Automated interaction may reduce emotional reactivity in VR: a randomized study with healthy participants

Julia Elisabeth Diemer[1,2]*, Magdalena Sich[1], Bastian Lange[3], Mathias Müller[3], Marius Koller[4], Philip Schäfer[4], Gerrit Meixner[4], Alexander Brunnauer[1,5] and Peter Zwanzger[1,5]

[1]Clinical Center for Psychiatry, Psychotherapy, Psychosomatic Medicine, Geriatrics and Neurology, kbo-Inn-Salzach-Klinikum gGmbH, Wasserburg am Inn, Germany, [2]Department of Psychology, LMU Munich, Munich, Germany, [3]VTplus GmbH, Würzburg, Germany, [4]UniTyLab, Heilbronn University, Heilbronn, Germany, [5]Department of Psychiatry and Psychotherapy, LMU Munich, Munich, Germany

Introduction: Virtual reality (VR) has been investigated as a psychotherapy tool for 3 decades, with the most recent developments introducing automated self-help programs without direct therapeutic assistance. In particular, several publications have presented therapist-free VR exposure therapy for anxiety disorders as a way of improving patient access to psychological therapy. However, it is unknown whether the removal of interaction with a real person during VR exposure changes the emotional experience. The aim of this study was to test the effect of automated interaction on presence and emotion in VR. Both are important processes in VR-supported exposure therapy, with stronger emotional arousal and greater presence considered necessary for successful therapy.

Methods: This is a monocentric, randomized study comparing automated (auto) interaction in VR with experimenter-led (live) interaction during emotionally relevant VR scenarios (spiders; public speaking). The presence or absence of a psychophysiological recording device (smartwatch) was included as a second between-subjects factor. N = 64 healthy participants gave ratings of subjective emotional arousal and presence in VR.

Results: In the spider scenario, arousal, our primary outcome measure, was significantly reduced in the auto condition, with no effect of the factor smartwatch. This effect was not seen in the public speaking scenario. For presence, in both scenarios, an interaction (condition × smartwatch) was significant, with participants in the auto condition reporting greater presence if they wore a smartwatch.

Discussion: We found that the design features of our VR system critically influenced presence and emotional effects. In particular, the reduced emotional arousal in the spider scenes means that automated setups might not be as suited for exposure therapy as therapist-led VR scenarios. To our knowledge, a direct comparison between automated and natural communication during emotional exposure in VR has not been made. However, our data indicate the importance of such comparisons. Our study thus addresses an important gap in the current VR literature. Future research will

need to establish whether similar effects can be found in the target populations of exposure therapy, that is, patients with anxiety disorders, and whether such emotional effects might impact therapeutic outcomes.

Introduction

Virtual reality (VR) has been investigated as a tool in psychotherapy for roughly 30 years (Diemer, 2019; Wechsler et al., 2019; Wang et al., 2024). The most prominent application of VR in psychotherapy is exposure therapy. This is because traditional exposure therapy, which is conducted "in vivo" (i.e., in the actual situations the patients fear), is time-consuming (frequently requiring travel), and in many therapeutic settings, the resources for exposure in vivo are not available (Pittig et al., 2019). Here, VR technology offers a convenient alternative to in vivo exposure as VR exposure can be conducted in the therapist's office.

Exposure is a key component of cognitive–behavioral treatment packages for anxiety disorders. It is an evidence-based treatment and is recommended in current international guidelines as a first-line approach in anxiety disorders (Andrews et al., 2018; Bandelow et al., 2021; National Institute for Health and Care Excellence [NICE], 2014). In exposure therapy, patients learn to confront fear-evoking situations and feelings of fear and anxiety as such (Craske et al., 2008; Foa and Kozak, 1986). The aims are for the patients to experience that feared consequences do not occur, that feelings of fear and anxiety are not harmful and abate with time, and that they can trust in their ability to cope with these situations and emotions (self-efficacy; Bandura, 1977; Craske et al., 2008; Craske et al., 2018). Different theories have been put forward to explain the effectiveness of exposure therapy (Foa and Kozak, 1986; McNally, 2007; Craske et al., 2008; Craske et al., 2018), and research is underway to explain the mechanisms of change that operate during exposure therapy (Rupp et al., 2017; Craske et al., 2018). While the theories differ with regard to the proposed therapeutic mechanisms, they have in common the assumption that fear and anxiety must be activated in order for therapeutic change to take place. This is inversely confirmed by a body of research demonstrating that avoidance, that is, efforts to keep fear levels down, is detrimental to exposure therapy (Cammin-Nowak et al., 2013; Smits et al., 2004). Although the necessity to focus exclusively on the fear-provoking stimuli and to maximize fear during exposure has been questioned (Diemer et al., 2023; Hofmann and Hay, 2018; Parrish et al., 2008; Shiban et al., 2017), research indicates that sufficient emotional engagement with the feared situation and stimuli is in fact a necessary condition for successful exposure therapy (Diemer et al., 2023; Hofmann and Hay, 2018). Thus, there is broad agreement that patients with anxiety disorders must experience feelings of fear and anxiety during psychotherapy so that they can change their expectations, attitudes, and tolerance toward these emotions (Craske et al., 2008).

One of the first research questions asked of VR was, consequently, whether a medium that is, at first sight, identifiable as a computer simulation can activate sufficient emotion for exposure therapy to work. By now, there is ample evidence that VR elicits a strong, multimodal fear response. For example,

numerous studies have registered psychophysiological arousal in response to VR fear stimuli (for a review, see Diemer et al., 2014), while subjective fear ratings during VR exposure consistently show significant fear levels (Diemer et al., 2016; Shiban et al., 2017). The efficacy of VR exposure therapy for anxiety disorders has been shown in several meta-analyses, with the best evidence for specific phobias (Opris et al., 2012; Wechsler et al., 2019). VR exposure therapy for other anxiety disorders, such as panic disorder, agoraphobia, social phobia, and generalized anxiety disorder, has so far been investigated less intensively but with promising results (van Loenen et al., 2022; Wechsler et al., 2019). For specific phobias, there is also convincing evidence that treatment gains generalize from VR to real-life situations (Morina et al., 2015).

To explain the emotional effects of VR, research has focused on the concept of presence. Presence describes the illusion of a person immersed in VR that they are actually inside the VR world (Slater and Wilbur, 1997). Intuitively, presence seems to be a pre-requisite for emotional experience in VR, but research has shown that emotions—in particular, arousing emotions like fear and anxiety—in turn increase the sense of presence in VR (Bouchard et al., 2008; Gorini et al., 2011). Furthermore, results of experimental studies of the time course of emotion and presence in VR have shown significant time-lagged correlations in both directions (Diemer et al., 2015; Peperkorn et al., 2015); that is, the relationship of emotion and presence is likely bidirectional in nature. In summary, these findings mean that for exposure therapy to work, patients must experience strong feelings of fear or anxiety during exposure. In a VR setting, high levels of presence contribute to this emotional arousal. Thus, VR systems for use in exposure therapy should be designed to elicit intense fear and anxiety and strong feelings of presence, and a VR setup that elicits more emotion and stronger presence can be expected to work better in therapy.

Recently, research into VR-supported exposure therapy has turned to automated self-help programs. Current approaches include innovative efforts to individualize exposure intensity automatically during VRET based on psychophysiological signals (Mahmoudi-Nejad et al., 2024; Weber et al., 2024). Therapist-free, automated self-help programs with a VR component for the exposure part of the treatment are increasingly discussed in the literature as a possible method of scaling up therapeutic interventions and as a means of improving patient access to psychological therapy (Bentz et al., 2021; Donker et al., 2023). Evidence from clinical studies shows that automated cognitive–behavioral treatment for anxiety disorders can be effective (Donker et al., 2019; Freeman et al., 2022; for a review, see Graham et al., 2025), but there has been hardly any research into the emotional processes during automated VR exposure. In view of the importance of emotional experience for exposure therapy, this is surprising. Previous VR systems used VR as a technological aid, but the interaction of therapist and patient unfolded as it usually would

in a one-on-one therapeutic setting. There is evidence that the interaction with a therapist has important emotional and therapeutic implications, suggesting that the therapeutic alliance enhances emotional coping (Hofmann and Hay, 2018) and the efficacy of exposure therapy (Buchholz and Abramowitz, 2020). Therefore, the removal of the therapist from an automated treatment package should be investigated in terms of its emotional effect on patients and possible effects on therapy outcomes. In the context of VR, and in light of the evidence reported above concerning the interrelatedness of presence and emotion, the possible impact of automation on presence should also be experimentally tested.

