validity after each experimental block. This approach allowed increasing the feature relevance without an association between the main task and a specific motor response.

Materials and methods

Participants

Sixteen participants (six female, mean age: 30.7 years, age range: 21–46, two left-handed) took part in Experiment 3. Seven of them had participated in the previous experiments, and one had participated in Experiment 2. All participants had a normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Normal color vision in all participants was assessed by pseudo-isochromatic color plates (Velhagen and Broschmann 2003). Before the experiment, written informed consent was obtained following the Declaration of Helsinki. The study was approved by the ethics committee of the German Society of Psychology, and participants were remunerated for their time.

Stimuli, design, and procedure

In general, procedures were very similar to Experiment 2. The participants' task was again to indicate whether the spatial frequencies of two gratings were identical or different, and the occurrence of different colors and orientations was manipulated for both gratings simultaneously. However, in contrast to Experiment 1 and 2, the probability for each color and orientation was identical, e.g., 50% red/green and 50% blue/yellow, as well as 50% horizontal and 50% tilted orientation. Each trial started with a verbal cue, indicating the most likely color and orientation for the next trial for 583 ms. Explicit cues comprised the written words of the likely upcoming color ["gelb" or "rot" (i.e., "yellow", "red")] and the likely upcoming orientation ("horizontal" or "diagonal"). To keep the information provided by the cue minimal, only one color of the color combinations was cued (e.g., "yellow" instead of "blue/yellow"). The cue validity was equivalent to the feature probabilities of Experiments 1 and 2. Hence, the cue was valid in 78.125% of the trials for both features; in 9.375%, the color information was invalid; likewise in 9.375% of the trials, the cue was invalid regarding the orientation, and in 3.125% of the trials, both the color and the orientation information were invalid. The number of valid and invalid cues was the same for all blocks. The cue and target stimuli were separated by a cue–target interval, which randomly varied between 250 and 500 ms.

Additionally, to increase the relevance of the color and orientation features, a secondary task was added to the experiment. Participants were asked to report their belief about the cue validity at the end of each block, using a rating scale ranging from 0 to 100% (in 5% steps). Participants were not aware that the overall validity was the same for all the blocks.

As in the previous experiments, participants performed a training session of 128 trials with 100% expected features before the experiment.

Analysis

In general, the analysis of Experiment 3 was identical to Experiment 2. Again, the results of Experiment 3 were compared with the previous experiment's results. Additionally, to test whether explicitly formed expectations affect behavior differently than implicit formed expectations, RT costs and error rates for ColPE [*(ColPE_high/OriPE_low)–(ColPE_low/OriPE_low)]* and OriPE [*(ColPE_low/OriPE_high)–(ColPE_low/OriPE_low)]* were compared between Experiment 2 and Experiment 3.

To test whether participants perceived the verbal cues as valid, two one-sample t tests were conducted. First, to test whether participants assigned any predictive value to the verbal cues at all, we tested whether the estimated cue validities were better than chance level (i.e., significantly different from 50%). Second, to test whether participants were able to infer the actual validity, estimated cue validity values were compared against the true cue validity of 78.125%.

Results

Participants estimated cue validity for all blocks on average at 67.46% (±2.33% SEM). A one-sample t test indicated that this value was significantly different from chance level (50%) ($t(15)=7.482$, $p<0.05$). However, cue validity as perceived by the participants was also lower than the actual cue validity of 78.125% as indicated by a second one-sample t test ($t(15)=-4.573$, $p<0.05$).

Similar to the previous experiments, the mean error rate was low, with an average of 4.65% (±0.65 SEM). The ANOVA of the error rates showed neither significant main effects of *ColPE* ($F(1,15)=4.028$, $p=0.0631$, $\eta_p^2=0.211$), with 4.23% for expected colors versus 7.59% for unexpected colors, nor a significant main effect of *OriPE* ($F(1,15)=1.219$, $p=0.287$, $\eta_p^2=0.075$), with 4.43% for expected orientations versus 6.14% for unexpected orientations. The interaction between the two factors was not significant ($F(1,15)=0.486$, $p=0.496$, $\eta_p^2=0.314$).

As in the previous experiments, the ANOVA of the mean RTs revealed significant main effects for *ColPE* ($F(1,15)=14.31$, $p<0.05$, $\eta_p^2=0.488$), with 612 ms for expected colors versus 676 ms for unexpected colors, and *OriPE* ($F(1,15)=10.08$, $p<0.05$, $\eta_p^2=0.402$), with 613 ms for expected orientations versus 676 ms for unexpected orientations, reflecting the cost for unexpected stimuli. Similar to the previous experiments, both unexpected color and orientation stimuli resulted in significantly higher RTs compared to the standard targets. Likewise, the interaction of *ColPE* × *OriPE* was again not significant ($F(1,15)=0.993$, $p=0.335$, $\eta_p^2=0.062$). The mean RTs and error rates are shown in Fig. 5.

A joint analysis of the combined data from Experiment 2 and Experiment 3 revealed no significant difference in error rates for color expectation between experiments [Experiment 2: 2.76% (expected) vs. 4.54% (unexpected); Experiment 3: 4.06% (expected) vs. 7.51% (unexpected); $t(18.5)=-1.65$, $p=0.116$] However, there was a significant difference in error rates for orientation expectation between experiments [Experiment 2: 2.76% (expected) vs. 4.69% (unexpected);

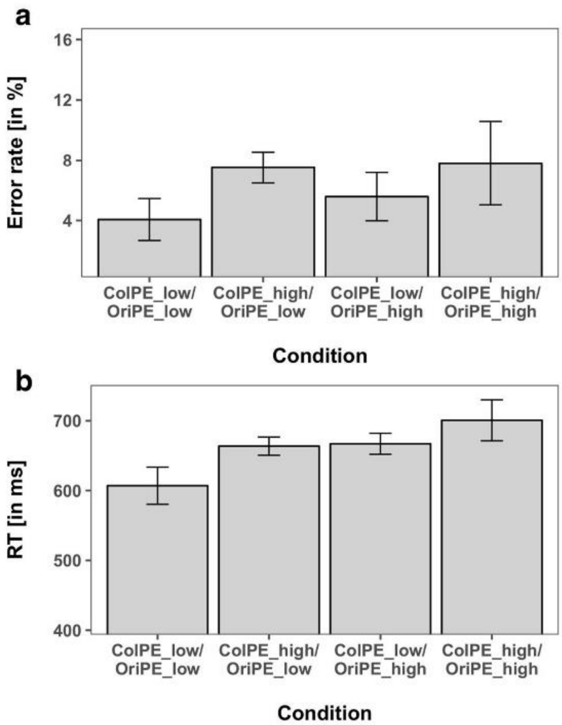

Fig. 5 Performance measures of the combination of color and orientation manipulations of Experiment 3. **a** Error rates. **b** Reaction times. Error bars reflect the 95% confidence intervals

Experiment 3: 4.06% (expected) vs. 5.58% (unexpected); $t(18.5) = 0.49$, $p < 0.05$]. Additionally, there was no significant difference in error rates between the experiments for simultaneous prediction errors of color and orientation ($t(18.5) = 1.20$, $p = 0.244$), with regard to the ISC of error rates between experiments, with an ISC of 1.79 in Experiment 2 and 1.22 in Experiment 3.

Comparing RTs from Experiment 2 with those from Experiment 3 revealed a significant difference in RT costs for unexpected colors between the experiments ($t(18.5) = -2.39$, $p < 0.05$), with unexpected colors resulting in an RT increase of 18.36 ms in Experiment 2 versus an increase of 56.72 ms in Experiment 3. The comparison of RTs for the orientation expectation between experiments was non-significant ($t(18.5) = -1.87$, $p = 0.077$), with unexpected orientation resulting in 21.65 ms higher RTs in Experiment 2 versus an increase of 60.08 ms in Experiment 3. The comparison of RT costs for simultaneous prediction errors of color and orientation between experiments was non-significant concerning the ISC-related RT costs ($t(18.5) = 0.92$, $p = 0.369$), with an ISC of -5 in Experiment 2 and −23 in Experiment 3.

Discussion

Experiment 3 was conducted to examine whether the lack of an interaction between the feature expectations observed in Experiment 2 was due to the implicitly learned probabilities of the task-irrelevant features that may have prevented object feature binding. To increase the demands for participants to attend to the color and orientation features, we manipulated feature expectations in Experiment 3 explicitly on a trial-by-trial basis through verbal cues. Furthermore, to make the features relevant, a secondary task was added to the experiment. Participants were asked to report their estimate about the cue validity after every experimental block.

The estimated validity differed from the true cue validity of 78.125%, indicating that participants underestimated cue validity. However, participants noticed the predictive value of the cue as indicated by validity estimates different from chance level. Not only that they perceived cues as valid, they also used the cue information to prepare for upcoming stimulus configurations, as can be seen from the reaction time pattern. In particular, participants responded faster in trials involving validly cued features rather than invalidly cued features, reflecting the RT costs for invalidly cued and hence unexpected features. This pattern is consistent with the reaction time pattern observed in the previous experiments where participants responded faster in trials with expected rather than unexpected features. Hence, the findings from Experiment 3 indicate that the verbal cues successfully manipulated feature expectations. Additionally, as in the previous experiments, RTs for two unexpected features increased additively rather than interactively, providing no evidence for joint expectations regarding both features. This result suggests that the effects observed in Experiment 2 were not specific to the implicit nature of feature expectancy.

To test whether the explicitly formed feature expectations affected the processing of prediction error concerning these expectations differently than the implicitly learned expectations of Experiment 2, we compared the results of both experiments. The analysis revealed significantly higher RT costs for unexpected colors in Experiment 3 than in Experiment 2, but not for unexpected orientations. Furthermore, unexpected orientations but not unexpected colors resulted in significantly higher error rates in Experiment 2 than in Experiment 3. This finding shows that explicitly formed expectations affect the behavior differently than implicitly formed expectations. This effect was not consistent across both visual dimensions, and the behavioral pattern indicated an asymmetry about explicitly formed expectations, such that the effects of prediction errors based on explicitly formed expectations increased for color and decreased for orientation. This result might either reflect an attentional bias towards color in Experiment 3, rendering prediction errors in the color domain more relevant. Alternatively, predictions concerning orientation may be formed more implicitly than explicitly.

Experiment 4

The previous experiments indicated that expectations regarding different visual features affect behavior additively rather than interactively. This was true even when these features belonged to the same object. Moreover, this effect was likewise observed for implicit and explicit expectations. The task performed in these experiments required participants to simultaneously attend to two objects, thereby reducing the attentional resources available for each single object. Since attention has been suggested to be a necessary prerequisite for feature binding (Treisman and Gelade 1980), divided attention might possibly decrease the degree of feature binding and accordingly potential interactions between feature expectations. This interpretation seems consistent with findings from previous studies. Evidence for a separate coding of feature expectations has been reported in a study using multiple non-foveated stimuli (Stefanics et al. 2019), whereas a study using a single central stimulus provided evidence in favor of an interaction between feature predictions (Jiang et al. 2016). In Experiment 4, we, therefore, tested whether an interaction between feature expectations would be observed when feature binding was maximized by manipulating feature expectations of a single central grating.

Materials and methods

Participants

Sixteen participants (seven female, mean age: 32.6 years, age range: 23–47, two left-handed) took part in Experiment 4. All participants had a normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Normal color vision in all participants was assessed by pseudo-isochromatic color plates (Velhagen and Broschmann 2003). Before the experiment, written informed consent was obtained following the Declaration of Helsinki. The study was approved by the ethics committee of the German Society of Psychology, and participants were remunerated for their time.

Stimuli, design, and procedure

In general, procedures were similar to Experiment 2. However, in contrast to Experiment 2, only a single central grating with a fixation point was presented in Experiment 4 (Fig. 6). Moreover, the task this time was to indicate whether the spatial frequency of the grating was either high or low. The probabilities of the different color and orientation features were identical to those in Experiment 1 and Experiment 2. As in the previous experiments, participants performed a training session of 128 trials with 100% expected features before the experiment.

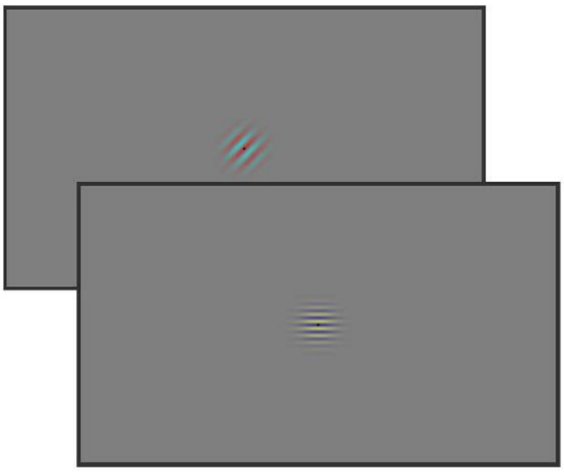

Fig. 6 Stimulus examples of Experiment 4. Participants were asked to respond to the spatial frequency of the grating, which could be high or low. The probabilities of occurrence of both color and orientation of the single grating were manipulated to induce feature expectations

Analysis

The analysis of Experiment 4 was identical to Experiment 1.

Results

Similar to the previous experiments, the mean error rate was low, with an average of 5.33% ($\pm$0.65 SEM).

The ANOVA of the error rates yielded a significant main effect for *ColPE* ($F(1,15)=5.811$, $p<0.05$, $\eta_P^2=0.279$), with lower error rates for expected colors (4.95%) compared to unexpected colors (8.04%), and a significant main effect for *OriPE* ($F(1,15)=8.356$, $p<0.05$, $\eta_P^2=0.358$) with lower error rates for expected orientations (4.91%) compared to unexpected orientations (8.31%). The interaction was not significant ($F(1,15)=0.908$, $p=0.356$, $\eta_P^2=0.057$).

Again, the ANOVA of the mean RTs revealed a significant main effect for *ColPE* ($F(1,15)=16.559$, $p<0.05$, $\eta_P^2=0.525$), with 427 ms for expected colors versus 444 ms for unexpected colors, and a significant main effect for *OriPE* ($F(1,15)=15.17$, $p<0.05$, $\eta_P^2=0.503$), with 427 ms for expected orientations versus 442 ms for unexpected orientations, reflecting the cost for unexpected features. Thus, both unexpected color and unexpected orientation stimuli resulted in significantly higher RTs compared to the more frequently presented standard targets. The interaction of *ColPE* × *OriPE* was not significant ($F(1,15)=0.101$, $p=0.75$, $\eta_P^2=0.007$), providing no evidence for a mutual influence of prediction errors for both features. Hence, both feature prediction errors were again processed independently, even when they were part of the same object. The mean RTs and error rates are shown in Fig. 7.

Discussion

Experiment 4 was conducted to examine whether the lack of an interaction between feature expectations observed in Experiments 1–3 can be explained by reduced object feature binding due to divided attention across multiple objects. Therefore, in Experiment 4, only one single central stimulus was presented, which was used to manipulate color and orientation expectations. Consistent with the previous experiments, the RT costs associated with one feature being unexpected were also evident in Experiment 4. Furthermore, as in the previous experiments and despite a higher degree of feature object binding, RTs for two unexpected features increased additively rather than interactively providing no evidence for joint expectations regarding both features. Accordingly, the lack of an interaction between prediction

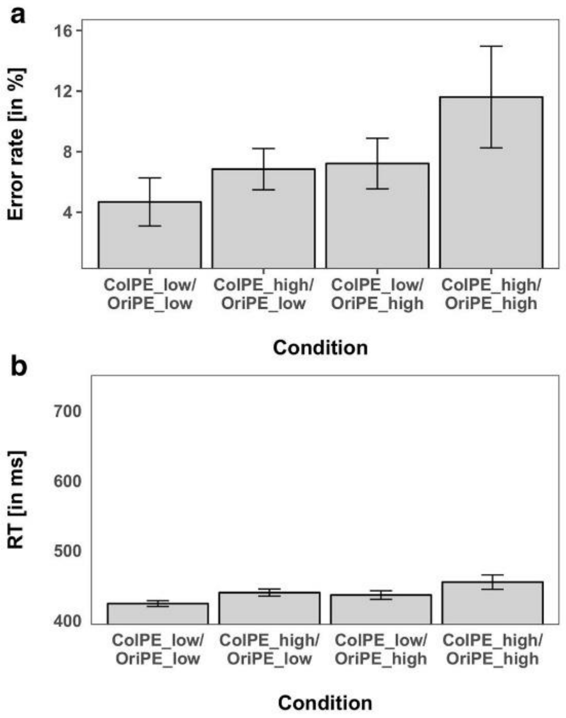

Fig. 7 Performance measures of the combination of color and orientation manipulations of Experiment 4. **a** Error rates. **b** Reaction times. Error bars reflect the 95% confidence intervals

errors in different visual features in the previous experiments cannot be accounted for by a need to divide the attentional focus.

General discussion

This study investigated whether expectations regarding multiple visual features of different dimensions (color and orientation) are formed independently when they refer to the same object, or whether feature expectations are combined to a joint object-level expectation. A previous study by Jiang et al. (2016) reported behavioral and functional imaging evidence supporting the latter assumption. In particular, they suggested that a prediction error about one feature spreads across other object features and marks the entire object as "unexpected". In the current study, four behavioral experiments were performed to systematically manipulate the distribution of feature expectations to the same or different objects and, hence, to extend the findings of Jiang et al. (2016).

The newly developed paradigm allowed inducing prediction errors in different visual dimensions for the same or different objects. Furthermore, the paradigm allowed investigating prediction errors emerging within two task-irrelevant feature dimensions and hence avoided any potential confounds with response-related prediction errors in the motor domain. In a first experiment, prediction errors were induced in the color and orientation dimension, and each type of prediction error was confined to separate objects. Both unexpected color and unexpected orientation increased RTs, indicating that the paradigm successfully elicited prediction error signals of task-irrelevant feature dimensions and that these predictive error signals, although task-irrelevant, altered behavior in an ongoing task. Prediction errors in both visual dimensions were comparable in magnitude and showed no signs of mutual influence as indicated by an additive rather than an interactive RT pattern. Therefore, they appear to be calculated separately when the features are distributed between separate objects.

We then tested whether this pattern changed when prediction errors in different dimensions co-occurred within the same object. Again, prediction errors in both dimensions reliably induced slower RTs when unexpected features were presented. As in the previous experiment and contrary to our initial hypothesis, RTs showed an additive rather than an interactive pattern. Accordingly, even when prediction errors were induced by features belonging to the same object, we could not find evidence for mutual influence and interference between dimensions. One possible explanation for this finding is that feature expectations in the present experiments were generated implicitly and may, hence, be formed before object binding occurs, a process that has critically been associated with focused attention (Treisman and Gelade 1980). To test this, an additional experiment was conducted where feature expectations were induced by verbal cues, hence increasing top-down aspects of feature expectations. Despite a more explicit representation of features expectations, no evidence for combined feature expectations was observed and the RT patterns found in Experiments 1 and 2 were replicated. These patterns persisted even when a single-object version of our task was used. This variant was introduced to increase the amount of attention allocated to a single target object and thereby the degree of object feature binding. No evidence for a mutual interaction of feature expectations was found, although visual features were part of an integrated object representation.

In sum, this series of experiments yielded no evidence for a mutual influence of prediction errors in different dimensions. This finding was independent of whether feature expectations were formed implicitly or explicitly and whether or not attention was fully engaged at a single target object.

Our results are thus in line with results reported by Stefanics et al. (2019) in the sense that task-irrelevant feature predictions do not interact. However, the finding of an independent coding of feature predictions seems at contrast with a study by Jiang et al. (2016) which reported a cross-blending of prediction errors. A crucial difference between the study by Stefanics et al. (2019) and our study on one hand and the study by Jiang et al. (2016) on the other hand is related to a feature's task relevance. Evidence for an interaction of prediction errors was observed only in the study by Jiang et al. (2016) where prediction errors in a task-irrelevant dimension affected predictions in a task-relevant dimension. In the present study, prediction errors were manipulated in two task-irrelevant dimensions and no interactions were found between the two. However, the present study involved spatial frequency as an additional task-relevant dimension. Expectations with regard to this dimension were held neutral, with both feature values being equally likely across the experiment. Prediction errors in the task-irrelevant dimensions interfered with spatial frequency judgements in the task-relevant dimension and increased reaction times. A possible explanation for this interference is that prediction errors emerging in the task-irrelevant dimension alter prediction errors in the task-relevant dimension, rendering the neutral but task-relevant feature unexpected. In this case, the results could be taken as support for the view that combined expectancies critically depend on task or response relevancy. However, it is unclear whether interference of task-irrelevant and task-relevant features is in fact based on a combination of prediction errors. Inconsistent prediction errors may also result in more general interference effects generating higher demands on attentional and cognitive control. In particular, prediction errors may render irrelevant features more salient and generate attentional capture that interferes with the ongoing task. For instance, interference could be based on a series of automatic and sequential attentional switches between salient feature dimensions before attention could then be deployed to the task-relevant dimension (spatial frequency). This interpretation corresponds to findings from visual search experiments, showing that search for a target with an unknown target-defining feature results in higher RT costs when the feature could change between different dimensions (e.g., color and orientation) compared to features within the same dimension (e.g., red and blue) (Müller et al. 1995; Treisman 1988). Alternatively, attentional capture by irrelevant but salient stimulus features may bind attentional resources on irrelevant feature dimensions and will, hence, decrease those available for the task performed.

In summary, the present results suggest that feature expectations for color and orientation are processed and resolved independently, and are unaltered by processes related to object binding. This finding is consistent with an early implementation of predictive coding within separate feature channels. Although the present findings cannot rule out that prediction errors for different object features might possibly be combined into an object-level expectancy, our results do not support the view that object feature binding leads to mutual influences of predictions errors of different object features.

Acknowledgements Open Access funding provided by Projekt DEAL.

References

Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617. https://doi.org/10.1038/nrn1476

Cheadle S, Egner T, Wyart V et al (2015) Feature expectation heightens visual sensitivity during fine orientation discrimination. J Vis 15:14. https://doi.org/10.1167/15.14.14

Derrick B, Toher D, White P (2017) How to compare the means of two samples that include paired observations and independent observations: a companion to Derrick, Russ, Toher and White. Quant Methods Psychol (2017) 13:120–126. https://doi.org/10.20982/tqmp.13.2.p120

Dombert PL, Fink GR, Vossel S (2016a) The impact of probabilistic feature cueing depends on the level of cue abstraction. Exp Brain Res 234:685–694. https://doi.org/10.1007/s00221-015-4487-2

Dombert PL, Kuhns A, Mengotti P et al (2016b) Functional mechanisms of probabilistic inference in feature- and space-based attentional systems. NeuroImage 142:553–564. https://doi.org/10.1016/j.neuroimage.2016.08.010

Gregory RL (1997) Knowledge in perception and illusion. Philos Trans R Soc B Biol Sci 352:1121–1127

Jabar SB, Filipowicz A, Anderson B (2017) Knowing where is different from knowing what: distinct response time profiles and accuracy effects for target location, orientation, and color probability. Atten Percept Psychophys 79:2338–2353. https://doi.org/10.3758/s13414-017-1412-8

Jiang J, Summerfield C, Egner T (2016) Visual prediction error spreads across object features in human visual cortex. J Neurosci 36:12746–12763. https://doi.org/10.1523/JNEUROSCI.1546-16.2016

Kok P, Jehee JFM, de Lange FP (2012) Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75:265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kuhns AB, Dombert PL, Mengotti P et al (2017) Spatial attention, motor intention, and bayesian cue predictability in the human brain. J Neurosci 37:5334–5344. https://doi.org/10.1523/JNEUROSCI.3255-16.2017

Mars RB, Debener S, Gladwin TE et al (2008) Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. J Neurosci 28:12539–12545. https://doi.org/10.1523/JNEUROSCI.2925-08.2008

Müller HJ, Heller D, Ziegler J (1995) Visual search for singleton feature targets within and across feature dimensions. Percept Psychophys 57:1–17. https://doi.org/10.3758/BF03211845

Peirce JW (2007) PsychoPy psychophysics software in python. J Neurosci Methods 162:8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017
Peirce JW (2008) Generating stimuli for neuroscience using PsychoPy. Front Neuroinformatics. https://doi.org/10.3389/neuro.11.010.2008
Richter D, Ekman M, de Lange FP (2017) Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. J Neurosci 38:7452–7461
Stefanics G, Heinzle J, Horváth AA, Stephan KE (2018) Visual mismatch and predictive coding: a computational single-trial ERP study. J Neurosci 38:4020–4030. https://doi.org/10.1523/JNEUROSCI.3365-17.2018
Stefanics G, Stephan KE, Heinzle J (2019) Feature-specific prediction errors for visual mismatch. NeuroImage 196:142–151. https://doi.org/10.1016/j.neuroimage.2019.04.020
Stein T, Peelen MV (2015) Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. J Exp Psychol Gen 144:1089–1104. https://doi.org/10.1037/xge0000109
Stojanoski BB, Niemeier M (2015) Colour expectations during object perception are associated with early and late modulations of electrophysiological activity. Exp Brain Res 233:2925–2934. https://doi.org/10.1007/s00221-015-4362-1
Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends Cogn Sci 13:403–409
Treisman A (1988) Features and objects: the fourteenth bartlett memorial lecture. Q J Exp Psychol Sect A 40:201–237. https://doi.org/10.1080/02724988843000104
Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12:97–136
Velhagen K, Broschmann D (eds) (2003) Tafeln zur Prüfung des Farbensinnes. Thieme, Stuttgart
Von Helmholtz H (1867) Handbuch der physiologischen Optik. Leopold Voß, Leipzig
Wyart V, Nobre AC, Summerfield C (2012) Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. Proc Natl Acad Sci 109:3593–3598. https://doi.org/10.1073/pnas.1120118109
Zhao J, Al-Aidroos N, Turk-Browne NB (2013) Attention is spontaneously biased toward regularities. Psychol Sci 24:667–677

Affiliations

Michael Wiesing[1] · Gereon R. Fink[1,2] · Ralph Weidner[1] · Simone Vossel[1,3]

Gereon R. Fink
g.r.fink@fz-juelich.de

Ralph Weidner
r.weidner@fz-juelich.de

Simone Vossel
s.vossel@fz-juelich.de

[1] Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre Juelich, Wilhelm-Johnen-Strasse, 52428 Juelich, Germany

[2] Department of Neurology, Faculty of Medicine, University Hospital Cologne, University of Cologne, Kerpener Strasse 62, 50937 Cologne, Germany

[3] Department of Psychology, Faculty of Human Sciences, University of Cologne, Gronewaldstrasse 2, 50931 Cologne, Germany

Chapter 2 - Timing

Wiesing, M., Fink, G. R., & Weidner, R. (2020). Accuracy and precision of stimulus timing and reaction times with Unreal Engine and SteamVR. *PloS one*, *15*(4), e0231152.

Author contributions

MW, G.R.F., and R.W. conceptualized and designed the research; MW collected the data; wrote software, analyzed and visualized the data; R.W. supervised the research project; MW, G.R.F., and R.W. wrote the manuscript.

