Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality

Conference Paper · April 2023

DOI: 10.1145/3544548.3580707


Simon Kimmel
OFFIS


Andrii Matviienko
KTH Royal Institute of Technology


Wilko Heuten
OFFIS


Susanne CJ Boll
Carl von Ossietzky Universität Oldenburg

Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality

Simon Kimmel
simon.kimmel@offis.de
OFFIS - Institute for Information Technology
Oldenburg, Germany

Frederike Jung
frederike.jung@offis.de
OFFIS - Institute for Information Technology
Oldenburg, Germany

Andrii Matviienko
matviienko.andrii@gmail.com
KTH Royal Institute of Technology
Stockholm, Sweden

Wilko Heuten
wilko.heuten@offis.de
OFFIS - Institute for Information Technology
Oldenburg, Germany

Susanne Boll
susanne.boll@informatik.uni-oldenburg.de
University of Oldenburg
Oldenburg, Germany

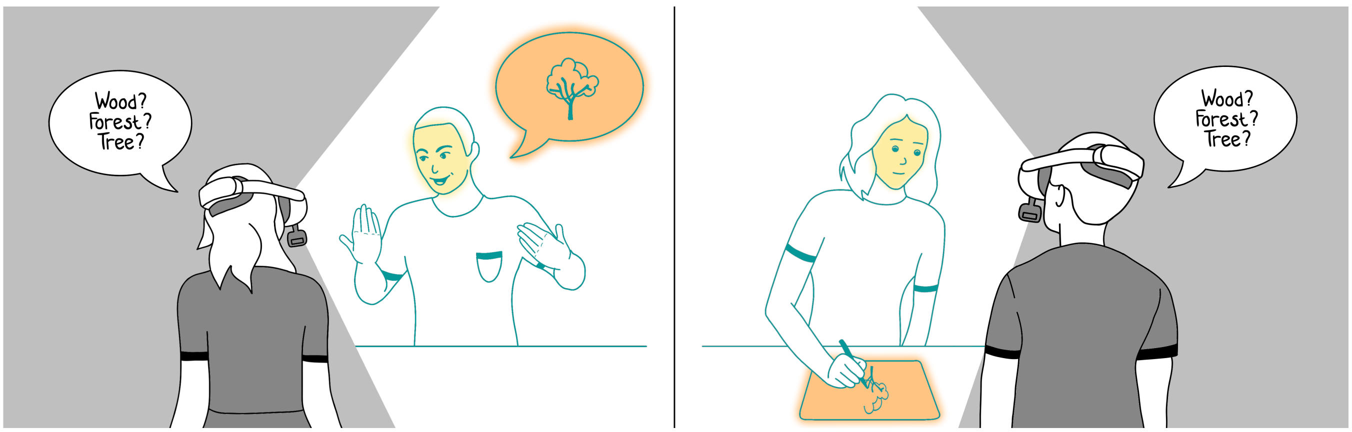

Figure 1: Overview of both collaborative VR explaining tasks (orange) and two exemplary facial expression types (yellow) employed in the evaluation: Verbal explaining of terms in Social VR while rendering eye and mouth movements (left), and graphical explanation of terms while rendering a neutral facial expression (right).

ABSTRACT

As the world becomes more interconnected, physical separation between people increases. Existing collaborative Virtual Reality (VR) applications, designed to bridge this distance, are not yet sufficient in providing a sense of social connection comparable to face-to-face interactions. Possible reasons are the limited multimodality of VR systems and the lack of non-verbal cues in VR avatars. We systematically investigated how facial expressions influence Social Presence in two collaborative VR tasks. We explored four types of facial expressions: eyes and mouth movements, their combination, and no expressions, for two types of explanations: verbal and graphical. To examine how these expressions influence Social Presence, we conducted a controlled VR experiment (N = 48), in which participants had to explain a specific term to their counterpart. Our results demonstrate that eye and mouth movements positively influence Social Presence in VR. Particularly, combining verbal explanations and eye movements induces the highest feeling of co-presence.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; **Virtual reality**; **User studies**.

KEYWORDS

virtual reality, collaboration, social presence, facial expressions

ACM Reference Format:

Simon Kimmel, Frederike Jung, Andrii Matviienko, Wilko Heuten, and Susanne Boll. 2023. Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3544548.3580707

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580707

1 INTRODUCTION

The physical distance between humans is increasing in both private and professional life. The ongoing flexibilization of employment allows more and more employees to work remotely, and globalization enables individuals to relocate their lives to other regions, countries, or continents. Additionally, the ongoing pandemic entails increasing spatial separation between people, with its social distancing and quarantine measures. Researchers introduced Social Virtual Reality (VR) - a shared multi-user VR experience - to bridge this distance and counteract a resulting decline in social connectedness. By combining immersive technologies and environments with the synchronous, interactive engagement of several users [12], Social VR facilitates joint immersion into an alternative virtual environment. Beyond merely looking at a computer screen or limiting the perspective of the communication partner to a headshot, as in videoconferencing, Social VR affords more embodied interaction in a life-like fashion, e.g., joint explorations of virtual worlds, sharing experiences, or playing games. Thus, Social VR offers great potential to create a sustainable solution to the increasing demand for mobility and frequent traveling [8]. However, despite this potential, current VR environments are often considered a “lonely escape” [48], where users may experience isolation from their social encounters and artificial interaction with communication partners. Therefore, VR developers and researchers face the question of how to enhance Social VR experiences to create a “sense of being with another” [5], so-called Social Presence.

As a recent advancement in Social VR, technology developers launched individual sensor systems such as body- or facial trackers with depth cameras and machine learning algorithms to enrich a user’s avatar with different non-verbal behaviors [7, 29, 75]. On the one hand, related research found promising indications that Social VR may benefit from such technologies to display the user’s gestures, body language, or spatial distance [43]. On the other hand, neuro-cognitive and behavioral psychology findings stress that facial expressions are crucial contributors to human real-life communication [2, 19]. So far, how facial expressiveness influences Social Presence has been explored in entirely computer-mediated communication scenarios [54, 55], or asynchronous VR collaborations between computer- and VR-users [24, 25]. While these studies indicate that facial expressions contribute to the quality of technology-mediated social interaction, two limitations emerge for the current research approaches. First, only the computer user’s avatar was enhanced with non-verbal cues. Consequently, the studies lack insights about Social Presence in entirely Social VR-based settings. Second, past work has not sufficiently examined how the types of collaboration within VR influence the impact of facial expressions. Therefore, we address these limitations by employing the recent advances in available face tracking technologies and building upon the reported importance of facial expressiveness in face-to-face communication.

