Pushing Yourself to the Limit – Influence of Emotional Virtual Environment Design on Physical Training in VR

SEBASTIAN OBERDÖRFER, SOPHIA C. STEINHAEUSSER, AMIIN NAJJAR, CLEMENS TÜMMERS, and MARC ERICH LATOSCHIK, University of Würzburg, Germany

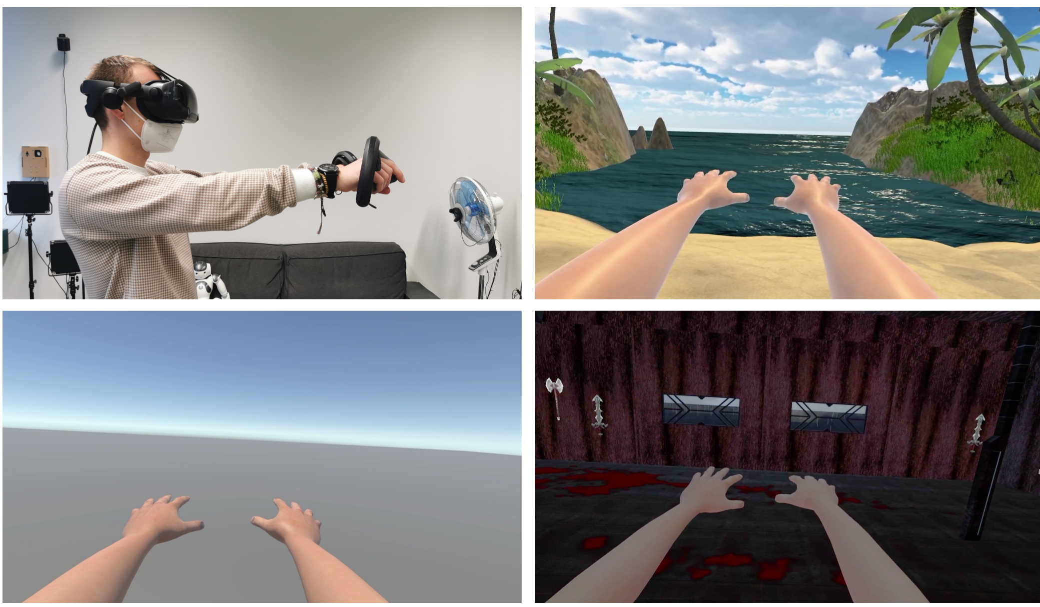


Fig. 1. The participants were immersed in three different VEs in counterbalanced order and performed a strength-endurance exercise.

The design of virtual environments (VEs) can strongly influence users' emotions. These VEs are also an important aspect of immersive Virtual Reality (VR) exergames – training system that can inspire athletes to train in a highly motivated way and achieve a higher training intensity. VR-based training and rehabilitation systems can increase a user's motivation to train and to repeat physical exercises. The surrounding VE can potentially predominantly influence users' motivation and hence potentially even physical performance. Besides providing potentially motivating environments, physical training can be enhanced by gamification. However, it is unclear whether the surrounding VE of a VR-based physical training system influences the effectiveness of gamification. We investigate whether an emotional positive or emotional negative design influences the sport performance and interacts with the positive effects of gamification. In a user study, we immerse participants in VEs following either an emotional positive, neutral, or negative design and measure the

Authors' Contact Information: Sebastian Oberdörfer; Sophia C. Steinhaeusser; Amiin Najjar; Clemens Tümmers; Marc Erich Latoschik, University of Würzburg, Germany.

ACM 1557-735X/2024/8-ART111
https://doi.org/XXXXXXX.XXXXXXX

duration the participants can hold a static strength-endurance exercise. The study targeted the investigation of the effects of 1) emotional VE design as well as the 2) presence and absence of gamification. We did not observe significant differences in the performance of the participants independent of the conditions of VE design or gamification. Gamification caused a dominating effect on emotion and motivation over the emotional design of the VEs, thus indicating an overall positive impact. The emotional design influenced the participants' intrinsic motivation but caused mixed results with respect to emotion. Overall, our results indicate the importance of using gamification, support the commonly used emotional positive VEs for physical training, but further indicate that the design space could also include other directions of VE design.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Virtual reality**.

Additional Key Words and Phrases: virtual environments, exergame, virtual reality, emotions

ACM Reference Format:
Sebastian Oberdörfer, Sophia C. Steinhaeusser, Amiin Najjar, Clemens Tümmers, and Marc Erich Latoschik. 2024. Pushing Yourself to the Limit – Influence of Emotional Virtual Environment Design on Physical Training in VR. *J. ACM* 37, 4, Article 111 (August 2024), 26 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Exergames are a special form of a serious game. Instead of fulfilling educative purposes, exergames inspire users to physically train in a highly motivated way [32]. When experienced in immersive Virtual Reality (VR), exergames put their users into surrounding virtual environments (VEs) potentially influencing the overall training motivation and performance [1]. In a similar way, VR enables patients to perform physical rehabilitation exercises in different environments than in the therapist's gym, thus achieving a higher motivation [11, 52]. Finally, VR allows elderly people who are restricted in their mobility to experience beautiful places [9]. This breaks the monotony of their daily lives and can even bring back memories.

Typically, VR physical training and rehabilitation applications use natural or sci-fi VEs [1, 6, 9, 11], thus achieving a relaxing and joyful experience. This stands in contrast with some gym interior designs that feature raw elements like grid fences, graffitis, and imbalanced light. Such a design can similarly improve an athlete's motivation and performance by evoking a rather tough and unforgiving atmosphere. Hence, the design of VEs for physical training is only partially explored. As the design of VEs is unlimited and can simulate any environment, it is important to investigate the effects of emotional positive and negative design on physical performance and motivation. Since the motivation of athletes can further be influenced by providing gamification [4], evaluating a potential interacting aspect of VE design and gamification effectiveness is similarly important.

To close this research gap, research needs to follow design recommendations for emotional positive and negative VE design [45] to create VEs of the respective effect. These VEs can be used to investigate a potential influence on an athlete's performance as well as motivation and an interaction with gamification. This could result in design recommendations for physical training applications and researchers in the area of sport.

Contribution

In this paper, we present the design of two VEs that distinctly follow an either emotional positive or negative design as displayed in Figure 1. The emotional positive VE represents a sunny beach with palm-trees and lush vegetation. The emotional negative VE is a dark and reddish arena featuring melee weapons as well as a thunderstorm. In a novel user study, we investigate the effects of our emotionally designed VEs as well as gamification on athletic performance and motivation within a simple static strength-endurance exercise. A growing plant gamification metaphor symbolizes the progress of the athletes and should provide an incentive to continue with the exercise. In addition, athletes receive points over the course of a training session. Our results show that the VEs' design did not influence the duration the exercise could be performed even though participants liked the emotional positive VE most. Similarly, we did not observe an influence of gamification with respect to performance. However, gamification caused a dominating effect on emotion and motivation over the emotional design of the VEs, thus indicating an overall positive impact. The emotional design influenced the participants intrinsic motivation but caused mixed results with respect to emotion during the exercise. This potentially can be an interesting insight for developers and researchers. It not only supports the typical emotional positive VEs, but could also indicate that the design space for VEs in VR-based training can go beyond this approach.

2 Theoretical Background

Immersive VR allows users to enter any VE and to perform any interaction supported within the boundaries of these artificial spaces. Thus, users can train in every imaginable VE. Immersion is "*the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant*" [44]. Immersion depends on a system's objective properties reducing real world sensory inputs and replacing them with computer-generated information. This, for instance, can be achieved by wearing a Head-Mounted Display (HMD). The objective characteristics further include possible actions of a user within a given system [41]. Immersion typically increases a user's visual angle on the VE in comparison to using a regular computer screen. With a higher visual angle, the *emotional responses* to audiovisual stimuli can be *increased* [7]. Immersion directly evokes and influences presence [43, 50]. *Presence* describes the subjective illusion of being in a real place despite being physically located in a different place [41]. Thus, presence indicates the perceived realness of a virtual experience [40]. Maintaining presence requires a support of sensorimotor contingencies [41] and a continuous stream of stimuli and experience [54]. Achieving such a high acceptance of the virtual experience can lead to the induction of higher emotions [34]. In addition, a higher emotional intensity of a predominant emotion in a VE can increase presence [20]. This can be of high importance when the goal of the VR simulation targets the improvement of a person's well-being by providing relief from emotional strain in simulated natural VEs [26]. At the same time, VR might also be ideal for experiencing emotional challenge [33] like when playing a horror game or being challenged to decide over the fate of companions.

2.1 Physical Training in VR

VR successfully has been used to realize training in various sport disciplines leading to better core aspects such as performance, decision making and complex motor tasks [5]. For instance, rowing on a rowing ergometer in VR experiencing a large virtual lake surrounded by mountains leads to a better performance and experience than without [1]. Similarly, a VR-exergame can lead to a higher enjoyment of High Intensity Interval Training (HIIT) workouts, a higher intensity when providing a competition against one's previous performance, and constant enjoyment even when the resistance of the ergometer is increased [6]. VR-based training further has the advantages of being independent of weather conditions, providing competitions with others all across the world, and full control over features of the VE [30]. Virtually competing against other athletes is perceived very motivating even when only Desktop-3D is used [31]. Another benefit of VR can be implicit learning enabling athletes to observe and learn correct execution of exercises of virtual agents [42]. In this way, VR shows a high potential to develop otherwise hard to obtain skills [5]. For instance, VR can enable effective training for racket sports [27].

A VR sport and training environment consists of the VR environment, sport exercise, athlete, and non-VR environment [30]. Besides setting the stage for VR-based training, the design of VEs can influence the felt emotions during exposure [20]. Since training and rehabilitation VR systems commonly use beautiful and overall positive and inspiring environments, potential benefits of negative and tough environments are rather unexplored. In fear-inducing VEs, an emotional detachment effect caused by sensation seeking of recipients can occur [20] and, while generally evoking negative emotions, even yield a certain degree of excitement and enjoyment. A recent publication discusses the influence of VE design on the emotional perception of VEs and presents design recommendations for emotional positive and negative VE [45]. Emotional positive VEs predominantly feature natural aspects such as lush vegetation, access to water, sunshine, and overall vast, colorful and open spaces. In contrast, emotional negative VEs are designed to confine a user in dark places that feature unbalanced light, rough and dirty elements, and even signs of past violence. Following their recommendation, joyful and fear inducing VEs can be designed and used for the purpose of investigating the influence of VE design on the performance of VR-based training applications. Since emotions not only are connected to athletic performance [35] based on the Individual Zones of Optimal Functioning model [12], but also predict changes in motivation regulations in the context of sports [38], this could result in design recommendations for VR-based training environments to maximize an athlete's performance.

