You Spin my Head Right Round: Threshold of Limited Immersion for Rotation Gains in Redirected Walking

Patric Schmitz, Julian Hildebrandt, André Calero Valdez, Leif Kobbelt, and Martina Ziefle

Abstract—In virtual environments, the space that can be explored by real walking is limited by the size of the tracked area. To enable unimpeded walking through large virtual spaces in small real-world surroundings, redirection techniques are used. These unnoticeably manipulate the user's virtual walking trajectory. It is important to know how strongly such techniques can be applied without the user noticing the manipulation—or getting cybersick. Previously, this was estimated by measuring a detection threshold (DT) in highly-controlled psychophysical studies, which experimentally isolate the effect but do not aim for perceived immersion in the context of VR applications. While these studies suggest that only relatively low degrees of manipulation are tolerable, we claim that, besides establishing detection thresholds, it is important to know when the user's immersion breaks. We hypothesize that the degree of unnoticed manipulation is significantly different from the detection threshold when the user is immersed in a task. We conducted three studies: a) to devise an experimental paradigm to measure the *threshold of limited immersion* (TLI), b) to measure the TLI for slowly decreasing and increasing rotation gains, and c) to establish a baseline of cybersickness for our experimental setup. For rotation gains greater than 1.0, we found that immersion breaks quite late after the gain is detectable. However, for gains lesser than 1.0, some users reported a break of immersion even before established detection thresholds were reached. Apparently, the developed metric measures an additional quality of user experience. This article contributes to the development of effective spatial compression methods by utilizing the break of immersion as a benchmark for redirection techniques.

Index Terms—Virtual reality, redirected walking, rotation gain, perceptual threshold, immersion, cybersickness

1 Introduction

With Virtual Reality (VR) systems becoming a commodity, new challenges occur for designing locomotion interfaces. While the available space in specialized lab environments is sufficient to explore large virtual scenes, tracked spaces in home or office environments are typically smaller. Consequently, a one-to-one mapping between the real and virtual world cannot be granted. Many VR systems solve this by using indirect navigation modes, such as flying or teleportation, with tracked hand controllers, gamepads, or other interaction devices. Some systems use motion tracking to enable walking-in-place, or treadmills, which give the illusion of real walking while staying at a fixed location. Yet, natural walking interfaces offer a higher perceived presence in the virtual environment (VE) compared to these more common metaphors [41].

To enable natural movement through a virtual scene that exceeds the available tracked area, different spatial compression methods have been proposed, which are outlined in Section 2.1. While varying in effectiveness and intrusiveness, all redirection methods face a common challenge: the degree of manipulation should not exceed a threshold at which the users' immersion diminishes or cybersickness is caused by the manipulated motion. To establish spatial compression methods in practical applications, it is crucial to determine these thresholds and to comply with them when designing natural locomotion interfaces.

- *Patric Schmitz is with Visual Computing Institute of RWTH Aachen University. E-Mail: patric.schmitz@cs.rwth-aachen.de.*
- *Julian Hildebrandt is with Human Computer Interaction Center (HCIC) of RWTH Aachen University. E-Mail: hildebrandt@comm.rwth-aachen.de*
- *André Calero Valdez is with Human Computer Interaction Center (HCIC) of RWTH Aachen University. E-Mail: calero-valdez@comm.rwth-aachen.de.*
- *Leif Kobbelt is with Visual Computing Institute of RWTH Aachen University. E-Mail: kobbelt@cs.rwth-aachen.de.*
- *Martina Ziefle is with Human Computer Interaction Center (HCIC) of RWTH Aachen University. E-Mail: ziefle@comm.rwth-aachen.de.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

Previous research on thresholds for the allowed degree of motion manipulation has measured perceptual detection thresholds in controlled psychophysical experiments [35]. An overview of this work is given in Section 2.2. While being a useful basic metric to establish a lower bound on the allowed degree of manipulation, we argue that additional metrics are required to judge redirection techniques for real-world applications. When the user is engaged in a task and not actively paying attention to the performed manipulations, the relevant threshold should be, in general, different from the psychophysical detection threshold (DT).

In this article, we propose the *threshold of limited immersion* (TLI) as a complementary metric to guide the development of redirection techniques. We present the results of three user studies, which were conducted in succession. The first study comprises the methodical development of an experimental paradigm to asses the TLI. In a second study, we applied the method to measure the TLI for slowly varying rotational gains during a search and collect task. The third study serves as a baseline measurement for the cybersickness induced by our experimental setup without any motion gains applied.

The main contributions of this research are the following:

- we propose the *threshold of limited immersion* as a new metric for the evaluation of spatial compression techniques,
- we devise an experimental paradigm to measure the TLI,
- we empirically validate the metric in a user study,
- and we determine the TLI for slowly varying rotation gains during a search and collect task in a low complexity scene.

The remainder of this article is structured as follows. We first provide an overview on research on spatial compression methods and the limits of human perception for different types of motion gains in Section 2. Our research hypothesis and the empirical procedure we chose are presented in Section 3. We then present the experiments that we conducted to devise and validate our proposed metric and the results in Section 4. We consolidate and discuss the key results, pointing out current limitations of our approach, and outlining future work in Section 5. Lastly, we summarize our contribution with closing remarks in Section 6.

2 RELATED WORK

To effectively compress space in a VE, two aspects must be addressed. First, it is necessary to find a method that achieves the necessary degree of compression. Second, the applied motion manipulation should not deteriorate the immersive user experience in the virtual world. Therefore, it is necessary to understand various spatial compression methods, the limitations of the human perception, and the influence of human factors on the resulting user experience.

2.1 Spatial Compression Methods

Redirected walking applies dynamic gains to the user's motion in the virtual scene. In turn, the user unconsciously compensates for the altered motion, which enables to steer the user onto a desired trajectory in the real world [25]. Since its inception, the idea of redirected walking has been generalized and improved in a number of ways. To allow free exploration of virtual scenes, generalized *steering algorithms* have been proposed based on universal heuristics [10, 12], human motion models [13, 48], and hybrid approaches performing path-planning on a set of mandatory and optional waypoints [2].

Redirection based on motion gains requires a minimum amount of user rotation to effectively steer the user. To this end, motion gains have been combined with distracting objects that induce a turning motion whenever the user is about to hit a wall or obstacle [22]. *Distractors* have further been used to avert users' attention from virtual scene manipulations that are either caused by amplified user motion [21], continuous rotation independent from the user [7] or a combination of both [22]. Similarly, the perceptual phenomenon of *change blindness* has been used to unnoticeably redirect the user, based on adaptive environment models [37], by creating self-overlapping virtual spaces [38, 42], or by manipulating the scene during saccadic eye movement [5] or blinking [15]. Other perceptual illusions have also been investigated for redirection [34].

Planar map folding has recently been proposed as an approach to the spatial compression problem [39]. Collisions are avoided by precomputing a preferably conformal and locally bijective static mapping between the virtual floor plan and the available real-world space.

2.2 Human Factors and Perception

When selecting a suitable method for spatial compression, it is critical to assess the effect of such methods on user experience. Therefore, we first develop a working definition of presence and immersion, then look at how changes in each have previously been measured, and lastly look at the daunting effect of cybersickness.

2.2.1 Terminology: Presence and Immersion

Both terms *immersion* and *presence* have been used to describe the exceptional quality of VEs to detach the user's (self-) perception from the immediate physical surroundings. Notable previous work regards immersion as an inherent property of the display system that can be objectively quantified and is expressly not a subjective reaction to the VE [30]. The degree to which the display and transformation of sensory information is similar to the real world has furthermore been described as *sensory fidelity* [6]. In accordance with a recent survey [19] we denote these intrinsic display properties as *system immersion*. Conversely, the term immersion is used to describe a perceptual response to the VE, as the psychological state of perceiving oneself as being part of the VE stimulus flow [46]. In this definition of immersion, natural modes of interaction and control, and particularly the perception of self-movement are influencing factors on the *subjective* degree of immersion. In our work, we adopt this latter definition of immersion as a subjective user response to VE stimuli.

