Does Voice Matter? The Effect of Verbal Communication and Asymmetry on the Experience of Collaborative Social XR

Christian Merz*
HCI & PIIS Group
University of Würzburg

Carolin Wienrich†
PIIS Group
University of Würzburg

Marc Erich Latoschik‡
HCI Group
University of Würzburg

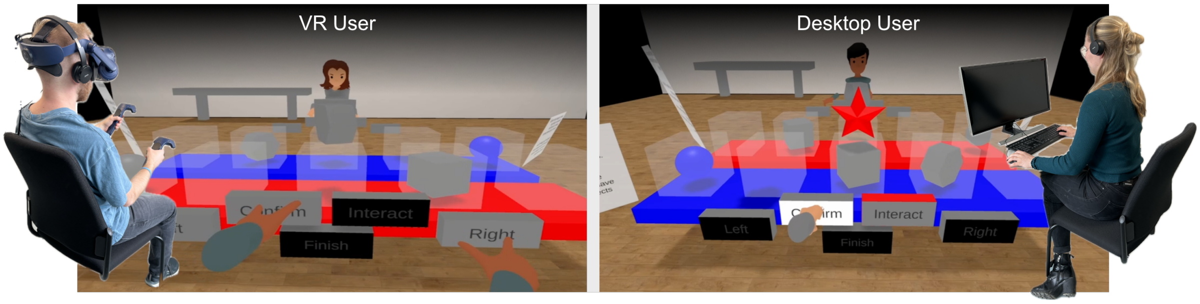

Figure 1: In the middle are the respective perspectives of a user with HMD and controllers vs. a user with a desktop screen and mouse during a collaborative sorting task in our dyadic study. On the sides are the device configurations used by participants.

ABSTRACT

This work evaluates how the asymmetry of device configurations and verbal communication influence the user experience of social eXtended Reality (XR) for self-perception, other-perception, and task perception. We developed an application that enables social collaboration between two users with varying device configurations.We compare the conditions of one symmetric interaction, where both device configurations are Head-Mounted Displays (HMDs) with tracked controllers, with the conditions of one asymmetric interaction, where one device configuration is an HMD with tracked controllers and the other device configuration is a desktop screen with a mouse. In our study, 52 participants collaborated in a dyadic interaction on a sorting task while talking to each other. We compare our results to previous work that evaluated the same scenario without verbal communication. In line with prior research, self-perception is influenced by the immersion of the used device configuration and verbal communication. While co-presence was not affected by the device configuration or the inclusion of verbal communication, social presence was only higher for HMD configurations that allowed verbal communication. Task perception was hardly affected by the device configuration or verbal communication. We conclude that the device in social XR is important for self-perception with or without verbal communication. However, the results indicate that the device configuration only affects the qualities of social interaction in collaborative scenarios when verbal communication is enabled. To sum up, asymmetric collaboration maintains the high quality of self-perception and interaction for highly immersed users while still enabling the participation of less immersed users.

Index Terms: VR, XR, Social VR, Verbal Communication, Immersion, Co-presence, Social Presence, Asymmetric Collaboration, Dyadic, Cross-Device.

1 INTRODUCTION

As the digital era evolves, mediated human-human interaction through video conferencing has become more and more prominent in both professional and personal spheres, particularly accentuated by the global shift towards remote work and computer-supported cooperative work (CSCW). Platforms like Zoom and Microsoft Teams have become ubiquitous, trying to provide a semblance of face-to-face interaction. Yet, despite their prevalence, conventional video conferencing tools often fail to fully replicate the nuanced dynamics of direct, in-person encounters, lacking in conveying essential nonverbal, bodily, spatial, and, consequently, social cues. To overcome these limitations, virtual, augmented, and mixed reality (VR, AR, MR: in short, XR) are prominent solutions because of their potential to enhance the user experience. They incorporate advanced sensory capabilities such as motion-, face-, and eye-tracking. These sensory capabilities aim to improve the 3D reconstruction fidelity of user avatars within shared virtual environments, thereby facilitating a richer, more nuanced form of human-human interaction within a social virtual environment. Using social virtual environments in embodied immersive VR leads to similar communication effects as face-to-face communication, showing its great potential. [48].

However, the accessibility of high-end XR equipment is often limited by various factors, such as spatial constraints, health considerations, and hardware availability, posing significant barriers to widespread adoption [50, 56]. To address these challenges, many commercial applications allow additional participation with varying device configurations, from desktops to smartphones, to engage in social interaction in XR environments. For example, Microsoft Teams enables an interaction between embodied head-mounted display (HMD) and desktop users. This approach aims to maintain the benefits of immersive setups for those equipped with HMDs while accommodating less immersive platforms. This asymmetry introduces new complexities into the collaborative experience, potentially impacting interaction quality, user satisfaction, and overall engagement across different participant's device configurations [9, 14].

Previous research has highlighted the importance of balanced interaction forms among symmetrically equipped peers for enhancing task collaboration and social presence [61]. Despite that, most

---

*e-mail: christian.merz@uni-wuerzburg.de
†e-mail: carolin.wienrich@uni-wuerzburg.de
‡e-mail: marc.latoschik@uni-wuerzburg.de

research in asymmetric virtual collaboration focuses on device-specific roles and interactions within asymmetric collaborations [6, 14, 49], calling for more systematic research regarding collaboration with different device configurations. This is particularly important for use cases in CSCW since remote meetings are often used to collaborate on a task. Here, the user's role should not be defined by the respective device but by the role based on their relationship with the interaction partner or their responsibility in the collaboration. One primary work focused on evaluating asymmetric virtual collaboration and found no impact of different immersion levels stemming from different devices on the social interaction [29]. However, this study examined a reduced social collaboration in which many social cues, such as speech or facial expressions, were excluded. Social signals, such as verbal communication, affect the interaction in symmetric and asymmetric collaborations [7, 19]. The **research question**, therefore, arises as to whether the use of different devices has a stronger influence on the user experience and the quality of the interaction when more social cues, such as speech, are added.

Contribution: To evaluate our research question, we developed a social XR application that enables collaboration between users with different devices. Two physically remote users collaborate in a virtual space while talking to each other using a desktop computer with a screen and a mouse, or an HMD with VR controllers. We conducted a user study with 52 participants, where we evaluated how the asymmetry and immersion of the used devices and the addition of social cues (i.e., verbal communication) affect self-perception (i.e., presence), other-perception (i.e., social presence), and task perception i.e., task load). Our contribution is twofold. First, we show that the device configuration and verbal communication affect self-perception. Second, our results indicate that different devices only affect other perceptions when social cues, like verbal communication, are included. Hence, we contribute to evaluating the potential of integrating highly immersive technologies with less immersive devices to enhance remote collaboration. Our work shows that device differences are perceivable with more social cues during the interaction, but only for the users with less immersive devices. For the design, the interplay between the immersion of the devices (i.e., the type of participation) and the richness of social signals must always be considered and, if necessary, designed differently for different participation conditions.

2 RELATED WORK

In this work, we follow the definition of immersion by Slater and Wilbur [47] as "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". We can objectively measure users' immersion based on the characteristics of the device they are using [45]. Hence, we can determine the immersion and experimentally manipulate it using different device configurations. Various studies evaluate the impact of immersion on the user experience [4].

Using highly immersive devices like an HMD with full body tracking and representing the user as an avatar leads to similar communication aspects in user interaction as in face-to-face interaction [48]. With the increasing spread of VR applications and scenarios, the relevance and benefits of the use cases of social virtual collaborations have never been more pronounced [21]. However, using virtual reality may come at a cost since various factors limit the accessibility to HMDs [50, 56]. Consequently, users invariably engage in shared virtual environments and applications, using devices that provide a range of different immersive capabilities and allow for more flexibility for participation.