In this work, we investigate how users’ facial expressions on virtual avatars influence Social Presence in collaborative synchronous VR tasks. We conducted an extensive VR experiment (N = 48), systematically comparing the effects of four different levels of facial expressiveness on perceived Social Presence in two Social VR collaboration tasks. The conditions included a) neutral facial expression, b) rendering of the eye and c) mouth movements, and d) a combination of both. The collaboration tasks consisted of a verbal and a graphical explanation task, inspired by two frequently employed types of collaboration in Social VR. Considering that the research and commercialization of Social VR are in their “infancy” or early development stages [18, 33], our approach contributes toward shaping future Social VR applications.

Our results reveal that facial expressions positively influence Social Presence in VR environments. This manifests in elevated co-presence levels and gaze duration, which are determinants of Social Presence. Regarding co-presence, eye movements elicit the highest levels in verbal explanations. In contrast, mouth movements more prominently evoke co-presence than neutral facial expressions when not specifying the explanation type. Generally, verbal explanations have a higher impact on co-presence than graphical ones. For gaze duration, when verbal explanations are combined with eye and mouth movements, VR partners look at each others’ faces the longest.

The contribution of our work is two-fold:

- We provide an empirical evaluation of facial expressions in verbal and non-verbal collaborative Social VR tasks, demonstrating that rendered eye and mouth movements increase Social Presence, and collaboration tasks modulate the impact of facial expressions, particularly for eye movements.
- We contribute design recommendations for developers and designers to create Social VR experiences and, ultimately, increase connectedness over distance.

2 RELATED WORK

In this section, we outline and connect our approach to the existing body of work about increasing Social Presence in VR, focusing on behavioral realism and highlighting current research developments on real-time VR facial rendering.

2.1 Increasing Social Presence in VR

Social Presence is one of the leading factors for effective Social VR simulations [64, 74]. In contrast to concepts of sociability or social space, Social Presence describes the unique psychological phenomenon of perceiving a technologically-mediated counterpart as “real” [36]. Thus, it describes how well a VR system elicits the illusion of sharing the virtual space with someone else and whether the user has access to another’s attitudes, feelings, or emotions [5, 64]. In addition, it is a predictor of a variety of positive communication outcomes, including trust, enjoyment, and attractiveness [53]. Beyond creating standardized measurements [5, 23, 58, 64], several initiatives have identified factors that impact Social Presence levels when using communication technologies [53, 74]. These include psychological aspects such as (1) user characteristics, contextual attributes such as (2) task type, or technological properties like (3) user representation.

Researchers determined users’ demographic and personality-related characteristics as essential for Social Presence. As an example, women and people with a greater intrinsic desire for social interaction generally perceive stronger Social Presence [20, 31, 32]. Beyond these psychological aspects, related work illustrated that the type of activity that users jointly engage in (e.g., independent vs. interdependent; collaborative vs. competitive) within the virtual environment can affect perceived Social Presence levels [53, 74]. Thus, sharing a global goal by collectively solving a task may significantly increase Social Presence [70, 74]. However, one of the most discussed aspects to potentially influence Social Presence in the context of Social VR is the representation of users. This refers to the visualization of self-embodiment and embodiment of the communication partner. In this regard, an embodiment reinforces the perception of Social Presence [1, 63, 66]. Yet, the findings on how this representation should appear are ambiguous. Photographically realistic avatars may enhance human-likeness rates and virtual body acceptance [38]. However, there are no conclusive reports of photo-realism increasing perceived Social Presence [38, 53, 74]. This is often attributed to the Uncanny Valley effect, describing an eerie sensation when an avatar imperfectly resembles a human [50]. Further, in a social setup, preferences regarding the photo-realistic representation of an avatar may highly depend on the relationship and attitudes the communicators have towards one another [59]. In contrast to photo-realism with its inherent difficulties, researchers have found that the behavioral realism of user representations is a more reliable attributing factor for inducing Social Presence [28, 56, 63, 65, 68].

In this paper, we aim to increase the Social Presence of Social VR by focusing on the technological characteristics of a system, in particular on the user representation. Thence, we explore the concept of behavioral realism in more detail.

2.2 Behavioral Realism and Facial Expressiveness

According to (socio-) psychological theories, non-verbal behavior is an integral part of how humans interact and communicate with one another [2, 9, 10, 69]. Non-verbal cues may lead to coordination and understanding among individuals [10]. Thus, elevating an avatar's behavioral realism through non-verbal cues is associated with positively affecting perceived Social Presence [28, 56, 63, 65, 68]. In their extensive review of Social Presence research, Yassien et al. defined self-embodiment and non-verbal cues as key design dimensions currently underexplored [74]. The literature distinguishes four modalities of non-verbal communication cues: (1) gestural behavior, (2) proximity or spatial behavior, (3) gaze, and (4) facial behavior [9, 43, 49].

Several researchers have built and evaluated systems for tracking and rendering bodily movements or gaze behavior [43, 60, 63, 65]. They found significant increases in Social Presence levels by displaying participants' movements in real-time onto VR avatars. This applied to VR settings in which participants portrayed adequate movement behavior [43, 63, 65] and perceived virtual eye contact [60]. Furthermore, Maloney et al. [43] explored specific types of non-verbal behavior in an observational and interview study. They showed that participants translated known non-verbal communication cues from the real-life human-human interaction into Social VR settings of avatar-avatar communication. Gestural behavior, such as dancing, 'talking to the hand', or blowing kisses, mainly affected social interaction beyond verbal content. Although the authors did not measure perceived Social Presence, they underlined the generally positive feedback regarding non-verbal gestural and proximity cues and their contribution to more realistic Social VR experiences. The presented body of work emphasized the benefits of implementing non-verbal cues. However, it overlooked what is considered "the most significant non-verbal language to communicate emotions since the beginning of human evolution" [2] – facial expressions.