2.2 Gamification

Besides being an exergame [32], a VR sport and physical training environment can further be enhanced by gamification. Gamification refers to the use of game elements in non-gaming contexts to increase the motivation for an extended use [4]. Gamification is not only used in VR-based fitness training [47] to increase motivation, healthy habits as well as physical performance [3], but also in various contexts such as education, health and wellbeing, and marketing [39]. When integrating gamification, it is important to design it to the target context as a gamification approach in one domain might not necessarily work in a different domain [39]. Tuveri at al. [47] used the game mechanics of levels, points, challenges as well as badges and prizes. They found that challenges by providing concrete tasks are the most relevant game mechanic as it motivated and informed the athlete about the own progress at the same time. Similarly, points were an integral element to affect fun and motivation during an exercise. While levels helped users to follow a progress plan, badges seemed only useful in a social environment and prizes indicated to have no effect. Since our approach targets the investigation of emotional VE design, only elements that motivate during and not between the training sessions can be tested. Thus, challenges and points should be integrated in our study.

3 System Design

Following the proposed design guidelines of Steinhaeusser et al. [45], we developed an emotional positive and an emotional negative environment. We refer to the fulfilled individual guidelines list in Table 1 by using GL_{pos} for positive and GL_{neg} for negative guidelines. To compare their effects to a baseline, we also developed a neutral VE. While ruling out potential effects of the real environment, the neutral VE also ensured that participants used VR devices and hence we avoided confounds of our measurements. Figure 1 provides an overview of our three VEs. In addition, we also developed a gamification approach. This allows for an investigation whether the emotional direction of the VEs affects the effectiveness of gamification elements or vice-versa. Figure 2 and Figure 3 provide an overview of the gamification.

3.1 Virtual Environment Design

The emotional positive VE immerses a user in a sunny beach (GL_{pos}4), thus fully omitting man-made elements (GL_{pos}2,3,5 not applicable). The VE features palm trees, lush vegetation, sunshine (GL_{pos}11), blue sky with scattered clouds, and blue water. The overall structure resembles a safe

Table 1. Overview of the design guidelines for emotional positive and negative VEs as proposed by Steinhaeusser et al. [45]. The order is adjusted to indicate potential contrasts between the two extremes.

	GL	Positive	GL	Negative
Design theme				
	1	Avoid restricting a player's movement.	1	Highlight boundaries of the VE to cause feelings of confinement.
	2	Adjust the general architecture of a settlement to its size.		
	3	Combine urban environments with natural spots to create a positive atmosphere.		
	4	Immerse the player in natural environments.	3	Use motion and alternation where it is not suspected.
	5	In man-made environments, add greenery and natural objects.		
	6	Use rounded shapes to create aesthetically pleasing VEs.		
	7	Design VEs conveying safety.	2	Use signs of passed violent acts to evoke fear.
Colors & Textures				
	8	Use bright or pastel colors to create aesthetically pleasing VEs.	4	Predominantly use dark colors in the environment.
	9	Yellow induces joy, whereas green evokes relaxation and comfort.	5	Reddish highlights should be added.
	10	Use soft textures with high fidelity.	6	Apply smooth textures together with rust and filth.
Lighting				
	11	Create illuminated daytime scenes by predominantly using sunlight.	8	Implement low-key lighting using soft blue, grey, or red colored light depending on the time of day.
	12	Use warm light tones and balance them.	10	Use unbalanced light temperatures.
			7	Adjust illumination to the currently induced emotional state with vanishing light intensifying the negative emotions.
			9	Warm light adds tension while cold light reduces pleasantness.
			11	Offer players interactive or dynamic light to navigate through the VE.
			12	Apply weather influences to limit the players' view.

holiday paradise ($GL_{pos}7$) that allows a visitor to freely explore a tropical island ($GL_{pos}1$). While aiming at a natural appearance, all objects and the overall landscape are rather rounded shapes with soft textures ($GL_{pos}10$) to create an aesthetically pleasing environment ($GL_{pos}6$). The predominant colors are yellow and green ($GL_{pos}9$) in bright as well as pastel shades ($GL_{pos}8$). The virtual sunshine creates a warm and balanced light ($GL_{pos}12$). The trees as well as other plants are animated as if they move in a gentle breeze. To avoid confounds caused by the integration of sound effects and music, we only used sounds directly afforded by the VE. In the beach VE, we added the sound of waves rolling ashore and lowered the volume to a barely audible minimum.

The emotional negative VE immerses a user in a dark arena-like prison cell (GL_{neg}1,4). The room only features narrow windows blocked by iron bars, thus not only targeting a feeling of confinement, but also resulting in a low-key lighting (GL_{neg}8). The walls and floor have a concrete and dirty texture (GL_{neg}6). While the walls feature flickering red lamps (GL_{neg}3,5,8,9) as well as large melee weapons like axes and hammers, the floor shows acts of passed violence in the form of blood stains (GL_{neg}2). Looking at the ceiling, a user can see a cloudy sky through a cupola-like skylight (GL_{neg}12). Since the participants are not supposed to walk while exercising, we did not offer interactive or dynamic light (GL_{neg}11). Also, we decided against further lights to achieve unbalanced light temperatures to avoid straining the eyes of the participants when performing a static exercise (GL_{neg}10). Occasionally, lighting strikes illuminate the sky and are accompanied by an afforded low volume thunder.

The emotional neutral VE consists of a featureless light-grey plane. The skybox features no clouds and is light blue. No sound effects are played in this VE.

3.2 Task and Embodiment

As sport exercise, we selected arms' hold, a static own-bodyweight strength-endurance task as displayed in Figure 1. Participants needed to stand upright and stretch out their arms in front of them as long as they can manage. We selected this exercise for various reasons. Executing a static own-bodyweight exercise can be performed by anybody without any previous experience with physical training as it is quick to learn and easy to carry out. Besides ensuring an easy reproducible study design and easy accessibility for every participant, it also reduces risks of injury due to a wrong execution of a dynamic or even weighted strength exercise. In addition, dynamic or weighted exercises would require a certain fitness level or the selection of a correct weight per participant. Also, since exergames often target an audience that lack motivation to exercise on a regular basis, we intended to make our study available for everyone. Embedding a HIIT session might have resulted in additional performance measures like heart rate and maximum power output. However, HIIT and cardio training in general would again require participants not only to be fit enough to complete an entire session but also to be familiarized with training at specific intensity levels. Lastly, a static strength exercise will eventually exhaust every participant independent of their fitness level and result in an ever increasing struggle to keep the arms up. This internal fight could be most influenced by differences in the VE or existence of gamification. Thus, by selecting the arms' hold, we aim at generating insights into the effects of VE design on a very broad population.

To assist the participants in correctly executing the arms' hold and to increase the overall believability of the experiment, we decided to embody the participants in a generic, realistic avatar of the participant's ethnicity and natural gender as displayed in Figure 1 and Figure 4. Since we did not display a virtual mirror nor provided a preview of the avatars before the start of the simulation, we approached the ethnicity of the avatars purely by their skin color and kept the face as well as hairs rather simplistic. We designed a dark skinned, light skinned, and olive skinned avatar for both natural genders. Embodiment refers to the representation of a user in a VE with an avatar that belongs to them, follows their motions and hence provides them with a virtual body [18, 21]. Embodiment is comprised of the concepts virtual body ownership (VBO), agency, and self-location. VBO is the subjective experience of assigning a virtual body to oneself, agency describes the subjective experience of controlling a body, and self-location is the perception of being in one place [21]. While a higher illusion of VBO can lead to a higher presence, a higher perceived agency can lead to stronger emotional reactions to bodily threats in a VE [20]. In this way, the provision of an avatar potentially even increases the emotional reaction to our VEs.

A user's avatar can further have an influence on athletic performance. While avatars associated with a higher fitness level can reduce perceived effort and heart rate during physical activity in VR [23] and avatars with a more pronounced muscular can increase grip strength of male users [24], it was also observed that idealized avatars can cause negative effects on physical performance [25]. Similarly, while research showed that sweaty avatars can reduce the perceived intensity and exertion of cycling task compared to non-sweaty avatars [22], it was also observed that avatars dressed in sportswear did not show higher physical activity than avatars dressed in business wear [28]. Since we only provided the same type of avatar in different shades of skin color, potential positive or negative effects would have been applied to every participant and hence caused no confound in our comparative study. Yet, we decided to not idealize the avatar appearance and dress the avatars in gym wear.

We animated the avatars using inverse kinematics. Using this approach also had the advantage of slightly increasing the demand of the exercise. Participants needed to hold the VR game controllers in their hands during an experimental trial.

3.3 Gamification

The implementation of a gamification system followed the recommendations of Tuveri at al. [47]. In particular, we decided to integrate the gamification elements of *challenge* and *points* to investigates the effects of gamification in each of our three environmental conditions. Naturally, even more complex and innovative implementations of gamification would be possible. These implementations could build on the overall design of the VE or reward every correct execution of an integral movement pattern. For instance, inspiration could be drawn from Just Dance 2023 Edition [48] that rewards users with points for the correct execution of certain dance moves. This, however, would result in a completely different research direction requiring the derivation of gamification design principles and individual testing. This would be out of scope of our research, especially as we intend to first investigate whether an interaction between VE design and gamification exists before improving existing forms of gamification used in physical training.

We integrated the challenge gamification element by placing a flower pot in each VE from which a flower started to grow and finally blossom up. Although flowers are typically used in positive VEs, they can be easily adjusted in terms of color and material to follow the overall emotional direction and guidelines of the VEs while ensuring consistency in concept and avoiding unintentional effects of elements that are too different. The emotional positive VE featured a sunflower with bright colors. The neutral VE provided a thin leafy plant. Finally, two cacti with reddish spikes and metal green stems grew in the emotional negative VE. After conducting preliminary measurements, we set the growing animation to four minutes. The plant grows as long as a user keeps their arms up in the air and stops growing when the user lowers their arms or the blossom is fully developed.