PLOS ONE

RESEARCH ARTICLE

Accuracy and precision of stimulus timing and reaction times with Unreal Engine and SteamVR

Michael Wiesing[1]*, Gereon R. Fink[1,2], Ralph Weidner[1]

1 Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre, Juelich, Germany, **2** Department of Neurology, University Hospital Cologne and Faculty of Medicine, University of Cologne, Cologne, Germany

* mi.wiesing@fz-juelich.de

Abstract

The increasing interest in Virtual Reality (VR) as a tool for neuroscientific research contrasts with the current lack of established toolboxes and standards. In several recent studies, game engines like Unity or Unreal Engine were used. It remains to be tested whether these software packages provide sufficiently precise and accurate stimulus timing and time measurements that allow inferring ongoing mental and neural processes. We here investigated the precision and accuracy of the timing mechanisms of Unreal Engine 4 and SteamVR in combination with the HTC Vive VR system. In a first experiment, objective external measures revealed that stimulus durations were highly accurate. In contrast, in a second experiment, the assessment of the precision of built-in timing procedures revealed highly variable reaction time measurements and inaccurate determination of stimulus onsets. Hence, we developed a new software-based method that allows precise and accurate reaction time measurements with Unreal Engine and SteamVR. Instead of using the standard timing procedures implemented within Unreal Engine, time acquisition was outsourced to a background application. Timing benchmarks revealed that the newly developed method allows reaction time measurements with a precision and accuracy in the millisecond range. Overall, the present results indicate that the HTC Vive together with Unreal Engine and SteamVR can achieve high levels of precision and accuracy both concerning stimulus duration and critical time measurements. The latter can be achieved using a newly developed routine that allows not only accurate reaction time measures but also provides precise timing parameters that can be used in combination with time-sensitive functional measures such as electroencephalography (EEG) or transcranial magnetic stimulation (TMS).

Citation: Wiesing M, Fink GR, Weidner R (2020) Accuracy and precision of stimulus timing and reaction times with Unreal Engine and SteamVR. PLoS ONE 15(4): e0231152. https://doi.org/10.1371/journal.pone.0231152

Editor: Stefano Triberti, University of Milan, ITALY

Received: October 15, 2019

Accepted: March 17, 2020

Published: April 8, 2020

Data Availability Statement: The complete raw dataset of all experiments described in this paper can be found in https://osf.io/jtyge/. All UE4 project files and project builds, i.e., compiled executables of the different tests that can be executed without the UE4 Editor, can be found on https://jugit.fz-juelich.de/inm3/timingvr. This includes the source code of a background application developed for reaction time measurements with UE4. Furthermore, a custom version of UE4 used for the reaction time measurements is available on https://github.com/INM3FZJ/UnrealEngine/tree/TimingVR. Note that an EpicGames account connected to a

Introduction

Over the last 20 years, Virtual Reality (VR) has been increasingly recognized as a powerful research tool in behavioral neuroscience [1,2]. VR enables researchers to study complex and naturalistic behavior in virtual environments while maintaining a high degree of experimental

GitHub account is necessary to get access to the UE4 source code commits.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

control. VR can help to increase the ecological validity of experiments, i.e., allowing to conduct experiments in a context that is closer to everyday life, which might lead to more generalizable and valid explanations regarding cognitive processes. For example, a virtual scenario of a classroom allows to investigate the attentional capacities of children with ADHD in a realistic but well-controlled environment [3].

Consumer-grade HMD based VR Systems like Oculus Rift, Valve Index, or HTC Vive are now making VR available for many researchers and will soon dramatically increase the impact of VR in the field of cognitive neuroscience. Most likely, VR will have major impact on studies of visual perception as one of the many advantages that VR offers is its ability to present stimuli in stereoscopic 3D with a large field of view and with the HMDs entirely covering the visual field.

Yet, the suitability and reliability of HMDs concerning the measurement of visual cognitive performance needs to be demonstrated. Recent studies examined whether HMDs, at least in principle, can be used to reliably investigate visual processing components. The results indicate, that the Oculus Rift Development Kit 2 (DK2) and the HTC Vive allow assessing visual processing as reliably as CRT displays [4,5].

In addition, due to a lack of established stimulus software for VR experiments, several recent studies relied on using game engines like Unity [6] or Unreal Engine [7] [e.g., 8, 9]. However, game engines do not contain certain key features that are critical for neuroscientific experiments. For example, data collection with build-in features, such as response time measurements, are tied to the software's frame rate, resulting in a sampling rate of 90 Hz when using an HMD such as Oculus Rift CV1, HTC Vive, or HTC Vive Pro.

Furthermore, modern VR systems use runtime software tools like SteamVR [10] or the Oculus Rift Software [11], which operate in the so-called "direct mode" that allows direct communication with the HMD hence bypassing the operating system's typical display communication pathways. These runtime software tools determine when a new frame will be presented. They thus have control over the exact timing of stimulus presentation, while at the same time limiting the accuracy with which stimulus events can be measured, at least when the game engine's build-in timing methods are used.

These limitations are well described in the literature [e.g., 12,13] and different approaches have been proposed to increase the precision and accuracy of time critical measurements in combination with game engines [e.g., 14–16].

However, to date, no study has systematically quantified the maximum precision and accuracy of stimulus timing and time measurements achievable with consumer VR hardware in combination with game engines.

The goal of the present study was (1) to assess timing errors of the HTC Vive combined with Unreal Engine 4 (UE4) and SteamVR, and (2) to present a new method that allows to measure the timing of visual stimulus events and response times with an accuracy and precision in the millisecond range.

Displays

Experimental neuroscience requires accurate and precise control of stimulus durations and is hence in need for display techniques with appropriate temporal properties. Display technologies differ in how accurate stimulus onsets and offsets can be determined. The display technology currently used in VR differs from the computer screens often used in experimental neuroscience. For instance, in contrast to cathode-ray tube (CRT) or liquid crystal displays (LCD) monitors, many modern HMDs such as Oculus Rift or HTC Vive use organic light-emitting diode (OLED) displays, and their accuracy remains to be tested. In particular, it is essential to determine how accurately stimulus onset times (i.e., the time point when a stimulus appears on the display) can be controlled and how precise and solid stimulus durations can be defined.

For a long time, CRT monitors were considered the gold-standard for visual stimulus delivery in experimental neuroscience. In CRT monitors every pixel is updated periodically to generate an image via an electron beam that scans all pixels in rows from top-left to bottom right, a process referred to as raster scanning [17]. An important characteristic here is a pixel's persistence, i.e., the time pixels are illuminated after stimulation by the electron beam. Each pixel is covered with phosphor. When hit by the electron beam, the phosphor illuminates and quickly reaches its maximum luminance and then starts to fall off again over a given time. This time depends on specific characteristics of the given phosphor and ranges from virtually no persistence (e.g., P15) to several seconds (e.g., P31) [18]. Thus, on a CRT monitor, the entire image cannot be updated simultaneously and does hence not deliver a continuous luminance pattern.

LCD monitors are based on a completely different technology. LCD monitors constitute so-called sample and hold displays in which a steady light source is positioned behind a layer of liquid crystals. The orientation of the crystals defines a pixel's luminance. Although there is no longer an actual electron beam in modern LCDs, and it is no longer necessary to update displays periodically, it is still common that LCDs are updated in scan like patterns similar to those in CRT monitors. However, in contrast to CRT monitors, LCD signals do not decay and remain constant during the display refresh (except for possible backlight modulations) [17].

Since in both LCD and CRT monitors the image is built up line by line from top to bottom, stimulus onset measurements are synchronized to the vertical blank interval, which is the time between two display refreshes. Therefore, the stimulus onset measurement is only perfectly synchronized to the actual stimulus onset, if the stimulus is presented in the leftmost pixel of the first row. Consequently, the measurement error of the stimulus onset is correlated with the position of the stimulus on the display [19,20].

In contrast, the HTC Vive uses two low-persistence OLEDs. OLEDs are comprised of a thin film of organic material that uses an electrical circuit to control the emission of light [21,22]. Unlike LCDs, OLEDs do not rely on time-consuming reorientations of liquid crystals and have, therefore, faster response times than LCDs. Furthermore, while luminance transition times of LCDs depend on the luminance of the previous frame, in OLEDs, transition times are supposed to be independent of the previous frame's luminance [23]. In general, OLEDs show precise temporal responses and appear to be suitable for visual neuroscientific research [21] Hence, the displays used in HMD's are unique with regard to timing and time measurements due to particularities of how the display is refreshed.

The whole display gets loaded before it illuminates so that all pixels illuminate simultaneously. This type of display is hence often referred to as "global displays" [24]. In particular, a display refresh of the HTC Vive's displays starts with a blank period, during which the displays stay black. The pixels then illuminate for about 2 ms at the end of the screen refresh period. This is done to minimize visual motion artifacts like smearing and judder. Thus, the updating behavior of the HTC Vive's displays should allow accurate measurements of the stimulus onset, independent of its position.

Stimulus software

For traditional experiments in visual neuroscience, a variety of commercial and open-source software tools are available that meet strict timing requirements, e.g., Presentation [25], PsychoPy [26], or the Psychophysics Toolbox [27]. These software tools are optimized for the presentation of two-dimensional stimuli and offer no or only limited support for the current state of the art VR systems. Concerning VR, established toolboxes for neuroscientific experiments are lacking. While a variety of commercial and open-source software is available for the creation of virtual environments, most of the software tools do not contain features specifically designed for neuroscientific experiments. One exception is WorldViz Vizard [28], a commercial solution, which allows researchers to create and conduct VR experiments.

However, in several recent studies, game engines like Unity or Unreal Engine were used and it remains to be tested whether these software packages provide precise and accurate stimulus timing and time measurements that allow inference on mental and neural processes.

For instance, a time-critical presentation of visual stimuli is required for visual masking paradigms [29] or attentional blink paradigms [30] as well as for almost any experiment involving visual psychophysics [31]. Similarly, a precise recording of response times is required for all studies involving mental chronometry, with reaction times as a dependent measure. Furthermore, a precise recording of events within an experiment is mandatory for the analysis of all time-sensitive functional measures, such as EEG, TMS, magnetoencephalography (MEG), or galvanic skin response (GSR), where cognitive and perceptual events are assigned to functional markers.

The current study investigated the timing errors of the HTC Vive combined with UE4 and SteamVR. In the first experiment, we assessed the precision and accuracy of stimulus presentations. In the second experiment, we explored the limitations of UE4's build-in timing procedures for reaction time measurements resulting in variable and inaccurate reaction time measures that are inappropriate for a wide range of reaction time experiments. Further, a newly developed method and its benchmarking results will be presented. This new method allows precise and accurate reaction time measurements with UE4 and SteamVR.

In the following, in order to increase readability, we will always refer to UE4, although, if not stated differently, it always refers to the combination of UE4, SteamVR, and the HTC Vive.

General methods

Unreal engine settings

We aimed to keep almost all default settings of UE4 in this study. However, due to the high rendering demands of VR, it is generally recommended to adjust some settings to optimize UE4's VR rendering.

First, we changed the rendering technique from the default Deferred Renderer to the Forward Renderer. Forward rendering is generally the preferable rendering method for VR since it usually improves performance and allows better anti-aliasing methods than deferred rendering. Forward rendering allows using multisample anti-aliasing (MSAA) that increases sharpness and leads to better visuals [24,32]. Additionally, we enabled Instanced Stereo rendering. By default, the geometry has to be drawn twice for VR applications as compared to non-VR rendering, once for the left eye and once for the right eye, which essentially doubles the number of draw calls, i.e., the rendering information that is send from the central processing unit (CPU) to the graphics processing unit (GPU). With Instanced Stereo rendering, the geometry has to be drawn only once and is then projected to the right-eye view and left-eye view of the geometry. This procedure halves the number of draw calls and thereby saves a substantial amount of CPU time.

All tests described in this paper were conducted with Unreal Engine 4.21.2 with the configuration described above.

SteamVR settings

In all experiments, we disabled supersampling anti-aliasing (SSAA) by setting the render resolution of the HTC Vive to 100%. SSAA works by rendering a scene with a higher resolution than the one displayed and then averages neighboring samples to create the image [33]. SSAA is often used for VR rendering to reduce aliasing and improve visual quality.

Additionally, we disabled "Motion Smoothing" for all experiments. Motion Smoothing is a method to stabilize the frame rate when an application starts to drop frames. As soon as the frame rate decreases below the HMD's refresh rate (90 frames per second (FPS) for the HTC Vive), SteamVR reduces the frame rate by half and instead extrapolates every second frame based on the last two presented frames [34]. However, while this method is helpful for the common usage of VR (e.g., gaming) to provide smoothly running VR experiences, it would corrupt all the timing precision and accuracy within a scientific experiment. Thus, it is recommended to disable this option whenever precise timing is required or when the timing of events (e.g., reaction times) needs to be determined precisely. However, besides Motion Smoothing another technique helps to smooth out dropped frames called "Asynchronous Reprojection". Like Motion Smoothing, Asynchronous Reprojection reduces the number of rendered frames by half. However, instead of extrapolating frames, it repeats the previous frame but reorients it based on the user's latest head rotation [35]. Since Asynchronous Reprojection cannot be disabled, a VR experiment should always be optimized about hard- and software to ensure stable frame timing.

All tests described in this paper were conducted with SteamVR 1.2.10 with the configuration described above.

Experiment 1

The aim of Experiment 1 was to test the precision and accuracy of visual stimulus presentations using UE4. The goal was to assess potential discrepancies between the stimulus timing defined by the researcher and the actual stimulus timing on the display. Please note, stimulus duration errors are observed even in software tools that, in contrast to UE4, are especially designed for behavioral experiments [20].

Material and methods

Tests were performed using the Black Box Toolkit v2 (BBTK) to test the timing precision and accuracy of simple black and white transitions with predefined durations. The BBTK is specially designed for benchmarking these type of tests [36]. We tested three different conditions to determine the precision and accuracy of UE4's stimulus timing under different rendering workloads. Furthermore, we ran every test on two different Windows computers to compare the timing precision and accuracy on systems with different hardware specifications.

Apparatus. All tests were conducted on two different desktop computers with Windows 10 as the operating system. The specifications of Computer 1) were Intel i7-8700, 32 GB DDR4 RAM, and an Nvidia GTX 1080 graphics card, and of Computer 2) Intel i7-7700K, 32 GB DDR4 RAM, and an Nvidia GTX 1080Ti graphics card. The tested VR System was the HTC Vive (1080 x 1200 pixels per eye).

General procedure. We used a well-established procedure to test the timing of visual stimulus presentations for PC [37–39]. For each test, a photo-sensor connected to the BBTK was attached to the middle of the HTC Vive's left lens. The photo-sensor was used to measure the stimulus timing under all conditions with a sampling resolution of 0.25 ms. Stimuli were filled-in squares repeatedly alternating their contrast from black to white. Contrast reversals were programmed to occur after durations of 11.111, 33.333, 66.666, 100, 200, 500, and 1000

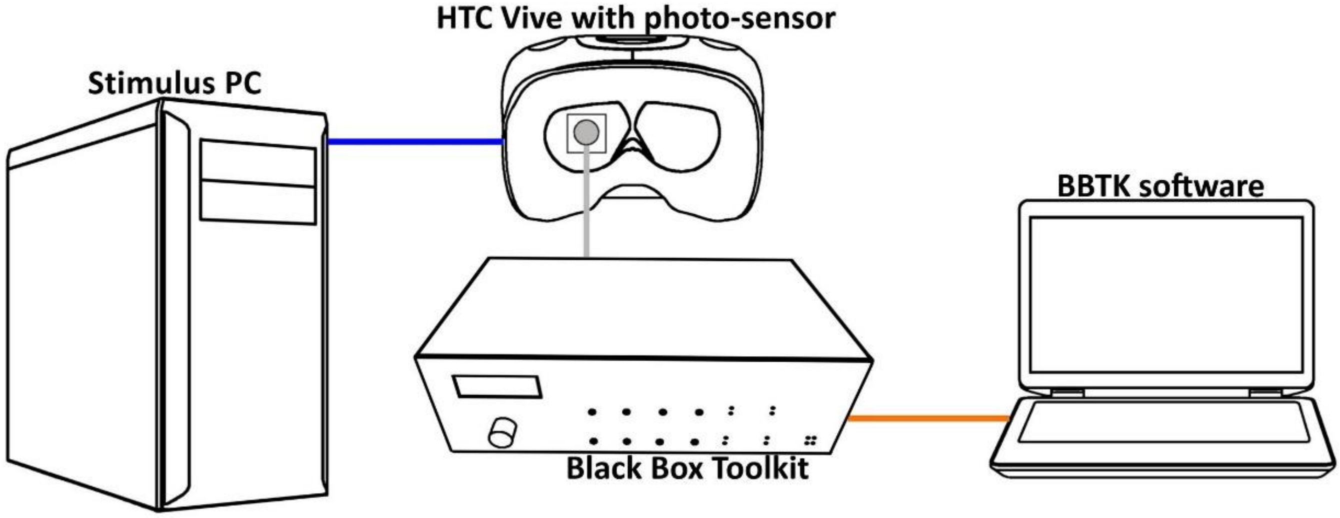

Fig 1. Presentation and measurement setup.

https://doi.org/10.1371/journal.pone.0231152.g001

ms (i.e., 1, 3, 6, 9, 18, 45, and 90 display refreshes at 90 Hz). All stimulus durations were defined in terms of the frames or ticks.

The BBTK was controlled with a second computer that was independent of the presentation computer to prevent interference of the stimulus delivery process and the measurements (Fig 1).

Each stimulus duration was tested in a single series with a duration of 10 minutes and hence allowed evaluating the stability of UE4's timing behavior over long durations. Furthermore, three different conditions were introduced, in which the rendering workload was systematically varied (Fig 2).

To test whether UE4 is generally able to provide sufficient timing precision and accuracy of visual stimulus presentations while the rendering workload is low, a first condition was introduced, which is from now on referred to as *Simple*. In this condition, we created an empty and completely black virtual environment that only consisted of a single square that was presented

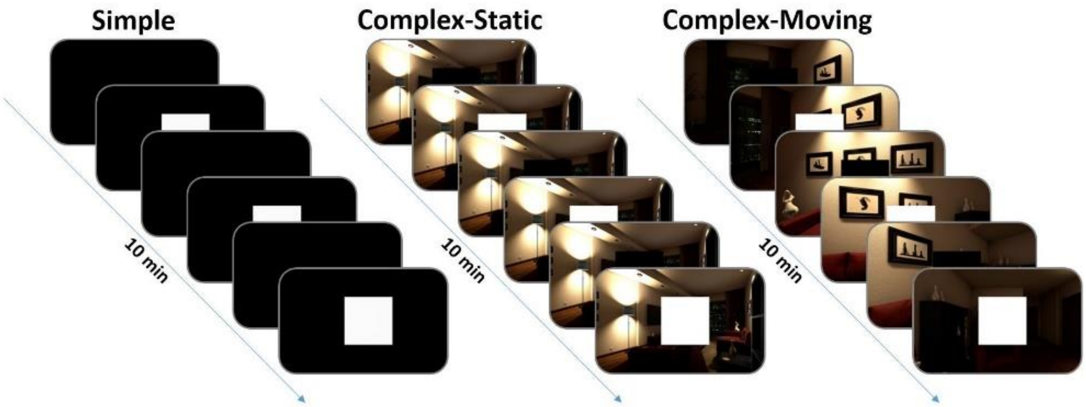

Fig 2. Stimulus conditions. Please note, the brightness of the illustrations was increased and does not represent the actual brightness presented.

https://doi.org/10.1371/journal.pone.0231152.g002

centrally of the left eye's image. The square's brightness alternated between black and white with the durations mentioned above.

This procedure allowed us to determine UE4's timing precision and accuracy of visual stimulus presentations with a minimal rendering workload and assured that presentation times were unaffected by possible performance constraints.

A second condition, further referred to as *Complex-Static*, aimed to evaluate the stimulus timing precision and accuracy with a high rendering workload. In a typical VR experiment, stimuli are not displayed in isolation but are embedded in a complex virtual environment.

The stimulus material in this condition still consisted of simple black and white transitions. In contrast to the *Simple* condition, the stimulus square was embedded in a virtual environment.

We used the *Realistic Rendering* sample project of UE4 as a virtual environment consisting of a highly realistic and detailed rendered living room [40] (S1 Fig).

A third condition *Complex Moving* aimed to evaluate the stimulus timing precision and accuracy while the movement of a user was simulated. We used the same virtual environment as in the *Complex Static* condition, but of a single static viewpoint, the first-person camera was placed in the center of the environment. The first-person camera was programmed to rotate around the yaw axis with a constant rate of 30° per second. This was done to simulate the user's head movements. Additionally, to provide a fixed location of the square relative to the camera, we attached the stimulus square to the camera. Apart from these adjustments, the general procedure was identical as in the other conditions.

The displays of the HMD are relatively small, so the photo-sensor could not be directly placed at the stimulus location to detect light from the stimulus square exclusively. Instead, it always included some parts of the surrounding background. Hence, for the *Complex-Static* and *Complex-Moving* conditions, it was necessary to calibrate the photo-sensor's sensitivity to ensure that the photo-sensor was driven only by the stimulus square and not by the environment's light.

Calibration consisted of two steps. First, we adjusted the threshold to set the activation point of the BBTK's photo-sensor to the highest possible value that still detected the white stimulus. In a second step, we adjusted the illumination of the environment. We first replaced the default post-processing of the environment with a custom post-processing. Post-processing describes techniques that change and add visual elements to a rendered scene after the rendering has been completed. This procedure allowed changing the luminance within the scene globally via a single setting, i.e., changing the light exposure, a post-processing effect that changes the overall brightness of a scene. For the calibration, we increased the light exposure until the environment's light activated the photo-sensor and then reduced it again until the photo-sensor was not activated by it anymore. This procedure ensured a proper illumination of the environment while ruling out any interference between the environment's light and the photo-sensor's light measurements.

Results

For each series, the mean durations of a single stimulus cycle (black and white stimulus interval), as well as the standard deviation and the overall longest and shortest measured durations were computed and reported for all tested stimulus durations. Furthermore, the mean duration of the white stimulus intervals was calculated. However, due to the low persistence displays, the display changed regularly between black and white even during the white stimulus intervals. Accordingly, the BBTK measured constant changes between black and white during white stimulus intervals. Hence, we defined the duration for the white stimulus interval as the time from the first white frame lighting up until the end of the last frame lightning up in that white stimulus interval.

Further, since the low persistence displays only light up for about 2 ms at the end of each display refresh, the white stimulus intervals started with a blank period during which the display was still black. Hence, the BBTK's photo detector was not able to separate the actual black stimulus interval from this first blank period of the white stimulus intervals first frame. Accordingly, the measured white stimulus interval durations are expected to be shorter than the expected stimulus duration by the duration of one frame's blank period, i.e., about 9 ms.

The overall variability of the stimulus intervals was consistently within the range of BBTK's sampling rate of 0.25 ms. However, the measured stimulus durations steadily exceeded the expected, i.e., programmed, stimulus duration. Further, the difference between expected and measured durations appeared to be linearly correlated with the expected stimulus duration. This result was confirmed by a simple linear regression between the expected and measured durations of the stimulus cycles ($F(1,5) = 3.62 \times 10^{16}$, $p < 0.05$), with an R^2 of 1.

Since the stimulus timing was defined in terms of ticks, the most parsimonious explanation for the difference in stimulus durations was that the true frame rate was lower than the expected 90 FPS. Hence, for all measured stimulus durations, the corresponding frame rate was determined as 89.53 FPS, which was consistent for all tested stimulus durations and computers.

A potential explanation for the lower frame rate could be performance problems affecting the frame rates. However, we systematically varied the rendering workload between conditions and any effect of performance problems on the frame rate should be reflected in different frame rates between conditions.

To test for potential performance problems between conditions, a two factorial Anova with the factors *Condition* (Simple, Complex-Static, Complex-Moving) and *Computer* (Computer 1, Computer 2) was performed. The ANOVA revealed no significant main effect for *Condition* ($F(1,2) = 0.001$, $p = 0.999$) indicating no difference between the FPS in the *Simple* (M = 89.528, SD = 0.370), *Complex-Static* (M = 89.528, SD = 0.370), and *Complex-Moving* (M = 89.528, SD = 0.370) conditions. In fact, the values observed in the different conditions were identical up to the fifth decimal place.

Furthermore, potential performance problems should be pronounced differently on the two different computer systems. However, the main effect *Computer* between *Computer 1* (M = 89.529, SD = 0.370) and *Computer 2* (M = 89.527, SD = 0.370) was not significant ($F(1,1) = 1.160$, $p = 0.281$). There was also no significant interaction between *Condition x Computer* ($F(1,2) = 0.001$, $p = 0.999$).

Besides, to exclude that the reduced frame rate (89.53 Hz vs. 90 Hz) observed was due to particularities of HMD tested, all tests of the *Simple* condition were repeated with an additional HTC Vive system on Computer 1. The mean results of these tests are presented in the supporting information (see S6 Table—S7 Table). The mean frame rate of 89.53 FPS was confirmed with this setup and a one factorial ANOVA with the Factor *HMD* (HMD 1, HMD 2) resulted in a non-significant difference of the mean FPS between *HMD 1* (M = 89.529, SD = 0.370) and *HMD 2* (M = 89.529, SD = 0.370); ($F(1,1) = 0.000$, $p = 0.99$).

Table 1 shows the results of the whole set of measurements for the three conditions and two computers. The individual results of the three conditions and both computers separately can be found in the supporting information (see S1 Table–S5 Table).

Discussion

Experiment 1 aimed to determine Unreal Engine's timing precision and accuracy for stimulus presentations. Stimulus timing was measured under different rendering workloads. The results

Table 1. Combined results across all conditions and computers (in ms).

expected duration	mean	sd	min	max	mean duration white
2000	2010.58	0.116	2010.50	2010.75	996.05
1000	1005.29	0.091	1005.25	1005.50	493.41
400	402.12	0.125	402.00	402.25	191.82
200	201.06	0.105	201.00	201.25	91.29
133.33	134.04	0.090	134.00	134.25	57.78
66.66	67.02	0.067	67.00	67.25	24.27
22.22	22.34	0.120	22.25	22.50	1.93

The first column represents the expected duration for the black and white stimulus cycles. The last column represents the measured durations of the white stimulus intervals.

https://doi.org/10.1371/journal.pone.0231152.t001

revealed that Unreal Engine in combination with the HTC Vive and SteamVR was able to achieve precise and accurate stimulus timings. Timing precision and accuracy were stable even with a high rendering workload. Furthermore, timing precision and accuracy remained stable when the user's head movements were simulated.

The HMD's empirical frame rate was determined as 89.53 FPS, slightly differing from the specifications given in the HTC Vive manual. Importantly, the frame rate was stable across different rendering workloads and when tested in combination with different computers, indicating that the observed frame rate was device-specific and not an artifact resulting from potential performance problems. Moreover, additional tests with a second HTC Vive system replicated this finding hence confirming that the reduced frame rate was not the results of a defective HMD.

The results suggest that the true refresh rate of the HTC Vive is 89.53 Hz rather than 90 Hz, at least when used in combination with UE4 running on a Windows 10 platform. Different frame rates may arise when using different software packages. Future research needs to investigate whether these findings hold for different rendering engines and operating systems.

Furthermore, the results of the present study emphasize that when using HMDs such as the HTC Vive, for precise stimulus timing, researchers need to take into account the peculiarities of low persistence displays with the so-called global update. In these displays, every frame starts with a blank period in which the displays stay black. Only after this blank period, the displays illuminate and the image appears. Therefore, black and white flickering stimuli as used in the present experiment involve longer black than white periods. Consequently, studies that rely on exact stimulus timing should consider this difference, when specifying stimulus durations.

Overall, we conclude that UE4, together with SteamVR and the HTC Vive, can present stimuli with precise and accurate stimulus durations.

Experiment 2

Introduction

Cognitive Neuroscience requires a mapping of internal neural states and cognitive processes. A prerequisite to successfully relate both is to infer a cognitive system's current state, which can only be done from behavior. Response times and differences in reaction times are one of the most important behavioral measures available to infer cognitive processes. Hence, a reliable and precise determination of participants' responses is essential for cognitive neuroscience since it allows inferring the mental processes to be related to neural events.