With the emergence of novel methods for facial motion tracking with advances in depth-camera technology [24, 42, 51, 61] and machine learning algorithms [7, 75], previous work examined VR settings that capture facial expressions. For example, researchers investigated different levels of manipulated facial expressions in a computer-mediated scenario and the impact of facial expressions on perceived Social Presence [25, 54, 55]. Despite most of the participants missing the manipulation, users experienced higher Social Presence and described their social interactions as more positive with enhanced avatar's facial expressions [54]. These findings underline that face animations may strongly, yet subconsciously, contribute to the social experience in technology-mediated communication. Further studies explored asymmetric VR experiments with one communication partner in VR and the other in front of a computer screen to estimate the role of face and upper body movements [25, 55]. The results highlighted the crucial role of facial expressions over body movements, which was also supported by Hart et al. [25]. Participants liked each other more, formed more accurate impressions about their partners [55], and felt the communication improved [25] when facial expressions were available.

Overall, the presented body of work underlines the research potential of facial expressions for Social VR settings. As experiments in this field primarily rely on computer-mediated or asymmetrical VR settings, little is known about the effects of facial cues in a purely VR-based collaboration. To our knowledge, no such evaluation of facial expressions regarding Social Presence has yet been conducted. Further, current findings on Social Presence lack the consideration of "boundary conditions" [53], referring to interactions among factors that influence Social Presence, which determine how strong an influence is and whether it comes into effect at all. For Social VR, such boundary conditions may comprise the type of activity the users partake in. In this context, the consideration of collaboration tasks is particularly interesting, as collaborations are associated with increased levels of Social Presence [74]. In this work, we aim to contribute to closing both of the above-identified research gaps through a controlled VR experiment.

3 EVALUATION

To systematically investigate real-time tracked facial expressions in Social VR applications for two types of collaboration tasks, we conducted a controlled VR experiment with pairs of participants physically situated in different locations. With this experiment, we addressed the following two research questions:

RQ1 *"How can we increase Social Presence in Social VR applications using real-time tracked facial expressions?"*

RQ2 *"How does the performed collaboration task influence the impact that facial expressions elicit on Social Presence?"*

3.1 Study Design

To answer the research questions, we designed a within-subject study with two independent variables: (1) facial expressions and, as a realization of a collaboration task, (2) types of explanation. We outline these in the following.

3.1.1 Facial Expressions. The participant is embodied in the virtual space via an avatar. When changing facial expressions in the real world, e.g., talking or smiling, the augmentations are directly mapped onto the avatar's face. Gaze and facial expressiveness are particularly important and well-studied for non-verbal interpersonal communication and Social Presence. In effect, in the Facial Action Coding System (FACS), a common standard to taxonomize human facial movements, eyes and mouth areas are highly represented [13, 14]. Since its adoption and publication in the late 1970s, the FACS has been commonly used by psychologists, animators, and developers of computer vision software to identify, code, and recreate human emotions [21, 41]. Given the prominent and detailed portrayal of eyes (i.e. example codes "upper lid raiser", "eyes turn left", "eyes turn right", "eyes up", "eyes down", etc.) and mouth (i.e. example codes "lips part", "lip stretcher", "tongue show", "mouth stretch", etc.)[14], we also focused on representing these facial areas. Therefore, we employed current sensor technologies to track eye and mouth movements in real time. In total, we implemented four levels of facial expressions: (1) neutral facial expression without any movements as a baseline, as commonly employed in current Social VR applications, (2) rendering eye movements, (3) mouth movements, and (4) a combination of both.

For *neutral facial expressions*, the faces of the participants' avatars are frozen in a neutral position. More specifically, the gaze is directed straight ahead, eyes are fully opened, the avatar's mouth is closed, and the corners of the mouth are neither lowered nor lifted. This neutral expression was based on the predefined neutral position of the employed avatar system (see Subsection 3.2). For the *rendering of real-time eye movements*, avatars' eye movements mirrored the actual eye movement of the participants. In particular, the avatar's eyes showed the real-life gaze direction, blinking, and the degree to which the participant's eyes are open or closed. Similar to the eye movements, *real-time mouth movements* were moving as the participant's lips. For the *combination of eye and mouth movements*, both approaches were combined.

3.1.2 Types of Explanation. As we aim for a systematic understanding of facial expressiveness in Social VR, we portray the expressions during two variations of a collaborative task.Specifically, (1) verbal and (2) graphical (non-verbal) explanation. Their selection was inspired by two commonly used types of collaborations implemented in Social VR. These include verbal communication in the form of audio chats on the one hand [22], and graphical communication in the form of freehand sketching, painting, or drawing [22, 27, 39, 57], during which often "a whiteboard is the center of focus" [39]. By relating our task selection to current VR collaboration practices, we aim to amplify the relevance of our research for designers and developers of Social VR applications.

As to validly contrast verbal and graphical collaboration, both were implemented in the study as individual yet complementary explanation tasks. These were modeled after well-known parlor games[1]. Doing so, we transferred the often group-oriented collaborations of current Social VR solutions into a 1-on-1 collaboration. This was required for the feasibility of this study. Furthermore, using these types of explanation ensured them being comparable in two characteristics. First, they can be performed at the same location in the virtual environment, which ensures consistent illumination of avatar faces. Second, both types of explanation only indirectly rely on facial expressions. Thus, facial expressions are neither necessary to complete the task nor to depict the explained content. Both types of explanation solely differ in the visual focus directed towards the counterpart's actions and explanation process. While for verbal explanation one's counterpart is located in the center of attention throughout the entire collaboration, for graphical interaction one's counterpart is mostly perceived peripherally.

For the *verbal explanation*, the participant explains the word verbally without using words with the same root or synonyms. For the *graphical explanation*, the participant draws the word on a virtual tablet displayed on the table using a virtual marker. As for the setting, two VR users meet at a virtual table (Figure 2), where they see each other's avatar, communicate and interact with one another to solve the explain-and-guess task collaboratively. One partner explains a word shown on a virtual card. If the other partner guesses correctly, according to the explaining party, a new card and term turn over. Upon an incorrect answer, the guessing continues for unlimited trials until the word is identified or the explaining participant decides to skip the card. This activity was chosen as a catalyst for social interactions. Furthermore, its collaborative and interdependent characteristics are associated with increased Social Presence [74]. The explanation game thereby assures a fundamental level of participant interaction and Social Presence for our experiment. Independent of the type of explanation, after one minute, the roles of the two participants are swapped, and the previously guessing participant assumes the explaining role and vice versa. Participants were encouraged to guess as many terms as possible in the given time frame for both explanations. We had 196 words in total to ensure no repetition. None of the participants went through all of the cards in our experiment. The terms were taken from other parlor games that employed graphical and verbal explanations to ensure that the terms could be explained using the selected explanation types. The occurrence of a specific emotional reaction of the participants upon viewing the terms could not be eliminated due to the individuality of such responses. However, to omit emotional reactivity as a confounding variable, the sequence in which the terms were presented was randomized across diads.