2.1 Asymmetric collaboration

Asymmetric collaboration in social XR involves collaborators using different device configurations, leading to varying levels of immersion. This phenomenon is described by Ens et al. [6], who define asymmetry in collaboration across users with different roles or interaction capabilities, and by Numan et al. [31], who describe users interacting with different devices as asymmetric interfaces. As Yassien et al. [65], we refer to asymmetric collaboration when users with different device configurations collaborate within a social XR environment.

Prior research has addressed challenges and solutions for engaging in remote asymmetric virtual collaboration, emphasizing compensatory strategies for variety in immersion levels [6, 35, 65]. Strategies include adjusting the interaction space for less immersed users or customizing their interaction possibilities to fit the use case [14, 13, 36]. Despite these approaches, commercial social XR platforms like Microsoft Teams, RecRoom, or Mozilla Hubs allow users to interact without predefined roles or restricted interaction possibilities, raising questions about the impact of immersion levels on user experience in such settings.

Merz et al. [29] compared different asymmetric device settings against a symmetric one without predefined user roles or restricted interactions. They deliberately left out verbal communication to identify the device-dependent differences. They conclude that self-perception is influenced by the other device but not the perception of the other. However, verbal communication enhances social interaction in XR for symmetric and asymmetric collaboration [7, 19, 42] and is a critical factor for social interaction and the perception of the other.

2.2 The Influence of Immersion on User Experience

When categorizing various user experience indicators, we focus on i) self-perception, ii) other-perception, and iii) task perception. These three constructs allow us to evaluate how the users experience i) being in the virtual environment and the user's self-representation, ii) the interaction partner and their collaboration in social settings, and iii) the task they are collaborating with their interaction partner.

2.2.1 Self-Perception

The degree of immersion is strongly related to the sense of presence [4, 43], meaning that higher immersion typically results in a stronger sense of presence for the user [44]. This feeling is often described as the "sense of being there" [47] or telepresence [4], which hinges on the sensorimotor contingencies the (VR) system provides. Recent theories support the link between immersion and presence, proposing that reduced immersion may cause a bottom-up incongruence, thereby diminishing the sense of presence [23, 43, 46]. Previous studies and theoretical frameworks have established presence as a key measure of varying levels of (self) immersion, suggesting that presence can be used as a check for manipulating self-immersion. A significant advantage of social XR compared to video conferencing is the ability to interact with a body in a virtual space. Almost all social XR applications feature an avatar that can create a sense of embodiment. The sense of embodiment is the sense of having a virtual body in a virtual environment consisting of the user's sense of self-location, agency, and body ownership [16]. Visuomotor synchrony and motor control are crucial for the sense of embodiment [8]. Therefore, less immersive device configurations offer lower visuomotor synchrony and will likely lead to a weaker sense of embodiment. Previous work confirms this since, similar to the sense of presence, the perceived sense of embodiment is influenced by the degree of (self) immersion [3].

2.2.2 Other-Perception

The feeling of "being there together" is called co-presence [39]. Kohonen-Aho and Alin described co-presence as the mutual awareness of others and their actions [17]. Biocca et al. [2] define social presence with co-presence and include psychological and behavioral engagement with the other. Higher social presence leads to positive communication outcomes and higher perceived trust and enjoyment in social interactions [11, 24]. Hence, social presence is a key factor for the perception of virtual collaboration, and it is also used to assess how successful a communication system is at emulating face-to-face interaction [33]. Co-presence and social presence also rely on the degree of immersion [1, 40]. Using higher immersive device configurations leads to higher co-presence [40]. Furthermore, social presence increases when higher immersive devices [1]. The perception of plausibility in a virtual environment is another crucial evaluation parameter for the quality of XR experiences [23, 46]. The plausibility of virtual humans is essential in evaluating the other-perception and their representation in social XR [26, 27]. Previous work showed that immersion affects virtual human plausibility ratings [27, 64].

2.2.3 Task Perception

Evaluating usability and task perception in XR gives insight into the effectiveness of the application [10, 25]. Using device configurations with lower immersion reduces the perceived usability when keeping the input metaphor between the configurations identical [58]. Devices with lower immersion lead to a more demanding and less satisfactory perception of the task [52, 54].

2.3 Summary and Present Work

Various studies on asymmetric collaboration focus on different user roles for different device configurations [65]. There is limited work on asymmetric collaboration with the same interaction possibilities in the virtual environment or examining these aspects with reduced social signals [29]. Building on this backdrop, our work aims to evaluate how verbal communication influences the collaborative experience. Our study addresses the research gap in the existing literature, focusing on the interplay between the immersion of the device configurations (i.e., the type of participation) and the richness of social signals (i.e., verbal communication) in collaborative settings. To fill this gap, we examine the influence of varying device configurations and, therefore, immersion levels on user experience in asymmetric social XR collaborations with verbal communication, focusing on factors of self-perception (presence [32, 51, 53], embodiment [34, 37]), other-perception (co-presence, social presence [66], plausibility [23, 46], perceived humanness [12]), and task perception (task load [10], task enjoyment, usability) as key experience indicators when manipulating immersion and for XR experiences in general [23, 55, 57, 59].

2.4 Hypotheses

Based on the related work, we deduce the following hypotheses for our user study.

2.4.1 Self-Perception

Since there is ample evidence that shows that immersion affects the self-perception [3, 4, 44, 29], especially presence and sense of embodiment, we hypothesize:

H1.1: Higher self-immersion leads to a higher level of self-perception.

Including verbal communication in social interaction increases the level of user experience [7, 19, 33]. Further, the perceived sense of presence is related to the level of user engagement [5]. Therefore, we hypothesize:

H1.2: Including verbal communication leads to a higher level of self-perception.

2.4.2 Other-Perception

Various studies and meta-analyses show that self-immersion affects how we perceive collaboration with others [1, 27, 40, 64].

H2.1: Higher self-immersion leads to a higher level of other-perception.

Prior research indicates that the appearance of the other might affect how we perceive the experience in VR [22]. Since the movement of users with HMD and controllers is tracked by the devices, and the movement of desktop users typically not, we hypothesize:

H2.2: Higher other-immersion leads to a higher level of other-perception.

Verbal communication is a critical cue for social interaction. This is supported by the work of Eynard et al. [7] since verbal communication increases the quality of the perception of the other. Further, verbal communication affects the experience in asymmetric collaborations [19, 42]. Hence, we hypothesize:

H2.3 Including verbal communication leads to a higher level of other perception.

2.4.3 Task Perception

Manipulating self-immersion affects the task perception while manipulating the other-immersion does not [29].

H3.1 Higher self-immersion leads to a higher quality of task perception.

Task perception is affected by manipulating whether verbal communication during an interaction is included or not [7].

H3.2 Including verbal communication leads to a higher quality of task perception.

3 METHOD

3.1 Study Design

In a between-subjects design, we evaluate the difference between one symmetric and one asymmetric level of immersion. As the symmetric setting, we selected two users with a standard VR setup consisting of an HMD and controllers as input modality (VR_{VR}). The asymmetric setting consisted of one user with a standard desktop setup with mouse input (D_{VR}) and one user with the same VR setup as in the symmetric setting (VR_{D}). Since both participants in the symmetric pairing (VR_{VR}) have the same device, our study setup consisted of three different codes (VR_{VR}, D_{VR}, VR_{D}), where the subscript defines the device of the interaction partner. To keep the group sizes equal, we conducted only half of the pairs for VR_{VR}. Figure 1 shows the study setup with the two different device configurations. To evaluate whether voice communication impacts the experience in asymmetric virtual collaboration, we used the raw data from the experiment of Merz et al. [29]. We followed the same study design but included verbal communication during the sorting task to compare our results to this study. Therefore, in our study, we compare the effect of immersion on oneself and the effect of immersion on the other, leading to one symmetric and one asymmetric dyad. Further, with the requested data, we can evaluate the impact of verbal communication on asymmetric collaboration. This leads to a between design with 2 (no verbal communication nVC, verbal communication VC) x 3 (VR_{VR}, D_{VR}, VR_{D}).