Furthermore, in order to relate functional data and ongoing cognitive processes, it is vital to accurately determine the points in time when specific sensory, cognitive, or motor events occur. For instance, studies dealing with time-sensitive functional measures, such as EEG, usually utilize the onset of a stimulus or the response time as markers to synchronize with a functional measure.

However, game engines such as UE4 are not designed for this kind of tasks and hence do not provide the necessary timing procedures allowing precise and accurate time measurements. For instance, precise reaction time measurements require that button events are registered in parallel to the stimulus presentation. UE4 registers button events only once per frame resulting in uncertainties in the range of one frame duration.

Furthermore, as described above, the actual presentation of stimuli is controlled by a VR runtime software such as SteamVR rather than by UE4 itself, which makes it impossible to accurately measure the onset time of a stimulus with UE4's build-in timing procedures.

Therefore, the aim of Experiment 2 was to determine the precision and accuracy of time measurements with UE4 and thereby to investigate the limitations of UE4's build-in timing procedures. Additionally, the precision and accuracy of reaction time measurements of PsychoPy and Presentation, two experimental software packages commonly used in cognitive neuroscience, were collected and compared against the results obtained with UE4.

Material and methods

All the settings of UE4 and SteamVR in Experiment 2 were identical to those used in Experiment 1.

Again, the BBTK was used to determine the precision and accuracy of time measurements, in particular, the precision of reaction time measurements. The BBTK's photo-sensor was used to measure the onset of a stimulus on the display, and a 1000 Hz USB response device, connected to the BBTK and the stimulus computer, was set to elicit a response when the photo-sensor signaled the onset of a stimulus (Fig 3). The BBTK software allowed generating response schedules in which a sequence of response times, as well as corresponding response durations (i.e., the duration of a button press) for a number of trials, was defined. This procedure allowed the BBTK to simulate a realistic reaction time pattern. The sequence of reaction times was obtained from a previously acquired data set collected in a behavioral experiment.

We used a well-established procedure to test the precision and accuracy of reaction time measurements [20,36,41]. Again, simple black and white transitions were used as visual stimuli. However, in contrast to Experiment 1, in which the black and white transitions were restricted to a square in the middle of the display, in Experiment 2, the black and white transitions involved the entire display. The latter was done because the tests involving PsychoPy and Presentation were conducted using an LCD monitor. As previously mentioned, in contrast to the HTC Vive's globally illuminated display, on LCD monitors the image is built up line-wise from top to bottom causing varying measurement errors depending on the position of the stimulus on the screen. Stimulus onsets were measured at the uppermost position in the left corner of the monitor to minimize hardware-related measurement errors when testing PsychoPy or Presentation. Furthermore, since LCD monitors have different and often inferior temporal properties as compared to OLEDs [42], the Samsung SyncMaster 2233 was used, which has previously been shown to provide sufficiently precise timing for vision research [43]. The tests of PsychoPy (v1.90.3) and Presentation (v20.3 02.25.19) were conducted with a refresh rate of 60 Hz.

In Experiment 2, the white stimulus was defined as the target stimulus eliciting a response by the BBTK. The target stimulus remained on the display until a response was given.

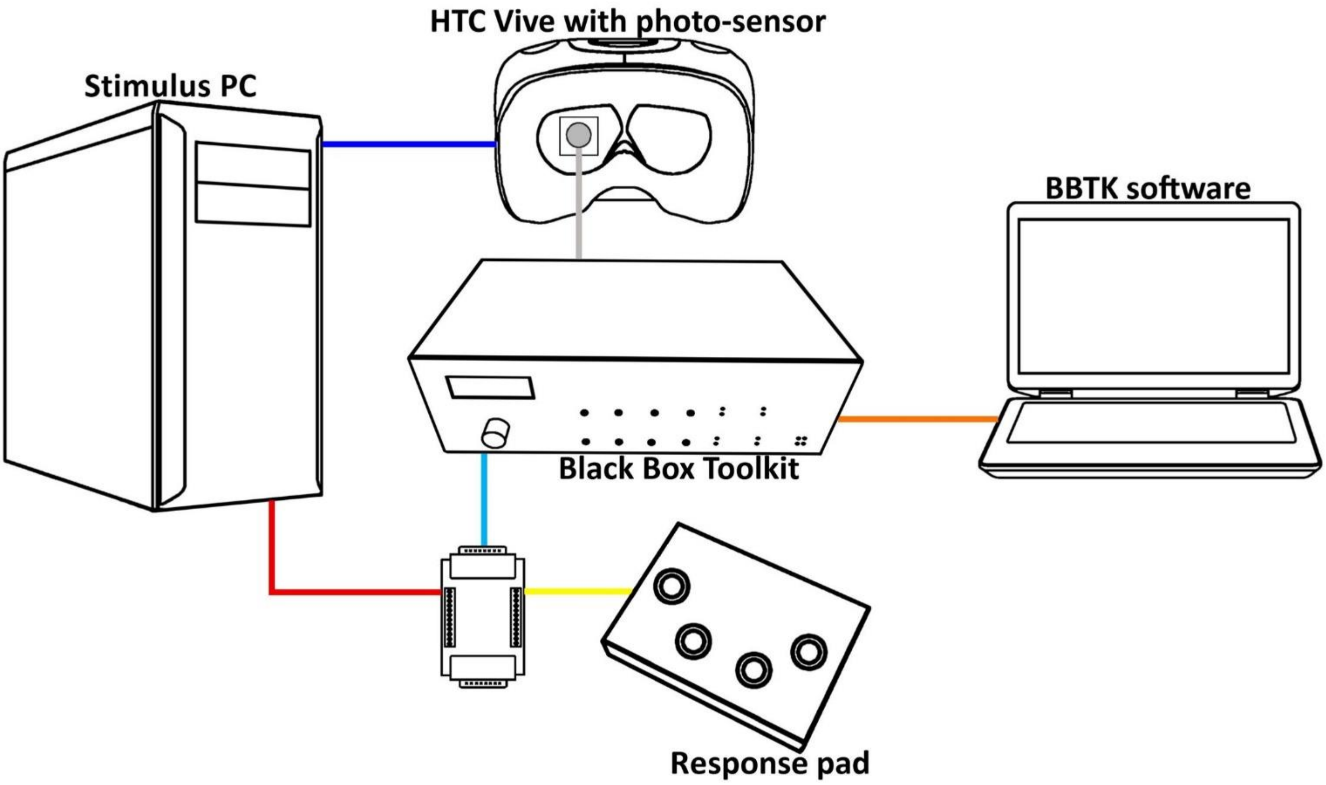

Fig 3. BBTK setup of the reaction time measurements.

https://doi.org/10.1371/journal.pone.0231152.g003

Following the response, the stimulus turned black for 1 second and turned white again until the following response was given (Fig 4). This procedure allowed to compare the true reaction

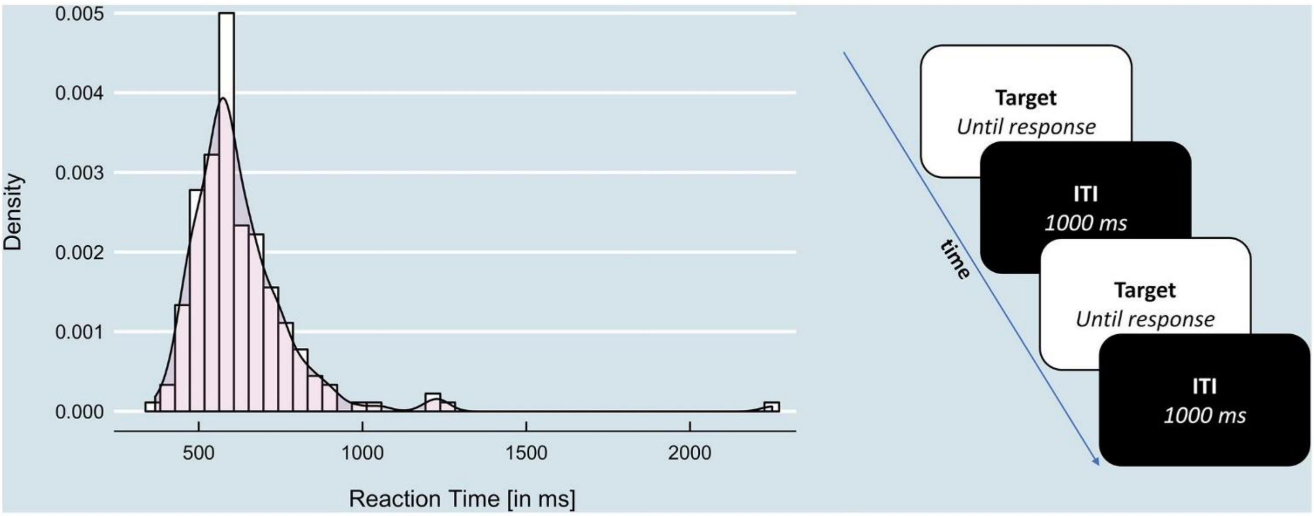

Fig 4. Example of one response schedule (left) and reaction time task of Experiment 2 (right).

https://doi.org/10.1371/journal.pone.0231152.g004

times as generated by the BBTK with the reaction times recorded by the stimulus software and to determine the reaction time errors as well as its variability.

All three software packages were tested in five runs of 200 trials each. For each of the five runs, a different response time schedule was applied. In order to represent realistic reaction time distributions, we assigned reaction times of human participants measured in a previously acquired dataset [44]. The reaction times of the first 200 trials of five randomly selected participants were used to generate five response time schedules (Fig 4). The response duration was kept constant and was set to 100 ms in all trials. For each software package, the same response time schedules and testing procedures were used.

All tests were conducted on a Windows 10 platform. The specification of the computer was an Intel i7-8700 CPU, 32 GB DDR4 RAM, with an Nvidia GTX 1080 graphics card.

Results

The mean results for the three tested software packages are presented in Table 2. Both PsychoPy and Presentation showed comparable and only small overall reaction time errors that led to a mean overestimation of the reaction times of approximately 4 ms. Furthermore, the variability of the reaction time errors was comparably small in both PsychoPy and Presentation.

In contrast, UE4 showed substantial errors, with highly variable reaction time measurements with a mean measurement error of +55 ms and variability in an error range of 12 ms. Two-sample Welch t-tests were conducted to compare the mean error of UE4 with each standard software package. The comparison of UE4 ($M = 55.073$, $SD = 3.326$) and Presentation ($M = 3.964$, $SD = 0.345$) resulted in significant different mean errors $t(1020) = 483.33$, $p < 0.05$). Similarly, the comparison between UE4 and PsychoPy ($M = 4.337$, $SD = 0.483$) was significant $t(1041) = 477.36$, $p < 0.05$). Additionally, F-tests resulted in significant differences of the variance between UE4 and Presentation ($F(1,999) = 93.103$, $p < 0.05$) as well as between UE4 and PsychoPy ($F(1,999) = 47.344$, $p < 0.05$).

Discussion

Experiment 2 aimed to quantify the precision and accuracy of reaction time measurements of UE4 and SteamVR in combination with the HTC Vive VR system. The tests illustrate the limitations of the build-in timing procedures of UE4. To obtain an estimate of the precision and accuracy provided by established software toolboxes for neuroscientific experiments in a non-VR context, we applied the same testing procedure to Presentation and PsychoPy. While both Presentation and PsychoPy provided precise and accurate reaction time measurements, reactions times measured with UE4's build-in timing procedures were highly variable with a high mean measurement error.

The variability of the reaction time measurements can be accounted for by the fact that button events in UE4 are not registered in parallel. Instead, response events are only recorded once per frame, hence resulting in an error range of one frame duration.

Table 2. Overview of measurement errors, standard deviation, minimum and maximum error for each software package (in ms).

Software	Mean error	SD	Min	Max
Presentation	3.964	0.345	3.00	5.00
PsychoPy	4.337	0.483	3.00	6.00
UE4	55.073	3.326	49.00	61.00

https://doi.org/10.1371/journal.pone.0231152.t002

In the tests reported above, stimulus onset times were determined after the command to change the stimulus color from black to white. The high overall mean error illustrates a delay between the moment the software executes the command to show a stimulus and the actual stimulus presentation time. This delay stems from several processing steps preceding the actual stimulus presentation, including a transfer of control over the stimulus presentation from UE4 to SteamVR.

However, these results were not surprising since these are known difficulties, and different frameworks have been suggested to simplify data collection for behavioral experiments [45,46] and to integrate and synchronize multiple data streams from different hardware devices, such as EEG or eye-tracking [14,16,47,48].

One framework mainly focused on the precision of reaction time measurements in combination with Unity and the HTC Vive [15]. In this study, the authors used an Arduino, a programmable microcontroller, to measure response times independently of Unity and with a sampling rate that is not limited by the refresh rate of the display. The benchmarking results illustrated that they were able to measure response times with a precision (standard deviation) of about 2.5 ms. Additionally, it was suggested that installing a photo-sensor in the HTC Vive could further improve time measurements. A small photo-sensor could measure the onset of a small peripheral stimulus appearing simultaneously with the target-stimulus. This additional stimulus could be registered directly with the Arduino.

Another recent study introduced the Unified Suite for Experiments (USE), a combination of soft- and hardware tools for the creation of experiments using Unity [14]. USE comes with a custom designed SyncBox, that is again based on an Arduino and allows to synchronize different data streams. A photo-sensor is used to measure the onset of an additional, peripheral stimulus that is presented simultaneously with the target. Benchmarking results demonstrated that the method was able to determine the stimulus onset with high precision for shorter durations. Instead, over a longer duration, the authors observed nonsystematic changes in accuracy. However, the authors provide pre-processing tools to account for these timing errors and to provide millisecond accurate stimulus onset measurements. Unfortunately, USE was not tested on HDMs directly. Instead, tests were performed on a computer monitor. Installing a photo-sensor in an HMD is technically challenging, as the photo-sensor needs to be small enough not to cause any discomfort to the participant. Nevertheless, even when small enough, the photo-sensor would be noticeable to the participant. Furthermore, this approach includes another peripheral flash stimulus that is correlated with the target. This additional stimulus could potentially confound some experiments, such as the second stimulus could distract from the target stimulus.

Therefore, in the next section, we will present a new software-based method that allows circumventing these limitations and measuring accurate and precise reaction times with UE4 and SteamVR.

A new method for reaction time measurements with Unreal Engine and SteamVR

Aims of the proposed method

The impact of VR on neuroscientific studies will critically depend on how easily VR experiments can be implemented and how precise control of stimulus delivery and response acquisition will be. Promising and important frameworks that aim to simplify the creation and control of VR experiments with game engines [e.g., 44, 45] are available already or are currently under development. However, so far, the progress with time-critical measurements, such as reaction times, is still insufficient and requires technically challenging setups, including additional hardware.

Here we aimed to develop a method for reaction time measurements that can provide precise and accurate measurements entirely on a software basis, which makes the need for additional hardware and complicated technical setups obsolete.

Although our method was developed mainly for UE4 and SteamVR, our aim was to ensure that the underlying principles are compatible with different game engines and the Oculus VR runtime.

However, our goal was not to develop a ready-to-use toolbox, as there are already promising toolboxes currently being developed [e.g., 45,48]. Instead, we aimed to develop and test our method as a proof-of-principle to provide a framework that can be integrated into other toolboxes.

Method

The concept of our method is based on separating reaction time measurements from UE4 by outsourcing the measurement procedure into a background application. Hence, our method follows the same idea as the ioHub of PsychoPy. In order to provide, e.g., framerate independent response times, the ioHub monitors response events in parallel of the PsychoPy main process by running a separate process in the background [26].

In the first section, we will describe how to obtain frame rate independent measures of response times.

In the second section, we will describe how this procedure can be extended to also obtain accurate measures of stimulus onsets.

In the third section, we will provide timing benchmarks of the proposed method.

Measuring frame rate independent response times. For the detection of responses, a global low-level keyboard hook is used. The Microsoft Developer Network describes a hook as "a point in the system message-handling mechanism where an application can install a subroutine to monitor the message traffic in the system and process certain types of messages before they reach the target window procedure."[49]. This procedure allows to intercept the keyboard message before Unreal receives the message and thus allows to process button input with a high sampling rate and independently of the displays frame rate (Fig 5). In particular, this

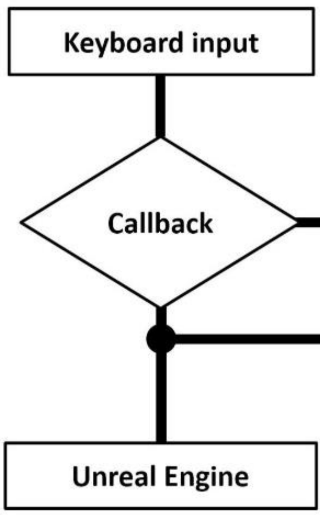

Fig 5. Illustration of the response time logging procedure.

https://doi.org/10.1371/journal.pone.0231152.g005

procedure allowed us to increase the sampling rate of the input processing from 90 Hz to the hardware limit of the response pad of 1000 Hz.

Furthermore, to determine the reaction time, i.e., the time between stimulus onset and keyboard response, an exact determination of when a stimulus appears on the screen is necessary in addition to the exact time of a keyboard event. Therefore, it is mandatory to establish an interface that allows synchronizing our background application with UE4. For this purpose, a trigger signal (onset-trigger) is sent from UE4 to the background application, marking the time point when a stimulus is presented by using Event Objects. Event Objects can be used to exchange signals between threads or processes, indicating that a particular event has occurred [50]. As soon as the function that initiates UE4 to present the target stimulus, the onset-trigger signal is sent to our background application determining the time of a stimulus' onset. The background application then registers the time, and the keyboard hook is activated. Once a response is given, the response time is logged, and the hook is stopped until the next onset-trigger is received from UE4.

We, therefore, ran another series of tests to assess whether this method reduces the variability of the measurement error previously observed. Again, five runs of 200 trials each, with the same response schedules used in Experiment 2 were tested.

The results demonstrated that our background application effectively reduced the variability of the response time errors down to a range of 2 ms. Conversely, the mean response time error was reduced by about half a frame duration to a mean error of 49 ms. A two-sample Welch t-test resulted in a significant difference of the mean errors between UE4 ($M = 55.073$, $SD = 3.326$) and UE4 + Hook ($M = 49.030$, $SD = 0.381$); $t(1025) = 57.081$, $p < 0.05$). A F-test comparing the variances also resulted in a significant reduced variance when measurements were conducted with the background application ($F(1,999) = 76.166$, $p < 0.05$).

The results are summarized in Table 3.

Measuring the stimulus onset. Accordingly, the variability of the reaction time errors can effectively be reduced. However, the mean absolute response time error of 49 ms remained relatively high compared to PsychoPy and Presentation. The remaining error indicated a delay between the moment UE4 receives the command to show a stimulus and the time when the stimulus appeared on the screen. A closer look at the rendering processes of UE4 and SteamVR revealed the origin of this delay and allowed compensating for it.

3D real-time rendering is accomplished within a pipelined architecture called graphics pipeline. Fig 6 coarsely depicts the conceptual stages for the graphics pipeline. The first stage consists of the game simulation and rendering preparation. Before the scene can be rendered, the system needs to determine the locations of objects to be rendered in the scene. The game simulation accomplishes this task. The game simulation calculates all the logic and transforms, i.e., the position, orientation, and scale of all objects in the scene, taking into account inputs by the user, animations, physics simulation, and AI. The rendering preparation then determines what is currently visible in the scene and creates a list of all the objects that need to be rendered and passes this list to the GPU. These steps are mostly processed on the CPU and controlled by UE4. The next stage is the actual geometry rendering, which is processed on the GPU. The final stage before the frame gets visible on the displays is referred to as scan-out. Scan-out

Table 3. Comparison of the reaction time errors, standard deviation, the minimum and maximum error of UE4 together with the hook procedure and the previous results obtained with UE4 build-in timing functions (in ms).

Software	Mean error	SD	Min	Max
UE4	55.073	3.326	49.00	61.00
UE4 + Hook	49.030	0.381	48.00	50.00

https://doi.org/10.1371/journal.pone.0231152.t003

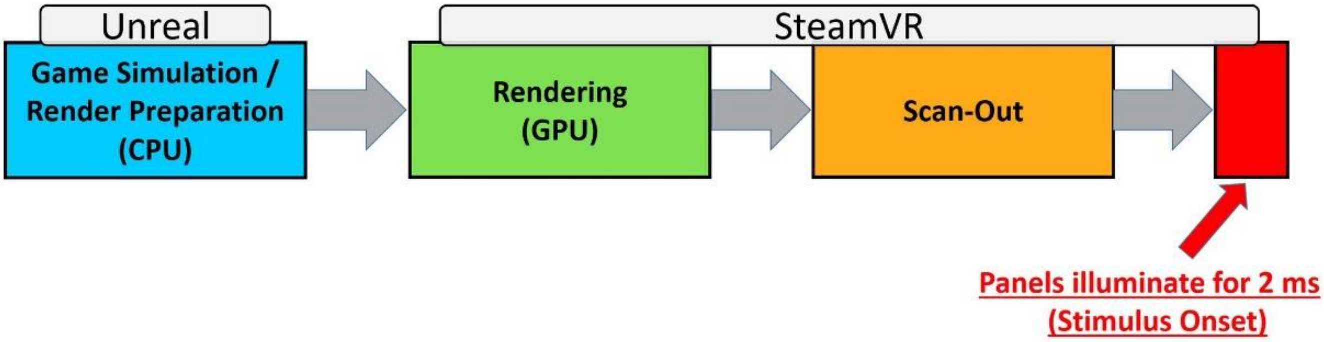

Fig 6. Basic illustration of the graphics pipeline.

https://doi.org/10.1371/journal.pone.0231152.g006

describes the transfer of the image via HDMI, thereby loading it onto the displays. After scan-out completion, the display panels illuminate, and the frame gets visible. The last two stages are controlled by SteamVR rather than UE4.

Calling the function to present a target stimulus in UE4 is just the first step of a whole cascade of processes required to display a stimulus. Timestamping a stimulus event at the beginning of this cascade (when the function to present the stimulus is called) is far too early, resulting in the measurement errors observed in Experiment 2.

However, the graphics pipeline illustrates that it is not possible to measure the stimulus onset directly via UE4. After submitting the rendering commands to the GPU, SteamVR determines when the frame is sent to the HMD. This is further complicated by the fact, that unlike in rendering to a normal computer monitor, the scan-out has to be finished before the displays light up and the stimulus is presented.

A more detailed illustration and description of the temporal characteristics of the graphics pipeline is shown in the supporting information (S2 Fig and S1 Text).

Prediction. The method that allows determining the true stimulus onset time is based on prediction. In particular, in order to predict the exact stimulus onset, a procedure can be implemented that is based on SteamVR's procedures to predict the user's pose. One of the critical determinants to provide a compelling VR experience is to reduce the so-called Motion-To-Photon latency. Motion-To-Photon latency describes the time that is required for a user's head movement to be fully reflected in the HMD. As described above, the system has to determine the layout of the scene before the scene can be rendered. One of the critical factors that determine what is currently visible and what hence needs to be rendered is the user's pose, i.e., the position and the orientation of the HMD. However, rendering takes time and induces undesired latencies, thereby reducing the quality of the VR experience. Furthermore, latency is one of the main drivers of motion sickness, a side effect that leads to symptoms like nausea, dizziness, or vertigo [51].

SteamVR's strategy to reduce the Motion-To-Photon latency is to predict the user's pose, giving the best estimate for the users's pose when the frame is displayed on the headset, to compensate for the latency in the system. To obtain an accurate prediction, the system needs to estimate when the currently rendered frame will finally be displayed. All necessary functions to predict when a frame will be displayed are provided within SteamVR's application programming interface (API) OpenVR. The proposed method is synchronized to SteamVR's pose prediction processes and further utilizes the provided functions to estimate the exact onset time of a frame.

SteamVR allows predicting when a frame will be presented. The prediction is purely temporal and does not contain any information on the frame's content. In other words, it holds information on when a frame will be presented but not on what will be presented in that frame. Trial specific information is only available through UE4. Hence, in order to be able to predict the onset time of a specific event such as the stimulus onset, it is necessary to combine information about the contents and temporal orders of trials from UE4 and the exact frame-timing from SteamVR. Our background application can meet this challenge.

In particular, our background application first synchronizes with the trial-timing of UE4. For this synchronization, UE4 sends a trigger-signal (onset-trigger) to the background application whenever a frame marking the stimulus-onset is about to be submitted to the GPU, right before the control over the rendering is handed over to SteamVR. Following the trigger-signal, the background application stores a timestamp and fetches the necessary information about the frame-timing from SteamVR, which is then used to predict the stimulus-onset (Fig 7).

A more detailed description of the processing steps involved in the prediction can be found in the supporting information, S3 Fig and S2 Text.

Assumptions. Onset prediction is based on a few assumptions that have to be met. The first assumption is that there are no additional frame buffers in the graphics pipeline. Moreover, the Experiment must run at a constant frame rate without dropped or reprojected frames and without the use of Motion Smoothing. We recommend turning off Motion Smoothing completely. However, Asynchronous Reprojection cannot be turned off, and to prevent frame drops and reprojected frames, optimization of the experiment is mandatory. Yet, even with such optimization, occasional performance problems cannot be ruled out entirely, and, hence, should be controlled for. Therefore, a control mechanism was implemented to detect trials affected by dropped or reprojected frames. The background application stores the number of dropped or reprojected frames of every experimental trial and stores it into a log file. Whenever a trial contains dropped or reprojected frames, the trial should be excluded from the analysis, since the validity of the measured reaction time cannot be guaranteed.

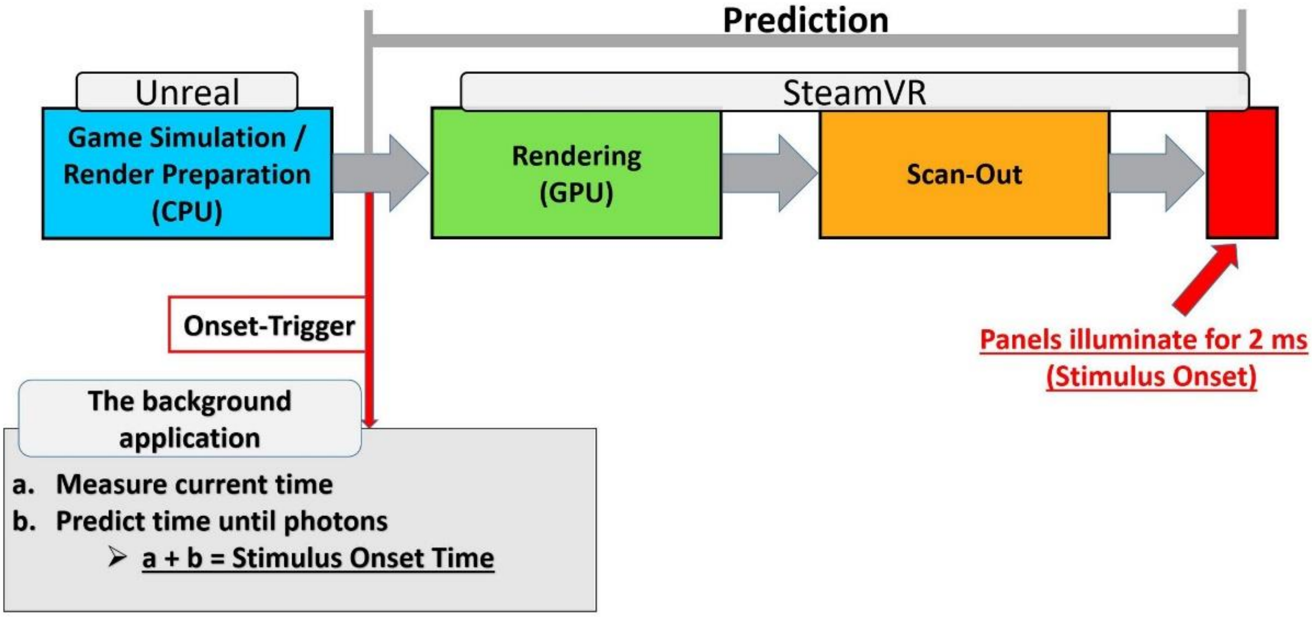

Fig 7. Illustration of the stimulus onset prediction.

https://doi.org/10.1371/journal.pone.0231152.g007

Validation of the method

To ensure that the above-described method can measure precisely and accurately reaction times, we repeated the testing procedure of Experiment 2. The three conditions of Experiment 1 were tested to determine the precision and accuracy of the method. Furthermore, the tests were conducted on both computers, previously used in Experiment 1, to determine the reliability of the method on different computer hardware. Overall, the method was tested in 6000 trials.