To systematically investigate these independent variables, we combined all levels of facial expressions and types of explanations, which resulted in eight experimental conditions. The order of conditions was randomized and each condition lasted two minutes in total. This duration was selected based on internal preliminary tests, showing that it was sufficient to perceive the manipulation of the facial animations and short enough to conduct the experiment without any intermissions.

[1] i.e., Taboo [26], Pictionary [44]

Figure 2: User in the real-world environment wearing VR glasses and the avatar embodying her in both types of explanation. The avatar portrays the user's eye movements (gaze direction, degree to which eyes are opened) and mouth movements (position of various mouth landmarks) while performing verbal explanation (left), and the avatar shows neutral facial expression in graphical explanation (right).

3.2 Apparatus

To enable the planned evaluation, we developed the subsequent apparatus. We designed an application using Unity 2020.3.25f1 that ran on two Pico Neo 3 Pro Eye. The application utilized the Photon Unity Networking framework to synchronize the current application state and transmit voice chat among both HMDs. Both HMDs integrate built-in Tobii VR4 Platform eye tracking with a sampling rate of 60Hz and sub-degree gaze accuracy [16]. For mouth tracking, the apparatus included two VIVE facial trackers, one of which was attached to each of the HMDs using a custom 3D printed mount, originally designed by Oliwier Krawczyk [35]. The utilized facial tracker consisted of a dual camera with infrared illumination to track facial movements with a sampling rate of 60Hz and a response time below 10 milliseconds [29]. The facial tracker was connected to a PC via USB running custom software. The software was employed to analyze the facial tracking data and transmit it wirelessly to the HMD via UDP. We used two hardware setups to run the software, depending on the experiment room: (1) a Windows10-based PC with 8GB memory, an Intel Core i7-8700K CPU, and an NVIDIA GeForce GTX 1080TI graphics card, and (2) a Windows10-based PC with 32GB memory, an Intel Core i7-10750H CPU and an NVIDIA GeForce RTX 2070 Super graphics card.

Mouth and eye movements were mapped onto full-body avatars designed using *Ready Player Me SDK v.1.7.0* [72]. *Ready Player Me* is a cross-game avatar platform used by several of the biggest Social VR applications currently on the market, such as *VRChat*, *Spatial*, and *MeetinVR* [73]. To further increase the accuracy of the mouth movements, especially when speaking, we used the Oculus Lip Sync SDK [46]. To facilitate synchronization between mouth movements and voice output, we rendered mouth movements in accordance with the latency of the voice chat. All questionnaires, integrated into the VR application, were implemented based on Feick et al.'s VR Questionnaire Toolkit [17].

3.3 Measures

We measured multiple dependent variables to compare all facial expressions and explanations. Following the suggestion of Sterna et al. for quantifying Social Presence, these consisted of several factors of the Networked Minds Social Presence Inventory (NMSPI) [5, 23] and behavioral measurements [64]. We selected the NMSPI as it is psychometrically validated and well-documented [4, 5, 64]. The following NMSPI factors were assessed using a 7-point Likert scale:

- *Co-Presence*: It describes a user's ability to identify the presence of another entity in a shared virtual environment [5]. To measure co-presence, the scale includes items that rate one's own sense of co-presence (perception of self), and one's perception of the co-presence of one's counterpart (perception of other).
- *Perceived Affective Understanding*: It is composed of two aspects: First, it refers to the user's ability to comprehend the emotions and attitudes of their counterpart during an interaction (perception of self). Second, it includes a user's

perception of their counterpart's ability to comprehend the user's emotions and attitudes [23].

- *Perceived Affective Interdependence*: It describes "the extent to which the user's emotional and attitudinal state affects and is affected by the emotional and attitudinal states of the interactant"[23]. It can be subdivided into perception of self and perception of other.

The explanation task presented in this study is assumed to strongly influence the other NMSPI factors *Attentional Allocation*, *Perceived Message Understanding*, and *Perceived Behavioral Interdependence*. For instance, the collaborative nature of explaining is likely to create a behavioral and attentional dependency among the participants. Consequently, these factors could not be evaluated validly and comparably and were discarded.

The behavioral measures comprised are the following:

- *Gaze Duration (s/min)*: We measured the duration for which a participant directs their gaze at their counterpart's face to assess eye contact. This definition of eye contact is in line with other research approaches [34]. It was derived from analyzing participant's eye movements via eye tracking with a sampling rate of 60Hz and is given in seconds per minute. Establishing and maintaining eye contact with another social entity is an important indicator of intimacy and interpersonal relationship [3, 30]. Although not yet fully validated as a measurement of Social presence, a variety of studies have found indications for a relation between eye contact and this phenomenon [6, 40, 52, 68]. Hence, tracking eye contact may provide valuable implications about perceived Social Presence [64, 74].
- *Gaze Frequency (num/min)*: In relation to participant's eye movements, we additionally measured the number of times that participants directed their gaze at their counterpart's face. It was measured from eye tracking data with a sampling rate of 60Hz and is depicted in number per minute. This measurement allows for gaining a deeper understanding of participants' gaze behavior.
- *Mouth Weight Changes (num/s)*: We evaluated 38 weight measurements collected by the employed VIVE facial tracker that each referred to one specific region of the participants' current mouth position. The measurements were examined for changes with a sampling rate of 12Hz. Finally, the mouth weight changes were summed up over all 38 mouth-related regions to acquire a final score. Measuring mouth movements based on changes in different regions of the mouth is oriented on a system of McIntyre et al. [45] which in turn is in line with the prominent FACS [13, 14]. Based on this measure we derived the level of participant's mouth movements, and whether it changed throughout the study.
- *Head Position Changes (num/s)*: We examined changes in the position of the participant's head. This measure was sampled with 12Hz from the HMD's positional sensors and is given in number of changes per second. We used it to assess whether the body movements of the participants changed.
- *Head Rotation Changes (num/s)*: Based on the same rationale, we analyzed head rotation data collected by the HMD with a sampling rate of 12Hz and depicted in number of changes per second.