3.2 Procedure

We invited a pair of participants to our study, arranging for them to be in separate rooms. Each participant was escorted to their designated experimental room by a researcher. The experimental process, detailed in Figure 2, began with participants completing a consent form and familiarizing themselves with the study's guidelines. Subsequently, they filled out the initial questionnaires before entering the virtual environment with the assigned devices. Within this private virtual environment, participants had the opportunity to learn the system's interface and customize their avatars based on personal preferences. Upon completing the avatar design, participants entered a shared social environment where they collaborated on a sorting task. Following the completion of this task, they provided feedback through post-experiment questionnaires. To ensure uniformity in the study's flow, we synchronized the start times of each activity for the participant pairs. On average, the study took about 35 minutes to complete.

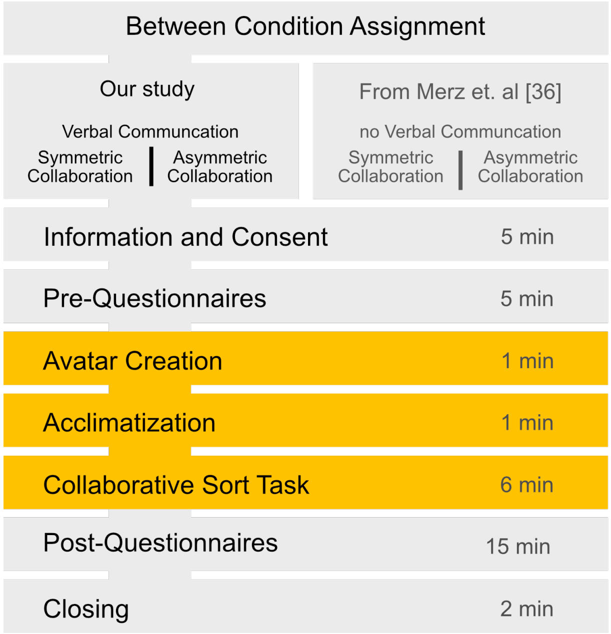

Figure 2: The figure shows the procedure of our study, where orange highlights the parts in the virtual environment.

3.3 Participants

We recruited $N = 52$ participants for our study. $N = 49$ were students who received credit points as part of their bachelor's degree, and $N = 3$ received money as compensation. The students had the same knowledge background of studying media and computer science. The condition D_{VR} and VR_{D} had $N = 17$ participants, and VR_{VR} had $N = 18$ participants. Our sample of $N = 52$ was $M = 20.29$ $(SD = 1.84)$ years old. 48 were female, 3 were male, and 1 did not report a gender. There were no significant differences between the prior VR experience and condition $\chi^2(8) = 8.86$, $p = .354$, or between the frequency of playing video games and condition $\chi^2(10) = 5.32$, $p = .869$, or between the frequency of using a desktop computer and condition $\chi^2(4) = 4.68$, $p = .321$. 80.8% of the participants had at least one hour of VR experience.

We used the data of participants with $N = 53$ from Merz et al. [29], with the same knowledge background. In their sample, $N = 34$ are female, and $N = 18$ are male, and $N = 1$ did not report a gender. They were, on average, $M = 24.92$ $(SD = 4.59)$ years old, and 90.6% of them had at least one hour of VR experience.

3.4 Measures

Table 1 presents an evaluation of constructs corresponding to specific hypotheses, detailing the questionnaires and individual items utilized for these measurements. We assessed immersive [63] tendency, composed of 18 items rated on a scale from 1 to 7, and simulator sickness [15], consisting of 14 items on a scale from 0 to 4. These variables were included as control variables due to their relevance in evaluating XR environments [28, 38].

Construct	Hypothesis	Variable	Measure
Control Measurement	-	Simulator Sicknesss Immersive Tendency	SSQ [15] ITQ [63]
Self-perception	H1.1 H1.2	Presence Embodiment	IPQ [41] VEQ [37]
Other-perception	H2.1 H2.2 H2.3	Co-/Social presence Plausibility Humanness	NMM [2] VHP [26] Single Item
Task Perception	H3.1 H3.2	Taskload Usability Task Enjoyment	RTLX [10] Single Item Single Item

Table 1: Summary of the constructs we assessed with the corresponding hypotheses and measures.

3.4.1 Self-Perception

For self-perception, we measured the perceived presence with the Igroup presence questionnaire (IPQ) [41] with the subscales spatial presence, involvement, and realism with a total of 19 items on a scale from 0 to 6. To measure the sense of embodiment, we used the virtual embodiment questionnaire (VEQ) [37], with the subscales of ownership, agency, and change consisting of 13 items in total on a scale ranging from 1 to 7.

3.4.2 Other-Perception

We evaluate the other perception with the networked mind measurement (NMM) [2] to measure co-presence and social presence. The questionnaire consists of four subscales: perception of self and of the other co-presence and perception of self and of the other psychological engagement. It consists of 34 items on a scale from 1 to 7. Further, we used the VHP questionnaire [26] and its subscales of virtual human appearance and behavior plausibility (ABP) and the virtual human's match to the virtual environment (MVE) consisting of 13 items on a scale from 0 to 6. Additionally, we used a single item that rated the interaction partner's human likeness on a scale from 1 to 7.

3.4.3 Task-Perception

To evaluate the task perception in our study, we measured task load with the Raw NASA TLX (RTLX) [10], with its six subscales on a scale from 0 to 20. Additionally, we used two single items, where participants rated their enjoyment of the task and the general usability of the system on a scale from 1 to 7.

3.5 System Description

We developed our system using the Unity3D game engine, specifically version 2020.3.21f1, and incorporated Photon's PUN2 framework for network functionalities. This system operates on a client-to-server model, which presents several benefits over the peer-to-peer approach. It efficiently sends data packets from the clients to the server at fixed intervals, which are then relayed to other clients for processing on their end. This setup enables the system to share updates regarding the avatars' appearances, joint positions, and orientations at a frequency of 20Hz and includes mechanisms for minimizing lag. Information on the user's progress within tasks or avatar appearance is also communicated to the server and, therefore, to all clients. Before the collaboration, users have the opportunity to personalize their avatars, selecting from stylized male or female avatars and adjusting attributes such as skin, hair, and shirt colors. In the virtual environment, avatars appear truncated from the hips and between the shoulders and hands, adopting a "ray-man style" to avoid the need for complex inverse kinematics for full-body motion due to the limited input from just three tracking devices [62]. An illustration of these avatars within our platform is provided as an example in Figure 1. In the VR condition, we utilized an HTC Vive Pro Eye HMD, featuring a field of view of 110 degrees, a refresh rate of 90Hz, and a resolution of 1440 x 1600 pixels per eye, accompanied by two Vive Pro Controllers for interaction. For the desktop D scenario, the setup comprised a 27-inch flat panel monitor with Full HD resolution (1920 x 1080) and a 60Hz refresh rate, along with a conventional office mouse for user input. Furthermore, we integrated SteamVR version 1.14.15 and its Unity plugin to facilitate the import of tracking data for the head-mounted display (HMD) and controllers into our system. The tracking of VR components is achieved through the use of two HTC Base Stations 2.0, employing an infrared-based tracking mechanism that ensures swift and precise positioning of devices, including HMDs and controllers with a latency of 22 milliseconds and sub-millimeter accuracy [30]. The system's tracking capability uses a high sampling rate of 1000 Hz, allowing for the precise capture of device movements. Each experimental room was equipped with a pair of base stations mounted on tripods to ensure stability, enabling comprehensive tracking in the VR conditions. We used the same headset (Lenovo 100 Stereo USB Headset) for the VR and D conditions to ensure consistent quality of the verbal communication across our conditions. The experimental setups were powered by high-end computers running Microsoft Windows 10, each equipped with an i7-11900K CPU, NVIDIA GeForce RTX 3080 graphics card, and 64 GB of DDR5-RAM, ensuring optimal performance and reliability for the study.