Results

The mean results for the three conditions are presented in Table 4. The mean measurement error was 1.44 ms within a range of 2 ms across conditions and computers. Performance remained good during all tests, as no dropped or reprojected frames were observed.

To test for differences in precision or accuracy between conditions or computers, a two factorial ANOVA with the factors *Condition* [Simple ($M = 1.446$ ms, $SD = 0.498$), Complex-Static ($M = 1.443$, $SD = 0.497$), Complex-Moving ($M = 1.439$, $SD = 0.496$)] and *Computer* (Computer 1, Computer 2) was performed. The ANOVA revealed no significant main effects—neither for the factor *Condition* ($F(1,2) = 0.102$, $p = 0.903$), nor for the factor Computer [Computer 1 ($M = 1.450$ ms, $SD = 0.4975$) and Computer 2 ($M = 1.435$, $SD = 0.4965$); ($F(1,1) = 1.305$, $p = 0.253$)]. The interaction between the factors *Condition* and *Computer* was also not significant ($F(1,2) = 0.345$, $p = 0.709$). The individual results of both computers separately can be found in the supporting information (see S8 Table and S9 Table).

Discussion

In Experiment 2, we observed highly variable reaction time measurements with inaccurate stimulus onset measurements with Unreal Engine, SteamVR, and the HTC Vive VR system, when using the built-in timing procedures.

Hence, we developed a new software-method to provide precise and accurate reaction time measurements with Unreal Engine and SteamVR. Instead of measuring reaction times in Unreal Engine directly, the measurement part was outsourced to a background application. Timing benchmarks indicate that the method allows precise and accurate reaction time measurements.

We made the source code of the background application and the project files as well as project builds of the UE4 projects as tested in this experiment available on https://jugit.fz-juelich.de/inm3/timingvr. The commits to the UE4 source code are available on https://github.com/INM3FZJ/UnrealEngine/tree/TimingVR. However, note that the background application was developed as a proof-of-principle rather than a ready-to-use toolbox, as the current implementation is limited in its functionality and was custom-made to fit the particular requirements of the current research project.

Table 4. Overview of mean reaction time errors, standard deviation, minimum and maximum error for each condition (in ms).

Condition	Mean error	SD	Min	Max
Simple	1.446	0.498	1.00	3.00
Complex-Static	1.443	0.497	1.00	2.00
Complex-Moving	1.439	0.496	1.00	2.00
Overall	1.442	0.497	1.00	3.00

https://doi.org/10.1371/journal.pone.0231152.t004

General discussion

The study aimed to investigate the precision and accuracy in stimulus presentation durations and reaction time measurements with Unreal Engine and SteamVR in combination with the HTC Vive VR system.

In Experiment 1, we tested the precision and accuracy of stimulus presentations to determine potential discrepancies between the stimulus timing defined by the researcher and the actual stimulus timing on the display. Our measurements revealed that the precise frame rate of the HTC Vive slightly differs from the information given in the HTC Vive manual. The HMD's empirical frame rate was determined as 89.53 FPS rather than the indicated 90 FPS. Importantly, the frame rate remained stable across conditions and in combination with different computer hardware. The same frame rate was confirmed in a second HTC Vive system, suggesting that the observed frame rate was device specific.

The results of Experiment 1 indicate that the stimulus timing was precise and accurate and remained stable across different computer hardware configurations as well as across different rendering workload requirements. UE4 in combination with SteamVR and the HTC Vive VR system, therefore, appears to be suitable for time-critical visual stimulus presentations. The so-called global-onset displays as used in the HTC Vive come with another significant advantage. Due to the nature of how the image is updated in LCD and CRT monitors, the measurement error of the stimulus onset is correlated with the position of the stimulus on the display. In contrast, the displays of the HTC Vive load the entire image before the displays illuminate. Consequently, when used in combination with precise time measurements, the global-onset displays allow determining the onset of a stimulus without variable measurement errors due to the stimulus position.

Experiment 2 aimed to investigate the limitations of the build-in timing procedures of UE4 for reaction time measurements. Timing benchmarks resulted in highly variable measurement errors more than ten times higher as those obtained with Presentation and PsychoPy.

Hence, a new software-based method as a proof of principle was developed that allows precise and accurate reaction time measurements with UE4 in combination with SteamVR. The new method measures reaction times as independent as possible from the procedures implemented in UE4. This is achieved by outsourcing the measurement procedure into a background application that allows circumventing the limitations of UE4's build-in timing procedures. In a first step, a subroutine implemented into the background application monitors and processes keyboard messages before they reach UE4 and hence allows the collection of frame rate independent response times. In a second step, a prediction algorithm, based on SteamVR's pose prediction procedures, was implemented to determine the true stimulus onset accurately. Timing benchmarks show that the method accurately and precisely determines stimulus onsets and hence in combination with veridical response time acquisition allows validly measuring reaction times.

The method presented here could help to simplify experimental procedures, as is allows to measure stimulus events and response times without a need for complex hardware and sensor setups.

Please note, the background application introduced here, is a proof-of-principle rather than a ready-to-use toolbox. Our aim was to provide an example of a framework for reaction time measurement that could be integrated into other toolboxes.

Although our method was tested with UE4, the underlying principles are, at least in principle, also applicable to other game engines such as Unity, as it is only necessary to provide a trigger signal to synchronize the reaction time measurements of the background application with the trial-timing of the game engine. Furthermore, to our understanding, the principles to predict the stimulus-onset should also apply to the Oculus VR runtime software. Oculus uses a similar mechanism to predict the user's pose as SteamVR, and the Oculus API also provides the tools necessary to predict the onset time of a frame and hence could be used to predict the stimulus onset. Additionally, the version of the background application here presented was already designed to also support different HMD's other than the HTC Vive.

Furthermore, the ability to accurately and precisely determine the time of stimulus events, such as stimulus onset, and response times, is not only essential for reaction time measurements. Instead, a precise recording of events within an experiment is mandatory for the analysis of all time-sensitive functional measures, such as EEG or MEG. Therefore, the method could also be extended, that stimulus and response events could be used as markers to synchronize with time-sensitive functional measures.

However, it should be noted that the method heavily relies on game engines and VR runtime software (here UE4 and SteamVR). Therefore, a limitation of this method is that future updates of these third-party tools might cause compatibility problems or result in incorrect measurements. Hence, it will be important to regularly validate the precision and accuracy of time measurements with new software releases.

Furthermore, we developed and tested the method only for standard keyboard input and did not include support for the HTC Vive's motion controllers. Motions controllers are increasingly used to investigate naturalistic behavior, as they allow to track the hand movements of the participant [e.g., 52,53]. Future research needs to assess timing errors related to motion controllers.

In sum, the HTC Vive, in combination with UE4 and our newly developed tool, constitutes a precise and valuable instrument for neuroscientific research and visual sciences. Besides its excellent performance, regarding accuracy and precision it allows taking advantage of all the benefits that VR offers. When used in combination with precise research tools, VR has the potential to implement new paradigms involving large fields of view of realistic and three-dimensional environments. It allows integrating a variety of behavioral responses, increasing the ecological validity of neuroscientific experiments, which may lead to more generalizable and valid explanations regarding cognitive processes.

Supporting information

S1 File.
(DOCX)

S1 Fig. Screenshot of UE4's realistic rendering sample.
(TIF)

S2 Fig. Illustration of the processing stages that a frame has to pass before it is presented to the display panels.
(TIF)

S3 Fig. Framework for the stimulus onset prediction.
(TIF)

S1 Table. Results across all conditions of Computer 1 (in ms).
(DOCX)

S2 Table. Results across all conditions of Computer 2 (in ms).
(DOCX)

S3 Table. Results of the simple condition across computers (in ms).
(DOCX)

S4 Table. Results of the Complex-Static condition across computers (in ms).
(DOCX)

S5 Table. Results of the Complex-moving condition across computers (in ms).
(DOCX)

S6 Table. Results of the simple condition of HMD 1 (in ms).
(DOCX)

S7 Table. Results of the simple condition of HMD 2 (in ms).
(DOCX)

S8 Table. Overview of mean reaction time errors, standard deviation, minimum and maximum error for each condition of Computer 1 (in ms).
(DOCX)

S9 Table. Overview of mean reaction time errors, standard deviation, minimum and maximum error for each condition of Computer 2 (in ms).
(DOCX)

S1 Text. Temporal characteristics of the graphics pipeline with UE4 and SteamVR.
(DOCX)

S2 Text. Stimulus onset prediction.
(DOCX)

Author Contributions

Conceptualization: Michael Wiesing, Gereon R. Fink, Ralph Weidner.

Data curation: Michael Wiesing.

Formal analysis: Michael Wiesing.

Funding acquisition: Gereon R. Fink.

Investigation: Michael Wiesing.

Methodology: Michael Wiesing, Ralph Weidner.

Project administration: Michael Wiesing.

Resources: Michael Wiesing, Gereon R. Fink, Ralph Weidner.

Software: Michael Wiesing.

Supervision: Ralph Weidner.

Visualization: Michael Wiesing.

Writing – original draft: Michael Wiesing.

Writing – review & editing: Michael Wiesing, Gereon R. Fink, Ralph Weidner.

References

1. Bohil CJ, Alicea B, Biocca FA. Virtual reality in neuroscience research and therapy. Nat Rev Neurosci. 2011 Dec; 12(12):752–62. https://doi.org/10.1038/nrn3122 PMID: 22048061
2. Loomis JM, Blascovich JJ, Beall AC. Immersive virtual environment technology as a basic research tool in psychology. Behav Res Methods Instrum Comput. 1999; 31(4):557–564. https://doi.org/10.3758/bf03200735 PMID: 10633974

3. Rizzo AA, Buckwalter JG, Bowerly T, Van Der Zaag C, Humphrey L, Neumann U, et al. The Virtual Classroom: A Virtual Reality Environment for the Assessment and Rehabilitation of Attention Deficits. Cyberpsychol Behav. 2000 Jun; 3(3):483–99.

4. Foerster RM, Poth CH, Behler C, Botsch M, Schneider WX. Using the virtual reality device Oculus Rift for neuropsychological assessment of visual processing capabilities. Sci Rep. 2016 Dec; 6(1):37016.

5. Foerster RM, Poth CH, Behler C, Botsch M, Schneider WX. Neuropsychological assessment of visual selective attention and processing capacity with head-mounted displays. Neuropsychology. 2019 Mar; 33(3):309–18. https://doi.org/10.1037/neu0000517 PMID: 30652888

6. Unity Technologies. Unity [Internet]. Unity. n.d. [cited 2019 Apr 22]. Available from: https://unity.com/frontpage

7. EpicGames. Unreal Engine [Internet]. Unreal Engine. n.d. [cited 2019 Apr 22]. Available from: https://www.unrealengine.com/en-US/

8. Laak K-J, Vasser M, Uibopuu OJ, Aru J. Attention is withdrawn from the area of the visual field where the own hand is currently moving. Neurosci Conscious [Internet]. 2017 Jan 1 [cited 2017 Dec 11]; 2017 (1). Available from: http://academic.oup.com/nc/article/doi/10.1093/nc/niw025/2970154

9. Lin J, Zhu Y, Kubricht J, Zhu S-C, Lu H. Visuomotor Adaptation and Sensory Recalibration in Reversed Hand Movement Task. 2017; 6.

10. Valve Corporation. SteamVR [Internet]. n.d. [cited 2019 Apr 24]. Available from: https://steamcommunity.com/steamvr

11. Oculus VR, LLC. Oculus [Internet]. n.d. [cited 2019 Jul 23]. Available from: https://www.oculus.com/

12. Le Chénéchal M, Chatel-Goldman J. HTC Vive Pro time performance benchmark for scientific research. In: ICAT-EGVE 2018 [Internet]. Limassol, Cyprus; 2018 [cited 2019 Feb 7]. Available from: https://hal.archives-ouvertes.fr/hal-01934741

13. Quinlivan B, Butler JS, Beiser I, Williams L, McGovern E, O'Riordan S, et al. Application of virtual reality head mounted display for investigation of movement: a novel effect of orientation of attention. J Neural Eng. 2016 Oct 1; 13(5):056006. https://doi.org/10.1088/1741-2560/13/5/056006 PMID: 27518212

14. Watson Marcus R, Benjamin V, Christopher T, Asif H, Thilo W. USE: An integrative suite for temporally-precise psychophysical experiments in virtual environments for human, nonhuman, and artificially intelligent agents [Internet]. Neuroscience; 2018 Oct [cited 2020 Feb 1]. Available from: http://biorxiv.org/lookup/doi/10.1101/434944

15. Wienrich C, Gross R, Kretschmer F, Muller-Plath G. Developing and Proving a Framework for Reaction Time Experiments in VR to Objectively Measure Social Interaction with Virtual Agents. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) [Internet]. Reutlingen: IEEE; 2018 [cited 2020 Feb 1]. p. 191–8. Available from: https://ieeexplore.ieee.org/document/8446352/

16. Doucet G, Gulli RA, Martinez-Trujillo JC. Cross-species 3D virtual reality toolbox for visual and cognitive experiments. J Neurosci Methods. 2016 Jun; 266:84–93. https://doi.org/10.1016/j.jneumeth.2016.03.009 PMID: 27015795

17. Elze T. Achieving precise display timing in visual neuroscience experiments. J Neurosci Methods. 2010 Aug; 191(2):171–9. https://doi.org/10.1016/j.jneumeth.2010.06.018 PMID: 20600318

18. Lagroix HEP, Yanko MR, Spalek TM. LCDs are better: Psychophysical and photometric estimates of the temporal characteristics of CRT and LCD monitors. Atten Percept Psychophys. 2012 Jul; 74 (5):1033–41. https://doi.org/10.3758/s13414-012-0281-4 PMID: 22359147

19. Elze T. Misspecifications of Stimulus Presentation Durations in Experimental Psychology: A Systematic Review of the Psychophysics Literature. Gribble PL, editor. PLoS ONE. 2010 Sep 29; 5(9):e12792. https://doi.org/10.1371/journal.pone.0012792 PMID: 20927362

20. Bridges D, Pitiot A, MacAskill MR, Peirce JW. The timing mega-study: comparing a range of experiment generators, both lab-based and online [Internet]. PsyArXiv; 2020 Jan [cited 2020 Jan 30]. Available from: https://osf.io/d6nu5

21. Cooper EA, Jiang H, Vildavski V, Farrell JE, Norcia AM. Assessment of OLED displays for vision research. J Vis. 2013 Oct 1; 13(12):16–16. https://doi.org/10.1167/13.12.16 PMID: 24155345

22. Geffroy B, le Roy P, Prat C. Organic light-emitting diode (OLED) technology: materials, devices and display technologies. Polym Int. 2006 Jun; 55(6):572–82.

23. Elze T, Taylor C, Bex PJ. An evaluation of organic light emitting diode monitors for medical applications: Great timing, but luminance artifacts: Organic light emitting diode monitors for medical applications. Med Phys. 2013 Aug 26; 40(9):092701. https://doi.org/10.1118/1.4818056 PMID: 24007183

24. Vlachos A. Advanced_VR_Rendering_GDC2015.pdf [Internet]. 2015 [cited 2019 Apr 22]. Available from: http://media.steampowered.com/apps/valve/2015/Alex_Vlachos_Advanced_VR_Rendering_GDC2015.pdf

25. Neurobehavioral Systems, Inc. Neurobehavioral Systems [Internet]. 2019 [cited 2019 May 23]. Available from: https://www.neurobs.com/
26. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: Experiments in behavior made easy. Behav Res Methods. 2019 Feb 1; 51(1):195–203. https://doi.org/10.3758/s13428-018-01193-y PMID: 30734206
27. Brainard DH. The Psychophysics Toolbox. Spat Vis. 1997; 10(4):433–6. PMID: 9176952
28. WorldViz. Vizard | Virtual Reality software for researchers [Internet]. 2019 [cited 2019 May 23]. Available from: https://www.worldviz.com/vizard-virtual-reality-software
29. Enns JT, Di Lollo V. What's new in visual masking? Trends Cogn Sci. 2000 Sep 1; 4(9):345–52. https://doi.org/10.1016/s1364-6613(00)01520-5 PMID: 10962616
30. Raymond JE, Shapiro KL, Arnell KM. Temporary suppression of visual processing in an RSVP task: An attentional blink? J Exp Psychol Hum Percept Perform. 1992; 18(3):849–60. https://doi.org/10.1037//0096-1523.18.3.849 PMID: 1500880
31. Zeng H, Kreutzer S, Fink GR, Weidner R. The source of visual size adaptation. J Vis. 2017 Dec 1; 17 (14):8–8. https://doi.org/10.1167/17.14.8 PMID: 29228141
32. EpicGames. Forward Shading Renderer [Internet]. n.d. [cited 2019 May 20]. Available from: https://docs.unrealengine.com/en-us/Engine/Performance/ForwardRenderer
33. Akenine-Möller T, Haines E, Hoffman N. Real-Time Rendering, Fourth Edition [Internet]. A K Peters/CRC Press; 2018 [cited 2019 Aug 1]. Available from: https://www.taylorfrancis.com/books/9780429225406
34. Vlachos A. Introducing SteamVR Motion Smoothing Beta [Internet]. 2018 [cited 2019 Aug 29]. Available from: https://steamcommunity.com/games/250820/announcements/detail/1696061565016280495
35. Vlachos A. Advanced VR Rendering Performance. 2015.
36. Plant RR, Hammond N, Turner G. Self-validating presentation and response timing in cognitive paradigms: How and why? Behav Res Methods Instrum Comput. 2004 May; 36(2):291–303. https://doi.org/10.3758/bf03195575 PMID: 15354695
37. Garaizar P, Vadillo MA, López-de-Ipiña D, Matute H. Measuring Software Timing Errors in the Presentation of Visual Stimuli in Cognitive Neuroscience Experiments. Hamed SB, editor. PLoS ONE. 2014 Jan 7; 9(1):e85108. https://doi.org/10.1371/journal.pone.0085108 PMID: 24409318
38. Schmidt WC. Presentation accuracy of Web animation methods. Behav Res Methods Instrum Comput. 2001 May; 33(2):187–200. https://doi.org/10.3758/bf03195365 PMID: 11447672
39. Stewart N. Millisecond accuracy video display using OpenGL under Linux. Behav Res Methods. 2006 Feb; 38(1):142–5. https://doi.org/10.3758/bf03192759 PMID: 16817523
40. EpicGames. Realistic Rendering [Internet]. n.d. [cited 2019 Aug 1]. Available from: https://docs.unrealengine.com/en-US/Resources/Showcases/RealisticRendering/index.html
41. Reimers S, Stewart N. Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. Behav Res Methods. 2015 Jun; 47(2):309–27. https://doi.org/10.3758/s13428-014-0471-1 PMID: 24903687
42. Elze T, Tanner TG. Temporal Properties of Liquid Crystal Displays: Implications for Vision Science Experiments. Krekelberg B, editor. PLoS ONE. 2012 Sep 11; 7(9):e44048. https://doi.org/10.1371/journal.pone.0044048 PMID: 22984458
43. Wang P, Nikolic D. An LCD Monitor with Sufficiently Precise Timing for Research in Vision. Front Hum Neurosci [Internet]. 2011 [cited 2019 Feb 5];5. Available from: https://www.frontiersin.org/articles/10.3389/fnhum.2011.00085/full
44. Wiesing M. Combined expectancies: the role of expectations for the coding of salient bottom-up signals. Exp Brain Res. : 13.
45. Brookes J, Warburton M, Alghadier M, Mon-Williams M, Mushtaq F. Studying human behavior with virtual reality: The Unity Experiment Framework. Behav Res Methods [Internet]. 2019 Apr 22 [cited 2020 Feb 1]; Available from: http://link.springer.com/10.3758/s13428-019-01242-0
46. Vasser M, Kängsepp M, Magomedkerimov M, Kilvits K, Stafinjak V, Kivisik T, et al. VREX: an open-source toolbox for creating 3D virtual reality experiments. BMC Psychol [Internet]. 2017 Dec [cited 2019 Feb 7]; 5(1). Available from: http://bmcpsychology.biomedcentral.com/articles/10.1186/s40359-017-0173-4
47. Sivanathan A, Lim T, Louchart S, Ritchie J. Temporal multimodal data synchronisation for the analysis of a game driving task using EEG. Entertain Comput. 2014 Dec; 5(4):323–34.
48. Jangraw DC, Johri A, Gribetz M, Sajda P. NEDE: An open-source scripting suite for developing experiments in 3D virtual environments. J Neurosci Methods. 2014 Sep; 235:245–51. https://doi.org/10.1016/j.jneumeth.2014.06.033 PMID: 25064189

49. Microsoft. Hooks [Internet]. 2018 [cited 2019 Apr 15]. Available from: https://docs.microsoft.com/en-us/windows/desktop/winmsg/hooks

50. Microsoft. Event Objects [Internet]. 2018 [cited 2019 Jul 22]. Available from: https://docs.microsoft.com/en-us/windows/win32/sync/event-objects

51. Behr K-M, Nosper A, Klimmt C, Hartmann T. Some practical considerations of ethical issues in VR research. Presence Teleoperators Virtual Environ. 2005; 14(6):668–676.

52. Gerig N, Mayo J, Baur K, Wittmann F, Riener R, Wolf P. Missing depth cues in virtual reality limit performance and quality of three dimensional reaching movements. Jan Y-K, editor. PLOS ONE. 2018 Jan 2; 13(1):e0189275. https://doi.org/10.1371/journal.pone.0189275 PMID: 29293512

53. Soret R, Charras P, Hurter C, Peysakhovich V. Attentional orienting in virtual reality using endogenous and exogenous cues in auditory and visual modalities. In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications—ETRA '19 [Internet]. Denver, Colorado: ACM Press; 2019 [cited 2020 Jan 31]. p. 1–8. Available from: http://dl.acm.org/citation.cfm?doid=3314111.3321490

Supporting information

S1 Text. Temporal characteristics of the graphics pipeline with UE4 and SteamVR

Figure S2 illustrates the different stages that a frame (devoted as frame n) must pass before it is finally presented on display. The following description will focus on the temporal aspects rather than on the functional aspects of the described stages.

Please note, that every stage in the pipeline is synchronized to the vertical sync (VSync) event. Frame n is initially processed in the Game Thread where game simulation takes place. The Game Thread constitutes a relatively early stage in this processing pipeline and represents the stage where the experimental routine is executed, e.g., where the command to present the target stimulus is completed. The timestamp marking a stimulus' onset as measured in Experiment 2 was determined at this level. After one frame duration, everything is transferred to the Draw Thread for rendering preparation. It also requires an additional full frame duration until the resulting rendering commands are passed to the GPU. The GPU then again requires another full frame duration before, finally, the frame can be scanned out to the displays. The scan-out occurs while the display is black, which requires about 9 ms until the displays light up, and the new frame is presented for about 2 ms. The last two stages are controlled by SteamVR rather than UE4, illustrated by the horizontal black line in Fig S2.

Note that both the Game Thread and the Draw Thread are shifted relative to the VSync events. This is a novelty of rendering with SteamVR. In a non-VR graphics pipeline, both the game and the draw thread would start right after a VSync event. The rendering commands would then be submitted to the GPU right after the VSync event, and everything would be buffered for one or more frame durations before the GPU would render the frame. In order to reduce latency, VR rendering occurs without frame buffering. However, submitting a frame to the GPU is time-consuming, and without frame buffering, a so-called "GPU bubble" of up to two milliseconds would be produced, in which the GPU is idle until rendering can be initiated, thus effectively reducing the total time available for rendering. The rendering commands of the Draw Thread are submitted to the GPU before the VSync event to ensure that the GPU has a full frame duration for rendering. This process is called "running start". To ensure that both the Game Thread and the Draw Thread still have the budget of a full frame duration, with SteamVR, the calculations of both threads start a few milliseconds before the VSync event.

The scan-out is the final stage before the frame is displayed. After loading the entire display, the pixels illuminate, and the frame gets visible for about 2 ms.

When considering the above-presented graphics pipeline involved in presenting a single frame, it becomes evident that calling the function to present a target stimulus in the Game Thread is just the first step of a whole cascade of processes required to display a stimulus. Timestamping a stimulus event at the beginning of this cascade (when the function to present the stimulus is called) is far too early, resulting in the measurement errors observed in Experiment 2. However, the graphics pipeline also illustrates that it is not possible to measure the stimulus onset directly via UE4. After submitting the rendering commands to the GPU, SteamVR determines when the frame is sent to the HMD. This is further complicated by the fact, that unlike in rendering to a normal computer monitor, the scan-out has to be finished before the displays light up and the stimulus is presented.

S2 Text. Stimulus onset prediction

The prediction is based on the fact, that all rendering steps for VR are synchronized to the VSync events, which allows us to use VSync events as reliable time markers when a screen refresh has finished.

An exact estimate is based on accurate knowledge when (i.e., during which frame interval) the prediction was started. This can be accomplished by sending the onset-trigger from UE4 to the background application later within the processing sequence of UE4 than in the previous tests. The onset-trigger is then sent to the background application from UE4's SteamVR plugin at the end of Draw Thread's processing a few milliseconds before the next VSync event. This is accomplished by sending the onset-trigger signal right after a function called WaitGetPoses() returns. The WaitGetPoses() function is responsible for pose prediction and is called by UE4 after the Draw Thread has finished processing. It blocks the Draw Thread until a few milliseconds before the VSync event and then returns the predicted poses to be used for the rendering. By sending the onset-trigger right after WaitGetPoses() returns, a synchronization point with VSync events is generated and ensures that the prediction of a stimulus onset always starts in the same stage of the graphics pipeline.

The current implementation of the synchronization of UE4 and the background application works as follows. The command to show a stimulus on display is called in the Game Thread. In order to generate a precise prediction when the stimulus appears on display, one has to determine the time point when the Draw Thread in UE4 hands over stimulus processing to SteamVR. Therefore, the current background application uses two different trigger signals, one marking the moment when the command to show a stimulus is called and later a second trigger signal marking the moment when the process is handed over to SteamVR. The latter is about the exact timing and for the synchronization with the frame intervals; the former is relevant since it indicates an upcoming visual event that is important for the experimental procedure. Both signals are relevant and need to be considered unison, which creates a need to integrate or relate both at some stage. This could be done by implementing a complicated direct exchange of timing signals between the Game Thread and the Draw Thread. Alternatively, this problem could be solved by using an external interface that registers and relates both trigger signals such as our background application. The latter approach requires only minimal changes in the UE4 source code and hence avoids changes that might affect the normal working of UE4. Therefore, we decided to use this option to develop a proof-of-principle method for precise timing measures in UE4. It involves sending two trigger signals to the background application. In Game Thread, a trigger, which indicates a relevant stimulus to be presented, is sent whenever the command to show the stimulus is called. The trigger signal marking the moment when the signal is handed over from the Draw Thread to SteamVR is sent on every frame. Hence, the background application can integrate both signals in the following way. On every frame the background application receives a signal from the Draw Thread, thereby providing an exact measure on when information is passed to SteamVR, this information is only valuable when it needs to be related to a relevant visual stimulus, i.e., when it is preceded by a trigger signal indicating an upcoming visual event. Otherwise, this information can be ignored. Accordingly, our background application waits for a start-trigger from UE4, indicating an upcoming stimulus onset, and only then further processes the signal indicating the hand-over to SteamVR. When this signal is received, it takes the current time and starts predicting the actual stimulus onset. Adding the measured time to the prediction then estimates stimulus onset. In the next step, the keyboard hook is started, and the application waits until it receives a keyboard message and takes the response time. After writing the timing information into a log file, it starts again to wait for the next start-trigger.