In line with previous work, we define presence as the mental state in which a user feels physically present in a computer-mediated environment, which has been designated as experiential telepresence [8]. While the precise definition of the term as well as experimental procedures to measure the degree of presence have been widely and controversially discussed, it is mostly agreed upon that presence can most effectively be measured by subjective reporting [31], and typically in conjunction with a questionnaire [46]. In our work, we consider immersion to be a necessary precondition for a sense of presence. Moreover, presence would require additional factors to be at play such as scene believability, a sense of inclusion, and the activation of other higher level cognitive processes (e.g. focused attention) to turn the VE into a location the user feels present in.

2.2.2 Detection Thresholds and Breaks in Presence

While human factors' influence on VR hardware design have been sufficiently discussed to provide authentic visual, auditory and haptic perception (i.e. *system immersion*) [32,45], it is not fully understood to what extent redirection techniques can be used. Even more important, there is no consensus on the criterion that defines this extent.

One such criterion is the number of *Breaks in Presence* (BIP), which occur when the user experiences a transition of presence from the VE to reality [31]. To detect BIPs, subjects are usually instructed to verbally report anything that feels unnatural or implausible [11,38]. Frequencies of reported BIPs have been mapped to a continuous presence scale. The probability of being in a high-presence state is estimated based on the equilibrium probabilities of a presumed stochastic process, modeled by a two-state Markov chain [31].

Previous work introduced a taxonomy of redirection techniques and differentiates methods for overt redirection (e.g. distractors with "follow me"-semantic) categorically from those for subtle reorientation (e.g. rotation gains) [36]. A user study revealed that users experienced more BIPs when (1) an overt method is applied or (2) the intensity of the subtle method is classified by the authors as *not optimal*, which means above the psychophysical detection threshold.

Instead of considering the number of reported BIPs, studies that estimate perceptual thresholds for e.g. rotation gains aim to keep redirection techniques fully *unnoticed*. They use isolated experimental settings where participants are rotating in place, while the rotation is manipulated. After each performed motion, participants report whether the virtually displayed motion was amplified or condensed. Typically, those studies use two-alternative force-choice (2AFC) paradigms and estimate the detection thresholds by fitting a sigmoid-shaped psychometric function to the portion of correct answers. The DT is then defined as the 75% quantile of this function, denoting the boundary at which the portion of correct answers is significantly different from chance. The measured perceptional thresholds for rotation gains lie at 0.67 and 1.24 respectively [35]. Participants who are actively paying attention to their head rotation can correctly determine if the rotation was weakened or enlarged in 75% of the trials if their physical rotation is multiplied by factors within this range. Subsequent experiments that used the same paradigm revealed detection thresholds of 0.82 and 1.2 for acoustic-based rotation gains, for which subjects walked through a virtual room in complete darkness [29]. Another study that compared different audio conditions for audiovisual input revealed thresholds from 0.77 to 0.8 and from 1.08 to 1.11, concluding that visual input is dominant [20]. The 2AFC paradigm is well established to reliably detect psychophysical detection thresholds in terms of signal detection. Nevertheless, since subjects need to be informed about the redirection beforehand, the paradigm may also lead to users' intensive scrutiny of stimuli [11].

2.2.3 Cybersickness

Besides detection thresholds and breaks in presence as human factors, *cybersickness* might be the first and foremost threat to a VR system [32]. Cybersickness is defined as symptoms of motion sickness, but occurring during or after the use of VEs [17]. The symptoms vary individually and include eye strain, headache, pallor, sweating, dryness of mouth, fullness of stomach, disorientation, vertigo, ataxia, nausea, and might even lead to vomiting [16]. Causes of cybersickness are not fully explained, but there are three main theories regarding the emergence of cybersickness: Cybersickness occurs (1) because of conflicts between the visual and vestibular system, (2) because the users do not (yet) possess mechanisms to maintain postural stability [27], or (3) because adverse input tricks the body into thinking that it ingested something poisonous [17]. Regardless of the particular explanation for cybersickness, there are several factors with negative impact on users' comfort. Those factors are high visual flow, degree of control over the application, time on task, blur level, and latency [24].

Furthermore, individual differences, such as gender, age, or prior exposure, account for the emergence of cybersickness. Women tend to be more susceptible to cybersickness than men. This could be due to a reporting bias regarding cybersickness [4], or due to physical differences in the perceptual field of view. Women, due to their larger fields of view, have to handle more visual input in comparison to men. Previous exposure to similar technology is another important factor, since tolerance towards sickness-inducing stimuli might be learnable [33]. Another influencing factor is age, since traditional motion sickness appears to be stronger on young people under 12 years and lowest on people over 50 years [26]. However, there is indication that this association might be the other way around regarding cybersickness [1].

However, entertaining VR applications might still be highly enjoyable, even if cybersickness arises [44], and even if cybersickness occurs more strongly, games and movies in stereoscopic 3D are rated higher in immersion and presence than their monoscopic equivalents [28, 47]. Consequently, the level of immersion becomes a third dimension that has to be considered besides detection and cybersickness thresholds. Overall, it turns out that it is still unclear how to predict if, and how strongly cybersickness occurs with respect to interindividual differences and virtual reality experiences. This is even more the case when incorporating spatial compression methods. Understanding the impact of such methods on user experience and the emergence of cybersickness in VR applications helps in designing VEs with better acceptability for all users in spatially limited environments.

3 Hypotheses and Logic of Empirical Procedure

When developing spatial compression methods, we ultimately want to use them in practical VR applications to allow real walking in spatially limited surroundings. Many different methods have been proposed and evaluated regarding their effectiveness, yet we do not see any real applications making use of them. It is still a long way before natural walking interfaces will see widespread use and establish themselves as an everyday technology. We must focus on providing virtual scenes in limited surroundings that are both believable and lead to a high level of user immersion. To achieve user acceptance, there are a number of obstacles to overcome. Besides devising methods that work safely and reliably for arbitrary virtual scenes, we have to ensure that the methods do not degrade the immersive user experience, or even cause cybersickness.

We believe that no single spatial compression method will be able to cover all application scenarios equally well. Thus, to achieve universal effectiveness, a synergetic combination of different redirection methods is necessary. The applied set of methods must be adaptively chosen based on a number of factors determining their applicability at any given moment. Accordingly, the individual and combined degree of manipulation must be dynamically scaled. This should be done in a manner that meets the requirements of the individual user in mind (e.g., resistance/habit to sickness, attentional focus, and different levels of scene/task complexity). Users that have a strong resistance to cybersickness could get away with stronger compression than users with no experience. The ideal natural locomotion interface would combine redirection techniques, existing and yet to be developed, in such a way that a small living room area suffices to explore virtual scenes of arbitrary size and shape, while maintaining the immersive experience and without causing noticeable cybersickness.

Two existing methods to assess the allowable degree of redirection are to estimate psychophysical detection thresholds (DT) and to capture breaks in presence (BIP). The DT of an isolated stimulus is measured in a controlled setting using the established 2AFC paradigm for measuring psychophysical thresholds. It pertains directly to the disturbing influence in question, and is a very useful basic metric that puts a lower bound on the allowed degree of manipulation. It presumes that undetectable stimuli are unlikely to have an impact on user experience. BIPs are typically measured via verbal exclamation. They capture the point in time at which presence degrades from a high- to a low-presence state. BIPs can be triggered for many reasons, such as degraded immersion, stimuli from the real world, or if the user scrutinizes the VE or loses attention.

The application of a redirection technique can be seen as a deliberate reduction of the faithfulness of the VE, hence a reduction of system immersion. If the effect is stronger than some threshold, this will lead to a degradation of the perceived immersion of the user, and causes a BIP when the user's attention shifts from the individual task to the applied manipulation. For a real application, estimating this threshold is of high importance, since it denotes when a manipulation will become disturbing for the VE experience, rather then being detectable. We call this threshold the *threshold of limited immersion* (TLI).

We hypothesize that the TLI is influenced notably by the degree of immersion into a task or real-world application scenario itself, since users are not and should not be focusing on the manipulation when experiencing VR. We propose the TLI as a complementary metric to the established detection threshold (DT). We do not ask whether users can reliably detect a manipulation, but rather when they actually feel that a manipulation interferes with the quality of their experience.