3.6 Task and System Interaction

In our study, participants collaborated on a sorting task, which is based on the study of Merz et al. [29]. However, in our study, participants were able to talk to each other in the social environment while collaborating together. Figure 1 illustrates how both participants viewed the task from their respective points of view. Standing in front of either a blue or red colored table, each participant had five semi-transparent containers hovering above their table. Their goal was to sort five objects by placing them into these containers in ascending order based on the number of corners each object had.

The task starts with an object appearing above the central semi-transparent container. Users can move the object left or right using the "Left" and "Right" buttons. Confirming the placement with a "Confirm" button drops the object into a container and triggers the appearance of the next object. One of the five objects matches the color of the partner's table, which participants have to transfer to the partner via an "Interact" button. Activating the "Interact" button on one side notifies the partner through a color change on their "Interact" button, indicating the object is ready to be transferred. Upon acceptance, the object is transferred to the partner's side for sorting. The cycle continues until all objects are sorted, culminating in the option to press a "Finish" button. To complete the task, both participants have to press the "Finish" button. Unusable buttons at a given point in time are marked in black to guide participants.

To maintain consistency across different interaction modalities — D and VR — we ensured that the task interaction remained uniform, irrespective of the immersion level or input mechanism. The interaction with the virtual environment was only through button presses to facilitate comparison. In the D scenario, distance from the screen was standardized to ensure equal visibility for all participants, with mouse clicks serving as the means of interaction with the buttons in the virtual environment. Clicking virtual buttons with the left mouse button executed pre-recorded animations simulating the virtual hand moving onto the virtual buttons. Holding the right button down allowed rotating the camera viewpoint. In contrast, the VR scenario utilized Vive controllers for direct interaction with the virtual buttons, and head movements facilitated by the HMD enabled participants to look around the virtual environment.

4 RESULTS

We used Python 3.9 for data aggregation, score computation, and plot generation. Table 2 shows the descriptive values for the measured dependent variables of our study and the data of the study from Merz et al. [29] that we used for the no verbal communication (nVC) conditions. We did our statistical analysis with R version 4.3.2. Our control measures showed no assumption violations and were tested with ANOVA tests and corresponding post hoc tests with Bonferroni corrections.

There were multiple violations for the normality and variance homogeneity assumptions for our different measurements. Hence, we decided to calculate a two-factor ANOVA with robust standard errors. A power analysis with G*Power 3.1.9.7 revealed that finding interaction effects with our sample size of $N = 105$ (52 of our study and 53 from Merz et al. [29]) was unlikely at about 30%, and we did not find any significant interaction effects. Therefore, for significant main effects of verbal communication or immersion we decided to calculate simple main effects. Our manipulation of device configuration has three dimensions that are not resolved by simple effects. Hence, after the simple effects, we calculate post hoc tests. Verbal communication has two dimensions, where the simple main effects already resolve the pairwise comparisons. For the post hoc tests, we always compare the D_{VR} to VR_{VR} and not to VR_{D} as the symmetric condition acts as our baseline. Since our study is almost exclusively female participants compared to the sample from Merz et al. [29], we added gender as a covariant. This did not lead to any different significant results.

4.1 Control measurements

We found no significant differences in immersive tendency between the experimental groups or in simulator sickness for the pre- and post-measurements between our conditions.

4.2 Self-Perception

4.2.1 Presence

As expected, the two-way ANOVA revealed for *spatial presence* a significant main effect for verbal communication, $F(1,99) = 50.02, p < .001, \eta_p^2 = 0.34$; and a significant main effect for immersion, $F(2,99) = 23.87, p < .001, \eta_p^2 = 0.36$. Simple effects for verbal communication are significant and showed that using verbal communication resulted in higher *spatial presence* for D_{VR}, $F(1,99) = 23.68, p < .001$, VR_D, $F(1,99) = 11.28, p < .01$, and VR_{VR}, $F(1,99) = 24.06, p < .001$. Simple effects for immersion are significant for VC, $F(2,99) = 26.03, p < .001$, and nVC, $F(2,99) = 37.75, p < .001$. The pairwise comparisons for immersion reveal that for VC, the *spatial presence* in the D_{VR} condition is significantly lower than in VR_{VR}, $t = 4.41, p < .001$. For nVC, the *spatial presence* in the D_{VR} condition is significantly lower than in VR_{VR}, $t = 4.38, p < .001$. Figure 3 includes the mean values and the significant pairwise comparisons.

The two-way ANOVA revealed for *realism* a significant main effect for verbal communication, $F(1,99) = 57.88, p < .001, \eta_p^2 = 0.37$; and a significant main effect for immersion, $F(2,99) = 4.55, p < .05, \eta_p^2 = 0.08$. Simple effects for verbal communication are significant and showed that using verbal communication resulted in higher *realism* for D_{VR}, $F(1,99) = 13.39, p < .001$, VR_D, $F(1,99) = 4.18, p < .001$, and VR_{VR}, $F(1,99) = 5.36, p < .001$. Simple effects for immersion are not significant for VC, $F(2,99) = 0.58, p = .563$, and significant for nVC, $F(2,99) = 4.18, p < .01$, indicating an interaction effect. However, the pairwise comparisons