To measure the potential influences and interactions of our gamification system, we included it in all of our VEs. The participants assigned to the gamification condition experienced the three individual VEs in counterbalanced our. Hence, we needed to adjust the gamification system to avoid completely overruling the effects of a given VE by providing a too clear goal. We adjusted the points gamification element to provide 73 points to make it harder for the participants to count the points and merely aim for a higher highscore without taking in the scenery. Also, we did not provide a points counter for that reason. However, we provided the participants with a summary of the achieved points at the end of their experimental trial. The bonus point asset was displayed two meters away in front of a user. For the first four minutes, the system rewarded the points every 20 seconds. After reaching the four-minute threshold, a user received new points every 15

Fig. 2. Comparison of our implementations of the points gamification element. The left image depicts the gamification element in the negative environment following central guidelines for emotional negative design. The center image shows the positive implementation of the gamification element following guidelines for emotional positive design. Finally, the right image demonstrates the neutral implementation following the design principles for the neutral environment.

seconds. We continued to shorten the reward times after eight minutes to ten seconds and after twelve minutes to five seconds.

3.4 Functions and Technology

We showed illustrated instructions of the sport exercise on a virtual billboard in every VE before the beginning of the exercise. We used two pictures of a wooden drawing doll to demonstrate the correct execution of the arms' hold exercise. This allowed participants to familiarize with the task. After the familiarization phase, the billboard got disabled to avoid blocking out aspects of the VE design as well as to eliminate distracting elements.

We added an internal timer to the system to accurately measure the endurance of the participants. After calibrating their body and determining the position of the game controllers when the hands are stretched out, the system measured the holding time once a participant stretched out their arms again. The timer stopped as soon as the hands deviated by more than 10 cm from the previously determined height. We did not display the timer inside of the VEs to prohibit participants from trying to beat their previous time in subsequent experimental trials. The participants were not allowed to walk through the VE while doing the sport exercise.

We implemented the system using unity in the version 2020.3.20f1 [49] targeting the HTC Vive [17]. In our study, we used the HTC Vive Index. We used three additional Vive Tracker 2.0 attached to a participant's feet and back directly above the hipbone to achieve a full embodiment. Inverse kinematic was realized with FinalIK plugin [37].

4 Pretesting the Virtual Environments

In order to check our manipulation of participants' emotions with the help of our emotionally designed VEs described in subsection 3.1, we conducted a pretest. Using a within-subjects design, participants were immersed into all of the three VEs following the approach of Steinhaeusser et al. [45].

4.1 Measures

Emotions were measured through self-reports using the *Positive and Negative Affect Schedule* (PANAS) questionnaire [2]. In detail, we used the scales *Positive Affect* (PA) and *Negative Affect* (NA) to measure participants' general positive respectively negative affect (10 items each). Furthermore, following Steinhaeusser et al. [45] we operationalized joy using the *Joviality* scale (8 items) as well as fear (6 items) using the *Fear* scale of the *PANAS-X* questionnaire [36, 51]. All items were answered on a 5-point Likert scale.

Fig. 3. Comparison of our challenge gamification element implementations.

Fig. 4. Overview of our avatars.

4.2 Procedure

First, participants provided written informed consent. Then they filled in the PANAS and PANAS-X scales as a pre-measurement. Afterwards, they were immersed into the neutral VE for two minutes using a *HTC Vive*, followed by filling in the PANAS and PANAS-X scales. Next, participants were randomly immersed into the positively or negatively designed VE for two minutes, again followed by administration of the PANAS and PANAS-X scales. Last, they spent two minutes in the remaining VE and subsequently filled in the PANAS and PANAS-X scales again. Thus, the order of the emotionally charged VEs was counterbalanced. In the end, participants provided demographic data and were thanked.

Table 2. Descriptive statistics from pretest.

	Pre-Measurement		Neutral VE		Positive VE		Negative VE	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
PA	2.77	0.83	2.07	0.67	2.96	0.83	2.52	0.63
NA	1.24	0.26	1.17	0.18	1.06	0.15	1.69	0.79
Joviality	2.95	1.01	2.05	0.90	3.54	1.00	1.98	0.82
Fear	1.34	0.45	1.36	0.35	1.10	0.21	2.20	1.26

Note. Calculated means from 1 to 5.

We recruited the participants from the students enrolled at the University of Würzburg using an online recruiting system. Students recruited via the recruiting system received credits mandatory for obtaining their final degree.

4.3 Participants

Twenty-three participants with a mean age of *M* = 20.52 (*SD* = 1.93) took part in our pretest. While the majority of 19 individuals self-reported being female (age: *M* = 20.32, *SD* = 1.38), only four participants self-indicated as male (age: *M* = 21.50, *SD* = 3.79). No one self-reported being diverse gender.

4.4 Results

All analyses were carried out using *JASP* [19] version 0.16 and an alpha level of .05. Descriptive values are displayed in Table 2. Checking assumptions, Shapiro-Wilk tests indicated a violation of the normality assumption for NA_{pre}, $Fear_{pre}$, $PA_{neutralVE}$; $NA_{neutralVE}$, $Fear_{neutralVE}$, $NA_{positiveVE}$, $Fear_positiveVE$, $NA_{negativeVE}$, $Joviality_{negativeVE}$, and $Fear_{negativeVE}$. Therefore, we calculated Friedman tests as a non-parametric alternative.

Concerning PA, a significant effect of time was revealed, $\chi^2(3) = 32.92, p < .001$. Conover's post-hoc test indicate that PA was significantly higher prior to the experiment compared to PA measurements after the neutral VE ($p < .001$). Further, PA was significantly higher after both the positive ($p < .001$) and negative ($p = .002$) VE compared to the neutral VE. Last, PA was significantly higher after the positive VE compared to the negative VE ($p = .025$). No significant differences were found between pre-measurement and positive VE ($p = .234$) as well as negative VE (.281).

For NA, again a significant effect of time was found, $\chi^2(3) = 34.83, p < .001$. Conover's post-hoc comparisons show a significant difference between NA values prior to the experiment compared to after the positive VE ($p < .001$) with higher NA values in the pre-measurement. Furthermore, NA after the neutral VE was significantly higher than after the positive VE ($p = .034$) and significantly lower than after the negative VE ($p < .001$). Last, NA values were significantly higher after being immersed into the negative VE compared to the positive VE ($p < .001$). No significant differences were found between pre-measurement and neutral VE ($p = .122$) as well as negative VE ($p = .058$).

Considering joy, we indicated a significant effect of time on *Joviality*, $\chi^2(3) = 33.60, p < .001$. Conover's post-hoc comparisons a significant decrease of *Joviality* values from pre-measurement to neutral VE ($p = .004$) as well as negative VE ($p = .018$), and a significant increase from pre-measurement to positive VE ($p = .036$). Moreover, *Joviality* values were significantly higher in after the positive VE compared to the neutral VE ($p < .011$), but not significantly lower after the negative VE compared to the neutral VE ($p = .604$). Last, *Joviality* was significantly higher after the positive VE compared to the negative VE ($p < .001$).

For fear, again a significant effect of time was observed, $\chi^2(3) = 25.91, p < .001$. Calculated Conover's post-hoc tests indicate a significant difference between pre-measurement and both positive and negative VE with lower values after the positive VE (p = .022) and higher values after the negative VE (p = .008). Comparing the emotionally charged VEs to the neutral VE, *Fear* significantly decreased after the positive VE (p = .016) and significantly increased after the negative VE (p = .012). Last, *Fear* values were significantly higher after the negative compared to the positive VE ($p < .001$).

Taken together, the results from our prestudy confirm the intended manipulation of participants' emotions due to the utilized VEs.

5 Methodology

To investigate the effects of emotional VE design on physical performance, we conducted a 3x2 design user study. The investigation of the effects of emotional VE design followed a within-groups study design. However, the analysis of the effects of gamification followed a between-subjects design with either gamification being displayed or not displayed. We made this decision to avoid physically exhausting the participants by only requiring them to complete three instead of six sessions. The participants were randomly assigned to one of the two between conditions. Hence, they either experienced the environmental conditions with or without gamification in counterbalanced order. We invited the participants on three consecutive days to counteract the effects of physical fatigue by a 24-hour recovery period between the trails. Participants gave informed consent to participate in the study at the beginning of the first experimental trial.

While the design of VEs can influence a user's emotions [20], Robazza et al. [35] revealed an emotion-performance link in sports. Athletes showed a better rowing performance when exercising on a rowing ergometer while being immersed in a lake VE [1]. Moreover, gamification demonstrated to increase physical performance [3]. Thus, we assume the following hypotheses:
H1: Participants show a better physical performance in the two emotionally designed VEs than in the neutral VE.
H2: There is an interaction between the effects of gamification and VE design on physical performance.

Besides research demonstrating an emotional influence of VEs when following an emotional design [45], our pre-study showed an emotional effect of our designed VEs on the participants. Hence, we postulate the following hypotheses:
H3: Influence of the VE on participants' emotions is higher in the two emotionally designed VEs than in the neutral VE.
H4: Participants' emotions are positively affected by the positive VE and negatively affected by the negative VE.

Emotion is known to influence motivation regulation in the context of sports [38]. Also, gamification showed an influence on motivation in the context of physical training [3, 47]. Therefore, we formulate the following hypotheses:
H5: Participants motivation is higher in the two emotionally designed VEs than in the neutral VE.
H6: There is an interaction between the effects of gamification and VE design on motivation.

Since the predominant emotion in a given VE influences the formation of presence [20], we assume the following hypothesis:
H7: Participants report higher presence for the emotionally designed VEs compared to the neutral VE.

5.1 Measures

We used the following measures to test our hypotheses.

5.1.1 Performance. We operationalized performance on an objective and subjective level. Measuring the maximum execution time is a commonly used objective measurement for the performance in a static strength exercise like in strongman competitions [46]. The longer an athlete executes an exercise, the better the performance is. Our system automatically measured the duration of the exercise per participant. As soon as the hands of a participant deviated by more than 10cm from the starting height, the timer stopped and the system exported the duration to a text file for later analysis.