All behavioral measurements that determine changes in data use tolerance values to minimize measurement distortion due to sensor noise. All employed tolerances were developed empirically through a series of preliminary tests.

We acquired additional feedback via a custom post-experiment questionnaire displayed on a tablet. It comprised of:

- *Manipulation Check*: Participants were asked whether they noticed changes in the animation of their opponent's avatar using a dichotomous yes/no-question. If changes were detected, participants were additionally asked, which changes were noticed using an open question format.
- *Ranking Questions*: Three ranking questions were included, in which participants were instructed to rank the used facial expression types in relation to subjective preference. More precisely, participants were asked to order the facial expression types from most to least favorable for each specific type of explanation (verbal & graphical) as well as for the Social VR application in general. In addition, participants were asked to justify their rankings in an open question format. The ranking questions allowed a direct subjective comparison of the employed levels of facial expressions from the participant's perspective. This allowed us to contextualize the behavioral measures and identify underlying causes for differences between them and subjective opinions. In addition, opinions about the application, in general, were included to detect whether preferences were specific to the employed explanation tasks.
- *Social Interaction Feedback*: Finally, participants were asked for general feedback about interacting with another social entity in the Social VR environment. For this, we used two open questions, about their likes and dislikes.

3.4 Participants

We recruited 48 participants (17 female, 30 male, and 1 divers) aged between 17 and 47 (*M*: 24.46, *SD*: 5.05). Most (N = 21) stated that they had no previous VR experience, second to most participants claimed to have had some VR experience (N = 15) (*Median Likert Score*: 2). All participants had normal or corrected-to-normal eyesight. Since for each run of the study two participants had to interact with each other, it was ensured that both were strangers. This ruled out possible existing relationships among participants as a confounding variable. Moreover, as the communication was entirely in German, it was ensured that all participants had sufficient German language proficiency (level C1 or higher). Each participant received compensation of 13€ for taking part in the experiment. We ensured the complete anonymization of participants by assigning a randomized participant ID ranging from 1 to 100. This procedure prevented participants from deducing the ID of their interaction partner from their own ID, thus invalidating the anonymization.

3.5 Procedure

For our experiment, we followed strict hygiene measures to limit the risk of infection with the COVID-19 virus. These included regular ventilation of rooms and disinfection of the equipment. For each

Table 1: Median (Md) and interquartile range (IQR) of the three measured factors of the NMSPI using 7-point Likert scales: Each factor is subdivided into scores portraying the perception of self, the perception of other, and the combination of both for respective factor (Verb.: Verbal, Graph.: Graphical, N.: Neutral Facial Expression, FE.: Eye Movements, M.: Mouth Movements, E.&M.: Eyes and Mouth Movements).

	Co-Presence						Perc. Affective Understanding						Perc. Emotional Interdep.					
	Self		Other		Combined		Self		Other		Combined		Self		Other		Combined	
	Md	IQR	Md	IQR	Md	IQR	Md	IQR	Md	IQR	Md	IQR	Md	IQR	Md	IQR	Md	IQR
Verb.-N.	7.00	1.00	6.67	1.08	6.67	1.00	4.67	2.00	4.50	2.00	4.58	2.08	5.00	2.42	4.67	1.75	4.92	2.13
Verb.-E.	6.67	1.33	6.67	1.67	7.00	0.33	4.67	2.33	4.50	1.67	4.67	2.04	5.00	2.50	5.00	2.42	5.00	2.00
Verb.-M.	7.00	0.33	7.00	0.67	6.92	0.71	4.67	2.00	4.67	1.83	4.58	1.88	5.00	2.00	5.00	2.00	5.00	2.21
Verb.-E.&M.	6.67	1.00	6.33	1.08	6.75	1.00	4.67	2.33	4.33	1.75	4.42	2.08	5.00	2.50	4.67	2.67	4.83	2.71
Graph.-N.	6.67	0.67	6.67	1.00	6.50	1.04	4.67	2.33	4.33	1.75	4.50	2.04	5.00	2.08	4.67	2.08	4.83	2.50
Graph.-E.	7.00	0.67	7.00	1.00	6.67	1.54	4.67	2.33	4.67	2.00	4.50	1.88	5.00	2.33	5.00	2.08	4.83	2.46
Graph.-M.	6.83	1.00	6.67	1.00	6.83	0.75	4.50	2.33	4.33	1.67	4.67	2.17	5.00	2.67	4.67	2.67	5.33	2.54
Graph.-E.&M.	7.00	0.67	6.67	1.00	6.83	1.04	4.67	2.67	4.67	2.33	4.58	2.00	5.33	2.67	5.00	2.33	4.83	1.96

session, two participants were invited simultaneously. Upon arrival, each was guided to a separate room, being each supervised by one instructor for the entire experiment. After obtaining informed consent, we briefly explained the experimental procedure and collected demographic data. Subsequently, the participants familiarized themselves with the control of their avatar in the VR environment during a short introductory sequence. Here, they selected one of four predefined avatars with different physical features to embody them throughout the experiment. Afterward, we calibrated the eye and face trackers for the participant's facial features. During calibration, we led participants to believe that the trackers were required for correct headset functioning. This ensured that they were unaware of the subsequent manipulation of facial expressions prior to the experiment. Following successful calibration, the participants entered the VR experiment environment, to get acquainted with their interaction partner. Subsequently, the experiment began with two short training rounds to practice first, the verbal and then the graphical explanation. During this training, the participants' avatars did not render any facial expressions. The main experimental procedure began after training completion. After each condition, the participants were instructed to fill in the Networked Minds Social Presence Inventory within the VR environment. This reduced study duration without affecting the validity of the measurement [62]. At the end of the study, the participants filled out the post-experiment questionnaire. The presented experimental procedure was reviewed for ethical compliance by the in-house study board and tested in a pilot study with 4 participants.