			D_{VR}		VR_{D}		VR_{VR}	
			nVC	VC	nVC	VC	nVC	VC
Measure	Range	Subscale	N=18	N=17	N=18	N=17	N=17	N=18
IPQ	0-6	Spatial Presence	2.18 (1.32)	3.82 (1.12)	4.11 (0.97)	5.25 (1.02)	3.74 (1.14)	5.40 (0.65)
		Involvement	2.62 (1.67)	3.56 (1.27)	3.75 (1.17)	5.15 (1.07)	3.54 (1.15)	5.08 (1.51)
		Realism	2.54 (0.61)	3.78 (0.55)	3.22 (0.53)	3.91 (0.74)	2.91 (0.62)	3.69 (0.53)
VEQ	1-7	Ownership	1.65 (1.08)	2.09 (1.22)	3.35 (1.28)	3.37 (1.47)	3.15 (1.23)	3.38 (1.51)
		Agency	3.31 (1.70)	3.71 (1.46)	5.50 (0.65)	5.50 (1.03)	5.87 (1.06)	5.74 (1.02)
		Change	1.44 (1.19)	2.26 (2.06)	2.35 (1.34)	3.03 (1.54)	2.28 (1.10)	2.60 (1.83)
NMM	1-7	CP self	3.88 (0.72)	3.79 (0.66)	3.86 (0.33)	4.00 (0.79)	3.82 (0.62)	3.94 (0.48)
		CP other	3.71 (0.85)	3.85 (0.73)	3.74 (0.66)	3.84 (0.36)	3.85 (0.50)	3.88 (0.42)
		PE self	2.74 (0.77)	3.26 (0.85)	3.00 (0.67)	3.67 (0.80)	3.09 (1.01)	4.23 (0.87)
		PE other	2.46 (0.95)	3.01 (1.16)	2.74 (0.79)	3.73 (1.00)	2.92 (1.20)	4.41 (0.96)
VHP	1-7	ABP	4.69 (0.85)	5.13 (0.94)	4.98 (0.56)	5.08 (0.93)	4.96 (0.49)	5.55 (0.68)
		MVE	5.74 (0.69)	5.82 (0.98)	6.07 (0.60)	5.85 (0.99)	5.57 (0.84)	6.18 (0.72)
Item	1-7	Humanlikeness	3.33 (2.03)	4.00 (2.26)	3.39 (1.82)	4.71 (1.53)	3.41 (1.73)	5.17 (1.65)
RTLX	0-100	Mental dem.	38.06 (19.94)	43.24 (21.79)	36.94 (23.02)	43.53 (22.41)	40.88 (21.16)	46.39 (18.61)
		Physical dem.	6.39 (5.89)	15.88 (15.83)	15.00 (11.88)	20.00 (16.11)	10.88 (11.21)	16.94 (12.85)
		Frustration	30.00 (23.51)	22.94 (12.38)	31.11 (24.89)	31.11 (30.34)	29.12 (23.00)	25.56 (16.79)
		Temporal dem.	32.50 (25.39)	22.94 (17.42)	28.89 (25.58)	30.59 (26.15)	25.00 (24.81)	29.72 (21.11)
		Performance	36.67 (25.50)	37.94 (26.70)	34.72 (20.47)	44.12 (27.85)	51.18 (23.15)	35.83 (24.99)
		Effort	25.28 (23.61)	26.76 (16.77)	28.89 (22.20)	27.06 (24.82)	23.82 (16.16)	34.44 (24.61)
Item	1-7	Usability	4.61 (1.42)	4.59 (1.50)	4.39 (1.50)	3.82 (1.78)	4.65 (1.37)	5.00 (1.19)
		Task enjoyment	5.28 (1.27)	5.12 (1.11)	5.17 (1.29)	5.82 (1.01)	5.41 (1.62)	6.00 (0.77)

Table 2: This table shows the descriptive data with mean and standard deviation for the three different device conditions with the no verbal communication (nVC) of Merz et al. [29] and the verbal communication (VC) of our conducted study.

for immersion reveal no significant difference. Therefore, there is no significant difference in *realism* when manipulating immersion.

The two-way ANOVA revealed for *involvement* a significant main effect for verbal communication, $F(1,99) = 26.22, p < .001, \eta_p^2 = 0.20$; and a significant main effect for immersion, $F(2,99) = 10.60, p < .001, \eta_p^2 = 0.18$. Simple effects for verbal communication are significant and showed that using verbal communication resulted in higher *involvement* for D_{VR}, $F(1,99) = 4.32, p < .05$, VR_D, $F(1,99) = 9.67, p < .01$, and VR_{VR}, $F(1,99) = 11.74, p < .001$. Simple effects for immersion are significant for *VC*, $F(2,99) = 27.72, p < .001$, and *nVC*, $F(2,99) = 12.86, p < .05$. The pairwise comparisons for immersion reveal that for *VC*, the *involvement* in the D_{VR} condition is significantly lower than in VR_{VR}, $t = 3.39, p < .01$. For *nVC*, there are no significant differences between the immersion conditions for *involvement*.

4.2.2 Embodiment

The two-way ANOVA revealed for *agency* no significant main effect for verbal communication, $F(1,99) = 0.01, p = .946, \eta_p^2 < 0.01$; and a significant main effect for immersion, $F(2,99) = 25.63, p < .001, \eta_p^2 = 0.43$; Simple effects for immersion are significant for *VC*, $F(2,99) = 14.70, p < .001$, and *nVC*, $F(2,99) = 23.57, p < .001$. The pairwise comparisons for immersion reveal that for *VC*, the *agency* in the D_{VR} condition is significantly lower than in VR_{VR}, $t = 4.99, p < .001$. For *nVC*, the *agency* in the D_{VR} condition is significantly lower than in VR_{VR}, $t = 6.30, p < .001$. Figure 4 includes the mean values and the significant pairwise comparisons.

The two-way ANOVA revealed for *ownership* a significant main effect for immersion, $F(2,99) = 15.88, p < .001, \eta_p^2 = 0.22$. Simple effects for immersion are significant for *VC*, $F(2,99) = 5.51, p < .01$, and *nVC*, $F(2,99) = 8.97, p < .001$. The pairwise comparisons for immersion reveal that for *VC*, the *ownership* in the D_{VR} condition is significantly lower than in VR_{VR}, $t = 2.91, p < .05$. For *nVC*, the *ownership* in the D_{VR} condition is significantly lower than in VR_{VR}, $t = 3.38, p < .01$.

The two-way ANOVA revealed for *change* a significant main effect for immersion, $F(2,99) = 3.46, p < .05, \eta_p^2 = 0.05$. Simple effects for immersion reveal no significant differences.

4.3 Other-Perception

The two-way ANOVAs revealed for both co-presence scales no significant effects.

4.3.1 Social Presence

The two-way ANOVA revealed for *psychobehavioral engagement self* a significant main effect for verbal communication, $F(1,99) = 19.78, p < .001, \eta_p^2 = 0.19$; and a significant main effect for immersion, $F(2,99) = 4.98, p < .01, \eta_p^2 = 0.10$. Simple effects for verbal communication are significant and showed that using verbal communication resulted in higher *psychobehavioral engagement self* for VR_{D}, $F(1,99) = 5.57, p < .05$, and VR_{VR}, $F(1,99) = 16.25, p < .001$. Simple effects for immersion are only significant for *VC*, $F(2,99) = 5.91, p < .01$, indicating an interaction effect. The pairwise comparisons for immersion reveal that for *VC*, the *psychobehavioral engagement self* in the D_{VR} condition is significantly lower than in VR_{VR}, $t = 3.42, p < .01$. Figure 5 includes the mean values and the significant pairwise comparisons.

The two-way ANOVA revealed for *psychobehavioral engagement other* a significant main effect for verbal communication, $F(1,99) = 24.28, p < .001, \eta_p^2 = 0.21$; and a significant main effect for immersion, $F(2,99) = 6.47, p < .01, \eta_p^2 = 0.13$. Simple effects for verbal communication are significant and showed that using verbal communication resulted in higher *psychobehavioral engagement self* for VR_{D}, $F(1,99) = 8.36, p < .01$, and VR_{VR}, $F(1,99) = 18.61, p < .001$. Simple effects for immersion are only significant for VC, $F(2,99) = 8.26, p < .001$, indicating an interaction effect. The pairwise comparisons for immersion reveal that for

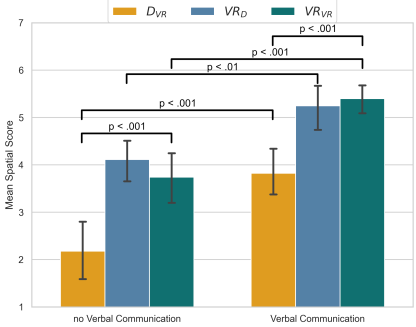

Figure 3: Mean values and CI($\pm 95\%$) for perceived spatial presence, with all significant pairwise comparisons.

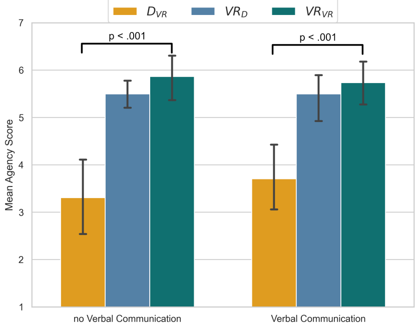

Figure 4: Mean values and CI($\pm 95\%$) for perceived agency, with all significant pairwise comparisons.