We measured the performance on a subjective level using the *NASA Task Load Index* (*NASA-TLX*) [14]. To facilitate the evaluation process, we used the *Raw NASA-TLX* [13]. It eliminates the weighting process and only includes the six subscales [29]. We calculated the score for each subscale as described by Hart and Staveland [14] leading to total scores ranging from 0 to 100. Low scores mean low task load and high performance. The participants rated the perceived task load after every experimental trial.

In particular, mental and temporal demand can reveal insights into the mental processes that began when the participants started to feel exhausted. As with any strength or endurance discipline, mental strength allows athletes to keep performing even when they start to reach their physical limits. Such a higher mental effort to keep going despite a fatiguing body also could result in a higher perceived temporal demand, A lower mental and temporal demand could indicate a perceived easier performance of the exercise. Similarly, physical demand, performance, and effort provide insights into the subjective perception of the performance in the task. A lower score indicates a perceived better performance. Finally, as with mental and temporal demand, a higher frustration would indicate a stronger struggle with the execution of the exercise and hence a subjective lower performance.

Taken together, the objective and subjective results allow for a test of H1 and H2.

5.1.2 Positive Affect and Negative Affect. To evaluate the emotional effects of the three VEs, we used the *Positive and Negative Affect Schedule* (*PANAS*) [2]. The PANAS consists of two 10-item 5-point Likert scales (5 = very much). Each scale measures one of the two primary dimensions of mood, i.e., positive (PA) and negative (NA) affects. We administered the PANAS before and after each experimental trial. Besides generally measuring the current affect, the PANAS proved to be effective for determining the influence of emotional VE designs [45]. In this way, the PANAS results allow for an investigation of H3 and H4.

5.1.3 Motivation. We measured the experienced motivation after each experimental trial using the *Situational Intrinsic and Extrensic Motivation Scale* (*SIMS*) [10]. The SIMS consists of 16 items measuring intrinsic motivation, identified regulation, external regulation, and amotivation on 7-point Likert scales (7 = corresponds exactly). The SIMS was completed after every experimental trial. We operationalized H5 and H6 by measuring motivation using the SIMS.

5.1.4 Presence. For the assessment of presence, we used the 19 core items of the Presence Questionnaire – version 3.0 (PQ) [53]. The participants rate their experienced presence on 7-point Likert scales, higher scores indicate a high feeling of presence. The participants rated the experienced presence after each experimental trial. Using the PQ allowed for an operationalization of H7.

5.1.5 Qualitative Feedback. After the end of the third experimental trial, we asked the participants about their favorite VE and the reason for their choice. Also, we asked them to name the VE that motivated them the most. In the gamification condition, we also asked the participants about qualitative feedback and preference concerning the gamification elements.

Table 3. Demographic data of participants.

	Gender			Age			VR Experience				
	m	f	d	all	m	f	never	exp.	weekly	monthly	yearly
Overall	21	26	0	23.73 (2.29)	24.76 (2.36)	22.89 (1.88)	21	14	2	4	6
Without GA	8	13	0	23.38 (1.99)	24.38 (2.33)	22.77 (1.54)	8	6	0	4	3
With GA	13	13	0	24.00 (2.51)	25.00 (2.45)	23.00 (2.24)	13	8	2	0	3

Note. Standard deviations in parentheses. VR Experience = prior Virtual Reality experience, m = male, f = female, d = diverse, exp. = in prior experiments, GA = Gamification. Age is not provided for diverse gender since no one self-reported as diverse gender.

5.2 Procedure

At the start of the first experimental session, the participants filled in the demography questionnaire followed by the PANAS. In each subsequent experimental session, they started with the PANAS, only. Subsequently, they completed a short warm-up session to reduce chances for an injury during the strength-endurance exercise. Once warmed up, the participants put on the VR devices. The participants were allowed to look around and take in the different VEs before the start of the exercise. In this way, participants were fully aware of their surroundings. We selected the avatar based on the participant's natural gender and skin color. After we adjusted the size of the avatar to the bodies of the participants, they began with the physical exercise as displayed in Figure 1. We instructed them to hold their arms up for as long as they can. This task should unmask emotional and motivational benefits for physical training when personal levels of exhaustion are reached. Once they lowered their arms, the system logged the duration of the exercise in seconds. After taking off the VR gear, the participants filled in the PANAS, NASA-TLX, SIMS, and PQ. Depending on the number of the session, we thanked them for participating and either reminded them to come back the next day or asked them which VE they liked most and wished them farewell. The order of the VEs was counterbalanced between the participants.

The study took place during the COVID-19 pandemic. To ensure for protection and hygiene, participants and experimenter wore masks and we cleaned all touched surfaces and used devices after each experimental trial.

We recruited the participants from the students enrolled at the University of Würzburg using an online recruiting system or personal connections. Students recruited via the recruiting system received credits mandatory for obtaining their final degree.

The institutional review board of Human Computer Media at the University of Würzburg approved our study.

5.3 Participants

Overall, 47 participants took part in our study. Being randomly assigned to a condition, 21 persons completed the virtual exercise without gamification, whereas 26 participants were assigned to the condition with gamification. Further demographic data are presented in Table 3.

6 Results

All analyses were carried out using *JASP* [19] version 0.16 and an alpha level of .05. For assumption checks Shapiro-Wilk tests were calculated to analyze normality of data and Mauchly's tests were carried out to analyze sphericity. Levene's tests were calculated to analyze homogeneity of variances. Descriptive data are displayed in Table 4.

Table 4. Descriptive statistics.

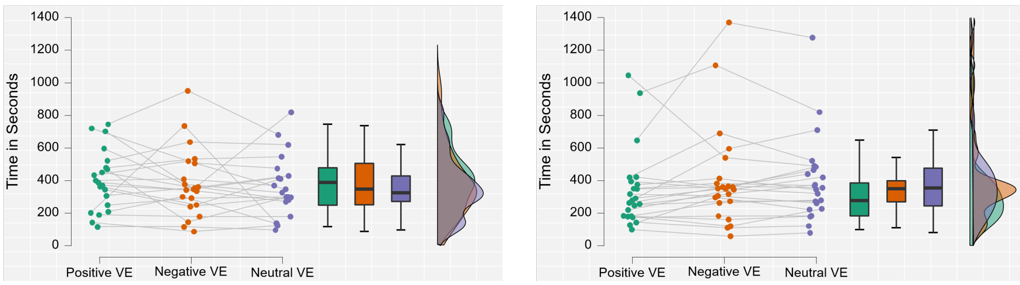
	Plain Environments			Gamification		
	Positive VE	Negative VE	Neutral VE	Positive VE	Negative VE	Neutral VE
Variable	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Ath. Performance[a]	399.52 (185.08)	381.43 (210.25)	359.43 (190.36)	343.13 (238.89)	404.70 (304.66)	400.17 (259.31)
Positive A Pre[b]	2.01 (0.37)	1.98 (0.39)	1.91 (0.37)	2.98 (0.71)	3.04 (0.53)	3.04 (0.59)
Negative A Pre[b]	2.23 (0.48)	2.16 (0.48)	2.14 (0.58)	1.39 (0.42)	1.27 (0.24)	1.35 (0.33)
Positive A Post[b]	1.94 (0.29)	1.94 (0.35)	1.74 (0.36)	3.19 (0.77)	3.20 (0.61)	3.23 (0.69)
Negative A Post[b]	2.34 (0.56)	2.28 (0.44)	2.00 (0.46)	1.21 (0.29)	1.22 (0.23)	1.24 (0.30)
Intrinsic M[c]	3.86 (1.23)	3.45 (1.33)	3.10 (1.43)	4.75 (1.15)	4.55 (0.97)	4.59 (0.96)
Identified R[c]	3.91 (1.30)	3.81 (1.20)	3.67 (1.37)	4.29 (1.04)	4.21 (0.94)	4.32 (0.98)
External R[c]	2.79 (1.43)	2.79 (1.34)	2.56 (1.39)	1.85 (1.10)	2.07 (1.09)	2.00 (1.21)
Amotivation[c]	2.27 (1.15)	2.31 (1.10)	2.37 (1.04)	2.14 (1.34)	2.05 (1.17)	1.76 (1.02)
Mental D[d]	15.62 (14.62)	13.43 (15.26)	9.07 (9.29)	13.46 (10.88)	11.44 (14.35)	8.13 (8.56)
Physical D[d]	56.50 (26.18)	48.79 (26.70)	52.98 (22.64)	40.87 (25.65)	44.83 (20.82)	43.24 (21.05)
Temporal D[d]	20.60 (24.43)	29.62 (30.49)	21.76 (21.73)	15.33 (17.59)	19.04 (19.07)	13.48 (18.02)
Performance[d]	36.14 (19.67)	48.17 (23.67)	46.48 (22.31)	40.41 (22.62)	37.70 (18.37)	40.65 (22.05)
Effort[d]	49.60 (22.54)	50.02 (23.96)	55.00 (22.96)	47.47 (24.95)	51.83 (21.76)	53.13 (26.21)
Frustration[d]	21.21 (20.61)	26.60 (23.13)	32.88 (27.14)	19.24 (22.33)	25.70 (24.70)	24.15 (23.85)
Presence[c]	5.21 (0.79)	5.05 (0.66)	4.73 (0.89)	5.25 (0.77)	4.93 (0.86)	4.84 (0.78)

Note. Ath. = Athletic, A = Affect, M = Motivation, R = Regulation, D = Demand.
[a] In seconds.
[b] Calculated values from 1 to 5
[c] Calculated values from 1 to 7.
[d] Calculated values from 0 to 100.

6.1 Performance – Time

To examine the effect of emotionally designed VEs and gamification on exercise duration, a mixed ANOVA was calculated. Concerning assumption checks, violation of the normal distribution was revealed for all three VEs, whereas sphericity and equality of variances were confirmed. Since the ANOVA is robust to violation of the normality assumption [8, 15], a two-way repeated measures ANOVA was calculated. No significant difference was observed for main effect of VE design ($F(2, 84) = 0.35$, $p = .709$) nor for main effect of gamification, $F(1, 42) = 0.00$, $p = .969$. Also, no significant interaction effect was observed, $F(2, 84) = 1.93$, $p = .152$. Last, a planned contrast comparing the positive and negative to the neutral VR also revealed no significant difference, $t = 0.13$, $p = .899$. The performance results are visualized in Figure 5.