To capture the TLI, a new experimental paradigm is required. The challenge in designing a measurement procedure is that we want participants to actively report an effect, even if we do not directly inform them about it. Further, users should report on an effect which they have probably never experienced before. Overall, we purposefully have to create immersion, then slowly break it, and get users to report on this—without them even knowing that something might break. In this article, we present three experiments in which both the measurement procedure and its validation was under study.

The first experiment was carried out to find a proper measurement procedure for reporting a perceived break of immersion ($N = 16$). Methodically, we approached this by iteratively refining our method and performing intermittent measurements on new groups of subjects. For this we built a testbed environment, consisting of a simple living room scene, in which a number of dynamically placed targets had to be collected. Details about the experimental setup are given in Sections 4.1–4.3. We applied continually varying rotational gains as our exemplary motion manipulation. To keep the gain constant between two collected targets we developed a target placement heuristic that minimizes the number of manual reorientations per trial. We refined the way in which subjects should report a break in immersion, as well as the specific instructions that were given. The goal was to match (a) the degree of manipulation at which subjects reported a break of immersion in the test environment with (b) the answers we obtained from a post-experimental interview. Details about the process iterations, the final measurement procedure, and a discussion of our results are given in Section 4.4.

In the second experiment we applied the developed experimental procedure to assess the TLI for continually varying rotational gains on a larger number of subjects ($N = 35$). The same scenario and task as in the first experiment were used. The purpose of the experiment is twofold. First, we put the experimental procedure to test and see if the reporting behavior of subjects matched our expectations. Second, we assessed the TLI for slowly varying rotational gains during a search-and-collect task in a static scene with low complexity. Results of the second study are presented in Section 4.5.

The final experiment was conducted to establish a baseline measurement of the cybersickness induced by our experimental setup when no rotational gains are applied ($N = 10$). Many previous studies on redirection techniques only measure the degree of cybersickness induced by the experiment. They do not perform a baseline measurement to isolate the effect of the applied manipulation on cybersickness levels.

4 Experiments

In this section we describe the three experiments. The aim of the first experiment (E1) was to set an experimental paradigm to measure the TLI. The second experiment (E2) actually measures the TLI for an exemplary type of motion manipulation, and the third one (E3) serves as a control group experiment to examine the effects of rotation gains on cybersickness.

Table 1. Abbreviations of variable names used in E2

Abbreviation	Meaning
INC	Threshold of limited immersion in increasing conditon
DEC	Threshold of limited immersion in decreasing conditon
SET	Self-efficacy towards technology
ITQ	Immersion Tendency Questionnaire
EXP	Experience with 3D technology
TOL	Tolerance for nauseous activites
VZ-2	Mental rotation ability
PD	Pupil distance
PRE	Presence
IMM	Immersion
ENJ	Enjoyment
ANX	Anxiety
ADO	Technology adoption
T-RDW	Trust in redirected walking
J-RDW	Judgement on redirected walking

4.1 Hardware Setup

The VE was rendered on a Samsung GearVR head-mount with a Samsung Galaxy S6 mobile phone. We chose this platform since the GPU requirements for our virtual scene are sufficiently low. Yet, being untethered and lightweight, the device is rather unobtrusive and allows for unhampered locomotion. The rendering performance of the device was continuously monitored to ensure a constant framerate above 60Hz at all times during our experiments. Our experiments were conducted in a tracked area of approximately 3m x 4m. Position tracking was performed at an update rate of 60Hz by a set of 4 A.R.T. DTrack1 cameras. Tracking data was transmitted via a dedicated 5GHz wireless connection. We repeatedly verified that the network was not prone to any package loss or considerable latency or jitter. The torso transformation was redundantly tracked with tracking markers attached to the front and back. To achieve a higher update rate and precision for the head orientation, which is crucial to our experiments, we fused the global tracker data with the orientation reported by the high-precision inertial sensor of the head-mount. The sound signal was generated by a Focusrite 2i4 USB soundcard and played back via a Sennheiser RS170 wireless headphone.

4.2 Target Placement Heuristic

The dynamic placement of targets had to be done such that, while keeping the rotation gain constant during movement, the number of times a user would hit a wall or leave the tracked space is minimized. This is not easily achieved since the turn direction is not known beforehand, and a left- or right turn with a given gain will result in a different real-world orientation of the user. A target placement heuristic was developed, which trades off the resulting worst case real-world position after either turn, the target position in the virtual world and the distance to the previous target. The walkable space in our virtual scene (a square of 6m x 6m) was densely sampled on a uniform grid, and potential target positions T were exhaustively tested to find the position $\vec{p}_{next}$ which minimizes our metric.

$$\vec{p}_{next} = \underset{\vec{p}\in T}{\operatorname{argmin}} \left(c_{real}E_{real} + c_{virt}E_{virt} + c_{dist}E_{dist} \right) \quad (1)$$

The penalty term E_{real} depends on the user's resulting real-world position $\vec{p}_{real}$ after taking a left or right turn, respectively, and then walking straight to the target.

$$E_{real} = \max_{d\in\{l,r\}} plat(\vec{p}_{real}(d), 0.8) \quad (2)$$

$$E_{virt} = plat(\vec{p}, 2.0) \quad (3)$$

$$E_{dist} = \max(0, 2 - |\vec{p} - \vec{p}_{last}|) \quad (4)$$

The plateau funcion $plat(\vec{p}, R)$ defines a linear penalty on distances larger than R in the form of a truncated cone.

$$plat(\vec{p}, R) = \max(|\vec{p}| - R, 0) \quad (5)$$

The penalty terms can be interpreted as follows. E_{real} puts a linear penalty on the worst case real-world position after either turn direction, proportional to the distance to a disk of 0.8m radius that is centered in the tracked area. We observed that using the distance to a disk, instead of the distance to the center, led to a behavior similar to the steer-to-orbit heuristic and reduced the number of reorientations in our scenario. Similarly, E_{virt} linearly penalizes target positions that are more than 2m away from the center of the virtual room. E_{dist} penalizes target positions which are closer than 2m to the previous target location, which spreads out the targets more evenly and assures that subjects walk a minimum distance before taking a turn. The relative weighting of the penalty terms was informally adapted to our experimental setup, giving the best results at $c_{real} = 1.0$, $c_{virt} = 0.4$, $c_{dist} = 1.5$.

The approach worked well for gains up to 2.0, with an average number of manual reorientations of 1.18 ($SD = 1.13$). We assumed 2.0 to be an upper bound on the rotation gains at which people would perceive their rotation as unnatural. For higher gains, the largest real-world rotation which can be achieved with a static target placement is below 90 degrees in either direction. Users will inevitably hit a wall or leave the tracked space for gains greater than 2.0.

4.3 General Experimental Design

The overall experimental design was constant across the three experiments: Participants had the task to find and collect the highest of five pillars (Fig. 1). To ensure that the whole room was explored, the height of the highest target varied. It was necessary to look at every pillar at least once to decide whether the focused target is the highest one. We created a simple living room scenario to provide a VE that is too plain for high distraction, but complex enough to occur in real applications. The participants wore headphones and were listening to ambient music. This was done to prevent biases caused by the reflection of their own voices from the lab walls. During pretesting, we found that we could "hear" how far we were from walls in the lab. A mismatch between virtual and real orientation was easily detectable without headphones. The music was played directly as a stereo signal in the first condition. In the second condition, a 3D sound source in the form of a radio was placed in the virtual scene from which the music seemed to originate. The binaural audio signal was computed by using direction dependent amplitude modulation and the application of a head-related transfer function for increased realism. The voice of the examiner was broadcast via the headphones, too.

An experiment consisted of four trials in randomized order: Decreasing or increasing rotation gain combined with ordinary stereo sound or locatable 3D sound. Every trial started with a gain of 1.0 and on every target collection the gain was increased (resp. decreased) in steps of 0.033. If participants were about to hit a wall or to leave the tracked space, they were manually turned towards the middle of the room and the next target appeared there. Subjects were not informed beforehand that their rotation is going to be manipulated. They were informed, however, that they may get nauseous and kindly advised to quit the experiment whenever they would feel uncomfortable. Participants were allowed to take a break at any time and got offered sweets and water. All subjects participated voluntarily and without compensation. Participants with a medical history of epilepsy and pregnant women were excluded.