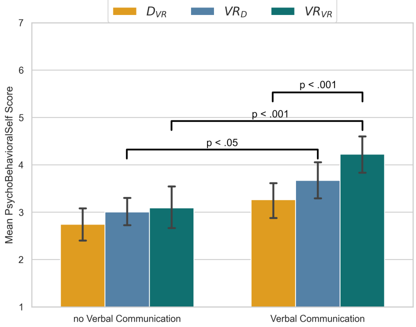

Figure 5: Mean values and CI($\pm 95\%$) for psychobehavioral engagement, with all significant pairwise comparisons.

VC, the *psychobehavioral engagement other* in the D_{VR} condition is significantly lower than in VR_{VR}, $t = 4.06, p < .001$.

4.3.2 Virtual Human Plausibility

The two-way ANOVA revealed for *appearance behavior plausibility* a significant main effect for verbal communication, $F(1,99) = 8.17, p < .01, \eta_p^2 = 0.06$. Simple effects showed that using verbal communication resulted in higher *appearance behavior plausibility* for VR_{VR}, $F(1,99) = 5.24, p < .05$.

The two-way ANOVA revealed no significant effects for *match to virtual environment*.

4.3.3 Humanlikeness

The two-way ANOVA revealed for *humanlikeness* a significant main effect for verbal communication, $F(1,99) = 13.40, p < .001, \eta_p^2 = 0.11$. Simple effects showed that using verbal communication resulted in higher *humanlikeness* for VR_{D}, $F(1,99) = 4.41, p < .05$, and VR_{VR}, $F(1,99) = 7.83, p < .01$.

4.4 Task Perception

For task perception, the two-way ANOVAs only revealed differences for *phsyical demand*. The two-way ANOVA revealed for *phsyical demand* a significant main effect for verbal communication, $F(1,99) = 7.64, p = .007, \eta_p^2 = 0.07$; and a significant main effect for immersion, $F(2,99) = 4.00, p = .021, \eta_p^2 = 0.04$. Simple effects showed that using verbal communication resulted in higher *phsyical demand* for D_{VR}, $F(1,99) = 4.88, p = .029$. Simple effects for immersion revealed no significant differences.

5 DISCUSSION

Our study evaluated the impact of verbal communication and asymmetry on self-perception, other-perception, and task perception. As expected, we can confirm our hypothesis H1.1 and, therefore, previous studies that higher self-immersion leads to a higher level of self-perception. This was evident for presence and embodiment and is in line with various studies [1, 44]. A prominent factor for increased self-perception with higher immersion is the visuomotor synchrony provided by the tracked devices of the VR conditions. We deduce that both of our used measurements of self-perception seem to be valid manipulation checks for varying immersion.

Including verbal communication leads only to higher self-perception for presence but not for the sense of embodiment. Hence, we can only partially accept H1.2. While prior research has already shown that verbal communication leads to higher user engagement [5] and user experience in general, [7, 19, 33], taking a closer look at the measurements leads to an interpretation of our found effect. Being able to speak to each other is an integral part of collaboration and social interaction. Including verbal communication makes the scenario more realistic and might draw people more into the virtual environment since speaking to the interaction partner increases the sensorimotor contingencies for the virtual environment [46]. Further, verbal communication in collaboration would be possible in face-to-face interaction, and talking to the interaction partner leads to a more semantic congruence of the task [23]. In comparison, the sense of embodiment depends more on the visuomotor synchrony provided by the device, which does not change when verbal communication is possible.

For other-perception we did not find consistent results over our measurements and we have to reject H2.1, H2.2, and H2.3. For the scales of co-presence, our results show no significant differences, neither for verbal communication nor immersion. Some studies in the literature found a positive relationship between presence and co-presence [40]. However, other studies did not find this relation [29]. For co-presence, the sole representation of another person seems to be the most prominent factor. This calls for a more systematic evaluation to find out in which scenarios the feeling of presence and co-presence correlate positively.

Our results for social presence indicate that there is an interaction effect between self-immersion and verbal communication. We theorize that only when verbal communication is included do the differences in the immersion of the own device become perceivable in the collaboration. Social presence places greater importance on interaction than co-presence because it involves understanding and engaging on a deeper level with another person or entity, which is more than only having the feeling of being there together [2]. Our results for appearance behavior plausibility and human likeness support the findings for social presence since they are higher when there was verbal communication for the VR conditions. Hence, the quality of social interaction is lower for less immersed users. However, asymmetric collaboration maintains the high quality of social interaction for highly immersed users while allowing others to participate and access the social collaboration.

Our study hardly found any significant differences in task perception since the task itself did not differ between our conditions, and we have to reject H3.1 and H3.2. Verbal communication may not have been crucial for the task that required handing over the objects to the partner because the task itself provided visual indicators when the partner wanted to initiate a swap. This could change for more complex tasks in asymmetric collaboration where verbal communication is an essential part of the task [19]. Nevertheless, our results reveal that participants experienced the task itself not differently, which in turn shows that our interaction paradigm with the task for the different device configurations worked as intended.

Putting our results in a broader picture, we again look at the definition of immersion by Slater and Wilbur [47]. We argue that verbal communication could be a factor of immersion, as it is another sense that the device delivers to the user to create a more realistic experience. Other important social cues for social interaction, like full body movement or facial expressions, can be tracked by body and face tracking solutions and fit into the definition of immersion as well. Hence, using these social cues increases immersion and, in turn, should increase self-perception and other-perception. For example, Kullmann et al. [20] have shown that other-perception (plausibility of the other avatar) increases when including facial expressions [20]. This aligns with current theories for virtual experiences that take a more holistic view of the user experience and go beyond sensorimotor contingencies [23, 60]. Including more social cues in virtual experiences that are used in real-world interaction increases semantic coherence. The inclusion of richer social cues comes with limitations in the user experience, especially for less immersed users, and research needs to answer the question of whether they can participate and with what limitations of the experience. However, as our results show, there is still a need to evaluate how specific social cues affect specific qualia, as different immersion factors might have systematic effects or not. Additionally, while there are arguments to treat the tracking of social cues as immersive factors, our work emphasizes the impact of social cues like verbal communication and, therefore, the need to clarify which aspects of immersion are manipulated.

5.1 Limitations and Future Work

Our sample size leads to missing power and a low probability of finding interaction effects of ANOVA. Further, our sample was homogenous and almost exclusively female. Our comparison sample had a different gender distribution. However, in our analysis, we did not find varying significant effects when including it as covariant. While the knowledge background is very specific, it is the same for the two samples. Hence, our sample is comparable to the one from Merz et al. [29], especially as we followed the same procedure and method. However, these limitations also potentially lessen the overall generalizability of our findings.

There are various social cues that we could have manipulated. We decided on verbal communication as this is an essential social cue for collaboration, where prior research has shown that it impacts the user experience in asymmetric collaboration [19]. However, facial expressions and full body language are important social cues as well [20, 46, 48]. Future research should evaluate how these other social cues would affect asymmetric collaboration. Furthermore, the question remains open as to whether even more prominent differences in immersion and social cues could lead to an effect in which the other person's device configuration can influence one's perception. Additionally, the question arises of how the requirements of a task can impact the user experience. Spatial tasks that require pointing would rely more strongly on full body language or eye movement. Hence, future research should evaluate the link between the task or type of collaboration and the social cues or communication channels.

Our results show that the quality of social interaction was lower for less immersed users. Future research can tackle this by trying to compensate for less immersion the device provides with AI-driven approaches. For example, AI's advances in gesture generation allow for real-time AI-generated gestures that are perceived as similar to human-made gestures [18]. This could provide human-like body language for users who have no or limited body tracking.