This study forms part of a larger project investigating the characteristics of VR systems to determine the optimal VR setup for exposure therapy in routine care. Our aim is to contribute to the identification of relevant design features that ensure the efficacy of a VRET system. We conducted a survey of n = 58 mental health professionals as part of this project; all indicated they would like access to monitoring of psychophysiological stress parameters like heart rate or heart rate variability (Sartory, 2012) as part of an ideal VR system (Koller et al., 2018). However, little is known about the possible effects of the presence of physiological assessment technology during exposure therapy on psychological measures like anxiety levels and, in VR, presence. The phenomenon of whitecoat hypertension (Khan et al., 2007), for example, indicates that effects due to context and wearing recording devices are possible.

The overall aim of the present study was to test the effect of automation on presence and emotion in VR. To our knowledge, a direct comparison between automated and natural communication during emotional exposure in VR has not been made. Our study thus addresses an important gap in the current VR literature. We chose one important aspect of automation, which we then manipulated experimentally, namely, the modality in which instructions were delivered. Here, we compared instructions given by a real human experimenter (live) with automated instructions delivered via previously recorded audio files (auto). Healthy participants gave ratings of subjective emotional arousal in two relevant exposure situations: (1) spiders of increasing size and number and (2) an impromptu speech in front of an audience. In view of the literature on therapeutic alliance and its emotional effects (Buchholz and Abramowitz, 2020), we hypothesized that there would be lower presence and lower emotional activation in the automated vs. life condition. As a further hypothesis, we expected a positive correlation between presence and emotional arousal. As a second independent factor, we assessed whether the presence of a smartwatch as a physiological recording device affected outcomes. Because we are unaware of any previous studies investigating the emotional effects of recording devices, we had no directional hypothesis but expected, as a non-directional hypothesis, differences in emotion and presence depending on the presence/absence of the recording device.

Materials and methods

Participants

A total of N = 65 participants were included in this study. The sample size was determined by power analysis (G*power; Faul et al., 2007) based on a power of 1-β = 0.80, α = 0.05, and an expected effect of $\eta^2 = 0.12$ for the primary outcome; this analysis returned an ideal sample size of N = 60, and a further n = 5 was added to account for possible dropouts. Recruitment was achieved via mailing lists and by word of mouth. Inclusion criteria were any gender, age between 18 and 70 years, and being free (by self-report) of any relevant somatic or psychiatric disorders. Exclusion criteria were the presence of any of the following (according to self-report): epilepsy, heart disease, balance problems, any psychiatric disorder, and—in women—pregnancy. All participants gave written informed consent prior to participation. The study was approved by the local ethics committee (Ethics Committee of the Medical School of the University of Munich) and conducted in accordance with the Declaration of Helsinki in its current form.

Study design and randomization

This is a monocentric, randomized study comparing automated (auto) interaction in VR with experimenter-led (live) interaction in VR during emotionally relevant VR scenarios (spiders; public speaking). The presence or absence of a psychophysiological recording device (smartwatch) was included as a second between-subjects factor. Blinding was not possible due to the type of experimental manipulations applied. The first author (JD), who was not involved in recruitment nor data collection, generated the randomization lists (ratio: 1:1:1:1) with the online tool Research Randomizer (https://www.randomizer.org/). Randomization results (assignment to experimental conditions) were concealed in closed envelopes, which were opened by the experimenters immediately before starting the VR paradigm. The study took place at a large community psychiatric hospital.

Measures

To characterize the sample, age, gender, and employment in a therapeutic profession (yes/no) were recorded. Furthermore, all participants filled in the affinity for technology interaction (ATI) scale (Franke et al., 2019). The question about the profession and the ATI scale were included to check whether there were any significant differences between conditions with regard to therapeutical or technological affinity (control variables).

Primary outcome

The primary outcome was subjective arousal, rated on a scale from 0 (no at all aroused) to 100 (maximal arousal). Importantly, arousal ratings were collected online during the VR experience (no fadeout or other interruption of the VR experience), 10 s after the start of each new step of the respective VR scenario (cf. Figure 1). Ratings were given by participants verbally upon being prompted by the experimenter (live condition) or by an audio file (auto condition). In both conditions, the experimenter noted the participants' responses manually.

Secondary outcomes

Analogous to the arousal rating, we asked for subjective presence ratings on a scale from 0 (not at all present in the VR scenario) to 100
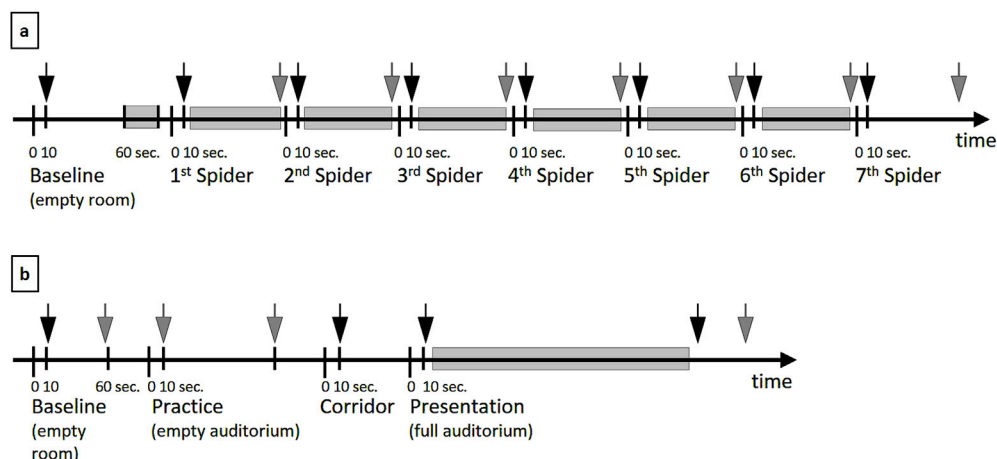
FIGURE 1
Schematic timeline of the different steps of the spider (a) and public speaking (b) scenarios, depicting events that were relevant for analysis. (a) Shaded time indicates the participant's approach to the spider(s) (duration varied between participants), ended by the participant's button press (no approach was possible in scene 7). (b) Shaded time indicates the duration of the participant's talk (variable duration). In both (a,b), arrows indicate time points of arousal and presence ratings (gray arrows denote ratings not included in analyses).

(maximal presence). Presence ratings were collected immediately after arousal ratings online during VR (cf. Figure 1).

To assure comparability to previous research, we asked participants to fill in the Igroup Presence Questionnaire (IPQ; Schubert et al., 2001) directly after experiencing the VR scene(s). This 14-item (7-point Likert scale, 0–6) measure assesses the presence on four scales (Spatial presence, five items; Involvement, four items; Experienced realism, four items; General presence, one item).

After the IPQ, participants filled in the German version of the Simulator Sickness Questionnaire (SSQ; Kennedy et al., 1993). Sixteen items (4-point Likert scale) form three subscales (Oculomotor, Disorientation, and Nausea).

Exploratory outcomes: approach and duration

In the spider scenario, participants were asked to approach the spider(s) as far as felt right for them before continuing to the next level in the spider hierarchy (or stopping the spider experience). Distance (approach) and time at this position were measured. In the public speaking scenario, we measured the duration of each participant's talk. For details, see Figure 1.

Experimental conditions

Live vs. automated (auto) interaction

(1) In the "live" condition, the experimenter gave all instructions and asked for all ratings (arousal and presence). A detailed, standardized script was used, and all prompts and instructions were read out. (2) In the "auto" condition, all instructions and rating prompts were given automatically at the appropriate moments by pre-recorded audio files (a different speaker from the experimenter). The instructions and prompts were identical in wording to the ones used in the live condition.