3.6 Data Analysis

The factors co-presence, perceived affective understanding, and perceived affective interdependence of the NMSPI were each analyzed concerning the participant's perception of self, perception of other, and by combining both, following the approach of the questionnaire guidelines [4]. All data we collected (including behavioral measures) proved to be non-parametric (*Shapiro-Wilk test*: $p < 0.05$) except for the combined perceived affective understanding scores. For this reason, we applied aligned rank transform for non-parametric factorial analysis for the assessment of the respective data [15, 71]. Combined perceived affective understanding data exhibited normality even after controlling for equality of variance (*Levene-Test*: $p > 0.05$), skewness (*estimate*: -0.24), and kurtosis (*estimate*: -0.54). Thus, they were analyzed using a two-way repeated measures ANOVA. For the main effect analysis for facial expressions, results were accumulated over the types of explanation. For the main effect analysis for types of explanation, on the other hand, results were accumulated over levels of facial expressions. For post-hoc analysis, we applied contrast tests using Tukey correction method. The qualitative data was evaluated using MAXQDA [67] and following Kuckartz's method for computer-aided content analysis [37].

4 RESULTS

We found that facial expressions positively influence Social Presence in VR environments. This manifests in elevated co-presence levels and gaze duration. Regarding co-presence, eye movements elicit highest levels in verbal explanations, while mouth movements more prominently evoke co-presence in comparison to neutral facial expressions, when not specifying the explanation type. Generally, verbal explanations have a higher impact on co-presence than graphical. For gaze duration, when verbal explanations are combined with both eye and mouth movements, VR partners look at each others' faces the longest. In the following, we outline these results and the qualitative findings in detail.

4.1 Networked Minds Social Presence Inventory

The NMSPI scores indicated statistically significant differences for co-presence. While verbal explanation generally increased the co-presence score, we found it to be the highest when combined with eye movements. Furthermore, mouth movements outperformed no facial expressions in terms of co-presence. An overview of all NMSPI results is shown in Table 1 and Figure 3. Subsequently, we elaborate on each recorded factor in detail.

4.1.1 Co-Presence - Combined Score. Participants felt highest co-presence when mouth movements were presented to a VR partner (*Md*: 6.83, *IQR*: 0.71), followed by both eye movements (*Md*: 6.83, *IQR*: 1), and the combination of both (*Md*: 6.83, *IQR*: 1). Moreover, the feeling of co-presence was the lowest without facial rendering

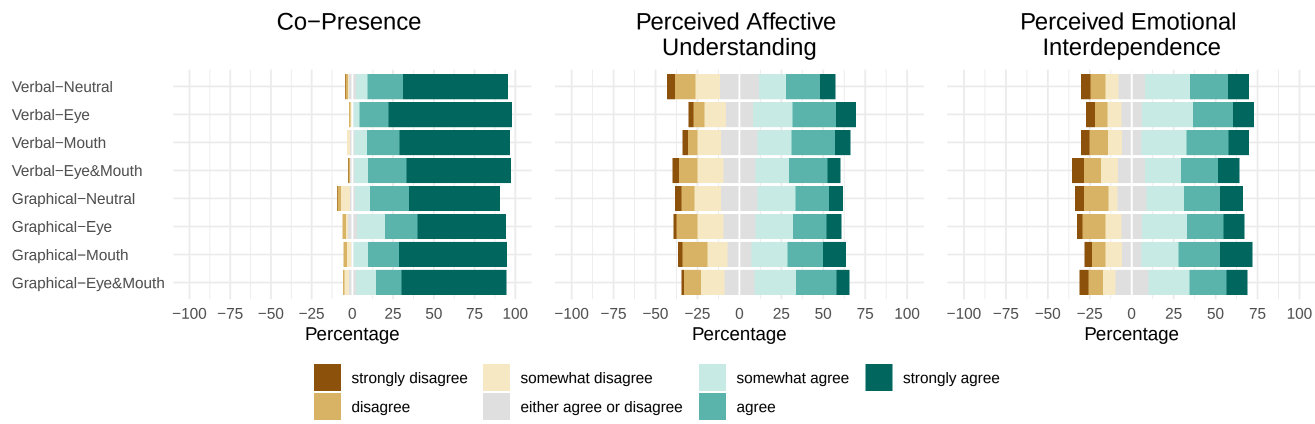

Figure 3: Overview of Likert data for each measured factor of the NMSPI: corresponding items were summarized for factors co-presence (left), perceived affective understanding (middle), and perceived affective interdependence (right). For each figure stacked bar charts are illustrated per experimental condition. All bars depict the distribution of Likert responses for all participants. The distributions combine items referencing perception of self and perception of other.

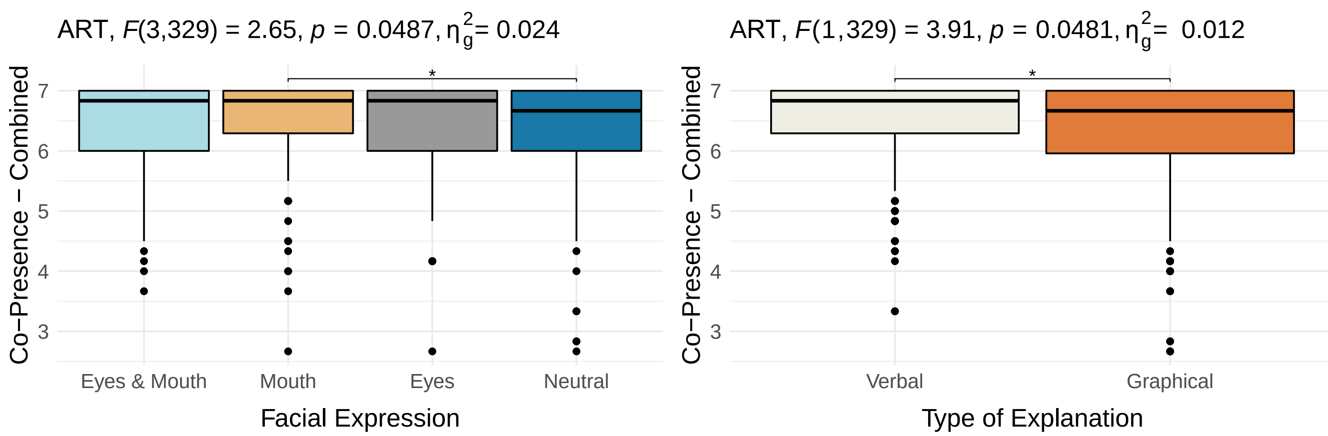

Figure 4: Results of inferential statistical analysis of combined co-presence scores using Analysis of Variance of Aligned Rank Transformed Data (ART) and post-hoc contrast tests with Tukey correction: analysis of combined co-presence scores averaged over facial expression method (left), and analysis of combined co-presence scores averaged over the types of explanation (right). All boxplots show median values, IQR, and outliers (outside 1.5 times the IQR above the upper quartile and below the lower quartile). Statistically significant differences among groups are highlighted by brackets and stars ($0 < {}^{***} \leq 0.001 < {}^{**} \leq 0.01 < {}^{*} \leq 0.05$).