In our study, we used more simplistic avatars, which are also often used in commercial social XR applications. Using different avatars like personalized photorealistic avatars leads to more experienced realism [22] and could lead to a higher incongruence in asymmetric collaboration [23]. Since avatars' movement is based on the user's tracking qualities, they could appear less human-like. This leaves a research gap regarding how photorealistic avatars would affect asymmetric collaboration in which one user has lower tracking qualities than the other user.

6 CONCLUSION

This study explored the impact of verbal communication and device asymmetry on user experience within social XR environments. Our findings reveal that both the device's immersion level and the presence of verbal communication significantly influence self-perception (presence and sense of embodiment). Other-perception (social presence) is significantly higher when verbal communication is included for users with highly immersive device configurations. Our manipulations did not markedly affect task perception. These results highlight the importance of integrating immersive devices and verbal communication in XR systems to enhance user engagement and interaction quality. Additionally, asymmetric collaboration is still effective in enabling access to social collaboration in XR since the interaction quality for highly immersed users is still high even when the interaction partner is less immersed. This study enriches the understanding of social XR, providing valuable insights for designing more inclusive and effective XR applications, thereby improving user satisfaction and engagement in virtual environments.

ACKNOWLEDGMENTS

We thank Tana Glemser for her help in conducting the study. This research has been funded by the German Federal Ministry of Education and Research (BMBF) in the project HiAvA (Grant 16SV8781) and by the Bavarian State Ministry for Digital Affairs in the project XR Hub (Grant A5-3822-2-16).

REFERENCES

[1] D. Ahn, Y. Seo, M. Kim, J. H. Kwon, Y. Jung, J. Ahn, and D. Lee. The effects of actual human size display and stereoscopic presentation on users' sense of being together with and of psychological immersion in a virtual character. *Cyberpsychology, Behavior, and Social Networking*, 17(7):483–487, 2014. doi: 10.1089/cyber.2013.0455 3, 7

[2] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*, pp. 1–9, 2001. 3, 4, 7

[3] F. Born, S. Abramowski, and M. Masuch. Exergaming in vr: The impact of immersive embodiment on motivation, performance, and perceived exertion. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1–8, 2019. doi: 10.1109/VS-Games.2019.8864579 2, 3

[4] J. J. Cummings and J. N. Bailenson. How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19(2):272–309, 2016. doi: 10.1080/15213269.2015.1015740 2, 3

[5] M. Deriu, F. Bachis, and M. Massa. Improving the user engagement in a fully immersive experience by the means of a conversational non-playable character used as a tourist guide. In *2021 IoT Vertical and Topical Summit for Tourism*, pp. 1–4. IEEE, 2021. doi: 10.1109/IEEECONF49204.2021.9604871 3, 7

[6] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131:81–98, 2019. doi: 10.1016/j.ijhcs.2019.05.011 2

[7] R. Eynard, M. Pallot, O. Christmann, and S. Richir. Impact of verbal communication on user experience in 3d immersive virtual environments. In *2015 IEEE International Conference on Engineering, Technology and Innovation/ International Technology Management Conference (ICE/ITMC)*, pp. 1–8, 2015. doi: 10.1109/ICE.2015.7438679 2, 3, 7

[8] D. Gall, D. Roth, J.-P. Stauffert, J. Zarges, and M. E. Latoschik. Embodiment in virtual reality intensifies emotional responses to virtual stimuli. *Frontiers in Psychology*, 12, 2021. doi: 10.3389/fpsyg.2021.674179 2

[9] J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio. Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 4021–4033, 2017. doi: 10.1145/3025453.3025683 1

[10] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. doi: 10.1177/154193120605000909 3, 4

[11] K. Hassanein and M. Head. Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *International journal of human-computer studies*, 65(8):689–708, 2007. doi: 10.1016/j.ijhcs.2006.11.018 3

[12] C.-C. Ho and K. F. MacDorman. Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness. *International Journal of Social Robotics*, 9:129–139, 2017. doi: 10.1007/s12369-016-0380-9 3

[13] R. Horst, R. Dörner, and M. Peter. Vr-guide: A specific user role for asymmetric virtual reality setups in distributed virtual reality applications. 10 2018. 2

[14] K. Jeong, J. Kim, M. Kim, J. Lee, and C. Kim. Asymmetric interface: user interface of asymmetric virtual reality for new presence and experience. *Symmetry*, 12(1):53, 2020. doi: 10.3390/sym12010053 1, 2

[15] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993. doi: 10.1207/s15327108ijap0303_3 4

[16] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012. doi: 10.1162/PRES_a_00124 2

[17] L. Kohonen-Aho and P. Alin. Introducing a Video-Based Strategy for Theorizing Social Presence Emergence in 3D Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 24(2):113–131, 05 2015. doi: 10.1162/PRES_a_00222 3

[18] N. Krome and S. Kopp. Towards real-time co-speech gesture generation in online interaction in social xr. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, pp. 1–8, 2023. doi: 10.1145/3570945.360731 8

[19] L. Kruse, J. Wittig, S. Finnern, M. Gundlach, N. Iserlohe, O. Ariza, and F. Steinicke. Blended collaboration: Communication and cooperation between two users across the reality-virtuality continuum. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544549.3585881 2, 3, 7, 8

[20] P. Kullmann, T. Menzel, M. Botsch, and M. E. Latoschik. An evaluation of other-avatar facial animation methods for social vr. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544549.3585617 8

[21] M. E. Latoschik, F. Kern, J.-P. Stauffert, A. Bartl, M. Botsch, and J.-L. Lugrin. Not alone here?! scalability and user experience of embodied ambient crowds in distributed social virtual reality. *IEEE transactions on visualization and computer graphics*, 25(5):2134–2144, 2019. doi: 10.1109/TVCG.2019.2899250 2

[22] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3139131.3139156 3, 8

[23] M. E. Latoschik and C. Wienrich. Congruence and plausibility, not presence: Pivotal conditions for xr experiences and effects, a novel approach. *Frontiers in Virtual Reality*, 3:694433, 2022. doi: 10.3389/frvir.2022.694433 2, 3, 7, 8

[24] K. M. Lee, Y. Jung, J. Kim, and S. R. Kim. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International journal of human-computer studies*, 64(10):962–973, 2006. doi: 10.1016/j.ijhcs.2006.05.002 3

[25] G. Makransky and G. B. Petersen. Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers & Education*, 134:15–30, 2019. doi: 10.1016/j.compedu.2019.02.002 3

[26] D. Mal, E. Wolf, N. Döllinger, M. Botsch, C. Wienrich, and M. E. Latoschik. Virtual human coherence and plausibility – towards a validated scale. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 788–789, 2022. doi: 10.1109/VRW55335.2022.00245 3, 4

[27] D. Mal, E. Wolf, N. Döllinger, M. Botsch, C. Wienrich, and M. E. Latoschik. From 2d-screens to vr: Exploring the effect of immersion on the plausibility of virtual humans. In *CHI 24 Conference on Human Factors in Computing Systems Extended Abstracts*, p. 8, 2024. doi: 10.1145/3613905.365077 3

[28] O. Merhi, E. Faugloire, M. Flanagan, and T. A. Stoffregen. Motion sickness, console video games, and head-mounted displays. *Human factors*, 49(5):920–934, 2007. doi: 10.1518/001872007X230262 4

[29] C. Merz, C. Göttfert, C. Wienrich, and M. E. Latoschik. Universal access for social xr across devices: The impact of immersion on the experience in asymmetric virtual collaboration. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 859–869. IEEE, 2024. doi: 10.1109/VR58804.2024.00105 2, 3, 4, 5, 6, 7, 8

[30] D. C. Niehorster, L. Li, and M. Lappe. The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception*, 8(3):2041669517708205, 2017. doi: 10.1177/2041669517708205 5