Smartwatch

A fully functional smartwatch (Microsoft Band 2) was applied. The watch was turned on in front of the participants and strapped to their left wrist. The watch was presented to participants in the "with smartwatch" condition as a device for monitoring heart rate and skin conductance. However, no actual physiological data were collected. In the "without smartwatch" condition, the smartwatch was neither shown nor used.

VR equipment

We used a VT+ VR system (Vtplus GmbH, Würzburg, Germany). The VR scenarios were rendered on a VR computer (compliant with EN 60601–1); the HTC Vive Business Edition was chosen as the head-mounted display (HMD). This choice is in line with recommendations for VR hardware for educational or medical applications (Mehrfard et al., 2019). The HMD used a tracking sensor to update the field of vision in accordance with head movements (refresh rate: 90 Hz). Further, the lighthouse tracking system (Vive Base Stations) was used. Two base stations communicate with a sensor in the HMD via infrared pulses (60 Hz). Between them, they cover an area of 1.6 m by 3.4 m, in which the VR user can move freely. The VR scenarios were developed with the Source SDK (Valve Corporation, Bellevue, Washington, USA) based modification (VrSessionMod 0.6, Vtplus GmbH) and controlled by the VR software CyberSession (CS 5.6, Vtplus GmbH, Würzburg, Germany; see www.cybersession.info). An Apple iPad APP (CS-Remote 1.1, Vtplus GmbH) was used as a remote terminal to operate the software. VR background soundscapes were delivered to participants' headphones (Vive Deluxe Audio Strap HS600).

VR scenarios

Spiders

Participants were invited to expose themselves to a series of seven spider scenes of increasing intensity. Each scene consisted of the same office scenario. First, the office was presented as a neutral baseline. Scenes 1–3 included one spider of different sizes (small, medium, and large); scenes 4–6 each included two to four spiders, again of increasing sizes. Arousal and presence ratings were taken 10 s into each scene. Then, participants received the following instruction: "Please look out for a spider. Move toward the spider and position yourself at the distance that is just about pleasant for you. Once you have reached this distance, please press the button of the controller." Scene 7 differed from the previous scenes in that the participants were surrounded by spiders of different sizes, so participants were instructed to stay at their position and observe the spiders until they wished to leave the scene. See Supplementary Material 1 for screenshots of all spider scenes.

Public Speaking

In the public speaking scenario, participants were asked to give an impromptu talk about their hometown. The scenario started in an empty lobby (baseline scenario). Participants were then immersed in a virtual, empty auditorium and asked to practice their talk (five virtual presentation slides with appropriate captions were provided). When participants indicated they had sufficiently practiced, they were immersed in a VR corridor that led to the auditorium (this time with an audience; see Appendix 2 for a screenshot of the auditorium). They were asked to give their talk, again with the slides. When the penultimate slide came up ("Any questions?"), a member of the VR audience raised his hand. Participants answered the question and were told that their talk was over, and the VR scene faded out. Ratings (arousal, presence) were taken 10 s after the start of baseline and corridor phases, upon entering the full auditorium, and immediately after a member of the audience raised his hand to ask a question.

Procedure

Individual appointments were made, and participants were invited to the VR laboratory on site. They were randomly assigned to either automated or live instructions and to whether the smartwatch was applied (see above). Participants were given a choice of experiencing the spider scenario, the public speaking scenario, or both. If they chose both scenarios, they were free to choose the order. After VR exposure, participants filled in questionnaires (IPQ, SSQ) and were briefly interviewed about their experience and debriefed.

Data preparation and analysis

The primary analytical strategy was mixed rm ANOVA with the between-subjects factors condition (auto vs. live) and smartwatch (with vs. without smartwatch), and the within-subjects factor time. Only data from participants who completed the respective scenario

were included in the analyses. Greenhouse–Geisser correction was applied where necessary. Follow-up ANOVAs were conducted to clarify the sources of significant main effects or interactions. For correlation analyses, mean scores of arousal and presence during the spider and public speaking scenarios (exposure) were calculated. To assess normality, we conducted Kolmogorov–Smirnov tests of the mean ratings of arousal and presence, separately for the spider and public speaking samples and separately in each condition. These tests returned no indications of a violation of the normality assumption (all p`s > 0.05).

VR log files were analyzed for the analysis of objective behavior data. These files logged time, participant position, and markers that indicated relevant events (arousal and presence ratings and participant button presses). Duration (seconds) and approach behavior (meters) were computed from the log files.

The software package IBM SPSS 25 was used for all analyses.

Results

Study participants

One participant abandoned the study before exposure to either scenario, so valid data are available from n = 64 participants. Demographic and baseline characteristics of this study sample are summarized in Table 1. There was a broad age range (min. 18, max. 61 years). There was a large overlap in participants who took part in and finished both scenarios (n = 52). There were n = 2 participants who only completed the spider scenario and n = 3 participants who only took part in the public speaking scenario. There were n = 7 participants who started the spider scenario but did not finish it. All n = 62 participants who started the public speaking scenario finished it.

Spider scenario

Complete data were available from n = 54 participants who exposed themselves to all seven spider scenarios. The baseline characteristics of this subsample are summarized in Table 2. There were no significant differences in baseline variables between subsamples, except for age (F-test in univariate ANOVA, interaction condition × smartwatch, p = 0.013). Because age was not correlated with outcome (see below, Table 5), no statistical correction was applied for age.

Primary outcome (arousal ratings, see Figure 2): rm ANOVA returned a significant main effect of time ($F_{2.459,122.95} = 44.706$; $p < 0.001$; partial $\eta^2 = 0.472$) and a significant main effect of the factor condition ($F_{1,50} = 7.411$; $p = 0.009$; partial $\eta^2 = 0.129$). The main effect of the factor smartwatch was not significant ($F_{1,50} = 0.237$; $p = 0.628$; partial $\eta^2 = 0.005$). The interaction effect of the two between-subject factors (condition × smartwatch) was not significant ($F_{1,50} = 0.247$; $p = 0.621$; partial $\eta^2 = 0.005$). The interaction of time × condition ($F_{2.459,122.95} = 0.998$; $p = 0.384$; partial $\eta^2 = 0.02$), the interaction of time × smartwatch ($F_{2.459,122.95} = 0.685$; $p = 0.535$.; partial $\eta^2 = 0.014$), and the triple interaction time × condition × smartwatch ($F_{2.459,122.95} = 2.211$; $p = 0.102$.; partial $\eta^2 = 0.042$) were not significant.

TABLE 1 Demographic and baseline characteristics of the study sample.

	Study sample (n = 64)		Condition: Live (n = 32)				Condition: Auto (n = 32)			
			w. smartwatch (n = 15)		w/o smartwatch (n = 17)		w. smartwatch (n = 15)		w/o smartwatch (n = 17)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	34.72	12.84	30.0	8.49	33.59	12.66	43.13	15.0	32.59	11.64
ATI score[a]	3.58	1.02	3.94	0.80	3.69	1.25	3.70	0.82	3.03	0.99
	N	%	N	%	N	%	N	%	N	%
Female	43	67.2	10	66.7	10	58.8	8	53.3	15	88.2
Therapeutic profession	38	59.4	8	53.3	10	58.8	10	66.7	10	66.7

[a]ATI, affinity toward technology interaction (n = 63), range, 1.0–6.0.

TABLE 2 Demographic and baseline characteristics of the spider sample.

	Spider sample (n = 54)		Condition: Live (n = 28)				Condition: Auto (n = 26)			
			w. smartwatch (n = 12)		w/o smartwatch (n = 16)		w. smartwatch (n = 13)		w/o smartwatch (n = 13)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	35.61	13.0	30.58	8.78	34.38	12.64	45.31	14.92	32.08	10.52
ATI score[a]	3.65	1.0	3.87	0.81	3.72	1.29	3.85	0.76	3.19	0.93
	N	%	N	%	N	%	N	%	N	%
Female	36	66.7	9	75.0	9	56.3	7	53.8	11	84.6
Therapeutic profession	34	63.0	7	58.3	10	62.5	9	69.2	8	61.5

[a]ATI, Affinity toward technology interaction (n = 53), range, 1.0–6.0.