4.4 Experiment 1 (E1): Methodical Paradigm for Measuring the Threshold of Limited Immersion

The aim of experiment 1 is to develop an experimental paradigm to measure the TLI by assessing the degree of manipulation at which a subject's immersion breaks.

4.4.1 Method

To devise the experimental paradigm, we chose iterative improvement of the experimental instruction as a method. We instructed the first participants to collect the targets and to press a button on the VR headset in case the VE "feels strange or unnatural". After the experiment we conducted a semi-structured interview to examine in detail why the button was pressed and informed the participant about the rotational

Fig. 1. Exemplary view of the subject: Pillars of different heights, two being out of sight.

gain. Based on these results we adjusted the instruction every 4 subjects. $N = 16$ participants (41.2% female) took part in E1, aged from 19 to 26 ($M = 22.83$, $SD = 2.31$).

4.4.2 Results

Up to participant five, none of the subjects ever pressed the button. All trials with increasing gain except one ended at the limit of our experimental setup of 2.0 ($M = 1.97$, $SD = 0.05$), while the trials with decreasing gain ended, on average, at $M = 0.28$ ($SD = 0.05$). According to the post-experimental interviews, participants reported that they were puzzled by the rotary manipulation, but they didn't report it for several reasons: First, they were not able to phrase their feelings and did not find this covered by the instruction. Second, participants erroneously attributed this feeling to themselves, and third, they simply forgot that they should press the button in case of unnatural behavior of the VE.

As a consequence participants five to eight were instructed to collect the highest target while reporting any inconsistency of their motion verbally (i.e., think-aloud technique) and as soon as possible—instead of pressing the button. Furthermore we stressed that whatever might feel wrong is not due to them, but the system, and asked them to repeat the instruction after they put on the VR headset. Subjects reported inconsistencies, on average, at $M = 1.62$ ($SD = 0.36$) resp. $M = 0.64$ ($SD = 0.28$). Overall, three trials ended at 2.0. Still, most participants recalled only the target collection instruction, but not the instruction to report inconsistencies.

From participant nine on, we stopped to instruct on target collection and inserted the description of the virtual room as a warm-up task for the think-aloud instruction instead. Once subjects reported an initial pillar in the center, we asked them to collect targets and keep thinking aloud. Subjects nine to twelve reported inconsistencies, on average, on $M = 1.81$ ($SD = 0.43$, five trials ended at 2.0), resp. $M = 0.43$ ($SD = 0.13$), although they still tended to attribute those inconsistencies to themselves. Furthermore, participants reported some non-gain related inconsistencies like that the graphics quality was not realistic or that the room's furniture was not looking real.

As final adjustment to the instruction we added the information, that participants should focus on problems with the motion capture system. Explicitly, we did *not* instruct to focus on rotation, in order to avoid biases. Even if only one trial in increasing condition ended not at 2.0 ($M = 1.98$, $SD = 0.04$, resp. $M = 0.54$, $SD = 0.16$), participants in general kept in mind to report inconsistencies and stated retrospectively that they just were not recognizing the manipulation of gains greater than 1. This adaptive procedure led to the following final instruction that was found appropriate for the consecutive experiments:

- *You are going to see a testbed environment. We want this environment to be as natural and realistic as possible. Today we focus on motion capturing.*
  - *Please think aloud during the whole task: Describe what you are seeing, thinking, and feeling.*
  - *Report immediately when something feels strange or unnatural to you.*
  - *We know that the motion capturing is defect, please report as soon as possible when defects appear.*
- *In case you are not sure if your impression originates from you or the system, be sure, it is definitively the system, so please report as soon as possible.*
- *You can communicate with the examiner during the whole experiment. Please stop if the examiner tells you to stop.*
- *You may get nauseous during the experiment and you can quit at any point, these are important results for us, too.*

Overall, the concrete immersion limiting effect of the rotation gain was described as "too fast or too slow rotation", "the feeling to wobble", "the feeling of drunkenness" or, in conditions with decreasing gain, "the impression to be under water".

4.4.3 Discussion

The results indicate that the participants' response behavior is heavily biased, since subjects tend to withhold the perceived effects of the rotary manipulation. We infer that three things are to be considered when TLI is to be measured: First, the threshold to admit that something feels strange in front of the examiner might be higher than the actual indisposition caused by the rotation gain. This might seem contradictory because this effect occurred even stronger when subjects responded non-verbally via button press. Secondly, users are very inexperienced when dealing with VR: they are easily fascinated by the VE itself, even if it is quite modest, and furthermore too unfamiliar with the hardware to deal with the attached buttons. Thirdly, the perception of rotary manipulation leads to novel stimuli that could not be achieved with any other technologic system before. We conclude that our experimental paradigm should address these limitations. Participants should not respond to low-fidelity of graphics and they were encouraged to mention any inconsistency in motion perception as soon as possible. This was done to reassure participants that they should not attribute motion inconsistencies to themselves.

4.5 Experiment 2 (E2): Measuring the Threshold of Limited Immersion

The purpose of experiment 2 is to quantify the threshold of limited immersion, using our experimental paradigm from E1.

4.5.1 Method

Experiment E2 consisted of a pre-questionnaire, the VR task and a post-questionnaire. The VR task was conducted with the same instruction as described in Sect. 4.4.2, as the termination criterion of a trial we defined any reported inconsistency that was not related to the VE's graphics (e.g. "the shadows don't look natural") or explicitly a singular event (e.g. "there was a small delay, but it's gone now"). Dependent variables were the rotation gain at which participants reported inconsistencies, and the audio-source (binaural or omnidirectional). Because we could not find any significant influence of the different sound sources in the scene, we condensed the four conditions into two dependent variables. DEC (Cronbach's $\alpha = 0.67$) for decreasing gains going down from 1.0 to 0 and INC (Cronbach's $\alpha = 0.73$) for increasing gains going from 1.0 to 2.0.

To understand whether individual factors impact the thresholds for rotational gains, we measured:

- demographical data
  - age, gender, and education
  - pupil distance (PD) using PD-ruler
- personality factors
  - self-efficacy towards technology (SET) [3]
  - the ability of mental rotation using the paper folding test (VZ-2) [9]
- VR-related factors
  - the tendency to perceive immersion (ITQ) [46]
  - the experience regarding 3D/VR/AR-Technology (EXP)
  - the tolerance towards nausea-inducing activities such as reading while being driven by car (TOL)

We look at these factors, as there are reasonable causational links between differences in these measures and possible differences in cybersickness and immersion. Previous research has shown that women show stronger symptoms of cybersickness than men, possibly due their larger field of view. We test the participants' age, as the sensory sensitivity of the organs responsible for the sense of balance changes with age. Very young and very old people suffer more strongly from vertigo. We measure the immersion tendency as a natural confounding factor in our experiments, as we want to measure a break in immersion. Previous experience with 3D technology could cause a learning and adaptation process. Similarly, a general tolerance for nauseating experiences could affect what users report on, therefore we control these variables. Lastly, we wanted to assess mental rotation capability to control for the user's ability to imagine rotations of the room.

As evaluation criteria we quantified presence (PRE) [46] and immersion (IMM) [40] as VE-specific criteria, enjoyment (ENJ) and anxiety (ANX) as perceived emotions [23], and behavioral intention of UTAUT2 model as measure of technology adoption (ADO) [43]. In addition we operationalized trust in redirected walking (T-RDW) (e.g. *"I feared to touch real objects or walls (negative item)"*) and judgment on redirected walking (J-RDW) (e.g. *"It felt strange to move around in the virtual environment (negative item)"*) in scales of 4 items each. We define T-RDW as ability to move confidently and casually, and J-RDW as perception of the VE as realistic and free of inconsistencies. Moreover, we conducted the simulator sickness questionnaire [14] before (SSQ_PRE), directly after (SSQ_POST1) and 10 minutes after (SSQ_POST2) the VE task.

All variables were measured on a Likert-scale from 0 to 5; the SSQ was rated on a scale ranging from 0 to 3, the overall SSQ score ranges from 0 to 235.26.

35 participants, aged between 19 and 36 years ($M = 24.97$, $SD = 3.71$) took part in the experiment. 54% of the sample were female. During the experiment eight participants dropped out by choice because of discomfort, one participant was excluded beforehand because of a medical history of epilepsy. Nevertheless, 86% of all possible trials were finished according to our defined criteria.