[31] N. Numan and A. Steed. Exploring user behaviour in asymmetric collaborative mixed reality. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2022. doi: 10.1145/3562939.3565630 2

[32] D. Nunez and E. Blake. Cognitive presence as a unified concept of virtual reality effectiveness. In *Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation*, pp. 115–118, 2001. doi: 10.1145/513867.513892 3

[33] C. S. Oh, J. N. Bailenson, and G. F. Welch. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5:409295, 2018. doi: 10.3389/frobt.2018.00114 3, 7

[34] T. C. Peck and M. Gonzalez-Franco. Avatar embodiment. a standardized questionnaire. *Frontiers in Virtual Reality*, 1:575943, 2021. doi: 10.3389/frvir.2020.575943 3

[35] P. Pérez, E. Gonzalez-Sosa, J. Gutiérrez, and N. García. Emerging immersive communication systems: overview, taxonomy, and good practices for qoe assessment. *Frontiers in Signal Processing*, 2:917684, 2022. doi: 10.3389/frsip.2022.917684 2

[36] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018. doi: 10.1145/3173574.3173620 2

[37] D. Roth and M. E. Latoschik. Construction of the virtual embodiment questionnaire (veq). *IEEE Transactions on Visualization & Computer Graphics*, 26(12):3546–3556, 2020. doi: 10.1109/TVCG.2020.3023603 3, 4

[38] S. Rózsa, R. Hargitai, A. Láng, A. Osváth, E. Hupuczi, I. Tamás, and J. Kállai. Measuring immersion, involvement, and attention focusing tendencies in the mediated environment: The applicability of the immersive tendencies questionnaire. *Frontiers in Psychology*, 13:931955, 2022. doi: 10.3389/fpsyg.2022.931955 4

[39] R. Schroeder. Copresence and interaction in virtual environments: An overview of the range of issues. In *Presence 2002: Fifth international workshop*, pp. 274–295. Citeseer, 2002. 3

[40] R. Schroeder. Being there together and the future of connected presence. *Presence*, 15(4):438–454, 2006. doi: 10.1162/pres.15.4.438 3, 7

[41] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001. doi: 10.1162/105474601300343603 4

[42] C. Simon, M. B. Hacene, S. Otmane, and A. Chellali. Study of communication modalities to support teaching tool manipulation skills in a shared immersive environment. *Computers & Graphics*, 117:31–41,

2023. doi: 10.1016/j.cag.2023.09.011 2, 3

[43] R. Skarbez, M. Smith, and M. C. Whitton. Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2:647997, 2021. doi: 10.3389/frvir.2021.647997 2

[44] M. Slater. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence*, 8(5):560–565, 1999. doi: 10.1162/105474699566477 2, 3, 7

[45] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009. doi: 10.1098/rstb.2009.0138 2

[46] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva. A separate reality: An update on place illusion and plausibility in virtual reality. *Frontiers in Virtual Reality*, 3, 2022. doi: 10.3389/frvir.2022.914392 2, 3, 7, 8

[47] M. Slater and S. Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616, 1997. doi: 10.1162/pres.1997.6.6.603 2, 7

[48] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. vol. 2018-April. Association for Computing Machinery, 4 2018. doi: 10.1145/3173574.3173863 1, 2, 8

[49] M. Sra, D. Jain, A. P. Caetano, A. Calvo, E. Hilton, and C. Schmandt. Resolving spatial variation and allowing spectator participation in multiplayer vr. In *Adjunct Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology*, pp. 221–222, 2016. doi: 10.1145/2984751.2984779 2

[50] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Latency and cybersickness: impact, causes, and measures. a review. *Frontiers in Virtual Reality*, 1:582204, 2020. doi: 10.3389/frvir.2020.582204 1, 2

[51] J. Takatalo, T. Kawai, J. Kaistinen, G. Nyman, and J. Häkkinen. User experience in 3d stereoscopic games. *Media Psychology*, 14(4):387–414, 2011. doi: 10.1080/15213269.2011.620538 3

[52] N. Talukdar and S. Yu. Breaking the psychological distance: the effect of immersive virtual reality on perceived novelty and user satisfaction. *Journal of Strategic Marketing*, pp. 1–25, 2021. doi: 10.1080/0965254X.2021.1967428 3

[53] R. Tamborini and P. Skalski. The role of presence in the experience of electronic games. In *Playing video games*, pp. 263–281. Routledge, 2012. 3

[54] Q. Tang, Y. Wang, H. Liu, Q. Liu, and S. Jiang. Experiencing an art education program through immersive virtual reality or ipad: Examining the mediating effects of sense of presence and extraneous cognitive load on enjoyment, attention, and retention. *Frontiers in Psychology*, 13, 2022. doi: 10.3389/fpsyg.2022.957037 3

[55] W. Tong, M. Xia, K. Wong, D. A. Bowman, T. Pong, H. Qu, and Y. Yang. Towards an understanding of distributed asymmetric collaborative visualization on problem-solving. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 387–397. IEEE Computer Society, Los Alamitos, CA, USA, mar 2023. doi: 10.1109/VR55154.2023.00054 3

[56] S. Vlahovic, M. Suznjevic, N. Pavlin-Bernardic, and L. Skorin-Kapov. The effect of vr gaming on discomfort, cybersickness, and reaction time. In *2021 13th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 163–168. IEEE, 2021. doi: 10.1109/QoMEX51781.2021.9465470 1, 2

[57] F. Welsford-Ackroyd, A. Chalmers, R. K. dos Anjos, D. Medeiros, H. Kim, and T. Rhee. Asymmetric interaction between hmd wearers and spectators with a large display. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 670–671. IEEE, 2020. doi: 10.1109/VRW50115.2020.00186 3

[58] N. Wenk, J. Penalver-Andres, K. Buetler, T. Nef, R. M. Müri, and L. Marchal-Crespo. Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment. *Virtual Reality*, pp. 1–25, 2021. doi: 10.1007/s10055-021-00565-8 3

[59] C. Wienrich and J. Gramlich. Appraisevr–an evaluation framework for immersive experiences. *i-com*, 19(2):103–121, 2020. doi: 10.1515/icom-2020-0008 3

[60] C. Wienrich, P. Komma, S. Vogt, and M. E. Latoschik. Spatial presence in mixed realities–considerations about the concept, measures, design, and experiments. *Frontiers in Virtual Reality*, 2, 2021. doi: 10.3389/frvir.2021.694315 8

[61] C. Wienrich, K. Schindler, N. Döllinqer, S. Kock, and O. Traupe. Social presence and cooperation in large-scale multi-user virtual reality-the relevance of social interdependence for location-based environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 207–214. IEEE, 2018. doi: 10.1109/VR.2018.8446575 1

[62] A. Winkler, J. Won, and Y. Ye. Questsim: Human motion tracking from sparse sensors with simulated avatars. In *SIGGRAPH Asia 2022 Conference Papers*, pp. 1–8, 2022. doi: 10.1145/3550469.3555411 5

[63] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, 06 1998. doi: 10.1162/105474698565686 4

[64] E. Wolf, D. Mal, V. Frohnapfel, N. Döllinger, S. Wenninger, M. Botsch, M. E. Latoschik, and C. Wienrich. Plausibility and perception of personalized virtual humans between virtual and augmented reality. In *2022 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 489–498. IEEE, 2022. doi: 10.1109/ISMAR55827.2022.00065 3

[65] A. Yassien, P. ElAgroudy, E. Makled, and S. Abdennadher. A design space for social presence in vr. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pp. 1–12, 2020. doi: 10.1145/3419249.3420112 2, 3

[66] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 547–556. IEEE, 2019. doi: 10.1109/VR.2019.8797719 3