Follow-up: Because none of the interaction effects reached statistical significance, post hoc F-tests were conducted as univariate ANOVAs on pooled data. This way, a comparison of arousal at baseline with all subsequent seven spider scenes revealed a highly significant increase of arousal in all spider scenes (rm ANOVA, all p-values <0.01). Likewise, the difference in arousal between conditions (live vs. auto) reached statistical significance at all time points (F-Tests in ANOVA, all p-values <0.05), with stronger arousal in the live condition (cf. Figure 2).

Secondary outcome

Presence ratings: rm ANOVA returned a significant main effect of time ($F_{4.459,222.938} = 6.761$; $p < 0.001$; partial $\eta^2 = 0.119$), while neither the factor condition ($F_{1,50} = 2.863$; $p = 0.097$; partial $\eta^2 = 0.054$) nor the factor smartwatch returned a significant main effect ($F_{1,50} = 0.528$; $p = 0.471$; partial $\eta^2 = 0.010$). The interaction of time × condition ($F_{4.459,222.938} = 0.844$; $p = 0.509$; partial $\eta^2 = 0.017$), the interaction of time × smartwatch ($F_{4.459,222.938} = 0.798$; $p = 0.539$.; partial $\eta^2 = 0.016$), and the triple interaction time × condition × smartwatch ($F_{4.459,222.938} = 1.320$; $p = 0.260$.; partial $\eta^2 = 0.026$) were not significant. However, the interaction condition × smartwatch was highly significant ($F_{1,50} = 12.743$; $p = 0.001$; partial $\eta^2 = 0.203$).

The follow-up test to clarify the sources of the main effect of time on presence was carried out in a manner analogous to that for arousal. These tests returned no significant differences for the first three spider levels vs. baseline and a significant increase of presence during spider scenes 4, 5, 6, and 7 over baseline (F-Test in rm ANOVA, p-values <0.05). To clarify the interaction of condition × smartwatch, we pooled the data over all time points and compared presence in the smartwatch group vs. no smartwatch group separately in the two conditions (live vs. auto). This ANOVA returned a significant effect of the factor smartwatch in both the live condition ($F_{1,26} = 7.017$, $p = 0.014$, partial $\eta^2 = 0.213$) and the auto condition ($F_{1,24} = 6.233$, $p = 0.020$, partial $\eta^2 = 0.206$). Interestingly, the effect had different directions: in the live condition, wearing a smartwatch was associated with significantly reduced presence ratings, while in the auto condition, the smartwatch was associated with significantly elevated presence ratings (cf. Figure 3).

Public speaking scenario

Complete data were available from n = 62 participants who exposed themselves to the entire public speaking scenario.
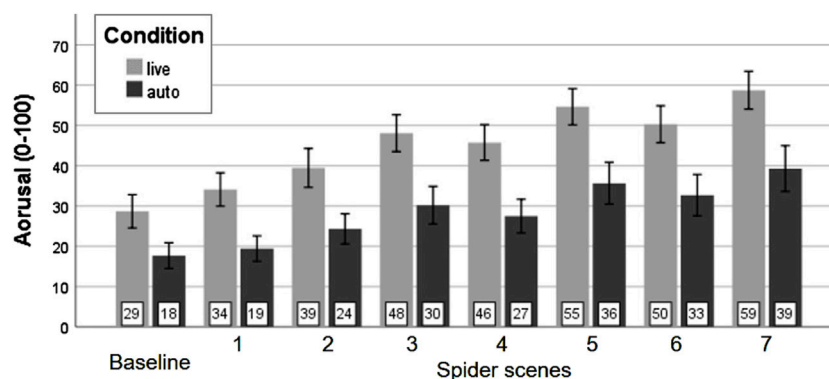
FIGURE 2
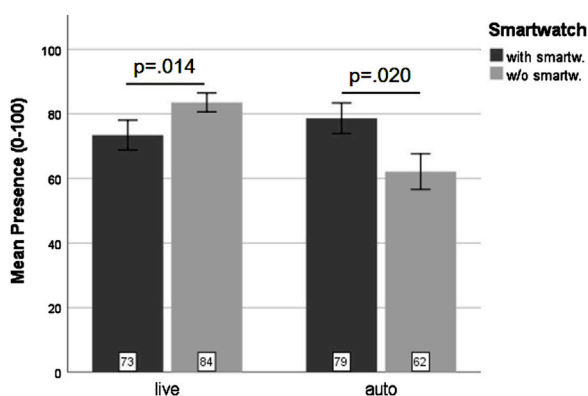Subjective arousal during VR spider scenes. Error bars = one standard error of the mean.

FIGURE 3
Mean presence ratings in the spider scenes collapsed across all time points. Error bars = one standard error of the mean.

Characteristics are summarized in Table 3. Both age (F-test in univariate ANOVA, main effect condition, p = 0.015) and ATI score (F-test in univariate ANOVA, main effect condition, p = 0.036) differed between the live and auto conditions. Again, due to a lack of correlation with outcome (see Table 6), no statistical correction was used for age or ATI. Gender distribution across conditions was not regular (Fisher's exact test, p = 0.049). Therefore, analyses were repeated with gender as an additional factor (see below).

Primary outcome (arousal ratings): rm ANOVA returned a significant main effect of time ($F_{2.173,126.038}$ = 35.271; p < 0.001; partial η^2 = 0.378). There were no other significant effects of the factor condition ($F_{1,58}$ = 2.375; p = 0.129; partial η^2 = 0.039), the factor smartwatch ($F_{1,58}$ = 0.177; p = 0.676; partial η^2 = 0.003), or the interaction condition × smartwatch ($F_{1,58}$ = 0.415; p = 0.522; partial η^2 = 0.007). Likewise, the interaction of time × condition ($F_{2.173,126.038}$ = 1.705; p = 0.183; partial η^2 = 0.029), the interaction of time × smartwatch ($F_{2.173,126.038}$ = 1.124; p = 0.332.; partial η^2 = 0.019), and the triple interaction time × condition × smartwatch ($F_{2.173,126.038}$ = 0.422; p = 0.673.; partial η^2 = 0.007) were not significant.

Follow-up: None of the interaction effects reached statistical significance, so we again conducted post hoc F-tests as univariate rm-ANOVAs on pooled data. The comparison of arousal at baseline with all subsequent time points of the public speaking scenario revealed a highly significant increase of arousal over baseline at all subsequent time points (F-tests in rm ANOVA, all p-values <0.001). Furthermore, arousal levels of the two ratings taken while the participants were in front of the VR audience were significantly higher than arousal during navigation along the corridor (F-tests in rm ANOVA, both p-values <0.01). There was no significant difference in arousal between the two ratings taken in front of the virtual audience. These post hoc tests confirm a marked increase in arousal after baseline, peaking during the actual public speaking scene (cf. Figure 4).

Secondary outcome

Presence ratings: rm ANOVA returned a significant main effect of time ($F_{2.061,119.565}$ = 10.440; p < 0.001; partial η^2 = 0.153), while no other main effects reached significance (condition: $F_{1,58}$ = 1.904, p = 0.173, partial η^2 = 0.032; smartwatch: $F_{1,58}$ = 1.258, p = 0.267, partial η^2 = 0.021). There was a significant interaction of the factors condition × smartwatch ($F_{1,58}$ = 4.770, p = 0.033, partial η^2 = 0.076), while none of the other interactions reached statistical significance (time × condition: $F_{2.061,119.565}$ = 0.743, p = 0.481, partial η^2 = 0.013; time × smartwatch: $F_{2.061,119.565}$ = 0.410, p = 0.671, partial η^2 = 0.007; time × condition × smartwatch: $F_{2.061,119.565}$ = 0.450, p = 0.645, partial η^2 = 0.008).