To examine the effect of training day and repeated practice but also exhaustion, we again calculated a repeated measures ANOVA comparing data from session one ($M_{WithoutGamification} = 369.47$, $SD_{WithoutGamification} = 208.16$, $M_{WithGamification} = 412.92$, $SD_{WithGamification} = 351.05$), two ($M_{WithoutGamification} = 378.83$, $SD_{WithoutGamification} = 186.05$, $M_{WithGamification} = 431.58$, $SD_{WithGamification} = 392.81$), and three ($M_{WithoutGamification} = 391.88$, $SD_{WithoutGamification} = 193.13$, $M_{WithGamification} = 451.54$, $SD_{WithGamification} = 359.74$) regardless of the VEs. Checking assumptions a violation of normality was indicated, whereas sphericity and equality of variances was confirmed. No significant differences were found in terms of training day ($F(2, 86) = 0.62$, $p = .540$) or condition ($F(2, 86) = 0.04$, $p = .947$). Also, no significant interaction was observed, $F(1, 43) = 0.38$, $p = .540$.



(a) Performance in group without gamification. (b) Performance in group with gamification.

Fig. 5. Results for performance. Boxplots display medians, error bars display 95% CI. Connected points belong to one participant, not indicating a temporal sequence.

6.2 Performance – Task Load

To examine effects of emotional VE design and gamification on task load, we calculated mixed ANOVAs again. Shapiro Wilk tests indicated violation of the normality assumption for mental demand, temporal demand, performance, and frustration. Sphericity was confirmed for all subscales. Equality of variances was indicated for all subscales except temporal demand. While being considered robust against violation of normal distribution [8, 15], unequal variances prohibit ANOVA calculation. Thus, only the post-hoc tests were interpreted for temporal demand [16].

First, we indicated a significant difference in mental demand between the three VEs, $F(2, 84) = 5.78$, $p = .004$, $\omega^2 = .03$. In contrast, no significant main effect of gamification was found, $F(1, 42) = 0.29$, $p = .591$. Similarly, the interaction effect was insignificant, $F(2, 84) = 0.07$, $p = .933$. Bonferroni-adjusted post-hoc tests reveal significant higher values in the positive than in the neutral VE, $p = .004$, $d = .51$. The difference between positive and negative VE ($p = .238$) as well as between negative and neutral VE ($p = .100$) was not significant.

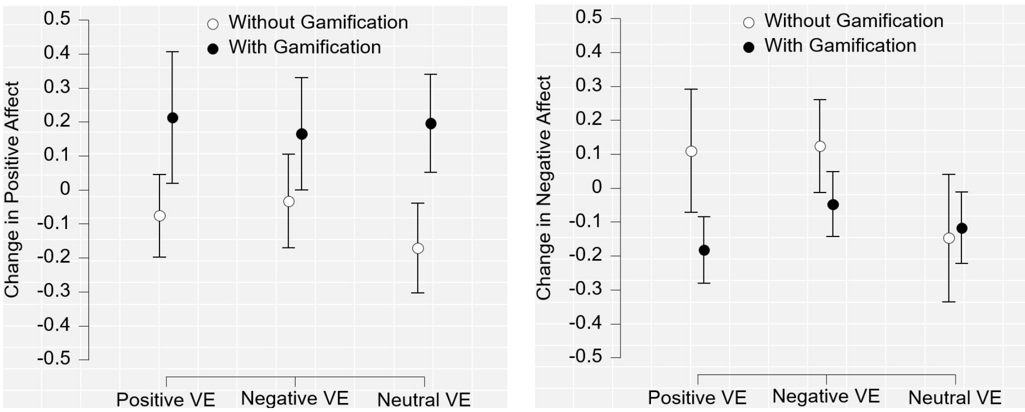
In contrast, no significant main effect of VE ($F(2, 84) = 0.21$, $p = .815$) or gamification ($F(1, 42) = 2.39$, $p = .130$) on physical demand was found. Also, the interaction effect was not significant, $F(2, 84) = 1.88$, $p = .158$.

Analyzing the post-hoc tests for temporal demand, no significant differences between the three VEs ($ps > .05$) or the presence and absence of gamification ($p = .142$) was found.

Concerning performance, no main effect of VE design ($F(2, 84) = 1.57$, $p = .215$) nor gamification $F(1, 42) = 0.58$, $p = .453$) was indicated. The interaction effect just missed significance, $F(2, 84) = 2.67$, $p = .075$, $\omega^2 = .01$. However, post-hoc tests revealed no significant differences for this hybrid interaction ($ps > .05$).

For perceived effort again no significant main effects of VE design ($F(2, 84) = 1.71$, $p = .187$) and gamification ($F(1, 42) = 0.01$, $p = .917$) were observed. Similarly, the interaction effect was insignificant, $F(2, 84) = 0.26$, $p = .770$.

Last, the same pattern was found for frustration with no significant main effect of VE design ($F(2, 84) = 2.26$, $p = .111$) and gamification, $F(1, 42) = 0.50$, $p = .482$. Again, no significant interaction effect was found, $F(2, 84) =$ ß.56, $p = .574$.



(a) Positive Affect. (b) Negative Affect.

Fig. 6. Results for PANAS changes. Error bars display 95% CI.

6.3 PANAS

We analyzed changes in positive and negative affect over the athletic exercise in terms of combining gamification and emotionally designed VEs depicted in Figure 6 where positive values indicates increases and negative values indicate decreases in affect. Checking assumptions, normal distribution was confirmed for all values except for PA in the neutral VE. Sphericity was confirmed for both PA and NA. Last, violation of equality of variances was indicated PA in the positive and negative VE and for NA in the negative VE. Although being considered robust against violation of normal distribution [8, 15], unequal variances prohibit ANOVA calculation. However, post-hoc tests may be interpreted [16].

For PA Bonferroni-adjusted post hoc comparisons indicated a significant effect of gamification, $p = .006$, $d = .44$. Concerning influence of VE design, no significant differences were found, ps > .05. A planned contrast comparing positive and negative VE to the neutral VE further indicated no significant difference, $t = 0.86$, $p = .391$. For NA Bonferroni-adjusted post hoc comparisons revealed a significant effect of gamification, $p = .015$, $d = .38$. Regarding emotional VE design, a significant difference between negative and neutral VE was identified, $p = .034$, $d = .39$. No significant difference was found comparing positive and negative VE ($p = .781$) and positive and neutral VE, $p = .445$. A planned contrast comparing positive and negative VE to the neutral VE also indicated a significant difference, $t = 2.34$, $p = .022$.

6.4 SIMS

To investigate the effect of emotionally designed VEs and gamification on exercise duration, again mixed ANOVAs were calculated. Normal distribution was only confirmed for intrinsic motivation, whereas this assumption was violated for identified regulation, external regulation, and amotivation. Sphericity and variance homogeneity were confirmed for all subscales. Due to the robustness of the mixed ANOVA, four two-way repeated measures ANOVAs were calculated including planned contrasts.

Concerning intrinsic motivation displayed in Figure 7a, a significant main effect of VE design was observed, $F(2, 84) = 5.04$, $p = .009$, $\omega^2 = .02$. Furthermore, a significant main effect of gamification was revealed $F(1, 42) = 13.74$, $p < .001$, $\omega^2 = .13$. The interaction effect was not significant, $F(2, 84) =$

2.12, p = .127. Bonferroni-adjusted post hoc comparisons indicated a significant difference between positive and neutral VE (p = .024, d = .42) and between presence and absence of gamification, p < .001, d = .56. No significant difference was found comparing positive and negative VE (p = .131) as well as negative and neutral VE, p = 663. A planned contrast comparing positive and negative VE to the neutral VE again showed a significant difference, t = 2.39, p = .022.

In contrast, no significant difference was found for identified regulation displayed in Figure 7b, neither for main effect of VE design ($F(2, 84)$ = 0.50, p = .610) nor for main effect of gamification, $F(1, 42)$ = 2.26, p = .140. Similarly, there was no significant interaction, $F(2, 84)$ = 0.80, p = .452. The planned contrast comparing the positive and negative VE to the neutral VE indicated no significant difference, t = 0.55. p = .587.

Analyzing external regulation displayed in Figure 7c, no significant main effect of VE design was observed, $F(2, 84)$ = 0.89, p = .415. Nevertheless, a significant main effect of gamification was identified, $F(1, 42)$ = 4.29, p = .045, ω^2 = .04. The interaction effect was not significant, $F(2, 84)$ = 1.40, p = .253. Bonferroni-adjusted post hoc tests confirmed the significant effect of gamification, p = .045, d = .31. Comparing positive and negative VE to the neutral VE the planned contrast indicated no significant difference, t = 0.91, p = .371.

Last, no significant main effect of VE design on amotivation displayed in Figure 7d was revealed, $F(2, 84)$ = 0.78, p = .463. Also, no significant main effect of gamification ($F(1, 42)$ = 1.11, p = .298) or interaction effect ($F(2, 84)$ = 1.05, p = .135) was observed. Similarly, a planned contrast comparing positive and negative VE to the neutral VE showed no significant difference, t = 1.17, p = .247.