As soon as the background application received the onset-trigger, it takes the current time and then starts predicting the time remaining until the target will be displayed.
The first step in this calculation is to determine the remaining time of the current frame interval accurately. This is achieved by calling a function that returns the time since the last VSync event. The difference between a frame duration and the time since the last VSync determines the remaining time of the frame interval. Instead of hardcoding the frame duration, we calculate the frame duration based on the HMD's refresh rate, which is retrieved from SteamVR. This allows using the background application with different HMD and different refresh rates.
In the next step, one frame duration for the GPU rendering is added to the prediction. The last step of the prediction adds the duration of the scan-out, which is again retrieved from SteamVR. Finally, the predicted time is added to the initially measured time, which in sum makes up the stimulus onset time. The complete framework for the prediction is illustrated in Fig S3.
Note that the last function returns the time from VSync until the midpoint of the stimulus presentation instead of the time of the actual onset, i.e., the very first moment when the displays light up. This results from the default behavior of SteamVR's pose prediction. Therefore, the predicted time is effectively another frame duration resulting in an over-prediction of about 1 ms.
Furthermore, the refresh rate returned by SteamVR, which forms the basis of the calculation for the frame duration, is 90 Hz for the HTC Vive. However, the actual refresh rate that we have measured in Experiment 1 (89.53 Hz) was slightly lower, which results in an under-prediction of each of the three steps of about 58 µs. Taken together, the proposed method does not predict the stimulus onset exactly but results in an over-prediction of about 826 µs, when used with the HTC Vive.
In principle, the over-prediction could be corrected by subtracting 826 µs from the result. However, we decided not to correct this marginal over-prediction due to the following reason. The algorithm for the prediction is completely based on functions provided by OpenVR without any hardcoding. This implementation, at least in principle, should enable experimenters to use the method also with HMDs other than the HTC Vive, as long as they use SteamVR as the runtime.

Fig S1. Screenshot of UE4's Realistic Rendering sample.
The screenshot shows the original environment without the modifications as used in this study.

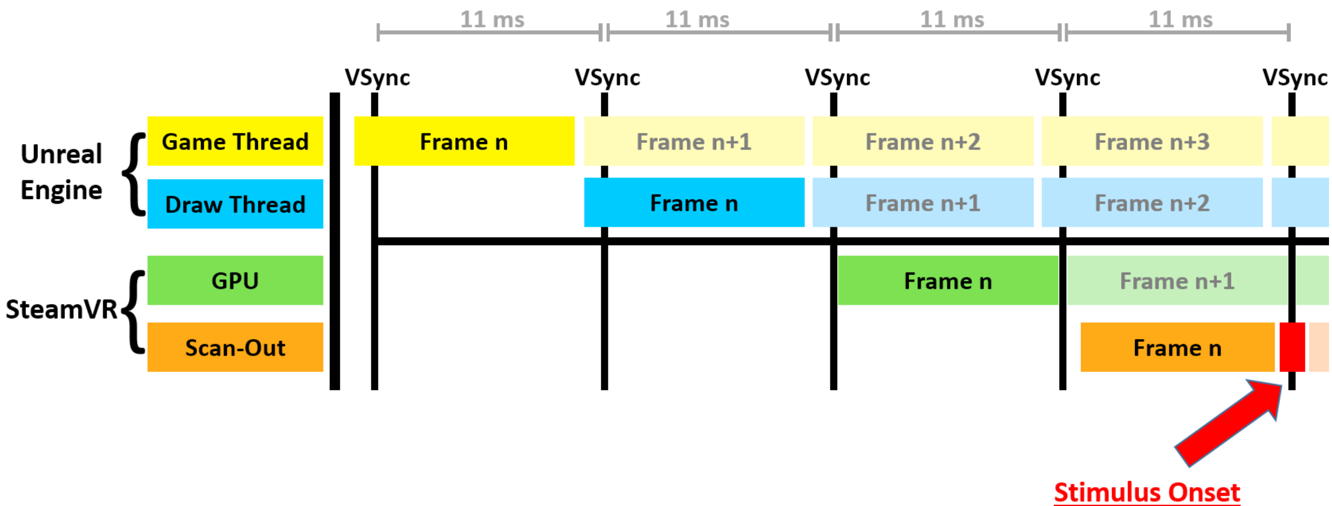

Fig S2. Illustration of the processing stages that a frame has to pass before it is presented to the display panels.

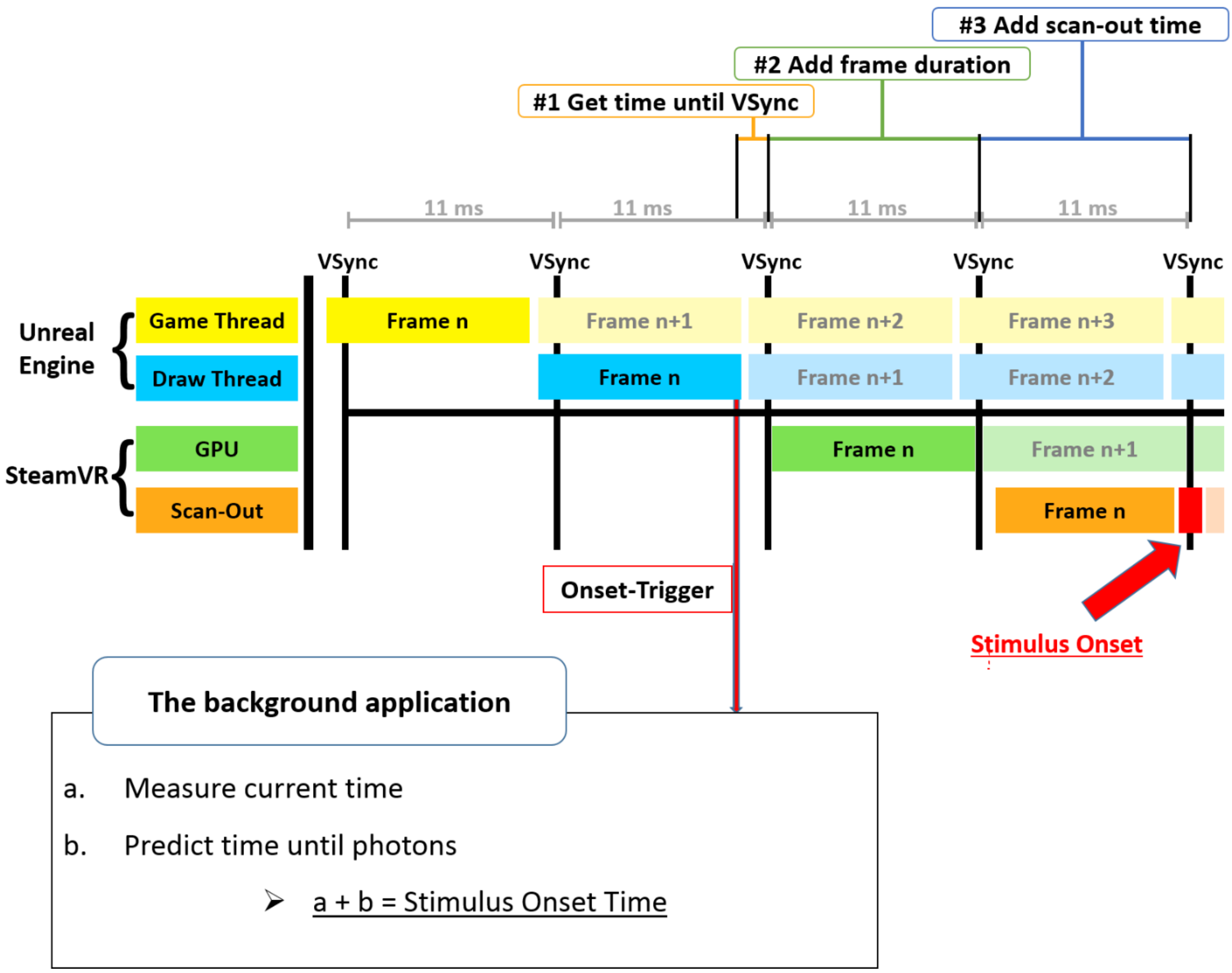

Fig. S3. Framework for the stimulus onset prediction.

S1 Table. Results across all conditions of Computer 1 (in ms).

expected duration	mean	sd	min	max	Mean duration white
2000	2010.56	0.107	2010.50	2010.75	996.04
1000	1005.28	0.081	1005.25	1005.50	496.40
400	402.11	0.124	402.00	402.25	191.82
200	201.06	0.104	201.00	201.25	91.30
133.33	134.04	0.089	134.00	134.25	57.78
66.66	67.02	0.066	67.00	67.25	24.27
22.22	22.34	0.120	22.25	22.50	1.93

The first column represents the expected duration for the black and white stimulus cycles. The last column represents the measured durations of the white stimulus (in ms).

S2 Table. Results across all conditions of Computer 2 (in ms).

expected duration	mean	sd	min	max	Mean duration white
2000	2010.59	0.121	2010.50	2010.75	996.06
1000	1005.30	0.098	1005.25	1005.50	493.41
400	402.12	0.125	402.00	402.25	191.82
200	201.06	0.106	201.00	201.25	91.29
133.33	134.04	0.091	134.00	134.25	57.80
66.66	67.02	0.068	67.00	67.25	24.27
22.22	22.34	0.120	22.25	22.50	1.93

The first column represents the expected duration for the black and white stimulus cycles. The last column represents the measured durations of the white stimulus (in ms).

S3 Table. Results of the Simple condition across computers (in ms).

expected duration	mean	sd	min	max	Mean duration white
2000	2010.58	0.116	2010.50	2010.75	996.04
1000	1005.29	0.091	1005.25	1005.50	493.39
400	402.12	0.125	402.00	402.25	191.81
200	201.06	0.105	201.00	201.25	91.28
133.33	134.04	0.090	134.00	134.25	57.77
66.66	67.02	0.067	67.00	67.25	24.26
22.22	22.34	0.120	22.25	22.50	1.92

The first column represents the expected duration for the black and white stimulus cycles. The last column represents the measured durations of the white stimulus (in ms).

S4 Table. Results of the Complex-Static condition across computers (in ms).

expected duration	mean	sd	min	max	Mean Duration white
2000	2010.58	0.116	2010.50	2010.75	996.10
1000	1005.29	0.090	1005.25	1005.50	493.41
400	402.12	0.125	402.00	402.25	191.82
200	201.06	0.105	201.00	201.25	91.30
133.33	134.04	0.090	134.00	134.25	57.79
66.66	67.02	0.067	67.00	67.25	24.28
22.22	22.34	0.120	22.25	22.50	1.93

The first column represents the expected duration for the black and white stimulus cycles. The last column represents the measured durations of the white stimulus (in ms).

S5 Table. Results of the Complex-moving condition across computers (in ms).

expected duration	mean	sd	min	max	Mean Duration white
2000	2010.58	0.116	2010.50	2010.75	996.10
1000	1005.29	0.091	1005.25	1005.50	493.41
400	402.12	0.125	402.00	402.25	191.82
200	201.06	0.105	201.00	201.25	91.30
133.33	134.04	0.090	134.00	134.25	57.79
66.66	67.02	0.067	67.00	67.25	24.28
22.22	22.34	0.120	22.25	22.50	1.93

The first column represents the expected duration for the black and white stimulus cycles. The last column represents the measured durations of the white stimulus (in ms).

S6 Table. Results of the Simple condition of HMD 1 (in ms).

Expected duration	mean	sd	min	max	mean duration white
2000	2010.56	0.107	2010.50	2010.75	996.04
1000	1005.28	0.082	1005.25	1005.50	493.39
400	402.11	0.124	402.00	402.25	191.81
200	201.06	0.104	201.00	201.25	91.28
133.33	134.04	0.089	134.00	134.25	57.77
66.66	67.02	0.066	67.00	67.25	24.26
22.22	22.34	0.120	22.25	22.50	1.92

The first column represents the expected duration for the black and white stimulus cycles. The last column represents the measured durations of the white stimulus.

S7 Table. Results of the Simple condition of HMD 2 (in ms).

expected duration	mean	sd	min	max	mean duration white
2000	2010.56	0.107	2010.50	2010.75	996.03
1000	1005.28	0.081	1005.25	1005.50	493.39
400	402.11	0.124	402.00	402.25	191.81
200	201.06	0.104	201.00	201.25	91.28
133.33	134.04	0.089	134.00	134.25	57.77
66.66	67.02	0.066	67.00	67.25	24.26
22.22	22.34	0.120	22.25	22.50	1.92

The first column represents the expected duration for the black and white stimulus cycles. The last column represents the measured durations of the white stimulus.

S8 Table. Overview of mean reaction time errors, standard deviation, minimum and maximum error for each condition of Computer 1 (in ms).

Condition	Mean error	SD	Min	Max
Simple	1.446	0.4973	1.00	2.00
Complex-Static	1.451	0.4978	1.00	2.00
Complex-Moving	1.452	0.4979	1.00	2.00
Overall	1.446	0.4973	1.00	2.00

S9 Table. Overview of mean reaction time errors, standard deviation, minimum and maximum error for each condition of Computer 2 (in ms).

Condition	Mean error	SD	Min	Max
Simple	1.445	0.4992	1.00	3.00
Complex-Static	1.435	0.4960	1.00	2.00
Complex-Moving	1.425	0.4946	1.00	2.00
Overall	1.445	0.4992	1.00	3.00

Chapter 3 - Transferring paradigms from physical to virtual reality

Wiesing, M., Fink, G. R., Vossel, S., & Weidner, R. (in prep). *Transferring paradigms from physical to virtual reality: Can reaction time effects be replicated in a virtual setting?*

Author contributions

M.W. and R.W. conceptualized the study and contributed to the design. M.W. and H.S. developed the virtual environment. MW collected the data; wrote software, analyzed and visualized the data. M.W., H.S., V.S., G.R.F., and R.W. authors contributed to the writing of the manuscript.

Transferring paradigms from physical to virtual reality: Can reaction time effects be replicated in a virtual setting?

Michael Wiesing[1], Hendrik Steinkönig[4], Simone Vossel[1,3], Gereon R. Fink[1,2], Ralph Weidner[1,3]

1 Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre, Juelich, Germany,
2 Department of Neurology, University Hospital Cologne and Faculty of Medicine, University of Cologne, Cologne, Germany
3 Department of Psychology, Faculty of Human Sciences, University of Cologne, Cologne, Germany
4 Independent Researcher, Paderborn, Germany

Corresponding author: Michael Wiesing, wiesing@hhu.de

Conflict of Interest: The authors declare no competing financial interests.

Author contributions

M.W. and R.W. conceptualized the study and contributed to the design. M.W. and H.S. developed the virtual environment. M.W. programmed the experimental procedures, tested the participants, and analyzed the data. All authors contributed to the writing of the manuscript.

Abstract

In the recent years, virtual reality (VR) has gained increasing popularity as a research tool in neuroscience and experimental psychology. However, whether the same cognitive processes are engaged in experiments conducted in front of a computer monitor or in immersive VR is still an open question. For example, acting in an VR environment may demand certain cognitive efforts, thereby reducing capacities available for other processes. Moreover, different findings may emerge due to more basic technical differences regarding, e.g., visual stimulation. So far, studies directly comparing monitor-based and VR experiments are rare. Here, we tested whether reaction time costs induced by violated expectations of basic visual features differ between visual stimulus presentation with a head-mounted display (HDMs) and a standard monitor setup.

We examined whether basic differences in stimulus generation of HMDs as compared to standard monitors affect early visual processing. Hence, a previously introduced experimental paradigm (Wiesing et al., 2020), investigating early processing of prediction errors of basic visual features, was replicated in both the original setup and a replica of the experiment within a virtual environment. In order to minimize dissimilarity between experiments, the entire laboratory was recreated in VR. The virtual replica matched the physical laboratory not only visually, but also in scale, allowing us to blend the virtual laboratory onto its real counterpart. Hence, the VR experiment closely corresponded to the non-VR experiment not only visually, but also with respect to auditory, tactile, or olfactory stimulation in the lab room.

A group of 16 participants performed the experiment in both experimental setups with the order counterbalanced. The results did not provide any evidence that the expectation-dependent processing of basic visual features is different when conducted in VR and showed no evidence for an additional binding of attentional resources. Instead, both experiments replicated results of the original study Wiesing, et al. (2020).

Introduction

To precisely relate behavior and brain activity to specific cognitive functions, neuroscientific studies are usually conducted under controlled laboratory conditions. In cognitive neuroscience and experimental psychology, the stimuli used are relatively simple compared to natural settings, with the big advantage that specific stimulus features can be precisely varied and controlled. Similarly, these experiments are often designed to obtain responses that are comparable between participants and experimental conditions which differ in only few parameters of interest. This restricts the behavioral options of the participants but allows for precise experimental control.

The approach to use simple but highly controlled stimuli and responses generated a vast amount of knowledge about the architecture of cognitive processes and their underlying neural implementation. However, testing this knowledge in real-life situations remains a challenge, since the richness of sensory information and the multitude of potential actions come at the expense of a lack of precise experimental control. Here, virtual real-life situations may provide a possible solution. Virtual Reality (VR) has the potential to bring realistic but well controlled environments into the laboratory. Highly controlled experimental paradigms could be integrated into realistic scenarios, thereby enabling the study of brain functions and complex behavior in completely new ways (Bohil et al., 2011; Loomis et al., 1999b; Wilson & Soranzo, 2015)

With the recent release of head-mounted displays (HMD) on the consumer market, such as Oculus Rift, HTC Vive or Valve Index, VR has increasingly been used in cognitive neuroscience and experimental psychology (Vasser & Aru, 2020b). One of the benefits of HMDs is its ability to present stimuli in stereoscopic 3D on a large field of view (FOV). HMDs cover the entire visual field, essentially giving full control over all visual input, which can be manipulated in real-time. Due to their relatively small size, HDMs are easily portable and an ideal tool for e.g. visual neuropsychological assessment, since they allow to control and maintain critical context factors such as illumination and screen sizes as well as distances, even when patients are examined in different rooms or in different institutions (Foerster et al. (2016, 2019).

However, VR not only allows to precisely control visual stimulation, but also to obtain accurate behavioral measurements as every current state of the art VR system comprises a sensitive motion-tracking system (Niehorster et al., 2017; Verdelet et al., 2019). Some systems are even already available with integrated eye-tracking cameras (Imaoka et al., 2020).

A critical difference between VR and standard monitor setups is the fact that the participants are not required to watch configurations on a 2-dimensional computer screen, but find themselves immersed in an interactive three-dimensional virtual environment, experienced from a first-person perspective. Hence, VR can be used to present stimuli at different viewing distances relative to the observer (Heber et al., 2008; Maringelli et al., 2001).

Although VR has the potential to allow studying perception and behavior within complex naturalistic environments, it is ultimately still an artificial setting which differs from non-virtual reality in various aspects, which have been demonstrated to sometimes cause perceptual and behavioral differences to the corresponding real-world setting. One of these differences is the field of view. While it is larger compared to usual standard 2D displays, it is still smaller than the human visual field, generating an impression that is often compared to the feeling of wearing diving goggles. The size of the FOV affects performance in walking or visual search tasks (Arthur, 2000) and can affect spatial judgments, such as judgement errors of azimuth (Nash et al., 2000). Furthermore, velocity and self-motion are perceived slower with a smaller FOV (Caramenti et al., 2019; Hopper et al., 2019) and head-eye coordination is altered. For instance, Pfeil et al. (2018) observed more head movements within VR than in reality. The latter might also be affected by the weight of current HMDs and their rather low pixel density. Furthermore, other optical artefacts can occur such as chromatic aberrations or spatial distortions, especially in the peripheral parts of the display, leaving only the central area of the display for clear vision. Similarly, especially earlier HMDs such as Oculus Rift CV1 or HTC Vive have a comparably low pixel-density as compared to a typical computer monitor. A side effect of this is the so-called *screen door effect*, which describes the visible empty space between pixels that is perceived as viewing through a mesh on the display.

It is well known that perceived distances in VR differ from real-world distance and that distances in stereoscopic displays are underestimated (e.g., Kelly et al., 2018; Witmer & Kline, 1998), possibly due to vergence-accommodation conflicts (VAC) (Bingham et al., 2001; Hoffman et al., 2008b). Such conflicts emerge due to the fact that in an HMD the image itself is located at a fixed distance but the perceived distance might vary (Batmaz et al., 2019). This can induce visual stress possibly leading to visual fatigue symptoms, such as eye strain or double vision (Guo et al., 2017; Iskander et al., 2019). In principle, VACs might pose a challenge to our visual system and it has been reported they can reduce the cognitive performance in sustained attention tasks (Poltavski et al., 2012) and increase Stroop interference (Daniel & Kapoula, 2019). Compensatory mechanisms necessary to maintain clear vision may bind visual attention, lowering processing resources available for other cognitive processes. However, it is important to note that the impact of VACs on cognitive processing has not been observed in HMDs, but in experiments using different prism and lens systems with more pronounced VACs as compared to HMDs. Hence, it remains to be examined whether these effects also hold when VACs are caused by HMDs.

These findings raise questions about the comparability of experimental findings obtained using non-immersive displays and HMDs. In principle, a failure to replicate findings in cognitive experiments with visual stimuli in VR can be attributed to different levels of processing. On the one hand, acting in an VR environment may demand certain cognitive efforts, thereby reducing capacities available for other processes. On the other hand, different findings may emerge due to more basic differences in visual stimulation.

However, so far, studies directly comparing monitor-based and VR experiment are rare and the results are heterogeneous. While some studies replicated previous findings in visual search (Olk et al., 2018) or flanker tasks (Roberts et al., 2019), others observed different behavioral patterns in VR than in a standard setting in mental rotation (Kozhevnikov & Dhond, 2012) or visuomotor adaptation tasks (Anglin et al., 2017). Findings of another study indicated that more attentional resources are allocated to stimuli in VR as compared to stimuli presented on a 2D monitor (Li et al., 2020b).

Here, we investigated the impact of the different modes of visual stimulus presentation in standard and VR setups on early visual processing of basic visual features. This was done by generating an experimental VR setup that was a faithful replica of a non-VR setup. Instead of a purely visual rendering of the environment, a multimodal virtual laboratory was created, providing also haptic feedback, sound and smell. This was achieved by a realistic and detailed rendering of the lab environment, with every physical object virtually represented in the correct scale and position. This allowed us to blend the visual rendering as an overlay onto its physical counterpart, resulting in a more complete and immersive simulation of the non-VR experiment.

A group of participants took part in two identical behavioral experiments, with the only difference being that one experiment was conducted in VR while the other involved the standard setup. Thus, any differences in task performance could be attributed to the different modes of visual presentation. The task used in the present experiment was based on a new experimental paradigm introduced by

Wiesing, Fink, Weidner, et al., (2020) testing the effects of feature-based predictions on reaction times. This task was chosen to investigate potential effects of an VR setting on relatively early levels of visual processing. In particular, this visual reaction time task involves two-dimensional stimuli and investigates how multiple simultaneous feature expectations are processed for the same object. Two objects were presented on a monitor and participants had to judge whether the spatial frequency of the objects was identical or different. Expectations of two task-irrelevant feature dimensions (color and orientation) were manipulated. The results of a series of four behavioral experiments consistently indicated that prediction errors of different object features are resolved independently before feature binding takes place. Hence, the findings suggest that the processing of the prediction errors happens on an early level of processing before attention comes into play. In particular, Experiment 2 of Wiesing, et al., (2020) was chosen for the current study. In that experiment, the expectations for two task-irrelevant object features, color and orientation, were manipulated while participants performed a discrimination task regarding another dimension.

Taken together, we tested in the current study whether differences in the visual stimulus presentation inherent to HDMs, such as vergence-accommodation conflicts, affect the processing of (violated) expectations about basic visual features when compared to a standard monitor setup.

Materials and methods

Participants

Sixteen participants (5 women, mean age: 34.25 years, age range: 20 - 47, one left-handed) took part in both experiments. All participants had normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Normal color vision was assessed by pseudo isochromatic color plates in all participants (Velhagen & Broschmann, 2003). Before the experiment, written informed consent was obtained following the Declaration of Helsinki. The study was approved by the ethics committee of the German Society of Psychology, and participants were remunerated for their time.

Apparatus

In both experiments, the stimuli were presented on a 22-in monitor at either a real or a virtual distance of 60 cm. In the non-VR experiment, a Samsung SyncMaster (resolution 1680 X 1050; refresh rate 60 Hz) was used. In the VR experiment, stimuli were presented on a virtual monitor that matched the real monitor both in size, shape and resolution, seen through an HTC Vive HMD (resolution 1080 X 1200 per eye, refresh rate 90 Hz). In both experiments, participants were seated at a table in a behavioral laboratory. Head movements were prevented by a chin rest. Stimulus presentation and response recording were controlled using Unreal Engine 4.22 (UE4) (EpicGames), in combination with a custom made software that allowed a precise recording of behavioral responses (see below). The modeling of the virtual environment for the VR-experiment was conducted in Cinema4D (MAXON, Germany) and Blender (Blender Foundation). Participants were provided with button response pads (NAtA Technologies) for each hand and responded by pressing the corresponding button on the response pad with the left and right index finger.

Stimuli and Task

A task previously used in a study by Wiesing et al. (2020) was used in both experiments. Visual stimuli consisted of two horizontally arranged gratings as target stimuli, each incorporating one of two possible spatial frequencies. All combinations of frequencies across stimuli were presented randomly and with an equal probability (i.e., 50 % same and 50 % different).

Additionally, grating stimuli could be colored in red/green or blue/yellow and could have two different orientations (45°, 90°). Both gratings were always identical in color and orientation in each trial. During piloting we found that, with the original stimulus parameters used in Wiesing et al., (2020), the spatial frequencies were difficult to be differentiated in the VR setup. This was resolved by increasing the size of the stimuli to 8° x 8° in both set-ups.

For each participant, a specific color combination (e.g., red/green) was defined as the “expected color” and the other combination as “unexpected color”. Likewise, one specific orientation (e.g., tilted by 45°) was defined as the “expected orientation” and the other orientation as “unexpected orientation”. For both, color and orientation, the expected feature was presented in 87.5% of the trials, while the rare feature was presented in 12.5% of all trials. The rare features were assumed to elicit a strong prediction error, while prediction errors were expected to be minimal in trials with the frequent feature combination.