Follow-up: No interaction effect including the factor time was significant, so we used pooled data for post-hoc F-tests of the time effect. Presence at baseline turned out to be significantly lower than at all subsequent time points of the public speaking scenario (F-tests in rm ANOVA, all p-values <0.01). Furthermore, there was no significant difference between presence ratings during the passage and at the beginning of the speech, but presence ratings at the end of the speech were significantly higher than during the passage and at the beginning of the speech (F-Tests in rm ANOVA, both p-values <0.05.). Thus, presence increased significantly over baseline during the entire scenario and reached a peak at the end of the speech (when the agent in the audience raised his arm to ask a question).

TABLE 3 Demographic and baseline characteristics of the public speaking sample.

	Public speaking sample (n = 62)		Condition: Live (n = 30)				Condition: Auto (n = 32)			
			w. smartwatch (n = 15)		w/o smartwatch (n = 15)		w. smartwatch (n = 15)		w/o smartwatch (n = 17)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	34.08	12.53	30.0	8.49	30.8	10.61	43.13	15.0	32.59	11.64
ATI score[a]	3.61	1.02	3.94	0.80	3.86	1.23	3.70	0.82	3.03	0.99
	N	%	N	%	N	%	N	%	N	%
Female	42	67.7	10	66.7	9	60.0	8	53.3	15	88.2
Therapeutic profession	36	58.1	8	53.3	8	53.3	10	66.7	10	58.8

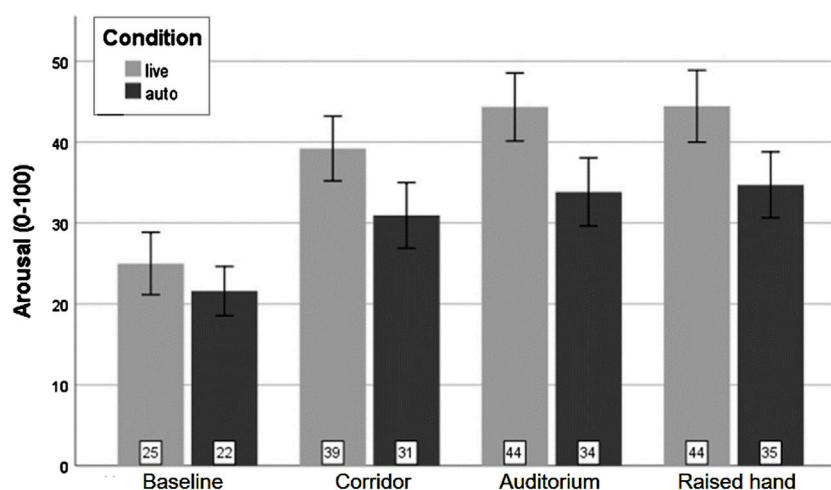[a]ATI, Affinity toward technology interaction, range, 1.0–6.0.

FIGURE 4
Subjective arousal during the VR public speaking scenario. Error bars = one standard error of the mean.

To clarify the interaction of condition × smartwatch, we pooled the data over all time points and compared presence in the smartwatch group vs. no smartwatch group separately in the two conditions (live vs. auto). This ANOVA returned a significant effect of the factor smartwatch in the auto condition ($F_{1,30} = 4.329$, $p = 0.046$, partial $\eta^2 = 0.126$) but not in the live condition ($F_{1,28} = 0.809$, $p = 0.376$, partial $\eta^2 = 0.028$). This means that in the auto condition, wearing a smartwatch was associated with significantly elevated presence ratings, while in the live condition, there was no effect of the smartwatch on presence (cf. Figure 5).
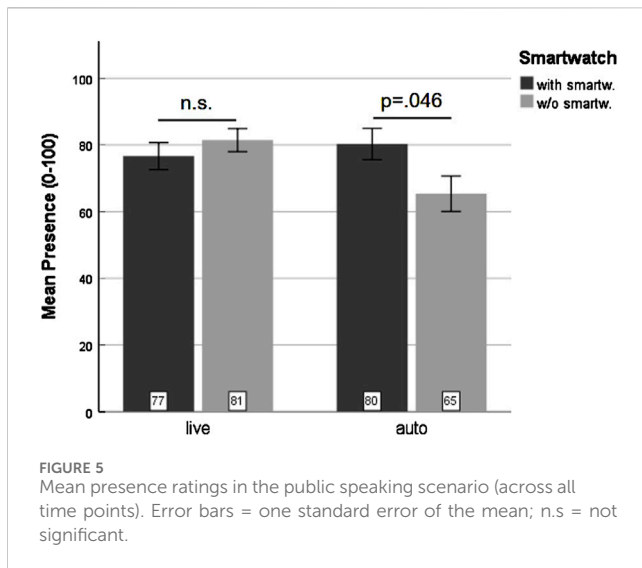
Repetition of these ANOVAs (arousal and presence ratings in the public speaking scenario) with gender as an additional factor did not change the pattern of results. For both arousal and presence, the time effect remained highly significant. The interaction condition × smartwatch in the presence ratings also remained significant. Thus, the results are robust despite differences in gender frequencies across the individual conditions.

Objective behavior data

Complete marker data were available for n = 49 participants in the spider scenario and n = 57 in the public speaking scenario.

Approach: rm ANOVA of approach in the six spider scenes returned a significant main effect of time ($F_{3.203, 147.341} = 22.062$, $p < 0.001$, partial $\eta^2 = 0.324$), while none of the other factors or interactions reached significance. Post hoc comparisons (ANOVA) of the different spider scenarios showed that the approach was significantly greater in spider scenario 1 than in any of the subsequent scenarios (all p-values <0.05). On average, participant approach distances were 3.06 m in scenario 1 and 2.92 m, 2.32 m, 2.81 m, 2.69 m, and 2.27 m in scenarios 2, 3, 4, 5, and 6, respectively. The maximum approach possible was 3.54 m.

Duration: rm ANOVA of the duration of the approach in the spider scenarios resulted in a small but significant main effect of the factor time ($F_{4.070, 183.149} = 3.098$, $p = 0.027$, partial $\eta^2 = 0.058$) and a significant main effect of the factor condition ($F_{1,45} = 8.067$, $p = 0.007$, partial $\eta^2 = 0.152$). Neither the main effect of the factor

FIGURE 5
Mean presence ratings in the public speaking scenario (across all time points). Error bars = one standard error of the mean; n.s = not significant.

smartwatch nor any of the interactions reached significance. Post hoc tests of the factor time (on pooled data) showed that few scenarios did, in fact, differ significantly; that is, scenario 3 differed from scenarios 1, 4, and 5 (all p-values >0.05), and scenario 6 differed from scenarios 1, 4, and 5 (all p-values >0.05). In essence, approach duration was shortest in scenarios 3 and 6 (mean 54.13 s and 55.24 s, respectively), with values in the other scenarios between 57.32 s (scenario 2) and 59.57 s (scenario 1). Post hoc univariate ANOVA of the factor condition (conducted on the means over time in both conditions) resulted in a mean difference between conditions of 8.09 s, with participants in the auto condition taking longer to approach the spiders (F = 7.272, p = 0.010). Univariate ANOVA of the duration in spider scenario 7 (no approach possible) returned no significant effects or interactions (all p-values >0.25). The mean duration was slightly less than 2 min (109.82 s, SD = 74.67 s).

In the public speaking scenario, univariate ANOVA of the duration of the talk returned no significant effects (all p-values >0.5). The mean duration was nearly 3 min (175.48 s, SD = 64.80 s).

Questionnaire data

Descriptive statistics for the IPQ (presence) and the SSQ (simulator sickness) are presented in Table 4. In contrast to online subjective presence ratings (see above), MANOVA of IPQ returned no significant differences by the factors conditions or smartwatch, nor an interaction of the two factors, on the IPQ scores (all p-values >0.05). MANOVA of the SSQ scores likewise resulted in no significant effects of either factor (all p-values >0.1).

Correlation analyses

See Tables 5, 6 for Pearson correlations of age, ATI Score, and presence/arousal ratings at baseline and during exposure. As expected, subjective ratings at baseline strongly predicted

subsequent ratings during exposure (presence and arousal, spiders, and public speaking). The correlation between presence and arousal ratings during exposure was significant only in the spider scenario but only of moderate strength.