4.5.2 Results

Table 2. Descriptive Statistics for TLI and Cybersickness

	α	min	max	M	SD
DEC	0.67	0.2	0.85	0.58	0.16
INC	0.73	1.3	2.0	1.85	0.2
SSQ_PRE	n/a	0	71.53	15.28	16.83
SSQ_POST1	n/a	2.47	222	48.92	45.92
SSQ_POST2	n/a	0	172.67	29.53	40.84

On average, participants reported limited immersion when in the decreasing condition at $M = 0.58$ ($SD = 0.16$) and in the increasing condition at $M = 1.85$ ($SD = 0.2$, see Table 3). Results exceeded our expectation to an extent that we had to end 58% of the trials for gains greater than 1 at a rotation gain of 2.0. This contradicts the previous assumption that the sensitivity to enlarged rotations is higher than to weakened ones and raises the assumption that the real TLI for gains greater than 1 may be even higher than 2.0. Remarkably, all participants reached their TLI for gains greater than 1 only if the detection threshold ($DT = 1.24$, see Fig. 2) was reached, whereas for gains smaller than 1 the TLI was in some cases even reached before crossing the established threshold ($DT = 0.67$, see Fig. 2).

Symptoms of cybersickness occurred, resulting in a simulator sickness score (SSQ) increase from an average of $M = 15.28$ ($SD = 16.83$) before the VR task to $M = 48.92$ ($SD = 45.92$) afterwards; symptoms cut back after 10 minutes ($M = 29.53$, $SD = 40.84$). This overall change in SSQ is significant by Friedmanns-ANOVA ($\chi^2(2) = 28.71$, $p < 0.01$), a post-hoc Wilkoxon-test with Bonferroni correction (significance level $p < 0.025$) yields that both changes SSQ_PRE to SSQ_POST1 ($T = 470, r = 0.52, p < 0.01$) and SSQ_POST1 to SSQ_POST2 ($T = 30, r = -0.53, p < 0.01$) were significant.

Table 3. Kendall-τ Correlations between TLI and Cybersickness

	DEC	INC	SSQ-PRE	SSQ-POST1	SSQ-POST2
DEC	/	$-.504^{**}$			
INC		/			
SSQ_PRE			/	$.288^{*}$	$.352^{**}$
SSQ_POST1				/	$.712^{**}$

$^{*}p < .05$, $^{**}p < .01$

We found a strong correlation between both thresholds DEC and INC (Kendall's $\tau = -0.504^{**}, p < 0.01$), but no correlation to any SSQ scores at any time of measurement; all SSQ scores were positively intercorrelated: SSQ_PRE is associated to SSQ_POST1 ($\tau = 0.288^{*}, p < 0.05$) and to SSQ_POST2 ($\tau = 0.352^{**}, p < 0.01$), and the correlation between SSQ_POST1 to SSQ_POST2 is even stronger ($\tau = 0.712^{**}, p < 0.01$) (Table 3). This means that higher cybersickness before the experiment is associated with higher cybersickness immediately after the experiment, and also with cybersickness 10 minutes after experiment. Or in other words, if you feel dizzy before VR, you feel dizzier afterwards, as well as after 10 minutes.

Table 4. Descriptive Statistics for Human Factors and Kendall-Tau Correlations with TLI

	α	M	SD	DEC	INC	SSQ-PRE	SSQ-POST1	SSQ-POST2
Age	n/a	24.97	3.71		$-.293^{*}$			
Gender	n/a	n/a	n/a				$.308^{*}$	$.322^{*}$
SET	0.79	4.25	0.84				$-.263^{*}$	$-.339^{*}$
ITQ	0.53	4.2	0.58			$-.415^{**}$		$-.298^{*}$
EXP	0.83	2.82	0.63					
TOL	0.73	3.86	1.15			$-.284^{*}$		
VZ-2	n/a	14.29	3.97					
PD	n/a	64.29	2.97					

$^{*}p < .05$, $^{**}p < .01$

Among all human factors measures, there was no significant correlation to any measured threshold of limited immersion (TLI) except one between age and INC ($\tau = -.293^{*}, p < 0.05$). However, a median split on age does not yield a significant difference between these groups ($Md = 22$ and $Md = 26$; $U = 91.5$, $p = 0.17$, *n.s.*)[1], nor is the age structure of our sample sufficient to analyze age effects. We assume this to be a statistical artifact.

[1]Mann-Whitney U Test requires Median (Md) report.

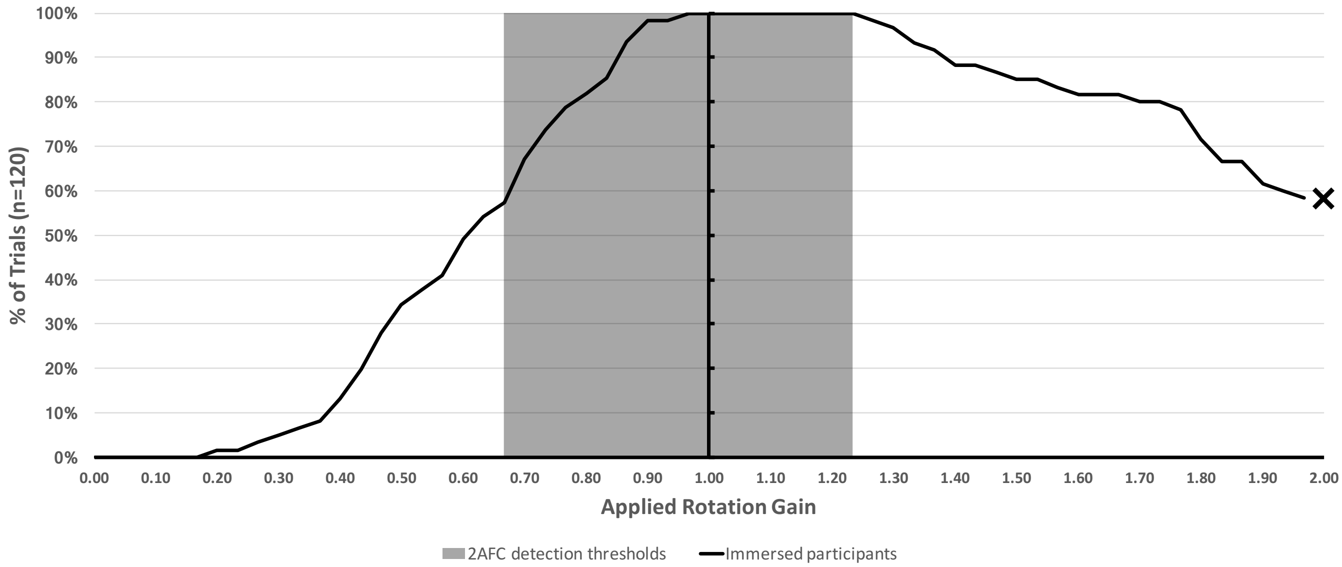

Fig. 2. Immersed sample of E2 and highlighted detection threshold according to [35]. Note: These metrics are not directly comparable and not measured under equivalent conditions. Detection threshold is a psychophysical measure, whereas we define limited immersion as a measure of user experience.

When considering cybersickness, correlations between gender (dummy-coded) and SSQ_POST1 ($\tau = .308^{*}, p < 0.05$) and also SSQ_POST2 ($\tau = .322^{*}, p < 0.05$), indicate that there might be a gender-effect. A Mann-Whitney-U-test indeed yields a significant difference between men and women regarding SSQ_POST1 ($U = 217^{*}, p < 0.05$) and SSQ_POST2 ($U = 195.5^{*}, p < 0.05$): Women ($Md = 49.33$ and $Md = 29.6$) were more susceptible to cybersickness than men ($Md = 22.2$ and $Md = 6.17$) in both times of measurement after the VR exposure. Cybersickness in these points in time furthermore correlates negatively with SET ($\tau = -.263^{*}, p < 0.05$ and $\tau = -.339^{*}, p < 0.05$), while ITQ has a negative association to SSQ_PRE ($\tau = -.415^{**}, p < 0.01$) and SSQ_POST2 ($\tau = -.298^{*}, p < 0.05$). This means that cybersickness occurs less in people that are confident using technology and cybersickness is associated with less immersion tendency.