(*Md*: 6.67, *IQR*: 1). This finding was supported by a statistically significant main effect for facial expressions ($F(3,329) = 2.65$, $p = 0.049$, $\eta_g^2 = 0.023$). However, the post-hoc analysis has shown that mouth movements induced a statistically higher feeling of co-presence compared to neutral facial expressions ($p = 0.034$) (Figure 4 (left)). As for the type of explanations, the verbal explanation elicited higher combined co-presence scores (*Md*: 6.83, *IQR*: 0.71) than the graphical one (*Md*: 6.67, *IQR*: 1.04). This finding was confirmed by a statistically significant main effect for type of explanation ($F(1,329) = 3.93$, $p = 0.048$, $\eta_g^2 = 0.011$) (Figure 4 (right)). Lastly, the statistical analysis revealed an interaction effect for facial rendering method*type of explanation ($F(3,329) = 3.91$, $p = 0.009$, $\eta_g^2 = 0.034$) (see Figure 6 (left)). The post-hoc analysis revealed that eye movements in the verbal explanation elicited higher combined co-presence compared to any facial movement with the graphical explanation ($p = 0.0001$), and neutral facial expressions in the verbal explanation ($p = 0.035$). Moreover, combined co-presence for mouth movements in the verbal explanation was statistically higher than with eye movements in the graphical task ($p = 0.031$). All remaining pairwise comparisons were not statistically different.

4.1.2 Co-Presence - Perception of Self. We did not observe a statistically significant main effect for the perception of self scores for the facial rendering methods ($F(3,329) = 1.77$, $p = 0.153$). We discovered that co-presence - perception of self with the verbal explanation (*Md*: 7, *IQR*: 0.67) elicited increased perception of self

Table 2: Summary of descriptive analysis per combination of rendered facial expression and type of explanation (Verb.: Verbal Explanation, Graph.: Graphical Explanation): The table includes means (M) and standard deviations (SD) for Gaze Duration (in seconds per minute), Number of Gazes (in number per minutes), Mouth Weight Changes (in number per seconds), Head Position Changes (in number per seconds), and Head Rotation Changes (in number per seconds).

	Gaze Duration (s/min)		Gaze Frequency (num/min)		Mouth Weight Changes (num/s)		Head Position Changes (num/s)		Head Rotation Changes (num/s)	
	M	SD	M	SD	M	SD	M	SD	M	SD
Verb.-N.	9.63	6.90	15.35	9.63	66.55	33.56	10.98	3.43	20.00	4.04
Verb.-Eyes	10.29	7.69	15.65	9.74	64.68	34.19	10.83	3.22	19.97	4.08
Verb.-Mouth	12.66	8.19	16.71	10.14	68.15	33.16	10.52	3.07	19.45	3.77
Verb.-Eyes&Mouth	14.19	8.50	16.19	9.49	62.86	35.35	10.63	3.05	19.37	3.63
Graph.-N.	0.57	0.87	1.56	1.56	56.24	30.32	9.15	1.92	15.89	2.69
Graph.-Eyes	0.84	0.97	1.98	2.06	55.43	29.25	9.33	1.92	16.02	2.69
Graph.-Mouth	1.00	1.14	2.13	2.23	54.34	30.22	9.53	1.68	16.22	2.33
Graph.-Eyes&Mouth	1.21	2.09	1.80	2.07	57.67	31.29	9.28	1.71	15.92	2.48

co-presence compared to the graphical task (*Md*: 7, *IQR*: 1.05). A statistically significant main effect supported this finding (F(1,329) = 6.33, p = 0.0123, η_g^2 = 0.019). Furthermore, the statistical analysis showed an interaction effect for facial rendering method*type of explanation (F(3,329) = 4.01, p = 0.008, η_g^2 = 0.035). Post-hoc contrast tests revealed that eye movements in the verbal explanation resulted in increased co-presence compared to eye movements (p < 0.0001) and neutral facial expressions in the graphical task (p = 0.0012). Moreover, the verbal task with only mouth movements statistically significantly increased co-presence compared to the graphical task with only eye movements (p = 0.034). All remaining pairwise comparisons were not statistically different.

4.1.3 Co-Presence - Perception of Other. As for the co-presence - perception of other scores, we observed that mouth movements (*Md*: 6.83, *IQR*: 1) have induced more perception of other than methods (*Md*: 6.67, *IQR*: 1). This result is confirmed by a statistically significant main effect for facial rendering method (F(3,329) = 2.64, p = 0.049, η_g^2 = 0.024). The post-hoc analysis concluded that there was a statistically significant difference between rendering solely mouth movements and neutral facial expression (p = 0.046). Considering the co-presence - perception of other score in terms of type of explanation, no differences between the levels are apparent (*Md*: 6.67, *IQR*: 1 for all levels). Consequently, no main effect was observed regarding type of explanation (F(1,329) = 2.21, p > 0.05). Finally, we found an interaction effect for facial rendering method*type of explanation (F(3,329) = 3.47, p = 0.016, η_g^2 = 0.031). The post-hoc analysis disclosed that the verbal task with just eye movements entails higher co-presence - perception of other scores than the graphical task with eye movements (p = 0.0006) and the graphical task with neutral facial expressions (p = 0.0003). Additionally, mouth movements in the verbal task outperform neutral facial expressions in the graphical task (p = 0.0322).

4.1.4 Perceived Affective Understanding & Interdependence. The scores were analyzed in terms of perception of self, perception of other, and in combination. Our analysis, however, did not reveal statistically significant main and interaction effects (p > 0.05), making the scores comparable across all facial rendering methods and types of explanation.

4.2 Behavioral Measures

From the behavioral measures, we discovered that the eye and mouth movements generally entailed longer gaze duration compared to other facial expressions. Furthermore, verbal task for both eye and mouth movement results in the longest time looking at the partner's face. However, the frequency of glances at the opponent's face did not change based on the facial expressions. Moreover, verbal expressions outperformed graphical ones in terms of gaze frequency, summed-up mouth weight changes, and head position and rotation. The summary of results is shown in Table 2. In the following, we present the results in detail.