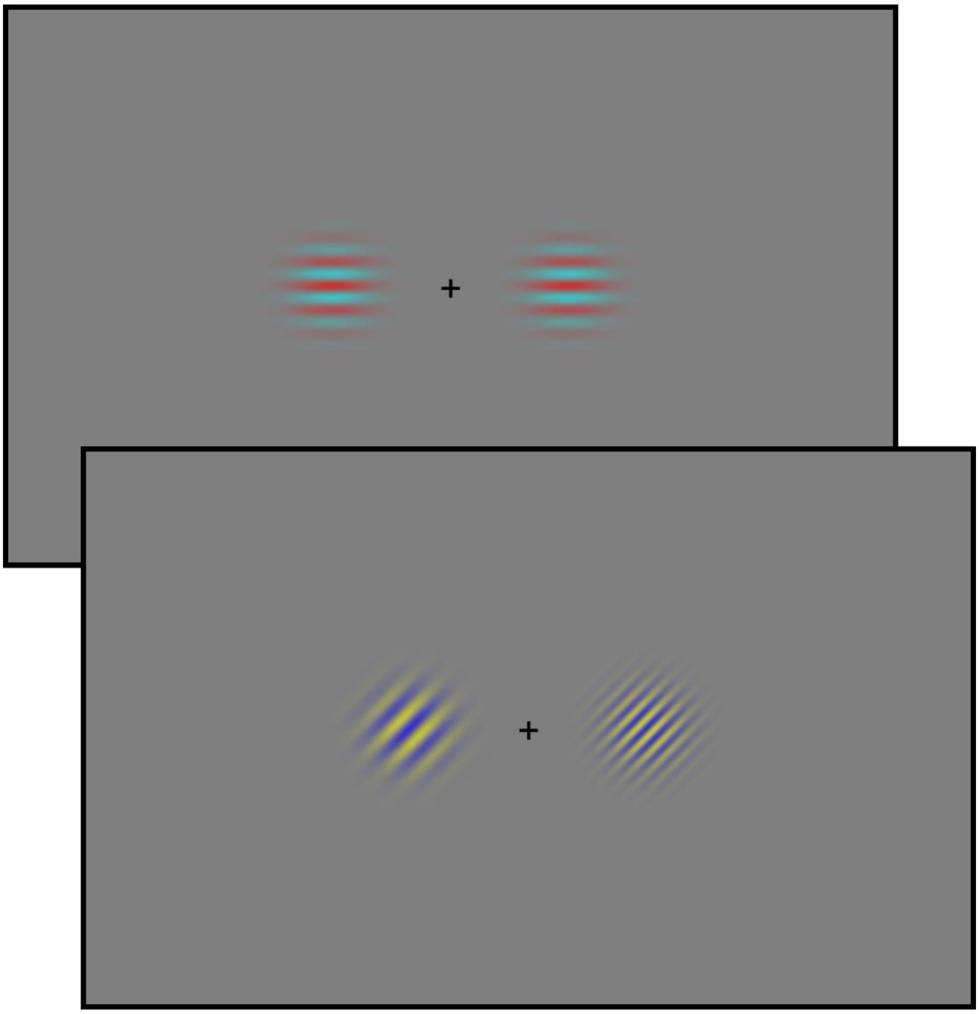

Figure 1. Stimulus example. Participants were asked to respond to the spatial frequency of the two gratings, which could be the same or different. One color scheme and one orientation were presented in the majority of the trials, but some trials were characterized by rare changes in color and/or orientation.

Hence, the amount of prediction error was manipulated separately for the two different features, resulting in a 2 x 2 factorial design with the factors *Color Prediction Error* (high, low) (*ColPE*) and *Orientation Prediction Error* (high, low) (*OriPE*). This yielded four experimental conditions: *ColPE_low/OriPE_low* (color expected and orientation expected), *ColPE_high/OriPE_low* (color unexpected and orientation expected), *ColPE_low/OriPE_high* (color expected and orientation unexpected), and *ColPE_high/OriPE_high* (both color and orientation unexpected).

Each experiment consisted of 14 blocks, comprising 64 trials, resulting in 896 trials. The experiment comprised 700 *ColPE_low/OriPE_low* trials (78.125 %), 84 *ColPE_high/OriPE_low* trials (9.375 %), 84 *ColPE_low/OriPE_high* trials (9.375 %), and 28 *ColPE_high/OriPE_high* trials (3.125 %).

Trials started with the presentation of the two target stimuli and lasted until a response was given. Trials were separated by an inter-trial interval, which randomly varied between 500 and 1000 ms. Each block ended with a break that could be terminated via button press.

The experimental task required participants to indicate by button press with the left or right index finger whether the target stimuli comprised identical or different spatial frequencies. Participants were instructed to respond as fast and accurately as possible. An erroneous response was followed by the message “Fehler” (i.e., German for “error”) displayed for 750 ms.

Game engines

Due to a lack of established stimulus software for VR experiments, several recent studies relied on game engines, such as Unity (e.g., Buckingham, 2019) or Unreal Engine 4 (UE4) (e.g., Lin et al., 2017), to develop their experiments. In a recent study, we observed that time measurements of stimulus and response events collected with UE4’s internal timing procedures are highly variable and inaccurate. Instead, we were only able to record reaction times with a sufficient precision and accuracy after implementing custom measurement procedures into the experimental setup (Wiesing, Fink, & Weidner, 2020). By objective measurements using the Black Box Toolkit (BBTK) (Plant & Turner, 2009), we have shown that UE4 in combination with our method was able to achieve reaction time measurements with a precision and accuracy similar to those of Presentation and PsychoPy.
For the current study, the same method was used to collect reaction time data in the VR experiment. For the non-VR experiment, a modified variant of the method was used that is suitable for non-VR rendering. The precision and accuracy of the time measurements was tested and validated prior to the experiment by using the procedure described by Wiesing, et al., (2020).
For the VR setup, we observed a precision (standard deviation) of 0.463 ms and an accuracy (mean error) of 1.43 ms, which is comparable to results reported by *Wiesing et al., (2020)*. While the precision of the non-VR setup was with 0.496 ms comparable to the VR setup, we observed a substantially reduced accuracy, with a mean error of +46.26 ms. The high mean error of the non-VR experiment can be partially explained due to measuring the stimulus onset in the middle of the screen. In this experiment, the stimuli were presented on an LCD. On LCD’s the pixels’ luminance is updated periodically from top to bottom over a duration of about one display refresh. However, the onset measurements were synchronized to the vertical blank interval, which is the time between two display refreshes. Therefore, the stimulus onset measurement was synchronized to the pixels in the first row and not to the pixels in the middle of the display representing the target stimuli, resulting in a measurement error of about half a frame duration (i.e., 8.333 ms).
The remaining error appears to be caused by frame buffering. However, for the current study, we decided to not investigate the reasons for the higher mean error any further. The results clearly showed that both the non-VR and the VR background applications provided RTs with a comparable high precision. Hence, for the behavioral analysis, the results of the timing validation were used to correct the RTs for the measurement errors (see below).

To further ensure that the experimental setup was not affected by differences in the software being used or differences in the implementation, also the non-VR experiment was created in UE4. Despite the modifications of the stimulus size, the only difference between the original experiment (Wiesing, Fink, Weidner, et al., 2020) and the non-VR experiment of this study was the software in which it was created. Consequently, any failure to replicate the findings of the original findings in the new non-VR experiment would indicate that behavioral effects are rather the result of difference of the software basis or implementation rather than caused by any particularities of the VR system.

Virtual Environment

The non-VR experiment was conducted in an electronically and acoustically shielded chamber. In order to transfer the non-VR Experiment into VR in the most realistic way, and to establish two identical experimental setups with the only difference being that one experiment takes place in VR, the virtual environment was designed to visually reflect the real test chamber as accurately as possible (Figure 2). Furthermore, the virtual environment, including furniture, was built with the same scale and spatial arrangement as the physical environment. This allowed to spatially map and align the virtual environment onto its physical counterpart. Hence, both experiments were conducted in the same test chamber, but in the VR experiment, we created a blending between the real and the virtual chamber (i.e., every object in the virtual chamber had a real-world counterpart, identical in shape, size and position).

As a result, the entire virtual environment was supported by real haptic feedback. Since participants were unable to see their own body while in VR, a pair of HTC Vive motion controllers was given to the participants representing the locations of their hands. 3D models of the controllers were rendered in the virtual scenario, exactly mimicking the motion of the physical controllers. This allowed participants to *touch* the environment with the controllers, i.e., they were able to experiment the feel and see when a controller collided with a surface. In addition, the blending between physical and virtual world spared the need to simulate other sensory modalities than the visual modality, such as the auditory stimuli. Since every interaction of the controllers took simultaneously place in the physical and the virtual environment, the participants were always able to hear spatially correct sounds caused in the physical environment even while being in VR. Similarly, there was no need to simulate any olfactory stimuli, but instead every virtual object had its own original smell provided by its physical counterpart.

In order to familiarize participants with the VR setup and immerse them in the virtual scenario, participants were encouraged to explore the virtual environment prior to the experiment for a few minutes.

During both the VR and non-VR experiment, participants placed their hands and arms on a second table, beneath the table on which the monitor was placed. This arrangement ensured that the participants were not able to see their own hands and arms during the experiment, and hence prevented participants from getting distracted due to their own invisibility within VR.

Similarly, the communication between participant and experimenter was handled through an intercom device on the table, to avoid that the participants had to talk to an invisible person in the room.

For an accurate overlap it was necessary to align the position and orientation of tracking origin defined in the physical lab with the corresponding predefined point in the virtual lab. In a first step, the tracking origin was defined using SteamVR's standard setup procedure, to define the tracking origin. In a second step, a motion controller was used as a gauge to test the fit of the overlap and the virtual lab was moved from within VR along the x-, y-, and z-axis until the overlap was achieved.

Figure 2.Comparison of the real test-chamber (left) and the virtual test-chamber (right) in the illuminated version.

General Procedure

The standard and the VR experiments were conducted on separate days. The order of experiments was counterbalanced across participants.

To familiarize the participants with the task, a training session of 128 trials was performed before each experiment. All training trials used the frequent feature combination of the main experiment (*ColPE_low/OriPE_low*), so that participants could generate expectations about the most likely color and orientation combination of the target stimuli. Participants were informed that the color and the orientation could change between trials during the main experiment. Furthermore, they were informed that color or orientation changes were entirely irrelevant to their task.

Procedure VR Experiment

Before the experiment, participants could adjust the height of both tables and the chin rest to their needs. However, since moving the physical objects would result in a misalignment between the environments, both the virtual tables and chin rest were programmed to be moveable via buttons presses. After adjusting the height of the physical objects, the experimenter adjusted the height of the virtual counterparts via button press until a perfect overlap was achieved.

Furthermore, before starting with the experiment, participants were familiarized with the virtual environment in a short exploration phase. Participants were given a pair of motion controllers as a basic hand representation, and they were encouraged to explore and touch the environment. This was done to convince participants that the visual environment indeed reflected the physical surrounding.

During exploration, the virtual environment was normally illuminated. During the experiment, the light was turned off and the environment was illuminated by the virtual monitor only.

To guarantee safety during the VR experiment, participants remained seated on a chair during the entire VR session, including the exploration phase to prevent increased postural sway in HMD-based virtual environments which have previously been reported (Cobb et al., 1999; Fransson et al., 2019; Slobounov et al., 2015).

Data availability statement

The complete dataset of all experiments, a video of the virtual environment and the UE4 project files can be found https://osf.io/8bm9r/ Please note that due to copyright reasons a few textures used for this project had to be replaced.

Behavioral Analysis

The free statistical software R (R Foundation for Statistical Computing, Vienna, Austria; https://www.r-project.org) was used for the data analysis.

For each participant, mean RTs and error rates for each condition and experiment were calculated. Error trials, trials following errors and trials with RTs differing more than two standard deviations from the mean were excluded from RT analysis.

The results of the timing validation resulted in substantial differences of the mean measurement errors between the experimental setups. To account for these differences, the collected RTs were corrected for both experiments, i.e., the observed mean RT error of 46.96 ms was subtracted from all RTs collected in the non-VR experiment, hence compensating the delay induced by synchronization of the LCD monitor. Similarly, the mean RT error of 1.43 ms observed with the VR setup was subtracted from every RT collected in the VR experiment.

Repeated-measures ANOVAs for the RTs and error rates were first conducted for each experiment separately, with the within-subject factors *ColPE* (high, low) and *OriPE* (high, low). The reported mean values for expected and unexpected colors and orientations were calculated by collapsing all trials with the specific feature being expected or unexpected (e.g., the mean values for the expected color reflects the mean of all *ColPE_low/OriPE_low* and *ColPE_low/OriPE_high* trials). To compare behavioral effects between experiments, an ANOVA with the within-subject factors *ColPE* (high, low), *OriPE* (high, low) and Experiment (*VR*, *nonVR*) was conducted.

Results

<u>Non-VR Experiment</u>

Overall, the number of incorrect responses was very low with on average 2.24 % (± 0.40 SEM) errors. The ANOVA of the error rates yielded no significant main effect of *ColPE* ($F(1,15) = 3.746$, $p = 0.072$, $\eta_P^2 = 0.200$), with 2.04 % errors for expected colors and 3.63 % errors for unexpected colors. Similarly the main effect for *OriPE*, with 2.02 % errors for expected orientation versus 3.79 % errors for unexpected orientations, was not significant *OriPE* ($F(1,15) = 3.911$, $p = 0.066$, $\eta_P^2 = 0.207$). The interaction between *ColPE* and *OriPE* ($F(1,15) = 0.273$, $p = 0.609$, $\eta_P^2 = 0.018$) was also not significant.

The ANOVA of the mean RTs revealed a significant main effect for *ColPE* ($F(1,15) = 18.77$, $p < 0.05$, $\eta_P^2 = 0.556$) with 516 ms for expected colors versus 543 ms for unexpected colors. Moreover, we observed a significant main effect for *OriPE* ($F(1,15) = 17.91$, $p < 0.05$, $\eta_P^2 = 0.544$), with 515 ms for expected orientations versus 548 ms for unexpected orientations. The interaction of *ColPE* X *OriPE* was not significant ($F(1,15) = 0.231$, $p = 0.638$, $\eta_P^2 = 0.015$). The mean RTs and error rates are shown in Figure 3.

VR Experiment

Similar to the non-VR experiment, the mean error rate was very low with an average of 2.15 % (±0.36 SEM).

The ANOVA of the error rates yielded a significant main effect for *ColPE* ($F(1,15) = 22.2$, $p < 0.05$, $\eta_P^2 = 0.567$) with lower error rates for expected colors (1.92 %) compared to unexpected colors (3.74 %). The main effect for *OriPE* was not significant ($F(1,15) = 1.633$, $p = 0.221$, $\eta_P^2 = 0.098$), with error rates 2.06 % for expected orientations and 2.73 % for unexpected orientations. The interaction was not significant ($F(1,15) = 0.307$, $p = 0.588$, $\eta_P^2 = 0.020$).

Again, the ANOVA of the mean RTs revealed a significant main effect for *ColPE (*$F(1,15) = 14.4$, $p < 0.05$, $\eta_P^2 = 0.490$), with 505 ms for expected colors versus 533 ms for unexpected colors, and a significant main effect for *OriPE (*$F(1,15) = 22.57$, $p < 0.05$, $\eta_P^2 = 0.601$), with 506 ms for expected orientations versus 526 ms for unexpected orientations. The interaction of *ColPE* X *OriPE* was not significant ($F(1,15) = 0.061$, $p = 0.809$, $\eta_P^2 = 0.004$). The mean RTs and error rates are illustrated in Figure 3.

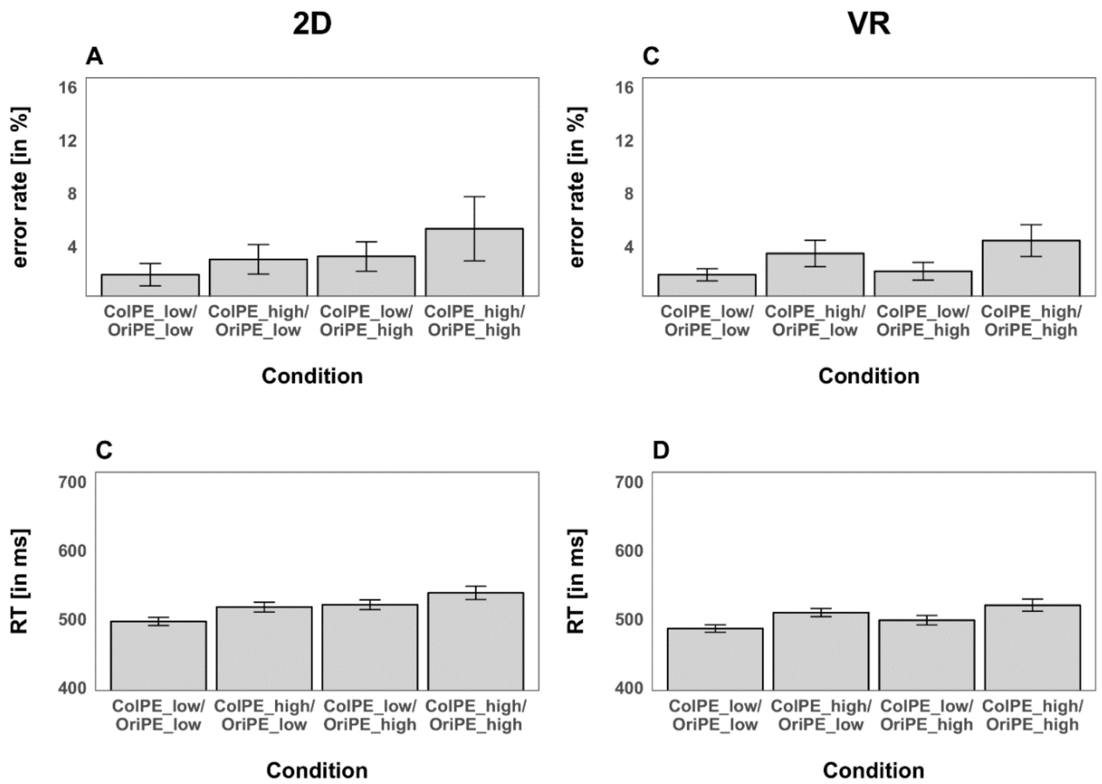

Figure 3. Performance measures of the combination of color and orientation manipulations for the non-VR and for the VR experiments. A: Error rates non-VR. B: Error rates VR. C: Reaction times non-VR. D: Reaction times VR. Error bars reflect the 95% confidence intervals.

In the next step, the results of both experiments were combined into one data set and compared within a single repeated measures ANOVA with the within-subject factors *ColPE* (high, low), *OriPE* (high, low) and *Experiment* (VR, non-VR). The analysis of the combined data set revealed a significant main effect for *ColPE* ($F(1,15) = 12.11$, $p < 0.05$, $\eta_P^2 = 0.447$) indicating that participants made more errors in trials with unexpected colors (mean: 3.26 %) than expected colors (mean: 1.98 %). The main effect for *OriPE,* with error rates of 2.04% for expected and 3.26 % for unexpected color, was not significant ($F(1,15) = 4.01$, $p = 0.063$, $\eta_P^2 = 0.211$). We found no evidence for a difference in task-difficulty in both experiments, as the main effect for *Experiment*, with an error rate of 2.24 % in the non-VR experiment and 2.15 % in the VR experiment, was not significant ($F(1,15) = 0.836$, $p = 0.375$, $\eta_P^2 = 0.053$). Consistent with the individual experiments, the interaction of *ColPE* X *OriPE* was not significant for the combined data set ($F(1,15) = 0.355$, $p = 0.56$, $\eta_P^2 = 0.023$). The interactions of *ColPE* X *Experiment* ($F(1,15) = 0.162$, $p = 0.693$, $\eta_P^2 = 0.011$) and *OriPE* X *Experiment* ($F(1,15) = 2.171$, $p = 0.161$, $\eta_P^2 = 0.126$) were not significant. Hence, we could not find evidence that error rates were different between experiments. Similarly, the three way interaction of *ColPE* X *OriPE* X *Experiment* was also not significant ($F(1,15) = 0.025$, $p = 0.877$, $\eta_P^2 = 0.002$).

The ANOVA comparing the RTs between experiments resulted in significant main effects for *ColPE*, with mean RTs of 511 ms for expected and 539 ms for unexpected colors ($F(1,15) = 23.77$, $p < 0.05$, $\eta_P^2 = 0.613$), and for *OriPE* ($F(1,15) = 43.1$, $p < 0.05$, $\eta_P^2 = 0.742$), with mean RTs of 537 ms for unexpected and 511 ms for expected orientations. The main effect for *Experiment*, with mean RTs of 531 ms in the non-VR and 518 ms in the VR experiment, was not significant ($F(1,15) = 1.32$, $p = 0.269$, $\eta_P^2 = 0.081$), indicating that the overall reaction times were not differently affected by the stimulus device. The interaction of *ColPE* X *OriPE* was not significant ($F(1,15) = 0.306$, $p = 0.588$, $\eta_P^2 = 0.020$). The interactions of *ColPE* X *Experiment* ($F(1,15) = 0.337$, $p = 0.57$, $\eta_P^2 = 0.022$) and *OriPE* X *Experiment* ($F(1,15) = 2.992$, $p = 0.104$, $\eta_P^2 = 0.166$) were also not significant, indicating that the RT effects were not different between experiments. Similarly, the three way interaction of *ColPE* X *OriPE* X *Experiment* was not significant ($F(1,15) = 0.047$, $p = 0.832$, $\eta_P^2 = 0.003$).

Discussion

The aim of the present study was to examine whether stimulus presentation via an HMD affects early processing of basic visual features when compared to a comparable stimulus presentation on a non-immersive display. In particular, we tested if the reaction time effects observed in a previous study (Wiesing et al. 2020), could be replicated in an VR setting. The results did not provide any evidence for significant effects of the VR setting.

In order to generate identical set-ups in VR and its non-VR equivalent, we rendered both the non-VR environment and the virtual computer screen within the environment in which the non-VR experimental stimulation was presented. Clearly, this posed a challenge for the display resolution of the HDM, and we had to slightly alter the stimuli used in the original study (Wiesing et al. 2020). Still, the stimuli used in the VR setting and the non-VR setting were matched in the current study. With the original stimulus configuration of Wiesing et al. (2020), the spatial frequencies of both Gabor patches were hard to differentiate in the VR setting. This problem was resolved after increasing the stimulus size. The most likely explanation for this limitation is the low pixel density of the HTC Vive. However, newer generations of HMDs, such as the Valve Index, have higher resolution displays and use different displays technologies, which help to increase the sharpness of the displays. Future research needs to investigate whether these problems persist when using HMDs with higher pixel densities.

Another technical issue that had to be addressed in the current study related to the accuracy of RT measurements in standard and VR setups, which differed substantially between both experimental setups. The most likely explanation for this difference is related to the frame-buffering in the non-VR and VR setting and affects the non-VR condition rather than the VR-setting. In particular, we used a method that was optimized and validated for the use with HDMs and collided with the buffering features of liquid crystal display (LCD) that was used in the non-VR setting. In game engines, the sampling rate to measure input such as responses is limited by the frame rate, which decreases the accuracy of RT measures substantially. Following the approach previously described in Wiesing et al. (2020), we circumvented this limitation in both experiments by collecting the response times via a software running in the background of UE4. Benchmarking tests confirmed a high precision and accuracy of reaction time measurement obtained in the VR setup, replicating the findings of our previous study. However, although we obtained comparably precise measurements, reaction times were highly inaccurate when obtained using the non-VR setup. Generally, a higher lag in the non-VR setup was not surprising, given the different refresh rates of both setups (non-VR = 60Hz vs VR = 90Hz) and superior temporal properties of OLED display (Cooper et al., 2013). Another factor that might have contributed to the increased lag are differences of the display refresh between the HTC Vive's OLED panels and LCD panels, as used in the non-VR setup. Typically, computer screens, such as the LCD used in this study, do not update the pixels all at once. Instead, the colors of the pixels change sequentially, line by line and from top to bottom. However, stimulus onset measurements are synchronized to the vertical blank event, i.e., the moment between two display refreshes. As a direct consequence, the measured timestamp does not perfectly correspond to the actual stimulus onset, but instead shows increasing measurement errors, the lower the position of the stimulus in the display. The panels used in the HTC Vive, on the other hand, have so-called *global-onset* displays, i.e., the entire frame appears at once instead of a sequential update. Consequently, with these displays, stimulus-onset measurements are unaffected by the location of the stimulus. In the current study, the stimuli were presented centrally within the display, causing a lag approximately half of a frame duration, i.e., ~8.33 ms, when using the LCD panel. The remaining observed lag is most likely explained by frame-buffering, which we did not properly account for during data collection. However, using beforehand obtained validation data allowed to correct the reaction time data prior to the analysis to account for the difference in measurement accuracy. This is essential since without careful consideration of different measurement errors, the observed differences could easily be mistaken for a genuine behavioral effect.

Overall, when display resolution and differences in frame buffering are properly considered, it is possible to successfully transfer standard behavioral paradigms into a VR-setting.

The RT pattern observed in the current study is well in line with and replicates the findings from previous experiments by Wiesing et al. (2020). Both studies demonstrated that prediction errors for different object features are resolved independently of each other. We further observed symmetrical RT costs and increased error rates associated with unexpected features in both the non-VR and VR experiment, providing no evidence that the stimulus presentation in current HMDs had a differential impact on the early processing of basic visual features when compared to stimulus presentation via a standard setup on a computer monitor. Similarly, the overall levels of accuracy and reaction times did not significantly differ between the two settings.

Our results are in line with the studies by Olk et al., (2018) and Roberts et al., (2019), which both were able to replicate well known behavioral effects in VR. On the other hand, our findings contrast with those studies, which observed behavioral differences between non-VR and VR. Anglin et al. (2017)

observed that participants followed different strategies in a visuomotor adaptation task in VR than in a standard setup. While the current study used a purely visual feature discrimination task, the study of Anglin et al. focused on spatial processing, which might be more easily disturbed by incorrect depth cues and spatial distortion caused by modern HMDs. This assumption could similarly explain the findings by Kozhevnikov & Dhond (2012), who observed that participants utilized different reference frames in mental rotation tasks when the task is conducted in immersive VR as compared the stimulus presentation on a computer monitor in both 2D and 3D.

Furthermore, previous research provided contradictory evidence regarding the allocation of attention within immersive VR. While findings from prism and lens induced VACs indicate that attentional resources might be reduced due to mechanisms to compensate VACs (Daniel & Kapoula, 2019), findings of a study comparing HMDs and normal computer screens found the exact opposite, that greater attentional resources are allocated to three-dimensional stimuli in VR as compared to stimuli presented on a monitor in 2D (Li et al., 2020). In the present study we expected differently allocated attention between both experiments to manifest in overall different RT cost between experiment, e.g., overall higher RT cost in the VR experiment, as well as different error rates, e.g., more errors in total in the non-VR experiment. However, we neither found any differences in overall RTs nor in error rates between experiments, providing no evidence that optical particularities of HMDs, such as VACs, have an impact on the allocation of attention.

A limitation of the current study in providing a maximum correspondence between experiments, was the lack of a full body tracked avatar. Instead, participants only had two motion controllers as basic hand representations, which might have resulted in a decreased feeling of embodiment (Pyasik et al., 2020; Seinfeld & Müller, 2020).

Overall, the results indicate that early processing of basic visual features does not differ when stimuli are presented on a computer monitor or within a modern HMD. Instead, within the context of recent literature comparing the different experimental setups, it appears that differences in cognitive processing might be related to spatial tasks, which are more easily affected by incorrect depth cues and spatial distortions caused by HMDs.

It is important to note that the present results are task-specific and therefore do not allow any conclusions about similar effects when examining the impact of VR on the performance in different paradigms. Despite the steadily growing body of literature, the impact of VR technology on cognition is still far from understood. Understanding the potential mechanisms underlying cognitive processes in physical and virtual environments further will be critical to validly transfer findings from VR experiments to real world scenarios.

Literature

Anglin, J. M., Sugiyama, T., & Liew, S.-L. (2017). Visuomotor adaptation in head-mounted virtual reality versus conventional training. *Scientific Reports*, *7*(1), 45469. https://doi.org/10.1038/srep45469

Arthur, K. W. (2000). *Effects of Field of View on Performance with Head-Mounted Displays*. *Unpublished doctoral dissertation, University of North Carolina, Chapel Hill*.