Discussion

In contrast to other research on automated VR exposure programs, which focused on the comparison between VR vs. *in vivo* (Miloff et al., 2019), our project contrasted different design features of VR (automation; psychophysiological measurement device) and analyzed in detail the effects of these features, alone and in combination, on critical emotional and presence responses. This perspective is vital for the identification of the necessary and sufficient design features for a VRET system to be effective.

Our manipulation of (1) interaction (live vs. auto) and (2) presence/absence of a psychophysiological recording device (smartwatch) resulted in an interesting pattern of results. In the spider scenario, arousal, our primary outcome measure, was significantly stronger—by approximately 50%—in the live condition, with no effect of the factor smartwatch. This effect was not found in the public speaking scenario. For presence, however, we found an interaction of both factors in either scenario, with the smartwatch increasing presence levels in the auto condition only. In the spider scenario, we additionally found the inverse effect in the live condition, where presence was significantly higher if no smartwatch was used. In both VR scenarios, arousal and presence levels increased over time. This effect was expected, as emotional triggers in both scenarios provided increasing intensity over time (cf. Diemer et al., 2015).

On the other dependent variables (objective behavior, IPQ, and SSQ), we found no further effects of the experimental manipulations, except that participants in the auto condition took longer to approach the spiders than participants in the live condition. On its own, this result is difficult to interpret, as a longer approach could imply either more careful/fearful approach behavior or, in contrast, less fear and greater leisure during spider exposure.

Taken together, our results imply that replacing natural communication with the experimenter (live) with previously recorded audio messages (auto) may reduce emotional involvement (arousal, presence) in emotionally relevant scenarios (spiders; public speaking). For presence, this negative effect might, in part, be offset by introducing psychophysiological monitoring, that is, a second interface between the participant and the VR system.

Previous studies of therapist-free VR exposure systems reported high presence levels but provided no comparison of varying degrees of automation (Donker et al., 2023; Lindner et al., 2020), thus falling short of a relative assessment of presence levels in comparison to less automated VR setups. However, our data indicate the importance of such comparisons. The effect of wearing a smartwatch on presence levels was unexpected and difficult to interpret. Further work, especially involving patient populations and replications, is needed.

The importance (for presence and emotion) of interaction with an experimenter or therapist warrants an explanation. In general, presence and emotion are known to correlate if the emotion in question is highly arousing (cf. Diemer et al., 2015). The fact that in

TABLE 4 IPQ and SSQ scores.

| | Study sample (n = 64) | | Condition: Live (n = 32) | | | | Condition: Auto (n = 32) | | | |
| | | | w. smartwatch (n = 15) | | w/o smartwatch (n = 17) | | w. smartwatch (n = 15) | | w/o smartwatch (n = 17) | |
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
IPQ Spatial[a]	22.28	5.44	23.07	4.85	23.82	4.14	22.33	5.55	20.00	6.58
IPQ Involv[b]	15.33	5.19	14.67	5.52	16.06	4.19	15.73	5.01	14.82	6.21
IPQ Realism[c]	11.69	4.90	12.53	5.54	12.71	4.63	11.2	4.51	10.35	4.94
IPQ General[d]	4.31	1.31	4.47	1.25	4.59	1.23	3.93	1.53	4.24	1.25
SSQ Nausea[e]	19.38	17.16	14.63	13.90	23.57	19.12	17.81	12.42	20.76	21.13
SSQ Oculom[f]	29.02	22.66	20.21	15.07	29.87	24.92	36.38	24.68	29.43	23.35
SSQ Disor[g]	44.59	40.16	35.26	31.51	40.94	35.99	57.54	49.04	45.04	42.77
SSQ Total[h]	33.89	25.27	25.18	16.75	34.98	26.34	40.39	26.79	34.76	28.87

[a]IPQ, spatial presence, range: 0–30.
[b]IPQ, involvement, range: 0–24.
[c]IPQ, experienced realism, range: 0–24.
[d]IPQ, general presence, range: 0–6.
[e]SSQ, nausea, range: 0–200.34.
[f]SSQ, oculomotor, range: 0–159.18.
[g]SSQ, disorientation, range: 0–292.32.
[h]SSQ, total score, range: 0–235.62.

TABLE 5 Correlation analysis (Pearson), spider scenario (n = 54).

	ATI score	BL arousal	Expo arousal	BL presence	Expo presence
Age	−0.026	−0.132	−0.169	0.080	0.106
ATI score		−0.046	−0.023	0.257	0.244
Baseline arousal			0.785**	0.049	0.119
Exposure arousal				0.140	0.298*
Baseline presence					0.871**

ATI, affinity toward technology interaction (n = 53).
**p < 0.01 (two-tailed).
*p < 0.05 (two-tailed).

TABLE 6 Correlation analysis (Pearson), public speaking scenario (n = 62).

	ATI score	BL arousal	Expo arousal	BL presence	Expo presence
Age	−0.001	−0.070	−0.194	0.212	0.020
ATI score		−0.083	−0.100	0.169	0.234
Baseline arousal			0.746**	0.016	0.084
Exposure arousal				−0.087	0.020
Baseline presence					0.790**

ATI, affinity toward technology interaction.
**p < 0.01 (two-tailed).
*p < 0.05 (two-tailed).

our sample, presence and arousal ratings correlated only moderately and only in the spider scenario might be due to the non-clinical nature of our sample, as the clearest evidence for a correlation of presence and emotion comes from patient samples with strong emotional arousal during exposure (Price and Anderson, 2007; Diemer et al., 2015). We previously found evidence that mildly distracting small talk during exposure therapy for specific phobia (compared to conversation focusing on fear-related stimuli) did not reduce but might even enhance processing of the exposure stimuli (Diemer et al., 2023), indicating a positive effect of natural communication/interaction irrespective of content on sensory (and possibly emotional) processing. Possibly, the interaction with a "real" person (i.e., the experimenter) during a VR experience might enhance emotional reactions, just as shared emotional experience enhances social connection (Cheong et al., 2023). Interestingly, we did not find a detrimental effect of automated interaction in the public speaking scenario. One might speculate that the public speaking task included a number of social stimuli, which might have buffered the effect of the experimental manipulation. Another possible explanation might be that the public speaking task was much more cognitively demanding than the spider scenario, thus reducing participants' attentional capacity and reactivity toward the kind of interaction provided. Further studies should investigate a range of exposure scenarios to test whether the emotional effects of automated vs. live interaction are, in fact, modulated by scenario characteristics like cognitive load.

To date, few studies have investigated the design features of (automated) VR scenarios for exposure therapy (Lindner et al., 2020; Wei et al., 2023). Given the increasing development of automated VR therapy programs (Graham et al., 2025), this is surprising, particularly because the few studies that did look into the effects of different setups show that design features matter. For example, Wei et al. (2023) reported that increasing the affective expression of an avatar therapist during VR exposure therapy for acrophobia resulted in a better therapeutic alliance. Our study adds to this literature by highlighting the need for more emotional interaction during automated VR exposure. Our results might also inform other medical VR applications, such as VR in medical training (Kim and Kim, 2023), pain control (Gopalan et al., 2025; Rooney et al., 2025), or neurorehabilitation (Kenea et al., 2025).

Limitations

Some limitations must be considered in this study. Most importantly, we included only healthy participants, so we do not know whether our results would generalize to patient samples. Further research is needed to clarify the possible effects of automation in psychiatric populations and whether there are detrimental effects on emotional experience and therapeutic outcomes in VR exposure therapy. While the instructions were identical in either condition, it should be noted that the experimenter in the live condition behaved naturally, leading to possible differences in tone, speed, and other aspects of language that we did not record or analyze. However, we consider this a necessary feature of the comparison (natural interaction vs. automated interaction) we

were interested in. The fact that the two scenarios (spiders; public speaking) had varying cognitive demands possibly introduced greater error variance, which might have contributed to the differing results; including a second scenario related to specific phobia (e.g., heights or air travel) rather than the public speaking task might have produced more uniform results. Likewise, we only assessed two specific aspects of automation, so our results are limited to the particular conditions we compared. Nonetheless, our observations underline the need to carefully test design elements of automated (and, indeed, any) VR setups. Furthermore, due to the combination of two between-subject factors, cell sizes are relatively small in our study, so some differences between subgroups (especially with regards to gender in the public speaking scenario) became apparent, which we countered with post hoc tests. While our results are stable, future research might like to specifically assess possible gender differences that we did not focus on.