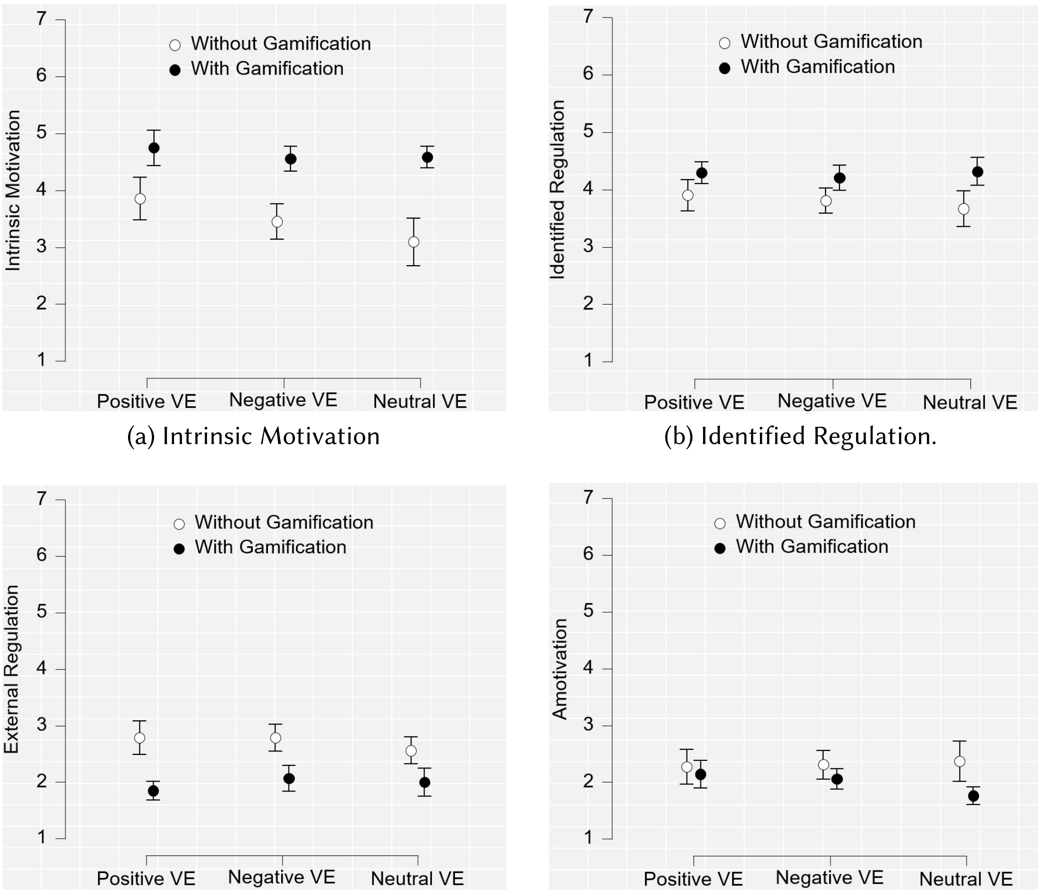
6.5 PQ

Presence was again analyzed in a mixed ANOVA. Regarding assumptions, normality of data and homogeneity of variances were confirmed. In contrast, Mauchly's test indicated a violation of the sphericity assumption. Thus, we applied Greenhouse-Geisser correction. Results showed a significant main effect of VE design, $F(2, 84)$ = 11.55, p < .001, ω^2 = .05. In contrast, both the main effect of gamification ($F(1, 42)$ = 0.00, p = 962) and the interaction effect ($F(2, 84)$ = 0.77, p = .436) were insignificant. Bonferroni-adjusted pairwise comparisons revealed a significant difference between the positive and neutral VE (p < .001, d = .72) as well as between the positive and negative VE, p = .033, d = .39. Last, the difference between negative and neutral VE just missed significance, p = .09, d = .33. In line, a planned contrast comparing positive and negative VE to the neutral VE indicated a significant difference, t = 4.05, p < .001.

6.6 Qualitative Feedback

The answers on preferred VE are displayed in Table 5. Since frequencies were less than five for negative and neutral VE, test calculation was not possible.

Regarding the gamification elements, 15 participants liked the sunflower best, whereas five persons preferred the cacti, four preferred the leafy plant, and only one person stated to like none of the plants. Rating the motivating factor of the plant after each session on a five-point scale participants reported an overall positive motivating factor of the plant gamification element (M = 3.72, SD = 1.27) regardless of its type. The data was not normally distributed according to Shapiro-Wilk tests and the sphericity assumption was violated as shown by Mauchly's test. Comparing between the sun flower (M = 3.96, SD = 1.19), cacti (M = 3.44, SD = 1.41), and leafy plant (M = 3.78, SD = 1.20) a Friedman test revealed no significant difference in reported motivating effect, $\chi^2(2)$ = 1.24, p = .538. After the last session, 22 of the participants reported the bonus points being a motivating gamification element, whereas only one person negated this question.



(a) Intrinsic Motivation (b) Identified Regulation.

(c) External Regulation. (d) Amotivation.

Fig. 7. Results for SIMS. Error bars display 95% CI.

Table 5. Preferred VE per condition.

	Positive VE	Negative VE	Neutral VE
Without Gamification	17	3	1
With Gamification	21	2	0

7 Discussion

In our study, we investigated the effects of three different VEs with gamification either being enabled or disabled on performance as well as perceived emotions, motivation and presence during a physical exercise. A pre-study demonstrated the intended emotional effect of experiencing the three tested VEs.

7.1 Influence of VE Design and Gamification on Physical Performance

Our first and second hypothesis focused on the physical performance of the participants. We hypothesized that participants will perform better in the emotionally-designed VEs compared to the neutral VE and that the existence of gamification will interact with the VE design. Looking at our study results, we found no significant main effect of VE design and of gamification with respect to objective exercise performance measured in duration of execution. We observed no significant interaction effect. Also, no effect of physical fatigue occurred over the course of the experiment. Besides a significant difference in mental demand between the positive and neutral VE, this pattern was also found for the subjective performance measured using the NASA-TLX. In particular, we found a significantly higher mental load for the positive VE compared to the neutral VE. This is especially interesting as the participants reported no significantly different physical demand and effort between the two conditions. A potential explanation could be that the participants wanted to perform well in the positive VE and hence were strongly focused to keep their arms up as long as they can. This assumption is backed by the significant difference in the intrinsic motivation between these two VEs measured on the SIMS. The participants reported a higher intrinsic motivation in the positive VE. Also, the participants liked the positive VE the most. This further supports our assumption. Being more mentally focussed while experiencing a higher intrinsic motivation might have reduced or distracted from the higher perceived physical demand and effort that comes along with a longer execution time. As a result, the execution of the arms' hold felt more mentally demanding in the positive VE while not causing a higher perceived physical demand and effort.

Although our objective performance measurements showed no significant difference, we still interpret the descriptive results. While a new personal best in a given sport might not result in a statistical significant difference, it still can be a very motivating experience for every type of athlete independent of elite, sub-elite or recreational status. Mainly focusing on inferential statistics would mask important aspects indicating a potential benefit of using digital simulations in sports. The participants yielded a longer exercise time when immersed in the emotionally designed VEs in contrast to the neutral VE. The exercise times were 40 seconds longer in the emotional positive VE and 22 seconds longer in the emotional negative VE. This supports our assumption that higher intrinsic motivation led participants to focus more mentally in order to perform well in the arms' hold exercise. It might also be a potential indication for the importance of using emotionally-charged VEs for physical training applications. Emotionally-charged VEs might lead users to achieve a better performance while feeling intrinsically more motivated. The addition of gamification indicates to dominate the emotional benefits of the VEs. A constant positive feedback throughout the exercise reduced the difference in performance between the emotional negative VE and neutral VE. Moreover, adding gamification resulted for the negative VE in a 23 second and for the neutral VE in a 60 second longer exercise time compared to the plain environments. The effect is supported by generally lower values for physical demand on the NASA-TLX for all of the VEs when gamification is present. Surprisingly, although not significantly different, the participants stopped the physical exercise 60 seconds on average earlier in the emotional positive VE than in the other two VEs when gamification was present and to the emotional positive VE without gamification. A potential explanation could be the implementation of the challenge gamification element in this VE. The sunflower followed the color-specific recommendations for positive emotional VE design. The flower further was rather flashy in comparison to the other two plants which might have caught the attention of the participants completely while it was growing. After the flower reached its final form, the participants might have ceased to draw motivation from watching it evolve. This sudden stop of motivation might have subconsciously influenced their mental focus in keeping their arms up. Since the other two plants might not have been perceived as visually interesting as the sunflower, the sudden stop of their growth did not influence the participants' performance. Hence, our results suggest that a sudden stop of a highly prominent challenge-based gamification element can potentially reduce the otherwise positive effect of an emotionally designed VE. Taken together, we must reject **H1** as we did not observe a significantly better physical performance in the two emotionally designed VEs. In addition, **H2** is rejected since no interaction with gamification was found.

7.2 Influence of Emotional VE Design on User Emotion

We expected an influence of the participants' emotions by the emotionally-designed VEs with H3 and H4. The analyses of the PANAS results indicated no significant influence of the VE design on PA. In contrast, NA changed significantly less in the emotional negative than in the neutral VE. This supports our theory-based design of the emotional negative VE [45]. Surprisingly, in strong contrast to the underlying guidelines and the results in the pre-study, the emotional positive VE yielded no such emotional benefit. A potential explanation might be the execution of the physical exercise and the immediate assessment of PA and NA after the exposure. The participants hold their arms up in the air as long as they could. The resulting extend of exhaustion might have masked the emotional effects of the individual VEs right after the end of the exercise. It might even explain the decrease in PA and increase in NA in the emotional positive VE. Similarly, PA significantly decreased while completing the exercise in the neutral VE. Taken together, our results are mixed, showing a significant difference for NA but not for PA, hence we need to reject **H3**. Also, we must reject **H4** as the emotional positive VE did not yield the expected results. Future work shall continue the investigation of emotional effects of VE design in physical training applications by adding a recovery phase before the assessment of the participants' emotions.

Our PA and NA results, however, indicate a significant benefit of adding gamification to the simulation. We found a significantly higher increase in PA and significantly higher decrease in NA when gamification was used. The positive effect of gamification might have dominated the effects of exhaustion by providing the participants with positive feedback about their progress. This feedback was absent in the VE-only conditions potentially leading the participants to only feel exhausted but not rewarded after the exercise. Overall, our study supports the overall positive effects of using gamification.

7.3 Motivational Influences by VE Design and Gamification

We also hypothesized that the participants' emotions are affected by the design of the VEs (H5) and the existence of gamification (H6). With respect to motivation, our analyses revealed a significant main effect of VE design on intrinsic motivation. In particular, exercising in the emotional positive VE resulted in a significantly higher intrinsic motivation compared to the neutral VE. However, the intrinsic motivation did not differ between the emotional negative and neutral VE. The higher intrinsic motivation might also explain the significantly higher mental demand in the positive VE over the neutral one. The participants potentially were more mentally focused to perform well and to keep their arms up in this VE in contrast to the neutral VE. Taken together, our results are mixed and hence we must reject **H5**.