Batmaz, A. U., Machuca, M. D. B., Pham, D. M., & Stuerzlinger, W. (2019). Do Head-Mounted Display Stereo Deficiencies Affect 3D Pointing Tasks in AR and VR? *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 585–592. https://doi.org/10.1109/VR.2019.8797975

Bingham, G. P., Bradley, A., Bailey, M., & Vinner, R. (2001). Accommodation, occlusion, and disparity matching are used to guide reaching: A comparison of actual versus virtual environments. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(6), 1314–1334. https://doi.org/10.1037/0096-1523.27.6.1314

Blender Foundation. (n.d.). *Blender*. https://www.blender.org/

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, *12*(12), 752–762. https://doi.org/10.1038/nrn3122

Buckingham, G. (2019). Examining the size–weight illusion with visuo-haptic conflict in immersive virtual reality. *Quarterly Journal of Experimental Psychology*, *72*(9), 2168–2175. https://doi.org/10.1177/1747021819835808

Caramenti, M., Pretto, P., Lafortuna, C. L., Bresciani, J.-P., & Dubois, A. (2019). Influence of the Size of the Field of View on Visual Perception While Running in a Treadmill-Mediated Virtual Environment. *Frontiers in Psychology*, *10*, 2344. https://doi.org/10.3389/fpsyg.2019.02344

Cobb, S. V. G., Nichols, S., Ramsey, A., & Wilson, J. R. (1999). Virtual Reality-Induced Symptoms and Effects (VRISE). *Presence: Teleoperators and Virtual Environments*, *8*(2), 169–186. https://doi.org/10.1162/105474699566152

Cooper, E. A., Jiang, H., Vildavski, V., Farrell, J. E., & Norcia, A. M. (2013). Assessment of OLED displays for vision research. *Journal of Vision*, *13*(12), 16–16. https://doi.org/10.1167/13.12.16

Daniel, F., & Kapoula, Z. (2019). Induced vergence-accommodation conflict reduces cognitive performance in the Stroop test. *Scientific Reports*, *9*(1), 1247. https://doi.org/10.1038/s41598-018-37778-y

EpicGames. (n.d.). *Unreal Engine*. Unreal Engine. https://www.unrealengine.com/en-US/

Foerster, R. M., Poth, C. H., Behler, C., Botsch, M., & Schneider, W. X. (2016). Using the virtual reality device Oculus Rift for neuropsychological assessment of visual processing capabilities. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep37016

Foerster, R. M., Poth, C. H., Behler, C., Botsch, M., & Schneider, W. X. (2019). Neuropsychological assessment of visual selective attention and processing capacity with head-mounted displays. *Neuropsychology*, *33*(3), 309–318. https://doi.org/10.1037/neu0000517

Fransson, P.-A., Patel, M., Jensen, H., Lundberg, M., Tjernström, F., Magnusson, M., & Ekvall Hansson, E. (2019). Postural instability in an immersive Virtual Reality adapts with repetition and includes directional and gender specific effects. *Scientific Reports*, *9*(1), 3168. https://doi.org/10.1038/s41598-019-39104-6

Guo, J., Weng, D., Been-Lirn Duh, H., Liu, Y., & Wang, Y. (2017). Effects of using HMDs on visual fatigue in virtual environments. *2017 IEEE Virtual Reality (VR)*, 249–250. https://doi.org/10.1109/VR.2017.7892270

Heber, I. A., Siebertz, S., Wolter, M., Kuhlen, T., & Fimm, B. (2008). Attentional Asymmetries in Virtual Space. *Nd ANNUAL MEETING*, 4.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3), 33. https://doi.org/10.1167/8.3.33

Hopper, J. E., Finney, H., & Jones, J. A. (2019). Field of View and Forward Motion Discrimination in Virtual Reality. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1663–1666. https://doi.org/10.1109/VR.2019.8797756

Imaoka, Y., Flury, A., & de Bruin, E. D. (2020). Assessing Saccadic Eye Movements With Head-Mounted Display Virtual Reality Technology. *Frontiers in Psychiatry*, *11*, 572938. https://doi.org/10.3389/fpsyt.2020.572938

Iskander, J., Hossny, M., & Nahavandi, S. (2019). Using biomechanics to investigate the effect of VR on eye vergence system. *Applied Ergonomics*, *81*, 102883. https://doi.org/10.1016/j.apergo.2019.102883

Kelly, J. W., Cherep, L. A., Klesel, B., Siegel, Z. D., & George, S. (2018). Comparison of Two Methods for Improving Distance Perception in Virtual Reality. *ACM Transactions on Applied Perception*, *15*(2), 1–11. https://doi.org/10.1145/3165285

Kozhevnikov, M., & Dhond, R. P. (2012). Understanding Immersivity: Image Generation and Transformation Processes in 3D Immersive Environments. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00284

Li, G., Anguera, J. A., Javed, S. V., Khan, M. ammad A., Wang, G., & Gazzaley, A. (2020). Enhanced Attention Using Head-mounted Virtual Reality. *Journal of Cognitive Neuroscience*, 1–18. https://doi.org/10.1162/jocn_a_01560

Lin, J., Zhu, Y., Kubricht, J., Zhu, S.-C., & Lu, H. (2017). *Visuomotor Adaptation and Sensory Recalibration in Reversed Hand Movement Task*. 7.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 557–564. https://doi.org/10.3758/BF03200735

Maringelli, F., McCarthy, J., Steed, A., Slater, M., & Umiltà, C. (2001). Shifting visuo-spatial attention in a virtual three-dimensional space. *Cognitive Brain Research*, *10*(3), 317–322. https://doi.org/10.1016/S0926-6410(00)00039-2

MAXON. (n.d.). *Cinema4D*. https://www.maxon.net/de/produkte/cinema-4d/cinema-4d/

Nash, E. B., Edwards, G. W., Thompson, J. A., & Barfield, W. (2000). A Review of Presence and Performance in Virtual Environments. *International Journal of Human-Computer Interaction*, *12*(1), 1–41. https://doi.org/10.1207/S15327590IJHC1201_1

Niehorster, D. C., Li, L., & Lappe, M. (2017). The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research. *I-Perception*, *8*(3), 204166951770820. https://doi.org/10.1177/2041669517708205

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

Pfeil, K., Taranta, E. M., Kulshreshth, A., Wisniewski, P., & LaViola, J. J. (2018). A comparison of eye-head coordination between virtual and physical realities. *Proceedings of the 15th ACM Symposium on Applied Perception - SAP ’18*, 1–7. https://doi.org/10.1145/3225153.3225157

Plant, R. R., & Turner, G. (2009). Millisecond precision psychological research in a world of commodity computers: New hardware, new problems? *Behavior Research Methods*, *41*(3), 598–614. https://doi.org/10.3758/BRM.41.3.598

Poltavski, D. V., Biberdorf, D., & Petros, T. V. (2012). Accommodative response and cortical activity during sustained attention. *Vision Research*, *63*, 1–8. https://doi.org/10.1016/j.visres.2012.04.017

Pyasik, M., Tieri, G., & Pia, L. (2020). Visual appearance of the virtual hand affects embodiment in the virtual hand illusion. *Scientific Reports*, *10*(1), 5412. https://doi.org/10.1038/s41598-020-62394-0

Roberts, A. C., Yeap, Y. W., Seah, H. S., Chan, E., Soh, C.-K., & Christopoulos, G. I. (2019). Assessing the suitability of virtual reality for psychological testing. *Psychological Assessment*, *31*(3), 318–328. https://doi.org/10.1037/pas0000663

Seinfeld, S., & Müller, J. (2020). Impact of visuomotor feedback on the embodiment of virtual hands detached from the body. *Scientific Reports*, *10*(1), 22427. https://doi.org/10.1038/s41598-020-79255-5

Slobounov, S. M., Ray, W., Johnson, B., Slobounov, E., & Newell, K. M. (2015). Modulation of cortical activity in 2D versus 3D virtual reality environments: An EEG study. *International Journal of Psychophysiology*, *95*(3), 254–260. https://doi.org/10.1016/j.ijpsycho.2014.11.003

Vasser, M., & Aru, J. (2020). Guidelines for immersive virtual reality in psychological research. *Current Opinion in Psychology*, *36*, 71–76. https://doi.org/10.1016/j.copsyc.2020.04.010

Velhagen, K., & Broschmann, D. (Hrsg.). (2003). *Tafeln zur Prüfung des Farbensinnes*. Thieme.

Verdelet, G., Salemme, R., Desoche, C., Volland, F., Farne, A., Coudert, A., Hermann, R., Truy, E., Gaveau, V., & Pavani, F. (2019). Assessing Spatial and Temporal Reliability of the Vive System as a Tool for Naturalistic Behavioural Research. *2019 International Conference on 3D Immersion (IC3D)*, 1–8. https://doi.org/10.1109/IC3D48390.2019.8975994

Wiesing, M., Fink, G. R., & Weidner, R. (2020). Accuracy and precision of stimulus timing and reaction times with Unreal Engine and SteamVR. *PLOS ONE*, *15*(4), e0231152. https://doi.org/10.1371/journal.pone.0231152

Wiesing, M., Fink, G. R., Weidner, R., & Vossel, S. (2020). Combined expectancies: The role of expectations for the coding of salient bottom-up signals. *Experimental Brain Research*, *238*(2), 381–393. https://doi.org/10.1007/s00221-019-05710-z

Wilson, C. J., & Soranzo, A. (2015). The Use of Virtual Reality in Psychology: A Case Study in Visual Perception. *Computational and Mathematical Methods in Medicine*, *2015*, 1–7. https://doi.org/10.1155/2015/151702

Witmer, B. G., & Kline, P. B. (1998). Judging Perceived and Traversed Distance in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *7*(2), 144–167. https://doi.org/10.1162/105474698565640

Chapter 4 – General discussion

In a typical experiment in cognitive neuroscience and experimental psychology, the objective is to record brain activity and behavior while an individual engages in a cognitive activity. To be able to precisely link brain activity to specific cognitive capabilities, neuroscientific studies have usually been conducted under controlled laboratory conditions. The need for experimental control is reflected in experimental paradigms providing only minimalistic sensory stimulation and allowing only for restricted and repetitive behaviors.
The approach of minimalistic but highly controlled experiments helped scientists to take tremendous strides in uncovering the neuronal basis of cognitive processing. However, it has been criticized that those minimalistic experimental paradigms fail to replicate the complexity of reality and lack ecological validity, casting doubts on the generalizability of findings from the laboratory to real-world situations (Schmuckler, 2001; Ulric Neisser, 1976). Since then it has been considered as a trade-off between ecological validity and experimental control and researchers found themselves in the dilemma of choosing the one or the other (Parsons, 2015).
Virtual reality has the promise to enable researchers achieving both ecological valid and precisely controlled experiments. The potential of virtual reality as a novel tool to study human behavior and underlying neural functioning has been recognized for decades (Bohil et al., 2011; Loomis et al., 1999a). However, due to the high cost and technical requirements to generate virtual environments, VR has been highly underutilized in the past. With the launch of consumer head-mounted displays, high quality but low-cost VR-systems, the technology became affordable for most research labs. Additionally, with the advances in computing power and widely available rendering engines, the interest in VR and the utilization of VR as a research tool has gained some momentum within the last years (Vasser & Aru, 2020).

Virtual Reality

The term virtual reality was originally coined in the 1970’s by Myron Krueger to describe computer applications that had to be considered as “responsive environments” (Woolley, 1993). However, it was Jaron Lanier, who popularized the term as “three-dimensional realities implemented with stereo viewing goggles and reality gloves” (Krueger, 1991). Lanier was the chief executive officer of VPL Research, Inc, founded in 1984 and one of the first companies developing and selling commercial VR systems.
In the introduction I defined VR as computer-generated worlds experienced via HMDs. This simplified view made sense in the context on the studies reported in chapter 2 and chapter 3. In chapter 2, I examined the accuracy and precision of stimulus timing and time measurements when stimuli were rendered for and presented within an HMD. In chapter 3, I specifically asked if HMDs might affect the early processing of basic visual features differently than a computer monitor.
However, if we want to exploit the full scientific potential of VR, it is better to understand VR in terms of the interplay between participant and content. For example, according to Slater (2018) the fundamental element of any VR system includes “a computer-generated world […] that perceptually surrounds the participant, and where perception is a function at least of head tracking” (Slater, 2018,

p. 431). Others define VR as a ‘‘computer-generated digital environment that can be experienced and interacted with as if that environment was real’’ (Jerald, 2015, p. 9) or as a ‘‘set of technologies that enable people to immersively experience a world beyond reality’’ (Berg & Vance, 2017, p. 1).
According to these views, VR is the idea of a computer-generated world which surrounds us, forming a parallel reality, with which we can interact naturally, e.g., looking around by turning our head, just like we would do in physical reality. Virtual reality is the idea to completely block any sensory input coming from physical reality and replacing it with an artificial world, which is experienced and can be interacted with, as if it was real.

Virtual Reality as a research tool

This perspective on VR also explains better why VR holds such an immense promise as a research tool in cognitive neuroscience and experimental psychology. VR enables researchers to present a wide range of stimulus conditions, which would be difficult or even impossible to create in physical reality. For example, Marek & Pollmann (2020) used VR to turn a classical two-dimensional contextual cueing task into a three-dimensional VR task, in which participants were surrounded by the stimuli. By using this novel stimulus setup, the authors showed that the search time reductions known from the classical version of the task, can also be observed when the stimuli were presented outside of the initial field of view. Mast & Oman (2004) were able to replicate a perceptual illusion in VR, which was previously only observed by astronauts in microgravity. Most astronaut experience so-called visual reorientation-illusions within microgravity, i.e., occasional changes of the perceived identity of environmental surfaces. A surface, which was perceived as a wall might become the ceiling a moment later. Mast & Oman recreated the visual ambiguity typically found in spacecrafts within VR, which often do not provide clear information about what is, for example, the floor or the ceiling. The authors showed that participants can cognitively manipulate the reorientation illusion effects within normal gravity on the ground. Others have used VR to create experiments, which would be ethically unacceptable when done in physical reality, because it involves placing participants in highly dangerous situations (e.g., Ramdhani et al. (2019); Patil et al. (2018)).
Importantly, since the scenarios are computer-generated, the researcher has control over essentially even the smallest detail of the virtual world, allowing them to optimize the entire scene just for the experiment in question. Furthermore, researchers can take advantage of the tracking capabilities of modern VR systems, allowing them to monitor complex behaviors of freely moving participants with great detail, a scenario which would be hard or even impossible to control in physical reality (Niehorster et al., 2017).

However, little is known if the virtual setting and the VR technology might engage different cognitive processes than physical reality, i.e., if participants respond to virtual stimuli as they would do to real-world stimuli (Kulik, 2018). Yet, the same can be asked about standard non-VR experiments, which also create artificial settings and participant behavior might not match with their real-world behavior (Pan & Hamilton, 2018b). This, however, raises questions about the transferability from VR to non-VR experiments and vice versa. So far, only a few studies have systematically examined if HMD based VR systems affects basic cognitive processes differently than a standard non-VR experiment, i.e., visual stimuli are presented on a monitor.

Findings from previous research indicate that participants rely on different frames of reference in mental rotation tasks (Kozhevnikov & Dhond, 2012) and follow different strategies in visuomotor adaptation tasks (Anglin et al., 2017) in VR as compared to a standard setup. Findings of another study suggest HMD might provide advantages for the allocation of attentional resources. Li et al. (2020) compared the task performance and functional EEG measurements during a visual discrimination task within an HMD and in front of a computer monitor. Their results indicate that participant allocated greater attentional resources to the stimulus material presented in the HMD condition as compared to the non-VR condition. On the other hand, Roberts et al. (2019) observed a similar task performance between the VR and non-VR version of a visual flanker task. Other findings indicate that HMDs increase cognitive load and reduce motor performance compared to a standard monitor setting (Juliano et al., 2021).

Yet so far, no research has examined on which level of processing the observed differences between HMD and monitor experiments arise. Here, my aim was to examine whether the processing of expectancies for basic visual features is different when the stimuli are presented in an HMD as compared to a standard computer monitor.

In a first step, I developed and tested a new behavioral paradigm. The paradigm investigated how prediction errors for two simultaneous unexpected features of the same object are formed and on which level of processing they arise. The results strongly suggest that unexpected but otherwise task-irrelevant colors or orientations result in increased reactions times. Furthermore, the results indicate that both prediction errors are resolved independently on an early level of processing when different feature dimensions are processed in parallel.

In a series of four experiments, I consistently observed the same pattern of results, showing main effects for the individual prediction errors but no interactions between them. This was important with respect for the replications planned for the study reported in chapter 3, as it clearly showed that the behavioral effects are robust and generally replicable when using the same combination of hardware and software.

Another critical factor is the precision and accuracy with which the relevant data can be obtained. In chapter 1, the dependent measure I was interested in were reaction times. Measuring reaction times consists of two simple time measurements: 1) the stimulus-onset, e.g., target appears on the screen, and 2) the response time, e.g., button press. The reaction time can be obtained by calculating the difference between the response time minus the time of the stimulus-onset. However, the precision and accuracy with which both timepoints can be measured, depends on both the software and hardware used for the stimulus presentation as well as for collecting input.

For example, many liquid crystal displays (LCDs) suffer from unreliable refresh rates and slow response times, which can affect both the precision and accuracy of the stimulus timing as well as reaction time measurements. In chapter 1 and in the non-VR experiment of chapter 3, stimuli were presented on a Samsung SyncMaster 2233. Previous research has shown that the temporal properties of the monitor are on par with tested cathode-ray tube (CRT) monitors (Wang & Nikolic, 2011). CRTs are usually considered the gold standard for visual stimulus presentation because of their precise and reliable timing.

To my best knowledge no study had yet examined if HMDs provide precise and reliable stimulus timing. Fortunately, the temporal properties of the HTC Vive turned out to be highly suitable for experiments in vision research. In fact, the so-called *global-onset* display used in the HTC Vive provide a clear advantage over typical LCDs and CRTs. On typical LCD and CRT screens, the image is built up line by line from top to bottom. As a result, the upper part of the image appears earlier than the lower part, with the most upper and lowest line of pixels typically being a bit less than a frame duration apart. However, the stimulus-onset is typically measured as the begin of the display refresh. Hence, the measurement error of the stimulus onset increases, the lower the location of stimulus on the display. In contrast, in the displays of the HTC Vive all pixels light up simultaneously, allowing to measure the onset of a stimulus independently of its position.

Furthermore, in chapter 1, I used PsychoPy, a standardized and well established toolbox for behavioral experiments (Peirce et al., 2011). PsychoPy has been proven to measure reaction times with a high accuracy and precision (Bridges et al., 2020). My own measurements, reported in chapter 2, confirm these findings.

However, PsychoPy does only provide limited support for modern HMDs and lacks the rendering capabilities required for realistic and interactive virtual worlds. Instead, many studies rely on game engines, such as Unity or Unreal Engine. However, game engines are not designed or optimized for behavioral experiments. In fact, game engines lack many basic features, required by most experiments. For example, PsychoPy provides several functions to create the most common visual stimuli used in experiments, such as Gabor patches or random-dot stereograms. On the other hand, in a game engine such as UE4, one will look in vain for a function to create a basic Gabor patch. In fact, the Gabor patches presented in both experiments of chapter 3 were created in PsychoPy and stored as textures, which were applied to the display of the virtual monitor. Similarly, game engines do not provide tools which help scientists to setup the trial structure and define stimulus events, and game engines suffer from constraints regarding data collection and data quality.

Fortunately, in the recent years, an increasing number of scientific toolboxes for game engines have been released and are in development. For example, Unity Experiment Framework (UXF) intends to provide a framework to setup and control experiments and to simplify the data collection for behavioral experiments using Unity (Brookes et al., 2020). Toggle Toolkit allows to setup triggers (e.g., collisions or button presses) and toggles (changing the state of an objects, e.g., turning light on or off) as well as to log the associated data for later analysis (Ugwitz et al., 2021). VREX is another example, which provide tools to setup experiments and come with various study protocols for attentional or memory tasks (Vasser et al., 2017). Other toolboxes provide solutions to combine the VE with simultaneous physiological and kinematic measurements (Grübel et al., 2017; Williams et al., 2019; Wolfel et al., 2021).

However, although a lot of progress has been made in various areas from setting up and controlling VR experiments, simplifying data collection, and synchronizing different data streams, to my knowledge no research had examined the precision and accuracy of stimulus timing and time measurements, when using a game engine in combination with a modern HMD-VR system.

Especially the limitations in obtaining the correct time of a stimulus-event like the stimulus-onset have not received much attention yet. However, to be able to relate functional or behavioral data to ongoing cognitive processes, it is crucial to determine when certain sensory events occurred.

While it is technically relatively easy to measure these events using game engines, doing so precisely and accurately is not that trivial. In fact, even software toolboxes which are developed explicitly for behavioral experiments have been observed to struggle with precise and accurate time measurements and stimulus timing (Bridges et al., 2020; Garaizar et al., 2014).

In chapter 2, I have demonstrated that both stimulus onset and response time measurements are imprecise and inaccurate when obtained using the standard API of UE4. The reasons can be explained by the underlying architecture of game engines and the graphics pipeline for VR rendering. The central component of any application created in a game engine is the game loop. The entire application runs in one loop, which handles any processes, from input over physics to drawing objects. The game loop iterates once every frame, limiting the sampling rate for any measurement within the game loop to the current frame rate. Consequently, the frame rate of an experiment directly determines the best possible precision with which data, such as timestamps, can be obtained. Furthermore, to prevent visual artifacts, such as tearing, each major VR runtime software synchronizes the framerate to the refresh rate of the display. According to my own measurements, the refresh rate of the HTC Vive is 89.53 Hz, i.e., about every 11.17ms a new frame. Hence, every time measurement will vary in a range of about ± 5.59 ms, which was also confirmed by my data.
A possible solution is to render the scene at a higher rate than the display is able to refresh. Quinlivan et al. (2016) aimed to render the visual scene at 1000 frames per second (FPS), allowing them to collect input and tracking data with a millisecond precision. However, the downside of this approach is the performance cost associated with the increased frame rate, making it unsuitable for realistic and complex rendered scenes. Although the environment and stimuli of Quinlivan et al. (2016) were minimalistic and probably causing only a low overhead, the frame rate eventually fluctuated in a range of 600 and 1000 FPS. The authors accounted for this by resampling the data at 500 Hz.
Others have suggested to work around the above discussed limitations by using a microcontroller as an external synchronization device and to separate the measurements from the rendering engine (Alsbury-Nealy et al., 2021; Watson et al., 2019; Wienrich et al., 2018). These studies were able to improve the precision of response time measurements, by collecting the button input externally via an Arduino. Watson et al. (2019) were also able to improve the accuracy of stimulus-onset measurements by using a photodiode to detect a small peripheral stimulus flashing simultaneously with the actual stimulus, when used on a normal computer screen. However, even a small photosensor and flash stimulus will probably be noticeable and distracting for participants.
Instead, I proposed an approach that solves both the stimulus-onset as well as the response time measurements on the software-level, without requiring additional hardware. Instead, I outsourced the measurements into another software, running in the background of UE4 and thereby circumventing the above-described limitations. Benchmarking data reported in chapter 2 confirmed the high precision and accuracy of my method.

Before conducting the experiments of chapter 3, I conducted some benchmark tests of the reaction times measurements for both the non-VR and the VR experiment. While the results of the VR experiment were basically identical to results reported in chapter 2, the accuracy of reaction times measured in the non-VR experiment was clearly off. The differences can partially be explained by the different temporal properties of the OLED displays of the HTC Vive and the LCD monitor. As mentioned above, the HTC Vive displays light up and present the new frame at an instance, while the LCD updates continuously from top to bottom. The test stimulus for the validation was placed at the center of the screen, while the stimulus onset was measured at the vertical sync, i.e., at the start of a new display refresh. However, this explains only a small amount of the observed lag of maybe 7-8 ms. However, most of the lag can be explained in that I did not properly account for additional frame buffering used for non-VR rendering. Fortunately, the precision was similarly high for both experiments, which allowed me to correct the reaction time data before running the analysis.

The timing results of chapter 3 clearly demonstrate that comparison between non-VR and VR experiments need to consider and correct for timing differences between HMDs and normal monitors as well as the associated rendering processes. Otherwise, it becomes arbitrary whether a behavioral difference results from the technology or is the mere product of different data quality between experiments.
As an additional sanity check I created both the VR and the non-VR experiment with UE4. The original study, on the other hand, was created in PsychoPy. Hence, apart from the modification in stimulus size, both the original and the new non-VR version differed only in the underlying software. Hence, any failure to reproduce the original effects in the non-VR UE4 version would suggest some software related issues.

The virtual environment, in which the VR experiment of chapter 3 took place, was a realistic and accurate replica of the real chamber in which both experiments took place. This was done to establish two identical experimental setups, which ideally only differ in the display device. The virtual replica corresponded not only visually with its real-world counterpart but also with respect to the scale, allowing me to use the visual rendering as an overlay which I blended onto the physical EEG chamber. As a result, the visual rendering was enhanced by the non-visual physical properties of the environment of the real physical environment itself. The sensory experience of, for example, touching the table with a controller provided not only the accurate visual feedback. Participants also felt the table blocking their movement and were able to hear the noises generated by the collision of controller and table.

What I did here, was to increase the immersion of the virtual EEG chamber. With immersion I refer to the so-called *system immersion*, which is an objective and theoretically quantifiable property of any VR system (Nilsson et al., 2016; Slater, 1999), describing the fidelity of the system to create vivid and interactive virtual environments, while shutting out physical reality (Cummings & Bailenson, 2016). For example, a high-resolution HMD is more immersive than a low-resolution but otherwise identical HMD. Similarly, a VR system that can simulate multiple sensory modalities has a higher immersion than a purely visual VR system. According to Mel Slater, immersion can be understood as the sum of *sensorimotor contingencies* (SCs) supported by a VR-system (Slater, 2009). SCs refer to all actions that can be carried out, in order to perceive the world, for example, by moving the head or body (O'Regan & Noë, 2001).
Previous studies indicate that more immersive virtual environments can improve cognitive functions, such as memory (Krokos et al., 2019; Sutcliffe et al., 2005) and elicit more intense emotional responses (Diemer et al., 2015; Visch et al., 2010). A recent study demonstrated how the lack of tactile and haptic feedback affects the task performance in an obstacle avoidance task (Giesel et al., 2020). The authors explained the differences in behavior by the disparity in expected consequences of actions between both conditions.
Furthermore, high levels of immersion are associated with a stronger sense of *presence*. The term is rooted in the concept of telepresence coined by Marvin Minsky, to describe the feeling that a person might have while controlling a robot remotely (Minsky, 1980). In the context of VR, however, it is commonly just referred to as *presence* or the *sense of presence*, which typically is used to describe the subjective feeling of *being there*, in the mediated or computer-generated environment rather than in physical reality (Weber et al., 2021). Presence or telepresence is regarded as one of the most crucial aspects of a VR experience, maybe even its defining feature (Slater & Wilbur, 1997; Steuer, 1992).