Another negative correlation can be found between SSQ_PRE and TOL: Subjects who report to be more resistant to potentially nausea-inducing activities report weaker symptoms of cybersickness even before the VR exposure ($\tau = -.284^{*}, p < 0.05$).

Table 5. Descriptive Statistics for evaluation criteria and Kendall-Tau Correlations with TLI and SSQ

	α	M	SD	DEC	INC	SSQ-PRE	SSQ-POST1	SSQ-POST2
PRE	0.52	3.02	0.52				$-.294^{*}$	$-.263^{*}$
IMM	0.54	3.51	0.59				$-.375^{**}$	$-.311^{*}$
ENJ	0.73	2.24	1					
ANX	0.71	1.43	0.92				$.451^{**}$	$.457^{**}$
ADO	0.84	2.8	1.57					
T-RDW	0.68	3.24	1.03					
J-RDW	0.54	2.42	1.03				$-.249^{*}$	$-.300^{*}$

$^{*}p < .05$, $^{**}p < .01$

Whereas no significant correlation between a TLI and any evaluation criteria could be found, the structure of correlations between those criteria and cybersickness scores is quite telling: Subjects who experience stronger symptoms of cybersickness at both times of measurement after the exposure feel less present ($\tau = -.294, p < 0.05$ resp. $\tau = -.263, p < 0.05$) and less immersed ($\tau = -.375, p < 0.01$ resp. $\tau = -.311, p < 0.05$). Furthermore they rate the sense of walking worse ($\tau = -.249, p < 0.05$ resp. $\tau = -.300, p < 0.05$) and even feel more anxious in the VE ($\tau = .451, p < 0.01$ resp. $\tau = .457, p < 0.05$).

4.5.3 Discussion

The results show that the threshold of limited immersion is not equivalent to the detection threshold. Furthermore, there is no indication to assume that there is an inter-individual TLI. Being a truly subjective metric highlights the importance of taking human factors into consideration. According to the DT, users should be indisposed or disturbed by gains greater than 1 more easily because they are capable to detect them earlier than gains lower than 1. In contrast, most users exceeded our experimental setup and did not get disturbed even if their rotation was doubled. In line with the detection threshold, some users started to perceive the rotation gain as disturbing as soon as they were able to detect it. Interestingly, this was not the case for gains smaller than 1: Most users already broke out of immersion before the detection threshold was reached. Both findings are strong arguments for using limited immersion as an additional benchmark for redirected walking techniques.

The individual TLI is indeed diverse but not yet explainable by any human factor or evaluation criterion that we took into account. However, underlying human factors must exist because both TLIs are strongly associated; there seems to be an underlying, individual robustness to rotary manipulation.

The finding that neither evaluation criterion is associated with TLIs is another strong indicator for the potential of rotation gains or even redirected walking techniques in general: If the user is capable to tolerate them, they can be applied without deteriorating his/her perceived quality, presence, immersion, emotions, trust or judgment. However, while the amount of applied rotational gain does not affect those criteria, cybersickness does, and it is noteworthy that a considerable part of our sample dropped out by choice.

As expected, cybersickness did occur since our experiment was designed to let users explore the very limits of their applicable rotation gains. But even if the extent of rotation gains is not connected to the intensity of cybersickness, the question of how much cybersickness is tolerable gets even more important. Our users stated: “None!”. Due to the result that women were more susceptible to symptoms of cybersickness than men, it is crucial to take user diversity into account. As women and men differ in both visual perception and spatial cognition, it could be possible to lower cybersickness symptoms via technical adoptions e.g. in the displayed field of view or the intensity of colors. However, it is still unknown to what extent those symptoms are caused by the plain application of rotation gains, since cybersickness occurs even in VEs without redirected walking.

4.6 Experiment 3 (E3): Baseline for Cybersickness in our Virtual Environment

As cybersickness turned out to be crucial for the application of redirected walking techniques, the aim of E3 is to set a baseline of cybersickness for our VE to classify the results of E2.

4.6.1 Method

We conducted a control group study with the exact same instruction as in E2, but in contrast to the previous experiment the rotation gain remained at 1.0. Participants collected as much targets as the average participant in E2. To mimic E2 as accurate as possible we terminated the trials manually after two conditions with 16 and two conditions with 26 targets in randomized order. 10 participants took part in the third experiment, 2 of them were female and the sample was aged between 19 and 27 ($M = 23.4$, $SD = 3.2$). No participant dropped out of the study, but one outlier was excluded because of outstandingly intense symptoms of cybersickness even before the experiment (SSQ scores 74.8, 86.02 and 74.8).

4.6.2 Results

Table 6. Descriptive Statistics for TLI and Cybersickness

	min	max	M	SD
SSQ_PRE	0	26.18	7.48	9.16
SSQ_POST1	0	33.66	8.31	12.51
SSQ_POST2	0	22.44	6.23	8.15

Within the control group, participants reported symptoms of cybersickness to an average intensity of $M = 7.48$ ($SD = 9.16$) before, $M = 8.31$ ($SD = 12.51$) right after, and $M = 6.23$ ($SD = 8.15$) 10 minutes after the VR exposure (Table 6). The scores did not change significantly ($\chi^2(2) = 0.73, p = 0.69, n.s.$) among the times of measurement (see Fig. 3).

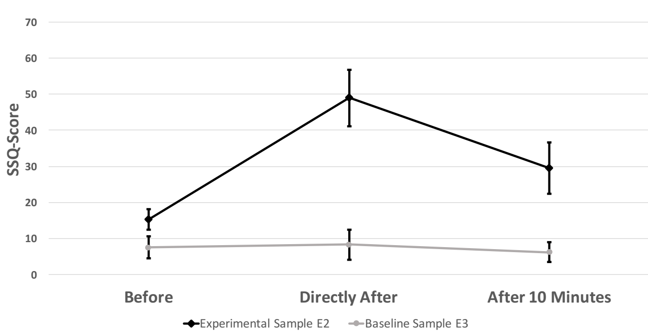

Fig. 3. Average SSQ scores among experimental groups. Error bars denote the standard error of the mean. The scale ranges from 0 to 235.62, but was shortened for reasons of legibility.

Compared to the experimental group of E2, cybersickness scores directly after ($Md = 3.74$ and $MD = 32.06$) and 10 minutes after ($Md = 3.74$ and $Md = 14.8$) the VR task were significantly lower ($U = 44, p < 0.01$ resp. $U = 80, p < 0.05$), whereas the SSQ scores did not differ significantly from the scores before the experiment ($Md = 3.74$ and $Md = 14.8$; $U = 103, p = 0.141, n.s.$).

4.6.3 Discussion

Results are indicating that indeed the plain use of rotation gain and not its extent causes the increase of cybersickness symptoms in E2. Other factors that would have been in line with theory (e.g. duration in VR or weight of the headset) do not seem to have an impact on cybersickness in our design at all, since SSQ scores did not change during the times of measurement.

5 General Discussion

In the following we discuss the results of our study. We summarize the key insights, point out some limitations of our approach, and present a number of open questions and necessary future research.

5.1 Summary of Key Results

We developed an experimental method to assess the TLI by adapting how a break of immersion is reported and successively improving the given instructions. With the final procedure, subjects reported the break in immersion as soon as they noticed it, which we validated by applying the method and interviewing participants about their reporting behavior after the experiment.

We measured the TLI for slowly decreasing and increasing rotation gains during search and collect task in a static scene with low complexity. For decreasing gains, the measured TLI of 0.58, is not significantly different from the established DT ($g_{R[u]} = 0.67$). However, 43% of participants reported a break in immersion before the DT was reached. For increasing gains greater than 1, the measured TLI (1.85) is significantly higher than the established DT ($g_{R[u]} = 1.24$). Due to limitations of our test scenario we were not able to measure gains larger than 2.0. Yet, 58.3% of trials ended at a gain of 2.0, indicating that the actual TLI might be even higher. This asymmetry is in line with verbal remarks of subjects during and after the experiments, which stated that gains greater than 1 were perceived as more pleasant than gains lesser than 1. These results corroborate that the TLI quantifies an additional factor in the perception of motion manipulations. While we could examine that the TLI differs substantially from the DT, the causes for inter-individual differences of the TLI are not yet discovered nor predictable. The variance can not be explained by any of the measured human factors.