Supporting the well known effect of gamification, we found a significant difference between gamified and non-gamified training. In particular, participants of the gamified condition reported significantly higher intrinsic and extrinsic motivation. In combination with the higher positive emotion, our results lead to the recommendation of including gamification in physical training. However, it is important to keep the gamification active until the end of an exercise to avoid a sudden drop in motivation. The benefits of it indicate to be independent of the environmental conditions thus leading to the rejection of **H6**.

7.4 Link between Presence and Emotional VE Design

Lastly, we hypothesized that the emotional design of the VEs will increase the perceived presence. The analysis of presence indicated an effect of emotional design of the VEs on the participants' perceived presence. Independent of gamification presence was highest in the positively followed by the negatively designed VE. As backed up by our results, participants felt higher presence while being immersed in the two emotionally-charged VEs compared to the neutral VE. Our results support the observations made by Jicol et al. [20] of a higher emotional intensity in VE design leading to a higher presence. Independent of the emotional valence, users feel in the VE and thus a higher emotional intensity is a major factor in building up presence. This leads to the acceptance of **H7**.

7.5 Preference and Feedback

Finally, the participants liked the emotional positive VE most. In combination with the other measurements, our results confirm the benefits of using emotional positive VEs for VR-based training applications. With respect to the addition of gamification, the participants reported to have liked the emotional positive VE the most, again. The participants also liked the sunflower the most which supports the assumed effects of the end of its growth. With respect to motivating effect, all plants were perceived as similarly motivating. The participants also agreed to have experienced a motivation from the points gamification element. These results support the overall benefits of including gamification to yield a higher motivation. Considering the potential negative effect of our sunflower, it is, however, important to avoid a sudden stop in a potentially assumed endless stream of feedback.

8 Implications

Ultimately, the results of our study arenrelevant to researchers and developers of VR training applications. With respect to VE design, our results support the common approach of using positive and natural VEs for VR-based physical training [1, 11]. When exercising without gamification, participants achieved the best performance and highest intrinsic motivation in the positive VE. *Hence, we recommend to surround a user in natural VEs when targeting physical training in VR to improve their performance and intrinsic motivation, thus potentially leading to a better training effect and experience in the long run.* To achieve such an emotional design, we propose to follow the design guidelines of Steinhaeusser et al. [45] as displayed in Table 1. These guidelines provide actionable recommendations to design a VE at the levels of overarching design theme, colors as well as textures, and lighting. At the same time, our results also indicate that even negative VEs can increase a user's performance compared to a neutral VE. Therefore, VR-based physical training can also be integrated in emotional negative VEs. Since VEs are easily interchangeable, *we recommend to include positive as well as negative environments to allow users to select a VE according to their personal preferences.* To achieve such a selection, developers can follow the guidelines presented in Table 1 to develop at least one positive and one negative VE. Here, we also recommend to respect the targeted physical exercise. While most strength exercises are performed within a rather small space, cardio training often comes along with moving over a greater distance. Hence, we recommend *(1) to design VEs that focus on a single spot as done in our study for strength training* and *(2) to provide VEs focussing more on the experience of an ongoing stream of new VE design aspects for cardio training*. This can be especially effective when targeting exergames. Here, the overall motivation and experience of the athletes could benefit from VEs that go into the positive and negative direction over the course of a training program. A first indication for the effectiveness of such an approach was observed in a HIIT exergame supporting the intervals with a more dramatic atmosphere [6].

Concerning the integration of gamification, our results indicate a significant positiv impact of gamification on affect and motivation as well as increased exercise times. *Hence, we recommend to include at least challenge and points gamification elements to maximize physical training performance and experience.* If an emotional VE design is used, the visual appearance of the gamification elements should follow design guidelines of the predominant emotion as done in our study. Again, the guidelines listed in Table 1 provide actionable recommendations. However, it is important to adjust the difficulty of a challenge to the performance of the users to avoid a potential negative impact on performance. An easy approach could be to compete against oneself by providing a ghost time [6]. The central guideline is to keep challenge and points gamification elements active for the complete time of a given workout.

9 Limitations

A major limitation of our study's results could be the sport exercise itself. While it provides a demanding but simple physical task suitable for every subject, it might be not sensitive enough in comparison to motion-based endurance exercises like biking, running or HIIT workouts. High intensity endurance sports allow for a collection of physiological measurements like heart rate, consumption of oxygen as well as energy, power output and changes thereof [6]. These factors could reveal further performance-affecting effects of emotional VE design. However, for this novel approach we prioritized the accessibility of the task also for untrained participants over covering all challenges of typical athletic training. Thus, future work should integrate more diverse sport exercises to validate our findings. Further, interval training allows for several power surges and "hang-on" moments over the course of a workout. Considering multiple repetitions, our measured factors might be more affected by an emotional VE design in comparison to a single effort.

A different limitation of our experiment could be the design of the two emotionally charged VEs. While fulfilling the underlying design guidelines for emotional VE design, they differed in their internal structure and provided two completely different settings. A potential solution could be to use a neutral version of a virtual gym and to adjust the gym with respect to the emotional design. This would result in three conditions of the same setting. Also, the design of the emotional negative VE has fulfilled the recommendations of the guidelines, but might have resulted in a strong detachment effect caused by the large melee weapons, arena-like style, and overall less realistic setting. This detachment effect might have resulted in a positive experience for sensation seeking participants similar to playing a first-person shooter.

In alignment with the previous limitation, we did not measure the participants' interest in sensation seeking and overall sensitivity to environmental conditions. Thus, we could not control for personal differences in the overall perception of VEs. Looking at the preference of the participants in Table 5, it is possible that the majority not only preferred but also performed better in the positive VE than in the other two VEs compared to a few participants who enjoyed the negative VE the most. Yet, despite this limitation, our results indicate that offering VEs beyond traditional positive VEs and giving the choice to the athletes might result in an overall higher satisfaction when training with VR-based exergames.

Our recruited sample in itself also represents a limitation for the generalizability of our results. We only recruited healthy and young university students. While the task was especially chosen to be suitable for both trained and untrained participants, this might have led to ceiling effects covering effects from our manipulations. Further, the influence of emotional VE design and gamification might have a stronger or even completely different effect for elderly people or people suffering from the longterm aftereffects of not exercising for their entire life. Hence, future research needs to expand the range of the sample to expand on our preliminary results.

We based our argumentation for an effect of emotional VE design on physical performance mainly on descriptive statistics and a significant difference in the perceived task load. We did not find significant differences with respect to the objective performance. Also, we did not conduct a study to investigate longterm effects of using emotionally design VEs for exergames. Hence, although promising, our results need to be confirmed by future and more extensive research.

Lastly, our qualitative questions might not have been investigative enough. We mainly assessed the participants' preference for the VEs and gamification elements. However, we could have also explored how participants directly felt in the VEs and what aspects of the design predominantly evoked their feelings. In addition, it could be of importance to investigate what elements in particular were important to them while exercising.

10 Conclusion

Following design guidelines for positive and negative emotion induction within VEs [45], we created an emotional positive and negative as well as a neutral VE to investigate the influence of the VEs' design on athletic performance. We further investigated the effects of gamification in emotionally charged VEs. Although longer performance times were achieved in the emotional-charged VEs in contrast to a neutral one, we did not find a significant difference between the tested VEs. The emotional design influenced the participants intrinsic motivation but caused mixed results with respect to emotion during the exercise. In contrast, gamification caused a dominating effect on emotion and motivation over the emotional design of the VEs, thus indicating an overall positive impact. However, our results also suggest that highly motivating gamification elements must be active for the entire duration of the training to avoid negative impacts on the performance. Our results are of interest to researchers and designers of VR-based physical training as our results provide a potential indication for the benefits of emotional positive VEs. At the same time, even emotional negative VEs might benefit VR-assisted physical training, thus increasing the overall design space.

Future research shall inspect the effects of emotional VE design on high intensity and repetition-based physical training. In addition, it is important to investigate other VE training settings like a virtual gym or stadium that are emotionally enhanced. Finally, future work needs to investigate whether the positive effects of VE design and gamification persist over a longer training period.

References

[1] Sebastian Arndt, Andrew Perkis, and Jan-Niklas Voigt-Antons. 2018. Using Virtual Reality and Head-Mounted Displays to Increase Performance in Rowing Workouts. In *Proceedings of the 1st International Workshop on Multimedia Content Analysis in Sports (MMSports'18)*. Association for Computing Machinery, New York, NY, USA, 45—50. https://doi.org/10.1145/3265845.3265848

[2] B. Breyer and M. Bluemke. [n. d.]. Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel). https://doi.org/10.6102/zis242

[3] Rubén Camacho-Sánchez, Ana Manzano-León, José Miguel Rodríguez-Ferrer, Jorge Serna, and Pere Lavega-Burgués. 2023. Game-Based Learning and Gamification in Physical Education: A Systematic Review. *Education Sciences* 13, 2 (2023). https://doi.org/10.3390/educsci13020183

[4] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11)*. ACM, Tampere, Finland, 9–15. https://doi.org/10.1145/2181037.2181040

[5] Oliver R L Farley, Kirsten Spencer, and Baudinet Livvie. 2020. Virtual reality in sports coaching, skill acquisition and application to surfing: A review. *Journal of Human Sport and Exercise* 15, 3 (2020), 535–548. https://doi.org/10.14198/jhse.2020.153.06

[6] Matthew Farrow, Christof Lutteroth, Peter C Rouse, and James L J Bilzon. 2019. Virtual-reality exergaming improves performance during high-intensity interval training. *European Journal of Sport Science* 19, 6 (2019), 719–727. https://doi.org/10.1080/17461391.2018.1542459

[7] Dominik Gall and Marc Erich Latoschik. 2020. Visual angle modulates affective responses to audiovisual stimuli. *Computers in Human Behavior* 109 (2020), 106346.

[8] Gene V Glass, Percy D Peckham, and James R Sanders. 1972. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of educational research* 42, 3 (1972), 237–288.