Lombard & Ditton (1997) described presence as the "perceptual illusion of nonmediation [which] involves continuous ('realtime') responses of the human sensory, cognitive and affective processing systems". According to the authors, the *illusion of non-mediation* occurs when the participant does only perceive the virtual content but not the delivering medium and responds as if it would not exist. They point out that, although all experiences are mediated through the sensory system, non-mediation explicitly refers to experiencing the virtual environment without experiencing the mediating technology.
Another influential account on presence distinguishes between two perceptual illusions, the Place Illusion (PI) and the Plausibility Illusion (Psi). While PI describes the feeling of *being there*, Psi describes the feeling of "is apparently happening is really happening (even though you know for sure that it is not)*"* (Slater, 2009)*.* A key component for Psi are events in the environment which directly refer to the participant, without them having control over it, such as a computer character smiling at the participant as soon they have eye contact (Slater, 2009). Slater describes both dimensions as illusions to point out that the participants have the sensation of *being there* and that things a *really happening* despite the knowledge that this is not the case. Furthermore, both PI and Psi are orthogonal factors. According to Slater, if both PI and Psi are experienced, participants will respond realistically, i.e., as in a comparable real situation.

A factor, for which I did not account for, was the reduced field of view in the VR experiment. While in the non-VR experiment, the entire visual field was provided with visual input, in the VR condition, the horizontal field of view was restricted to approximately 110° (Al Zayer et al., 2019). In the study by Li et al. (2020), participant were wearing the empty frame of an HMD to reduce the FOV in the non-VR condition. Before conducting the experiments reported in chapter 3, I tried the same approach and wore the frame of an old and disassembled HMD watching at my stimuli on a computer screen. However, the light of the display caused a few but clearly visible reflections on the inside of the frame and I was concerned that this might cause distractions interfering with the task. Furthermore, an empty frame can only be seen as an approximation of the HMD's FOV, which depends on different factors, such as the distance between lenses and the eyes.

Another factor for which I did not account for, was the lack of a visible body in the VR experiment. Since participants cannot see their own body while in HMD-VR, it is common to include a self-avatar, i.e., a virtual representation of a body that is experienced from a first-person perspective and provides a substitute to their real body. This can give rise to so-called embodiment-illusions or virtual embodiment, in which the participants experiences the virtual body as their own (Gonzalez-Franco & Peck, 2018; Slater, 2009; Spanlang et al., 2014; Yuan & Steed, 2010).
Self-avatars have several benefits as they for example provide visual cues about the participant's location and immediate feedback about one owns actions. Another advantage is non-verbal communication in shared virtual environments (Y. Pan & Steed, 2019). Self-avatars have been found to improve distance judgements in VR (Mohler et al., 2010; Ries et al., 2008), which is more pronounced the more embodied the participants were (Gonzalez-Franco et al., 2019). Others have shown that self-avatars can reduce cognitive load and improve memory and cognitive processing within VR (Steed et al., 2016).
Here, I decided to not include a self-avatar but instead to only show floating controller models while participants familiarized with the VE. This was mainly done due to the technical requirements for the appropriate implementation of a full-body self-avatar, which are substantially higher than floating

controllers. Current off-the-shelf VR systems, such as the HTC Vive used in chapter 3, only provide motion-tracking for the head and both hands (3-point tracking), while feet and the rest of the body are not tracked. This makes it impossible to accurately to track and replicate the participant's movement onto the avatar. As a possible consequence, participants would experience a conflict between vision and proprioception, which could make it difficult to interact naturally with the VE or result in a reduction of presence (Slater & Steed, 2000). Incongruencies between the participant's and their self-avatars movement and size have also been found to increase simulator-sickness (Kim et al., 2020).
To ensure that the missing body did not serve as a distraction during the experiment, the experimental setup was arranged that the body of the participants was also in the non-VR experiment out of view. Participants placed their arms on a second table, hidden from view by the table on which the monitor was placed.
Furthermore, during the familiarization phase, prior to the VR experiment, participants were handed a pair of tracked motion controllers, as a simple hand representation. Presenting controller models instead of hands or a full-body self-avatar is comparably easy, given that the controllers are rigid bodies which already provide all necessary sensors for the tracking.

During the piloting phase, I noticed that comparably low pixel density of the HTC Vive made it almost impossible to differentiate between the high and low spatial frequency of the Gabor patches, when using the original stimulus parameters, as reported in chapter 1. Although the issue was easily resolved by increasing the stimulus size, it clearly demonstrates limitations for present small details. Fortunately, the resolution of newer HMDs has seen a massive increase. For example, the HTC Vive used in my studies provides 1080x1200 pixels per eye. The latest Vive HMD released by HTC, the HTC Vive Pro 2, already provides a resolution of 2448x2488 pixels per eye, making it unlikely to observe the same difficulties to clearly present small details as I have observed.

Testing the same group of participants in each experiment in counterbalanced order, both the non-VR and the VR experiment replicated the behavioral effects of the study reported in chapter 1. Again, the results clearly indicated that both the prediction errors for color and orientation can be manipulated independently of each other, regardless of whether the features belonged to the same objects or not. Consequently, the no evidence for differences between the two new experiments were found. This clearly indicates that the processing of early visual feature expectations as well as their violations does not differ between both experimental setups.
The results are in line with Roberts et al. (2019), who also replicated the results previously observed in a standard non-VR setup. On the other hand, the results by Li et al. (2020) suggest that selective attentional abilities might be enhanced within HMDs. Here, I expected to find different overall reaction times or errors rates between the experiments, if one of the experimental setups provides attentional advantages over the other. However, I neither observed differences in overall reaction times or error rates between the experiments, providing no evidence that the HMD caused differences in the distribution of attention. Importantly, the experiments in chapter 3 were not designed to compare the distribution of attention in VR and non-VR setups. Therefore, the paradigm might have failed to provide the sensitivity to detect differences in the allocation of attention, explaining the discrepant results of my own experiments and the study by Li et al. (2020). Another explanation could be the differences in the stimulus material. Li et al. (2020) used a three-dimensional scene, in which target and distractor stimuli were presented at different locations, including differences in depth. However, in contrast to the HMD, the monitor used in the non-VR condition was not capable to present the

stimuli in stereoscopic depth. Hence, the differences in depth information between the experiments might explain the observed attentional advantages. Furthermore, in the non-VR condition, participants saw only a fraction on the virtual environment on a monitor, while they found themselves surrounded by the same environment in the VR condition. Previous research indicates that this difference in perspective engages different spatial reference frames and spatial encoding. Kozhevnikov & Dhond (2012) compared mental rotation in three different display conditions, a stereoscopic HMD, a stereoscopic monitor (anaglyph glasses) and a traditional monitor. Their results suggest that participants utilized scene-based reference frames and allocentric encoding for both, the traditional and the stereoscopic monitor conditions. In contrast, only in the HMD condition, participants employed an egocentric frame of reference.

In chapter 3, I presented in both experiments two-dimensional stimuli on an ordinary computer screen, either a physical or virtual one. Hence, a difference between my experiments and the study by Li et al. (2020) which might explain the different findings, is the spatial encoding in the VR conditions of both studies.

Generally, the results of chapter 3 indicate that early visual processing is not different when stimuli are presented in the HTC Vive as compared to a standard computer monitor. Differences between VR and non-VR reported by previous studies appear to be related to spatial processing. More research is needed to investigate potential different mechanisms underlying cognitive processing in virtual environments and both standard monitor-based experiments and real-world scenarios. Understanding how the technology used for stimulus presentation and behavioral measurements affect cognitive processing will be critical to validly compare findings obtained with different technology and to eventually transfer findings from experiments to real-world scenarios.

Limitations

A limitation of the measurement method presented in chapter 2 is clearly the usability. While the method is already able to provide precise and accurate time measurements, the actual implementation, and its integration into UE4 is still in a premature state, e.g., the synchronization with UE4 depends on a custom build of UE. Moreover, the current implementation is limited in its functionality and was custom-made to fit the requirements of my research projects, making it complicated to adapt the method for other research projects.

Another, more general limitation if the approach is the dependency on third-party tools. In principle, every update of UE4 or SteamVR, as well as the driver of the HTC Vive, could potentially break the correct functioning of the background application. Hence, the method needs to be regularly revalidated and to be adapted to software changes.

To get the method future-proof, a logical next step would be to rewrite the entire logic using OpenXR, a new industry-wide standard API, which works across different platforms and VR hardware brands. Soon, OpenXR will completely replace the individual APIs from the different hardware vendors, including SteamVR's own API OpenVR, which was used for the background application. However, this is actually a good development and makes it easier to develop software tools, like the background application, or entire toolboxes based on the same standard and supporting all platforms and devices.

Furthermore, in all my reaction time measurements reported chapter 2, I used a standard response pad by the same company behind the BBTK. The main reason was the usability, since the BBTK response pad provides already the necessary plugs to control it via the BBTK. However, this did not stop me from opening one of the Vive motion controllers and to solder some cable on the mainboard to connect it to the BBTK. However, unfortunately SteamVR's API OpenVR does not provide the necessary functions to easily intercept the input signals of the controllers in the way, I was able to intercept the input provided by the BBTK response pad. Hence, without further tests of the controllers, studies should prefer traditional input devices, if possible, to ensure precise input data.

In chapter 2, I only tested the HTC Vive. Recently, Tachibana & Matsumiya (2021) examined the accuracy and precision of visual and auditory stimulus presentation with Python 2 and Python 3 and two different HMDs, the Oculus Rift CV1 and the HTC Vive Pro. Their study revealed some issues of stimuli with short stimulus durations. For auditory stimuli, they observed lags for auditory stimuli, when they were presented for short durations of one or two frames with both the Oculus Rift CV1 and the HTC Vive Pro. Interestingly, they also observed lags of visual stimuli with a duration of a single frame, but only when using the Oculus Rift.

Furthermore, it is important to point out, that results obtained in chapter 3 are based on only one experimental paradigm and only tested the processing of feature expectations for color and orientation. Hence, based on these results it is not possible to draw conclusions about other tasks or even visual features. For example, previous research has repeatedly shown that distance and size perception is distorted in VR when compared to real-world environments (e.g., Kelly et al., 2017; Maruhn et al., 2019; Phillips et al., 2009) and that spatial processes also differ between VR and standard setups (Anglin et al., 2017; Kozhevnikov & Dhond, 2012). Hence, it is conceivable that manipulations of the expected stimulus size might result in different findings between both setups.

Summary

In standard lab-based psychological experiments, the stimulus material is usually minimalistic, and the behavior of participants is measured in repetitive and simple responses such as simple button presses. Virtual reality has the potential to study brain function and behavior of freely moving participants in realistic and ecologically valid environments, without sacrificing experimental control. However, little is known yet if VR engages the same cognitive processes as equivalent real-world situations (Kulik, 2018; Pan & Hamilton, 2018a).

Here, I asked whether the simple fact that an experiment takes place in VR, changes the behavior of participants. From previous research it is well known that spatial perception is distorted in VR, as egocentric distances are usually underestimated in VR (Interrante et al., 2006; Kelly et al., 2017), when compared to real-world estimates. Also, when comparing the task performance between VR and a standard monitor setup, differences in visuomotor adaptation (Anglin et al., 2017) and mental rotation (Kozhevnikov & Dhond, 2012) have been observed.

My dissertation was concerned with the question if the VR technology affects already early visual processing differently than a typical computer monitor. If differences in visual processing exist, it is important to know on which level of processing these differences arise. Yet, to my knowledge, no previous research has investigated if VR already affects early processing of basic visual features.

First, a new behavioral paradigm was developed for a standard monitor-based setup and extensively tested to provide robust and replicable behavioral effects. The paradigm investigated how prediction errors of basic visual stimulus features are formed and on which levels of processing they arise. In a series of four experiments, participants consistently responded slower, when the color or the orientation of the targets were different than expected. This was irrespective of the fact that both the color and the orientation were completely task irrelevant in three of the four experiments and whether the features belonged to the same object (Exp 2 – Exp 4) or whether the features were separated on different objects (Exp 1). Increasing the relevance of the features and switching to an explicit manipulation of the expectations, in Experiment 3 of chapter 1, did not change results.

All in all, the behavioral effects turned out to be robust and replicable, when tested with the same combination of software and hardware, which was a critical premise for the comparison between a standard experiment and the same experiment in VR.

A critical premise for such a comparison is that both experimental setups provide the same high precision and accuracy for controlling stimuli and to collect behavioral data. In chapter 2 and chapter 3, I demonstrated that the accuracy and precision of reaction time measurements is highly dependent on the hardware and software used for the experiment. Here, the HTC Vive HMD system turned out to be well suited for experiments, which require displays with reliable timing parameters. Stimuli presented in the HTC Vive in combination with Unreal Engine were highly precise and accurate, when taking the true refresh rate of 89.53 Hz into account. Furthermore, especially the *global onset* display provides advantages over common LCDs for measuring the stimulus-onset.

Nevertheless, my experiments revealed that reaction times obtained with UE4's programming interface were highly inaccurate and imprecise. This was explained by the underlying architecture of game engines and the graphics pipeline for VR rendering. Furthermore, I proposed a novel software-based method to circumvent these limitations, by recording the data in a background process separate

from the engine’s internal logic. This approach turned out to provide reaction time measurements with a comparable precision and accuracy as provided by standard toolboxes such as Presentation and PsychoPy. The proposed method has not only proven to provide highly accurate and precise reaction time measurements, but also provide a basis for other time-sensitive measurements, such as EEG.

For the VR experiment of chapter 3, I aimed to design the experimental setup of the VR experiment as close as possible to the non-VR experiment. This included an accurate model of the entire EEG chamber, in which the experiments were conducted. In the end, both the non-VR and the VR experiment replicated the original findings obtained with PsychoPy in chapter 1. Hence, the results did not provide any evidence that basic visual feature and unexpected changes of these features is different just because the stimuli are presented in an HMD.

Bibliography

Al Zayer, M., Adhanom, I. B., MacNeilage, P., & Folmer, E. (2019). The effect of field-of-view restriction on sex bias in vr sickness and spatial navigation performance. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Alsbury-Nealy, K., Wang, H., Howarth, C., Gordienko, A., Schlichting, M. L., & Duncan, K. D. (2021). OpenMaze: An open-source toolbox for creating virtual navigation experiments. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01664-9

Anglin, J. M., Sugiyama, T., & Liew, S.-L. (2017). Visuomotor adaptation in head-mounted virtual reality versus conventional training. *Scientific Reports*, *7*(1), 45469. https://doi.org/10.1038/srep45469

Berg, L. P., & Vance, J. M. (2017). Industry use of virtual reality in product design and manufacturing: A survey. *Virtual Reality*, *21*(1), 1–17. https://doi.org/10.1007/s10055-016-0293-9

Birckhead, B., Khalil, C., Liu, X., Conovitz, S., Rizzo, A., Danovitch, I., Bullock, K., & Spiegel, B. (2019). Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group: Iterative Study. *JMIR MENTAL HEALTH*, 14.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, *12*(12), 752–762. https://doi.org/10.1038/nrn3122

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414.

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2020). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*, *52*(2), 455–463. https://doi.org/10.3758/s13428-019-01242-0

Cummings, J. J., & Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, *19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Daniel, F., & Kapoula, Z. (2019). Induced vergence-accommodation conflict reduces cognitive performance in the Stroop test. *Scientific reports*, *9*(1), 1–13.

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & MÃ¼hlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00026

Garaizar, P., Vadillo, M. A., López-de-Ipiña, D., & Matute, H. (2014). Measuring software timing errors in the presentation of visual stimuli in cognitive neuroscience experiments. *PloS one*, *9*(1), e85108.

Giesel, M., Nowakowska, A., Harris, J. M., & Hesse, C. (2020). Perceptual uncertainty and action consequences independently affect hand movements in a virtual environment. *Scientific Reports*, *10*(1), 22307. https://doi.org/10.1038/s41598-020-78378-z

Gonzalez-Franco, M., Abtahi, P., & Steed, A. (2019). Individual Differences in Embodied Distance Estimation in Virtual Reality. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 941–943. https://doi.org/10.1109/VR.2019.8798348

Gonzalez-Franco, M., & Peck, T. C. (2018). Avatar Embodiment. Towards a Standardized Questionnaire. *Frontiers in Robotics and AI*, *5*, 74. https://doi.org/10.3389/frobt.2018.00074

Grübel, J., Weibel, R., Jiang, M. H., Hölscher, C., Hackman, D. A., & Schinazi, V. R. (2017). EVE: A Framework for Experiments in Virtual Environments. In T. Barkowsky, H. Burte, C. Hölscher, & H. Schultheis (Hrsg.), *Spatial Cognition X* (Bd. 10523, S. 159–176). Springer International Publishing. https://doi.org/10.1007/978-3-319-68189-4_10

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, *8*(3), 33–33.

Interrante, V., Ries, B., & Anderson, L. (2006). Distance Perception in Immersive Virtual Environments, Revisited. *IEEE Virtual Reality Conference (VR 2006)*, 3–10. https://doi.org/10.1109/VR.2006.52

Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.

Juliano, J. M., Schweighofer, N., & Liew, S.-L. (2021). *Increased cognitive load in immersive virtual reality during visuomotor adaptation is associated with decreased long-term retention and context transfer* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-1139453/v1

Kelly, J. W., Cherep, L. A., & Siegel, Z. D. (2017). Perceived Space in the HTC Vive. *ACM Transactions on Applied Perception*, *15*(1), 1–16. https://doi.org/10.1145/3106155

Kim, S.-Y., Park, H., Jung, M., & Kim, K. (Kenny). (2020). Impact of Body Size Match to an Avatar on the Body Ownership Illusion and User's Subjective Experience. *Cyberpsychology, Behavior, and Social Networking*, *23*(4), 234–241. https://doi.org/10.1089/cyber.2019.0136

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340. https://doi.org/10.1348/000712607X251243

Kothgassner, O. D., & Felnhofer, A. (2020). Does virtual reality help to cut the Gordian knot between ecological validity and experimental control? *Annals of the International Communication Association*, *44*(3), 210–218. https://doi.org/10.1080/23808985.2020.1792790

Kozhevnikov, M., & Dhond, R. P. (2012). Understanding Immersivity: Image Generation and Transformation Processes in 3D Immersive Environments. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00284

Kramida, G. (2015). Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics*, *22*(7), 1912–1931.

Krokos, E., Plaisant, C., & Varshney, A. (2019). Virtual memory palaces: Immersion aids recall. *Virtual reality*, *23*(1), 1–15.

Krueger, M. W. (1991). *Artificial reality II*.

Kulik, A. (2018). Virtually the ultimate research lab. *British Journal of Psychology*, *109*(3), 434–436. https://doi.org/10.1111/bjop.12307

Li, G., Anguera, J. A., Javed, S. V., Khan, M. A., Wang, G., & Gazzaley, A. (2020). Enhanced Attention Using Head-mounted Virtual Reality. *Journal of Cognitive Neuroscience*, *32*(8), 1438–1454. https://doi.org/10.1162/jocn_a_01560

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of computer-mediated communication*, *3*(2), JCMC321.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999a). Immersive virtual environment technology as a basic research tool in psychology. *Behavior research methods, instruments, & computers*, *31*(4), 557–564.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999b). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 557–564. https://doi.org/10.3758/BF03200735

Marek, N., & Pollmann, S. (2020). Contextual-Cueing beyond the Initial Field of View—A Virtual Reality Experiment. *Brain Sciences*, *10*(7), 446. https://doi.org/10.3390/brainsci10070446

Maruhn, P., Schneider, S., & Bengler, K. (2019). Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains. *PLOS ONE*, *14*(10), e0224651. https://doi.org/10.1371/journal.pone.0224651

Mast, F. W., & Oman, C. M. (2004). Top-Down Processing and Visual Reorientation Illusions in a Virtual Reality Environment. *Swiss Journal of Psychology*, *63*(3), 143–149. https://doi.org/10.1024/1421-0185.63.3.143

Minsky, M. (1980). *Telepresence*. https://web.media.mit.edu/~minsky/papers/Telepresence.html

Mohler, B. J., Creem-Regehr, S. H., Thompson, W. B., & Bülthoff, H. H. (2010). The Effect of Viewing a Self-Avatar on Distance Judgments in an HMD-Based Virtual Environment. *Presence: Teleoperators and Virtual Environments*, *19*(3), 230–242. https://doi.org/10.1162/pres.19.3.230

Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology*. W. H. Freeman and Company.

Niehorster, D. C., Li, L., & Lappe, M. (2017). The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research. *I-Perception*, *8*(3), 204166951770820. https://doi.org/10.1177/2041669517708205

Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Immersion Revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, *12*(2), 108–134. https://doi.org/10.17011/ht/urn.201611174652

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939–973. https://doi.org/10.1017/S0140525X01000115

Pan, & Hamilton, A. F. de C. (2018a). Understanding dual realities and more in VR. *British Journal of Psychology*, *109*(3), 437–441. https://doi.org/10.1111/bjop.12315

Pan., & Hamilton, A. F. de C. (2018b). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Pan, ., & Steed, A. (2019). How Foot Tracking Matters: The Impact of an Animated Self-Avatar on Interaction, Embodiment and Presence in Shared Virtual Environments. *Frontiers in Robotics and AI*, *6*, 104. https://doi.org/10.3389/frobt.2019.00104

Parsons, T. D. (2015). Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00660

Patil, I., Zanon, M., Novembre, G., Zangrando, N., Chittaro, L., & Silani, G. (2018). Neuroanatomical basis of concern-based altruism in virtual environment. *Neuropsychologia*, *116*, 34–43.

Peirce, J., Gray, J., Halchenko, Y., Britton, D., Rokem, A., Strangman, G., & others. (2011). *PsychoPy–A Psychology Software in Python*.

Pfeil, K., Taranta, E. M., Kulshreshth, A., Wisniewski, P., & LaViola, J. J. (2018). A comparison of eye-head coordination between virtual and physical realities. *Proceedings of the 15th ACM Symposium on Applied Perception*, 1–7. https://doi.org/10.1145/3225153.3225157

Phillips, L., Ries, B., Interrante, V., Kaeding, M., & Anderson, L. (2009). Distance perception in NPR immersive virtual environments, revisited. *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization - APGV '09*, 11. https://doi.org/10.1145/1620993.1620996

Quinlivan, B., Butler, J. S., Beiser, I., Williams, L., McGovern, E., O'Riordan, S., Hutchinson, M., & Reilly, R. B. (2016). Application of virtual reality head mounted display for investigation of movement: A novel effect of orientation of attention. *Journal of Neural Engineering*, *13*(5), 056006. https://doi.org/10.1088/1741-2560/13/5/056006

Ramdhani, N., Akpewila, F., Faizah, M., & Resibisma, B. (2019). It's so Real: Psychophysiological Reaction towards Virtual Reality Exposure. *2019 5th International Conference on Science and Technology (ICST)*, 1–5. https://doi.org/10.1109/ICST47872.2019.9166212

Ries, B., Interrante, V., Kaeding, M., & Anderson, L. (2008). The effect of self-embodiment on distance perception in immersive virtual environments. *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology - VRST '08*, 167. https://doi.org/10.1145/1450579.1450614

Roberts, A. C., Yeap, Y. W., Seah, H. S., Chan, E., Soh, C.-K., & Christopoulos, G. I. (2019). Assessing the suitability of virtual reality for psychological testing. *Psychological Assessment*, *31*(3), 318–328. https://doi.org/10.1037/pas0000663

Schmuckler, M. A. (2001). What Is Ecological Validity? A Dimensional Analysis. *Infancy*, *2*(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02

Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *8*(5), 560–565. https://doi.org/10.1162/105474699566477

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

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, *109*(3), 431–433. https://doi.org/10.1111/bjop.12305

Slater, M., & Steed, A. (2000). A Virtual Presence Counter. *Presence: Teleoperators and Virtual Environments*, *9*(5), 413–434. https://doi.org/10.1162/105474600566925

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Spanlang, B., Normand, J.-M., Borland, D., Kilteni, K., Giannopoulos, E., PomÃ©s, A. s, GonzÃ¡lez-Franco, M., Perez-Marcos, D., Arroyo-Palacios, J., Muncunill, X. N., & Slater, M. (2014). How to Build an Embodiment Lab: Achieving Body Representation Illusions in Virtual Reality. *Frontiers in Robotics and AI*, *1*. https://doi.org/10.3389/frobt.2014.00009

Steed, A., Pan, Y., Zisch, F., & Steptoe, W. (2016). The impact of a self-avatar on cognitive load in immersive virtual reality. *2016 IEEE Virtual Reality (VR)*, 67–76. https://doi.org/10.1109/VR.2016.7504689

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, *42*(4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Sutcliffe, A., Gault, B., & Shin, J.-E. (2005). Presence, memory and interaction in virtual environments. *International Journal of Human-Computer Studies*, *62*(3), 307–327. https://doi.org/10.1016/j.ijhcs.2004.11.010

Tachibana, R., & Matsumiya, K. (2021). Accuracy and precision of visual and auditory stimulus presentation in virtual reality in Python 2 and 3 environments for human behavior research. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01663-w

Takac, M., Collett, J., Conduit, R., & Foe, A. D. (2021). Addressing virtual reality misclassification: A hardware-based qualification matrix for virtual reality technology. *Clinical Psychology & Psychotherapy*, *28*(3), 538–556. https://doi.org/10.1002/cpp.2624

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.

Ugwitz, P., Šašinková, A., Šašinka, Č., Stachoň, Z., & Juřík, V. (2021). Toggle toolkit: A tool for conducting experiments in unity virtual environments. *Behavior Research Methods*, *53*(4), 1581–1591. https://doi.org/10.3758/s13428-020-01510-4

Ulric Neisser. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology*. W. H. Freeman and Company.

Vasser, M., & Aru, J. (2020). Guidelines for immersive virtual reality in psychological research. *Current Opinion in Psychology*, *36*, 71–76. https://doi.org/10.1016/j.copsyc.2020.04.010

Vasser, M., Kängsepp, M., Magomedkerimov, M., Kilvits, K., Stafinjak, V., Kivisik, T., Vicente, R., & Aru, J. (2017). VREX: an open-source toolbox for creating 3D virtual reality experiments. *BMC psychology*, *5*(1), 1–8.

Visch, V. T., Tan, E. S., & Molenaar, D. (2010). The emotional and cognitive effect of immersion in film viewing. *Cognition and Emotion*, *24*(8), 1439–1445.

Wang, P., & Nikolic, D. (2011). An LCD monitor with sufficiently precise timing for research in vision. *Frontiers in human neuroscience*, 85.

Watson, M. R., Voloh, B., Thomas, C., Hasan, A., & Womelsdorf, T. (2019). USE: An integrative suite for temporally-precise psychophysical experiments in virtual environments for human, nonhuman, and artificially intelligent agents. *Journal of neuroscience methods*, *326*, 108374.

Weber, S., Weibel, D., & Mast, F. W. (2021). How to Get There When You Are There Already? Defining Presence in Virtual Reality and the Importance of Perceived Realism. *Frontiers in Psychology*, *12*, 628298. https://doi.org/10.3389/fpsyg.2021.628298

Wienrich, C., Gross, R., Kretschmer, F., & Müller-Plath, G. (2018). Developing and proving a framework for reaction time experiments in VR to objectively measure social interaction with virtual agents. *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*, 191–198.

Williams, H. E., Chapman, C. S., Pilarski, P. M., Vette, A. H., & Hebert, J. S. (2019). Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol. *PLOS ONE*, *14*(12), e0219333. https://doi.org/10.1371/journal.pone.0219333

Wolfel, M., Hepperle, D., Purps, C. F., Deuchler, J., & Hettmann, W. (2021). Entering a new Dimension in Virtual Reality Research: An Overview of Existing Toolkits, their Features and Challenges. *2021 International Conference on Cyberworlds (CW)*, 180–187. https://doi.org/10.1109/CW52790.2021.00038

Woolley, B. (1993). *Virtual worlds: A journey in hype and hyperreality*. Benjamin Woolley.

Yuan, Y., & Steed, A. (2010). Is the rubber hand illusion induced by immersive virtual reality? *2010 IEEE Virtual Reality Conference (VR)*, 95–102. https://doi.org/10.1109/VR.2010.5444807