One must note that DT and TLI are not directly comparable. One measures psychophysical detection in an isolated setting, the other measures a break in immersion, a user experience measure, in a complex setting. Both measures are complementary to each other, making it even more astonishing, that the thresholds differ asymmetrically.

We could show that, in our experimental setup, the high levels of cybersickness that were experienced by subjects (pre = 15.28; post = 45.92) were primarily caused by the applied rotation gain. In a control experiment with a constant gain of 1.0 we measured no significant symptoms of cybersickness (pre = 7.48; post = 8.31). In line with previous findings, female participants reported higher sickness levels than did male participants. This underlines the importance of consideration of user diversity for the effective application of redirection methods.

5.2 Limitations

The empirical research approach provided valuable insights into the thresholds of limited immersion for rotation gains in redirected walking. Still, there are several context-related and methodological considerations that should be critically taken into account.

Our current approach to measuring the break of immersion has several drawbacks which we were able to identify. We do not claim that the developed method is final, but rather aim at continuously improving the experimental method in further studies, alongside its application to measure factors that affect the TLI.

The fact that we ended more than half of the trials at a gain of 2.0 is unsatisfying. Although we were able to show that there is a significant difference between the DT and the the TLI, the measured value is strongly biased by this cutoff.

Our choice of slowly decreasing and increasing rotational gains as the exemplary motion manipulation does not allow for a direct comparison with the established detection thresholds, which used randomized rotation gains within a given interval ($0.5 \leq g_{R[u]} \leq 1.5$) [35]. We chose continuously varying gains because we expected the TLI to be higher than the DT, and the range of possible values was not known beforehand. We used slowly varying gains to find a reliable upper bound on the TLI, assuming that the slow change will cause some degree of habituation.

Our proposed experimental method for assessing the TLI is arguably time-consuming and costly. The think-aloud protocol requires much attention of the examiner, compared to fully automated measurements where subjects report by pressing a button after being exposed to the stimulus. Likewise, there is a certain examiner bias introduced by the procedure. Comments given by the subjects need to be judged, and there is some amount of ambiguity as to whether a subject refers to the manipulated motion or other properties of the system. We found that this can be easily resolved in the majority of cases by asking subjects to rephrase the disturbing effect they observe. In the remaining cases we took a note whenever an ambiguous comment was made, and asked the subject at a later point, when the manipulation was more pronounced, if their earlier comment related to the same effect.

Given the proposed instructions, subjects are not entirely uninformed about the manipulation. Although we do not make them aware of the specific manipulation, we found it necessary to instruct that some aspect of the system behavior might feel unnatural. Otherwise, subjects failed to attribute the effect to the VE at all.

One could argue that the TLI is not a dependable metric since it varies with a number of confounded factors such as the scene and task complexity, user attention, and expectation of mismatches, which can not be reliably quantified in itself, yet. While this also holds true for the DT as previously shown [18], it is the very idea of the TLI to incorporate all relevant influences in an application scenario and describe their summative effect. A metric which depends on the influencing factors of interest is the prerequisite for exploring the parameter space and developing predictive models for the break of immersion as a function of human factors and the virtual scenario properties.

Finally, another limitation refers to the fact that even though a considerable number of known individual factors impacting VR performance have been assessed, neither of them was able to explain the user diversity—which definitively was present. Apparently, our sample was still too homogeneous in the respective measurements. In order to understand the nature of the user diversity, future experiments should contrast more extreme user groups, e.g. a group of highly trained persons with VR environments in contrast to a group that has nearly no experience with VR exposure. This would possibly help to maximize the differences and the underlying human factors.

5.3 Future Work

To accurately measure the TLI for slowly varying rotation gains we need to devise a different testing scenario. Static target placement is inherently limited to gains below 2.0, and we have seen that many subjects exceeded this threshold. If the upper bound on rotation gains has been established in this way, we need to perform the same study but with randomized gains. Apart from making the TLI more comparable to previous studies, this will enable us to isolate the expected habituation effects. Similarly, we need to apply the same procedure to translation, curvature and time-dependent gains as well as combinations thereof. The latter is particularly important, since an effective redirection method will combine the different types of gains. For some methods, such as planar maps, a separation of the different types of gains is not even achievable since they are implicitly defined by the mapping. Apart from looking at absolute gain values, we need to investigate the dynamics of changing gains and its impact on the TLI. Randomized gains and continuously varying gains are the two extreme ends of this spectrum. Between abrupt changes of arbitrary magnitude and continuous changes that are as small as possible, there is a whole design space of possible first and higher order derivatives that might influence the TLI.

A quantitative evaluation of our captured head and torso tracking data needs to be performed. We want to investigate whether the different frames of reference in which a rotation can take place (legs, torso, head, eyes) have an effect on the TLI or cybersickness levels. If a user rotates the whole body instead of only the head, the subjective degradation of immersion might differ significantly.

In the long term, there is a vast area of psychological research on the applicability of redirection techniques. After identifying the relevant factors that influence the TLI, we need to find reliable metrics that quantify them. Task and scene complexity, user attention, motion patterns, and other subjective measurements which need to be reliably operationalized in future studies. Identifying relevant factors and defining the right metrics is, however, a prerequisite for the development of predictive models that also generalize between different virtual environments. Exploring this space of interrelated technical and psychological factors is a huge challenge that offers many research opportunities. Another line of research might be directed to the reliabale profiling of user groups. Alongside the parameters threshold of limited immersion, detection threshold and proneness to cybersickness user profiles could be generated and validated. In a second step those profiles could be utilized to individually tailor VR environments.

Besides the TLI, other quantities for estimating the applicability of a redirection technique need to be considered. Apart from quantifying the effectiveness of the method, i.e. the degree of spatial compression, it is crucial to establish a threshold at which cybersickness occurs. Other metrics are similarly important, such as the navigational ability and the transfer of spatial knowledge from the virtual to the real world. When considering this variety of metrics together with their influencing factors, it will become possible to combine redirection methods in an optimal way: yielding maximal spatial compression in arbitrary virtual scenes, while not causing cybersickness or a degraded user experience.

6 Conclusion

In this article we proposed a new metric for the evaluation of spatial compression methods: the threshold of limited immersion (TLI). We argue that established psychophysical detection thresholds are too conservative to be directly applicable in practical applications. Instead of measuring when a manipulation is detectable, the TLI estimates when the user's immersion breaks in a real application scenario.

To measure the TLI we devised an experimental paradigm that makes users report the subjective break in immersion without being explicitly informed about the applied manipulation. To achieve this, we refined the experimental procedure until the reported break of immersion matched the answers in a post-experimental interview.

We conducted three user studies in this context. The first study comprised the systematic development of the experimental procedure. In a second study, we applied the method to measure the TLI for slowly decreasing and increasing rotation gains. In a final study with a control group we isolated the effect of the applied motion manipulations on cybersickness. Our findings contribute to spatial compression by proposing an additional quality of user experience. This helps understanding the limitations of effective spatial compression with respect to user diversity and enables larger virtual environments to be explored in limited spaces.

Acknowledgments

Authors would like to thank Sarah Hilker, Luca Liener, Christian Mattes, Kaspar Scharf, Kevin Wegener as well as Pascal Fries for their help in the experimental work. Also, thanks to the participants for their willingness to spend their time in novel VR environments. This work has been funded by the strategy funds of the RWTH Aachen University within the project house "ICT Foundations of a Digitized Industry, Economy, and Society".

References

[1] L. L. Arns and M. M. Cerney. The relationship between age and incidence of cybersickness among immersive environment users. In *Virtual Reality Conference (VR)*, pages 267–268. IEEE, 2005.

[2] M. Azmandian, R. Yahata, M. Bolas, and E. Suma. An enhanced steering algorithm for redirected walking in virtual environments. In *Virtual Reality Conference (VR)*, pages 65–66. IEEE, 2014.

[3] G. Beier. Kontrollüberzeugungen im Umgang mit Technik: Ein Persönlichkeitsmerkmal mit Relevanz für die Gestaltung technischer Systeme. *Report Psychologie*, 9:684–693, 1999.