[9] Linda Graf, Stefan Liszio, and Maic Masuch. 2020. Playing in Virtual Nature: Improving Mood of Elderly People Using VR Technology. In *Proceedings of the Conference on Mensch und Computer (MuC '20)*. Association for Computing Machinery, New York, NY, USA, 155–164. https://doi.org/10.1145/3404983.3405507

[10] Frédéric Guay, Robert J. Vallerand, and Céline Blanchard. 2000. On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion* 24, 3 (2000), 175–213. https://doi.org/10.1023/A:1005614228250

[11] Negin Hamzeheinejad, Daniel Roth, Daniel Götz, Franz Weilbach, and Marc Erich Latoschik. 2019. Physiological Effectivity and User Experience of Immersive Gait Rehabilitation. In *The First IEEE VR Workshop on Applied VR for Enhanced Healthcare (AVEH)*. IEEE, 1421–1429. https://downloads.hci.informatik.uni-wuerzburg.de/2019-ieeevr-workshop-vr-gait-preprint.pdf

[12] Yuri Hanin. 2003. Performance Related Emotional States in Sport: A Qualitative Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 4, 1 (2003). https://doi.org/10.17169/fqs-4.1.747

[13] Sandra G. Hart. 2006. NASA-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. Santa Monica, 904–908.

[14] Sandra G. Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, P. A. Hancock and N Meshkati (Eds.). Elsevier Science Publishers B.V., Amsterdam, 139–183.

[15] Michael R Harwell, Elaine N Rubinstein, William S Hayes, and Corley C Olds. 1992. Summarizing Monte Carlo results in methodological research: The one-and two-factor fixed effects ANOVA cases. *Journal of educational statistics* 17, 4 (1992), 315–339.

[16] Jason Hsu. 1996. *Multiple comparisons: theory and methods*. CRC Press.

[17] HTC Corporation. 2011–2022. HTC Vive. available at https://www.vive.com.

[18] Wijnand A IJsselsteijn, Yvonne A. W de Kort, and Antal Haans. 2006. Is This My Hand I See Before Me? The Rubber Hand Illusion in Reality, Virtual Reality, and Mixed Reality. *Presence: Teleoperators and Virtual Environments* 15, 4 (aug 2006), 455–464. https://doi.org/10.1162/pres.15.4.455

[19] JASP Team. 2021. JASP. https://jasp-stats.org/

[20] Crescent Jicol, Chun Hin Wan, Benjamin Doling, Caitlin H Illingworth, Jinha Yoon, Charlotte Headey, Christof Lutteroth, Michael J Proulx, Karin Petrini, and Eamonn O'Neill. 2021. Effects of Emotion and Agency on Presence in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445588

[21] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (nov 2012), 373–387. https://doi.org/10.1162/pres_a_00124

[22] Martin Kocur, Johanna Bogon, Manuel Mayer, Miriam Witte, Amelie Karber, Niels Henze, and Valentin Schwind. 2022. Sweating Avatars Decrease Perceived Exertion and Increase Perceived Endurance While Cycling in Virtual Reality. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology (VRST '22)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3562939.3565628

[23] Martin Kocur, Florian Habler, Valentin Schwind, Paweł W. Woźniak, Christian Wolff, and Niels Henze. 2021. Physiological and Perceptual Responses to Athletic Avatars While Cycling in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445160

[24] Martin Kocur, Melanie Kloss, Valentin Schwind, Christian Wolff, and Niels Henze. 2020. Flexing Muscles in Virtual Reality: Effects of Avatars' Muscular Appearance on Physical Performance. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 193–205. https://doi.org/10.1145/3410404.3414261

[25] Jordan Koulouris, Zoe Jeffery, James Best, Eamonn O'Neill, and Christof Lutteroth. 2020. Me vs. Super(Wo)Man: Effects of Customization and Identification in a VR Exergame. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3313831.3376661

[26] Stefan Liszio, Linda Graf, and Maic Masuch. 2018. The Relaxing Effect of Virtual Nature - Immersive Technology Provides Relief in Acute Stress Situations. In *Annual Review of Cybertherapy and Telemedicine (ARCTT)*, Brenda K.

Wiederhold, Giuseppe Riva, and Stéphane Bouchard (Eds.). Interactive Media Institute, San Diego, CA, USA, 87–93.
[27] Huimin Liu, Zhiquan Wang, Christos Mousas, and Dominic Kao. 2020. Virtual Reality Racket Sports: Virtual Drills for Exercise and Training. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 566–576. https://doi.org/10.1109/ISMAR50242.2020.00084
[28] David Mal, Erik Wolf, Nina Döllinger, Carolin Wienrich, and Marc Erich Latoschik. 2023. The Impact of Avatar and Environment Congruence on Plausibility, Embodiment, Presence, and the Proteus Effect in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* (2023). https://doi.org/10.1109/TVCG.2023.3247089
[29] William F Moroney, David W Biers, F Thomas Eggemeier, and Jennifer A Mitchell. 1992. A comparison of two scoring procedures with the NASA task load index in a simulated flight task. In *Proceedings of the IEEE 1992 National Aerospace and Electronics Conference (NAECON)*. 734–740.
[30] David L Neumann, Robyn L Moffitt, Patrick R Thomas, Kylie Loveday, David P Watling, Chantal L Lombard, Simona Antonova, and Michael A Tremeer. 2018. A systematic review of the application of interactive virtual reality to sport. *Virtual Reality* 22, 3 (2018), 183–198. https://doi.org/10.1007/s10055-017-0320-5
[31] Mateus Nunes, Luciana Nedel, and Valter Roesler. 2014. Motivating People to Perform Better in Exergames: Competition in Virtual Environments. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing (SAC '14)*. Association for Computing Machinery, New York, NY, USA, 970–975. https://doi.org/10.1145/2554850.2555009
[32] Yoonsin Oh and Stephen Yang. 2010. Defining Exergames & Exergaming. In *Proceedings of Meaningful Play 2010*.
[33] Xiaolan Peng, Jin Huang, Alena Denisova, Hui Chen, Feng Tian, and Hongan Wang. 2020. A Palette of Deepened Emotions: Exploring Emotional Challenge in Virtual Reality Games. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3313831.3376221
[34] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. 2007. Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior* 10, 1 (2007), 45–56. https://doi.org/10.1089/cpb.2006.9993
[35] Claudio Robazza, Laura Bortoli, and Yuri Hanin. 2013. Perceived Effects of Emotion Intensity on Athletic Performance. *Research Quarterly for Exercise and Sport* 77, 3 (2013), 372–385. https://doi.org/10.1080/02701367.2006.10599371
[36] Christina Röcke and Daniel Grühn. 2003. German Translation of the PANAS-X. (2003). Unpublished manuscript, Free University Berlin.
[37] RootMotion. 2014–2022. Final IK. available at https://assetstore.unity.com/packages/tools/animation/final-ik-14290.
[38] Montse C Ruiz, Claudio Robazza, Asko Tolvanen, Saara Haapanen, and Joan L Duda. 2019. Coach-Created Motivational Climate and Athletes' Adaptation to Psychological Stress: Temporal Motivation-Emotion Interplay. *Frontiers in Psychology* 10 (2019). https://doi.org/10.3389/fpsyg.2019.00617
[39] Katie Seaborn and Deborah I Fels. 2015. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies* 74 (2015), 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006
[40] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *Comput. Surveys* 50, 6 (2017). https://doi.org/10.1145/3134301
[41] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B* 364 (2009), 3549–3557. https://doi.org/10.1098/rstb.2009.0138
[42] Mel Slater. 2017. Implicit Learning Through Embodiment in Immersive Virtual Reality. In *Virtual, Augmented, and Mixed Realities in Education*, Dejian Liu, Christopher J Dede, Ronghuai Huang, and John Richards (Eds.). Springer, Singapoore.
[43] Mel Slater, Vasilis Linakis, Martin Usoh, and Rob Kooper. 1996. Immersion, Presence, and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '96)*. ACM, Hong Kong, 163–172. https://doi.org/10.1145/3304181.3304216
[44] Mel Slater and Sylvia Wilbur. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence* 6, 6 (1997), 603–616. https://doi.org/10.1162/pres.1997.6.6.603
[45] Sophia C Steinhaeusser, Sebastian Oberdörfer, Sebastian von Mammen, Marc Erich Latoschik, and Birgit Lugrin. 2022. Joyful Adventures and Frightening Places - Designing Emotion-Inducing Virtual Environments. *Frontiers in Virtual Reality* (2022). https://doi.org/10.3389/frvir.2022.919163
[46] The World's Strongest Man. 2023. Hercules Hold. https://www.theworldsstrongestman.com/events/hercules-hold/.
[47] Elena Tuveri, Luca Macis, Fabio Sorrentino, Lucio Davide Spano, and Riccardo Scateni. 2016. Fitmersive Games: Fitness Gamification through Immersive VR. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '16)*. 212–215. https://doi.org/10.1145/2909132.2909287
[48] Ubisoft Entertainment. 2022. Just Dance 2023 Edition. available at https://www.ubisoft.com/en-us/game/just-dance/2023.
[49] Unity Technologies. 2022. Unity. available at https://unity.com.

[50] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1643–1652. https://doi.org/10.1109/TVCG.2018.2794629

[51] David Watson and Lee Anna Clark. 1994. *The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form*. University of Iowa.

[52] Carla Winter, Florian Kern, Dominik Gall, Marc Erich Latoschik, Paul Pauli, and Ivo Käthner. 2021. Immersive virtual reality during gait rehabilitation increases walking speed and motivation: a usability evaluation with healthy participants and patients with multiple sclerosis and stroke. *Journal of NeuroEngineering and Rehabilitation* 18, 1 (2021). https://doi.org/10.1186/s12984-021-00848-w

[53] Bob G Witmer, Christian J Jerome, and Michael J Singer. 2005. The Factor Structure of the Presence Questionnaire. *Presence* 14, 3 (2005), 298–314. https://doi.org/10.1162/105474605323384654

[54] Bob G Witmer and Michael J Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence* 7, 3 (1998), 225–240. https://doi.org/10.1162/105474698565686

Received 11 July 2023; revised 21 June 2024; accepted 30 July 2024