Conclusion

VR design features in our study strongly influenced presence and emotion. Our most relevant finding in this regard is the drastically reduced experience of emotional arousal when facing VR spiders in the automated vs. live conditions. This means that automated setups might not be as suited for exposure therapy as therapist-led VR scenarios. To our knowledge, a direct comparison between automated and natural communication during emotional exposure in VR has not been made. However, our data indicate the importance of such comparisons. Our study thus addresses an important gap in the current VR literature. Future research will need to establish whether similar effects can be found in the target populations of exposure therapy, that is, patients with anxiety disorders. In general, our results show that novel VR approaches should be carefully and critically evaluated. The increasing variety of design features of VR systems, together with the growing importance of automated VR programs, calls for comparative studies to develop robust and clinically effective VR setups. This way, research will be able to identify the key ingredients of a "best practice" VR and to guide future developments and implementations of VRET in routine care.

Data availability statement

The datasets presented in this article are not readily available because of privacy or ethical restrictions but are available from the corresponding author upon reasonable request. Requests to access the datasets should be directed to julia.diemer@kbo.de.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the Medical School of the University of Munich. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JD: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft. MS: Conceptualization, Data curation, Investigation, Methodology, Writing – review and editing, Validation. BL: Conceptualization, Software, Writing – review and editing. MM: Conceptualization, Funding acquisition, Resources, Software, Supervision, Writing – review and editing. MK: Writing – review and editing. PS: Writing – review and editing. GM: Funding acquisition, Supervision, Writing – review and editing. AB: Writing – review and editing. PZ: Resources, Supervision, Writing – review and editing.

Funding

Acknowledgments

Conflict of interest

MM is a shareholder and executive officer of VTplus GmbH. BL was employed by VTplus GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2025.1536968/full#supplementary-material

References

Andrews, G., Bell, C., Boyce, P., Gale, C., Lampe, L., Marwat, O., et al. (2018). Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the treatment of panic disorder, social anxiety disorder and generalised anxiety disorder. *Aust. N. Z. J. Psychiatry* 52 (12), 1109–1172. doi:10.1177/0004867418799453

Bandelow, B., Aden, I., Alpers, G. W., Benecke, A., Beutel, M. E., Deckert, J., et al. (2021). Deutsche S3-Leitlinie Behandlung von Angststörungen. *Version 2*. Available online at: http://www.awmf.org/leitlinien/detail/ll/051-028.html (Accessed September 13, 2024).

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84 (2), 191–215. doi:10.1037//0033-295x.84.2.191

Bentz, D., Wang, N., Ibach, M. K., Schicktanz, N. S., Zimmer, A., Papassotiropoulos, A., et al. (2021). Effectiveness of a stand-alone, smartphone-based virtual reality exposure app to reduce fear of heights in real-life: a randomized trial. *NPJ Digit. Med.* 4 (1), 16. doi:10.1038/s41746-021-00387-7

Bouchard, S., St-Jacques, J., Robillard, G., and Renaud, P. (2008). Anxiety increases the feeling of presence in virtual reality. *Presence* 17 (4), 376–391. doi:10.1162/pres.17.4.376

Buchholz, J. L., and Abramowitz, J. S. (2020). The therapeutic alliance in exposure therapy for anxiety-related disorders: a critical review. *J. Anxiety Disord.* 70, 102194. doi:10.1016/j.janxdis.2020.102194

Cammin-Nowak, S., Helbig-Lang, S., Lang, T., Gloster, A. T., Fehm, L., Gerlach, A. L., et al. (2013). Specificity of homework compliance effects on treatment. Outcome in CBT: evidence from a controlled trial on panic disorder and agoraphobia. *J. Clin. Psychol.* 69 (6), 616–629. doi:10.1002/jclp.21975

Cheong, J. H., Molani, Z., Sadhukha, S., and Chang, L. J. (2023). Synchronized affect in shared experiences strengthens social connection. *Commun. Biol.* 6, 1099. doi:10.1038/s42003-023-05461-2

Craske, M. G., Hermans, D., and Vervliet, B. (2018). State-of-the-art and future directions for extinction as a translational model for fear and anxiety. *Philos. Trans. R. Soc. B* 373, 20170025. doi:10.1098/rstb.2017.0025

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., and Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behav. Res. Ther.* 46, 5–27. doi:10.1016/j.brat.2007.10.003

Diemer, J. (2019). Das Potenzial der virtuellen Realität in der Verhaltenstherapie. *Nervenheilkunde* 38, 553–559. doi:10.1055/a-0928-0734

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., and Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Front. Psychol.* 6, 26. doi:10.3389/fpsyg.2015.00026

Diemer, J., Lohkamp, N., Mühlberger, A., and Zwanzger, P. (2016). Fear and physiological arousal during a virtual height challenge – effects in patients with acrophobia and healthy controls. *J. Anxiety Disord.* 37, 30–39. doi:10.1016/j.janxdis.2015.10.007

Diemer, J., Mühlberger, A., Pauli, P., and Zwanzger, P. (2014). Virtual reality exposure in anxiety disorders: impact on psychophysiological reactivity. *World J. Biol. Psychiatry* 15 (6), 427–442. doi:10.3109/15622975.2014.892632

Diemer, J., Mühlberger, A., Yassouridis, A., and Zwanzger, P. (2023). Distraction versus focusing during VR exposure therapy for acrophobia: a randomized controlled trial. *J. Behav. Ther. Exp. Psychiatry* 81, 101860. doi:10.1016/j.jbtep.2023.101860

Donker, T., Cornelisz, I., van Klaveren, C., van Straten, A., Carlbring, P., Cuijpers, P., et al. (2019). Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized clinical trial. *JAMA Psychiatry* 76 (7), 682–690. doi:10.1001/jamapsychiatry.2019.0219

Donker, T., Fehribach, J. R., van Klaveren, C., Cornelisz, I., Toffolo, M. B. J., van Straten, A., et al. (2023). Automated mobile virtual reality cognitive behavior therapy for

aviophobia in a natural setting: a randomized controlled trial. *Psychol. Med.* 53, 6232–6241. doi:10.1017/S0033291722003531

Faul, F., Erdfelder, E., Lang, A. –G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi:10.3758/BF03193146

Foa, E. B., and Kozak, M. J. (1986). Emotional processing of fear: exposure to corrective information. *Psychol. Bull.* 99, 20–35. doi:10.1037/0033-2909.99.1.20

Franke, T., Attig, C., and Wessel, D. (2019). A personal resource for technology interaction: development and validation of the Affinity for Technology Interaction (ATI) scale. *Int. J. Hum. Comput. Interact.* 35 (6), 456–467. doi:10.1080/10447318.2018.1456150

Freeman, D., Lambe, S., Kabir, T., Petit, A., Rosebrock, L., Yu, L.-M., et al. (2022). Automated virtual reality therapy to treat agoraphobic avoidance and distress in patients with psychosis (gameChange): a multicentre, parallel-group, single-blind, randomised, controlled trial in England with mediation and moderation analyses. *Lancet Psychiatry* 9, 375–388. doi:10.1016/S2215-0366(22)00060-8

Gopalan, R., Pande, H., Narayanan, S., and Chinnaswami, A. (2025). Virtual reality as a nonpharmacological tool for acute pain management: a scoping review. *Innov. Clin. Neurosci.* 22 (1–3), 28–50.

Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., and Riva, G. (2011). The role of immersion and narrative in mediated presence: the virtual hospital experience. *Behav. Soc. Netw.* 14 (3), 99–105. doi:10.1089/cyber.2010.0100

Graham, W. M., Drinkwater, R., Kelson, J., and Kabir, M. A. (2025). Self-guided virtual reality therapy for anxiety: a systematic review. *Int. J. Med. Inf.* 200, 105902. doi:10.1016/j.ijmedinf.2025.105902

Hofmann, S. G., and Hay, A. C. (2018). Rethinking avoidance: toward a balanced approach to avoidance in treating anxiety disorders. *J. Anxiety Disord.* 55, 14–21. doi:10.1016/j.janxdis.2018.03.004

Kenea, C. D., Abessa, T. G., Lamba, D., and Bonnechère, B. (2025). Immersive virtual reality in stroke rehabilitation: a systematic review and meta-analysis of its efficacy in upper limb recovery. *J. Clin. Med.* 14, 1783. doi:10.3390/jcm14061783

Kennedy, R. S., Lane, N. B., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3 (3), 203–220. doi:10.1207/s15327108ijap0303_3

Khan, T. V., Khan, S. S., Akhondi, A., and Khan, T. W. (2007). White coat hypertension: relevance to clinical and emergency medical services personnel. *MedGenMed* 9 (1), 52.

Kim, H.-Y., and Kim, E.-Y. (2023). Effects of medical education program using virtual reality: a systematic review and meta-analysis. *Int. J. Environ. Res. Public Health* 20, 3895. doi:10.3390/ijerph20053895

Koller, M., Schäfer, P., Sich, M., Diemer, J., Müller, M., and Meixner, G. (2018). "Next generation virtual reality exposure therapy systems – a study exploring design implications," in *9th international conference on intelligent systems (IS)*.

Lindner, P., Rozental, A., Jurell, A. l., Reuterskiöld, L., Andersson, G., Hamilton, W., et al. (2020). Experiences of gamified and automated virtual reality exposure therapy for spider phobia: qualitative study. *JMIR Serious Games* 8 (2), e17807. doi:10.2196/17807

Mahmoudi-Nejad, A., Guzdial, M. J., and Boulanger, P. (2024). Spiders based on anxiety: how reinforcement learning can deliver desired user experience in virtual reality personalized arachnophobia treatment. *ArXiv*, 2409. doi:10.48550/arXiv.2409.17406

McNally, R. J. (2007). Mechanisms of exposure therapy: how neuroscience can improve psychological treatments for anxiety disorders. *Clin. Psychol. Rev.* 27, 750–759. doi:10.1016/j.cpr.2007.01.003

Mehrfard, A., Fotouhi, J., Taylor, G., Forster, T., Navab, N., and Fuerst, B. (2019). A comparative analysis of virtual reality head-mounted display systems. *ArXiv 1912.02913v1 [cs.HC]*. doi:10.48550/arXiv.1912.02913

Miloff, A., Lindner, P., Dafgård, P., Deak, S., Garke, M., Hamilton, W., et al. (2019). Automated virtual reality exposure therapy for spider phobia vs. *in-vivo* one-session treatment: a randomized non-inferiority trial. *Behav. Res. Ther.* 118, 130–140. doi:10.1016/j.brat.2019.04.004

Morina, N., Ijntema, H., Meyerbröker, K., and Emmelkamp, P. M. G. (2015). Can virtual reality exposure therapy gains be generalized to real life? A meta-analysis of studies applying behavioral assessments. *Behav. Res. Ther.* 74, 18–24. doi:10.1016/j.brat.2015.08.010

National Institute for Health and Care Excellence (2014). *Anxiety disorders. Quality standard*. London: The British Psychological Society and The Royal College of Psychiatrists. Available online at: www.nice.org.uk/guidance/qs53 (Accessed September 13, 2024).

Opris, D., Pintea, S., Garcia-Palacios, A., Botella, C., Szamoskozi, S., and David, D. (2012). Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. *Depress. Anxiety* 29 (2), 85–93. doi:10.1002/da.20910

Parrish, C. L., Radomsky, A. S., and Dugas, M. J. (2008). Anxiety-control strategies: is there room for neutralization in successful exposure treatment? *Clin. Psychol. Rev.* 28, 1400–1412. doi:10.1016/j.cpr.2008.07.007

Peperkorn, H. M., Diemer, J., and Mühlberger, A. (2015). Temporal dynamics in the relation between presence and fear in virtual reality. *Comput. Hum. Behav.* 48, 542–547. doi:10.1016/j.chb.2015.02.028

Pittig, A., Kotter, R., and Hoyer, J. (2019). The struggle of behavioral therapists with exposure: self-reported practicability, negative beliefs, and therapist distress about exposure-based interventions. *Behav. Ther.* 50 (2), 353–366. doi:10.1016/j.beth.2018.07.003

Price, M., and Anderson, P. (2007). The role of presence in virtual reality exposure therapy. *J. Anxiety Disord.* 21, 742–751. doi:10.1016/j.janxdis.2006.11.002

Rooney, T., Sharpe, L., Winiarski, N., Todd, J., Colagiuri, B., Van Ryckeghem, D., et al. (2025). A synthesis of meta-analyses of immersive virtual reality interventions in pain. *Clin. Psychol. Rev.* 117, 102566. doi:10.1016/j.cpr.2025.102566

Rupp, C., Doebler, P., Ehring, T., and Vossbeck-Elsebusch, A. N. (2017). Emotional Processing Theory put to test: a meta-analysis on the association between process and outcome measures in exposure therapy. *Clin. Psychol. Psychother.* 24 (3), 697–711. doi:10.1002/cpp.2039

Sartory, G. (2012). "Physiologische Maße der Angst und Vermeidung," in *Angstdiagnostik*. Editors J. Hoyer and J. Margraf (Berlin/Heidelberg, Germany: Springer), 55–75.

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: factor analytic insights. *Presence Teleoperators Virtual Environ.* 10 (3), 266–281. doi:10.1162/105474601300343603

Shiban, Y., Diemer, J., Müller, J., Mühlberger, A., Pauli, P., and Mühlberger, A. (2017). Diaphragmatic breathing during virtual reality exposure therapy for aviophobia: functional coping strategy or avoidance behavior? A pilot study. *BMC Psychiatry* 17, 29. doi:10.1186/s12888-016-1181-2

Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence* 6, 603–616. doi:10.1162/pres.1997.6.6.603

Smits, J. A. J., Powers, M. B., Cho, Y. R., and Telch, M. J. (2004). Mechanism of change in cognitive– behavioral treatment of panic disorder: evidence for the fear of fear mediational hypothesis. *J. Consult. Clin. Psychol.* 72, 646–652. doi:10.1037/0022-006X.72.4.646

van Loenen, I., Scholten, W., Muntingh, A., Smit, J., and Batelaan, N. (2022). The effectiveness of virtual reality exposure–based cognitive behavioral therapy for severe anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder: meta-analysis. *J. Med. Internet Res.* 24 (2), e26736. doi:10.2196/26736

Wang, P., Ai, X., Zhang, X., Ma, F., Zhuang, Y., and Wang, S. (2024). Evaluating virtual reality technology in psychotherapy: impacts on anxiety, depression, and ADHD. *Front. Psychiatry* 15, 1480788. doi:10.3389/fpsyt.2024.1480788

Weber, R., Dash, A., and Wriessnegger, S. C. (2024). "Design of a virtual reality-based neuroadaptive system for treatment of arachnophobia," in *2024 IEEE international conference on metrology for eXtended reality, artificial intelligence and neural engineering (MetroXRAINE)*, 255–259. doi:10.1109/MetroXRAINE62247.2024.10796452

Wechsler, T. F., Kümpers, F., and Mühlberger, A. (2019). Inferiority or even superiority of virtual reality exposure therapy in phobias? – a systematic review and quantitative meta-analysis on randomized controlled trials specifically comparing the efficacy of virtual reality exposure to gold standard *in vivo* exposure in agoraphobia, specific phobia, and social phobia. *Front. Psychol.* 10, 1758. doi:10.3389/fpsyg.2019.01758

Wei, S., Freeman, D., and Rovira, A. (2023). A randomised controlled test of emotional attributes of a virtual coach within a virtual reality (VR) mental health treatment. *Sci. Rep.* 13, 11517. doi:10.1038/s41598-023-38499-7