[4] F. Biocca. Will simulation sickness slow down the diffusion of virtual environment technology? *Presence: Teleoperators & Virtual Environments*, 1(3):334–343, 1992.

[5] B. Bolte and M. Lappe. Subliminal reorientation and repositioning in immersive virtual environments using saccadic suppression. *IEEE Trans. Vis. & Comp. Graphics (TVCG)*, 21(4):545–552, 2015.

[6] K.-E. Bystrom, W. Barfield, and C. Hendrix. A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 8(2):241–244, 1999.

[7] H. Chen and H. Fuchs. Supporting free walking in a large virtual environment: imperceptible redirected walking with an immersive distractor. In *Proceedings of the Computer Graphics International Conference*, page 22. ACM, 2017.

[8] J. V. Draper, D. B. Kaber, and J. M. Usher. Telepresence. *Human factors*, 40(3):354–375, 1998.

[9] R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen. Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational testing service*, 1976.

[10] T. Field and P. Vamplew. Generalised algorithms for redirected walking in virtual environments. In *AISAT2004 International Conference on Artificial Intelligence in Science and Technology*, pages 58–63.

[11] T. Grechkin, J. Thomas, M. Azmandian, M. Bolas, and E. Suma. Revisiting detection thresholds for redirected walking: Combining translation and curvature gains. In *Proceedings of the ACM Symposium on Applied Perception*, pages 113–120. ACM, 2016.

[12] E. Hodgson and E. Bachmann. Comparing four approaches to generalized redirected walking: Simulation and live user data. *IEEE Trans. Vis. & Comp. Graphics (TVCG)*, 19(4):634–643, 2013.

[13] C. Hutton and E. Suma. A realistic walking model for enhancing redirection in virtual reality. In *Virtual Reality Conference (VR)*, pages 183–184. IEEE, 2016.

[14] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[15] E. Langbehn, G. Bruder, and F. Steinicke. Subliminal reorientation and repositioning in virtual reality during eye blinks. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, pages 213–213. ACM, 2016.

[16] J. J. LaViola, Jr. A discussion of cybersickness in virtual environments. *SIGCHI Bulletin*, 32(1):47–56, Jan. 2000.

[17] M. Mousavi, Y. H. Jen, and S. N. B. Musa. A review on cybersickness and usability in virtual environments. In *Advanced Engineering Forum*, volume 10, pages 34–39. Trans Tech Publications Ltd., 2013.

[18] C. T. Neth, J. L. Souman, D. Engel, U. Kloos, H. H. Bülthoff, and B. J. Mohler. Velocity-dependent dynamic curvature gain for redirected walking. *IEEE Trans. Vis. & Comp. Graphics (TVCG)*, 18(7):1041–1052, 2012.

[19] N. C. Nilsson, R. Nordahl, and S. Serafin. Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, 12, 2016.

[20] N. C. Nilsson, E. Suma, R. Nordahl, M. Bolas, and S. Serafin. Estimation of detection thresholds for audiovisual rotation gains. In *Virtual Reality Conference (VR)*, pages 241–242. IEEE, 2016.

[21] T. C. Peck, H. Fuchs, and M. C. Whitton. Evaluation of reorientation techniques and distractors for walking in large virtual environments. *IEEE Trans. Vis. & Comp. Graphics (TVCG)*, 15(3):383–394, 2009.

[22] T. C. Peck, H. Fuchs, and M. C. Whitton. Improved redirection with distractors: A large-scale-real-walking locomotion interface and its effect on navigation in virtual environments. In *Virtual Reality Conference (VR)*, pages 35–38. IEEE, 2010.

[23] R. Pekrun, T. Goetz, A. C. Frenzel, P. Barchfeld, and R. P. Perry. Measuring emotions in students learning and performance: The achievement emotions questionnaire (aeq). *Contemporary educational psychology*, 36(1):36–48, 2011.

[24] T. M. Porcino, E. Clua, D. Trevisan, C. N. Vasconcelos, and L. Valente. Minimizing cyber sickness in head mounted display systems: design guidelines and applications. In *Serious Games and Applications for Health (SeGAH), 2017 IEEE 5th International Conference on*, pages 1–6. IEEE, 2017.

[25] S. Razzaque, Z. Kohn, and M. C. Whitton. Redirected walking. In *Proceedings of EUROGRAPHICS*, volume 9, pages 105–106. Manchester, UK, Eurographics Association, 2001.

[26] J. T. Reason and J. J. Brand. *Motion sickness.* Academic press, 1975.

[27] G. E. Riccio and T. A. Stoffregen. An ecological theory of motion sickness and postural instability. *Ecological psychology*, 3(3):195–240, 1991.

[28] J. Schild, J. LaViola, and M. Masuch. Understanding user experience in stereoscopic 3d games. In *Proc. ACM Conf. Human Factors in Computing Systems (CHI)*, pages 89–98. ACM, 2012.

[29] S. Serafin, N. C. Nilsson, E. Sikstrom, A. De Goetzen, and R. Nordahl. Estimation of detection thresholds for acoustic based redirected walking techniques. In *Virtual Reality Conference (VR)*, pages 161–162. IEEE, 2013.

[30] M. Slater. A note on presence terminology. *Presence connect*, 3(3):1–5, 2003.

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

[32] K. Stanney. Realizing the full potential of virtual reality: human factors issues that could stand in the way. In *Virtual Reality Annual International Symposium, 1995. Proceedings.*, pages 28–34. IEEE, 1995.

[33] K. M. Stanney, K. S. Hale, I. Nahmens, and R. S. Kennedy. What to expect from immersive virtual environment exposure: Influences of gender, body mass index, and past experience. *Human Factors*, 45(3):504–520, 2003.

[34] F. Steinicke and G. Bruder. Using perceptual illusions for redirected walking. *IEEE Computer Graphics and Applications*, 33(1):6–11, 2013.

[35] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe. Estimation of detection thresholds for redirected walking techniques. *IEEE Trans. Vis. & Comp. Graphics (TVCG)*, 16(1):17–27, 2010.

[36] E. A. Suma, G. Bruder, F. Steinicke, D. M. Krum, and M. Bolas. A taxonomy for deploying redirection techniques in immersive virtual environments. In *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*, pages 43–46. IEEE, 2012.

[37] E. A. Suma, S. Clark, D. Krum, S. Finkelstein, M. Bolas, and Z. Warte. Leveraging change blindness for redirection in virtual environments. In *Virtual Reality Conference (VR)*, pages 159–166. IEEE, 2011.

[38] E. A. Suma, Z. Lipps, S. Finkelstein, D. M. Krum, and M. Bolas. Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture. *IEEE Trans. Vis. & Comp. Graphics (TVCG)*, 18(4):555–564, 2012.

[39] Q. Sun, L.-Y. Wei, and A. Kaufman. Mapping virtual and physical reality. *ACM Transactions on Graphics (TOG)*, 35(4):64, 2016.

[40] K. Tcha-Tokey, O. Christmann, E. Loup-Escande, and S. Richir. Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *The International Journal of Virtual Reality*, 16(1):33–48, 2016.

[41] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr. Walking > walking-in-place > flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 359–364. ACM Press/Addison-Wesley Publishing Co., 1999.

[42] K. Vasylevska, H. Kaufmann, M. Bolas, and E. A. Suma. Flexible spaces: Dynamic layout generation for infinite walking in virtual environments. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 39–42. IEEE, 2013.

[43] V. Venkatesh, J. Y. Thong, and X. Xu. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1):157–178, 2012.

[44] S. von Mammen, A. Knote, and S. Edenhofer. Cyber sick but still having fun. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pages 325–326. ACM, 2016.

[45] J. Wann and M. Mon-Williams. What does virtual reality need?: human factors issues in the design of three-dimensional computer environments. *International Journal of Human-Computer Studies*, 44(6):829–847, 1996.

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

[47] S.-n. Yang, T. Schlieski, B. Selmins, S. C. Cooper, R. A. Doherty, P. J. Corriveau, and J. E. Sheedy. Stereoscopic viewing and reported perceived immersion and symptoms. *Optometry and vision science*, 89(7):1068–1080, 2012.

[48] R. Zhang and S. A. Kuhl. Flexible and general redirected walking for head-mounted displays. In *Virtual Reality Conference (VR)*, pages 127–128. IEEE, 2013.