4.2.1 Gaze Duration. Our results indicated that participants directed their gaze the longest at their partner's face when both eye- and mouth movements were rendered (*M*: 7.7, *SD*: 8.97), followed by sole mouth (*M*: 6.83, *SD*: 8.25) and eye movements (*M*: 5.56, *SD*: 7.23). Moreover, participants spent a shorter time looking at their opponents with neutral facial expressions (*M*: 5.1, *SD*: 6.67). These findings are supported by a statistically significant main effect for facial expressions (F(3,329) = 22.11, p < 0.0001, η_g^2 = 0.168). Post-hoc analysis indicated that eye- and mouth movements entailed an increase in gaze duration compared to eye movements (p < 0.0001), and neutral facial expressions (p < 0.0001). Furthermore, mouth movements alone proved to be significantly distinct from rendering just eye movements (p = 0.0038) and neutral facial expressions (p < 0.0001) (Figure 5 (left)). Over all expressions, the gaze duration in the verbal explanation (*M*: 11.7, *SD*: 7.99) is higher than in the graphical explanation (*M*: 0.90, *SD*: 1.36). The main effect was statistically significant (F(1,329) = 1224.77, p < 0.0001, η_g^2 = 0.788). Additionally, we identified a statistically significant interaction effect for facial expression method*type of explanation (F(3,329) = 15.420, p < 0.0001, η_g^2 = 0.123) (Figure 6 (right)). Post-hoc analysis indicated that verbal explanation entails increased gaze duration for all facial expressions compared to the graphical explanation task (p

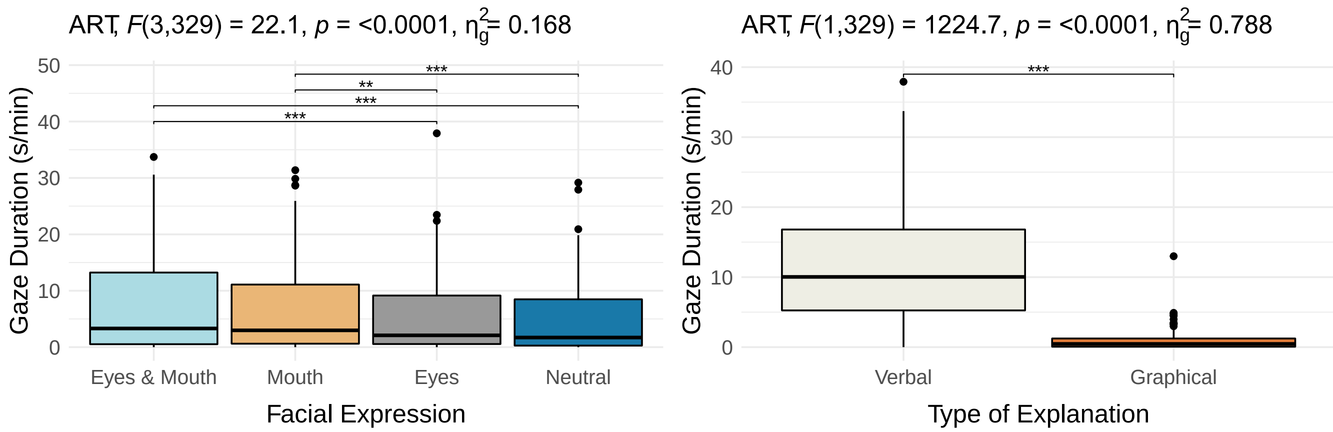

Figure 5: Results of inferential statistical analysis of gaze duration data using ART and post-hoc contrast test with Tukey correction: analysis of duration of gaze on opponents face averaged over facial expression methods (left), and analysis of duration of gaze on opponents face averaged over type of explanation (right). All boxplots show median values, IQR, and outliers (outside 1.5 times the IQR above the upper quartile and below the lower quartile). Statistically significant differences among groups are highlighted by brackets and stars ($0 < {}^{***} \leq 0.001 < {}^{**} \leq 0.01 < {}^{*} \leq 0.05$).

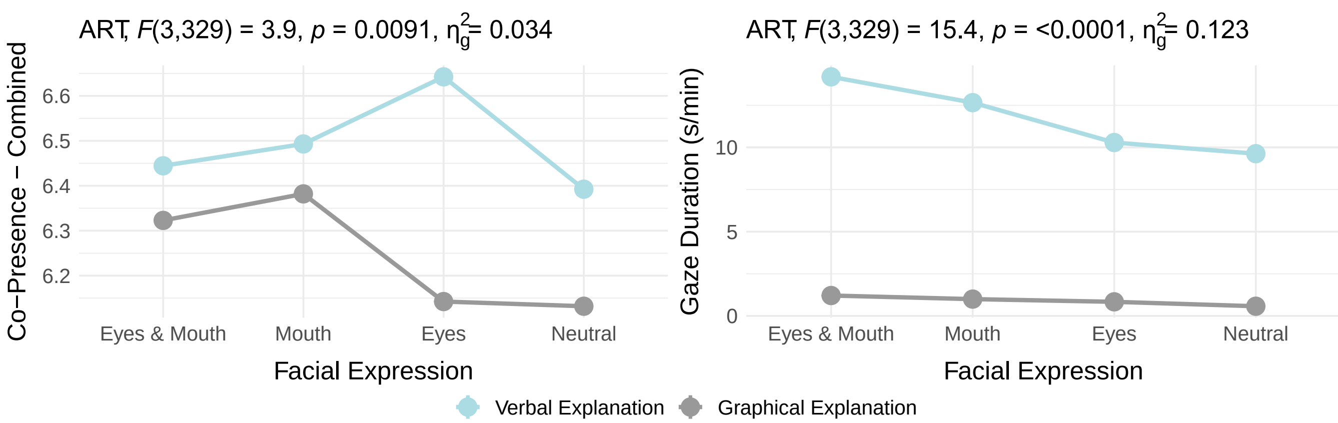

Figure 6: Line graph visualizations of interaction effects for facial expression*type of explanation: Interaction for combined co-presence score (left), and interaction for gaze duration (right). Both figures include one line depicting the respective measure for verbal explanation (blue) and one for graphical explanation (grey) per each facial expression method. Effects were evaluated using ART and post-hoc contrast tests with Tukey correction.

< 0.0001). Furthermore, the performance of the verbal task with eye- and mouth movement elicited significantly increased gaze duration compared to the verbal explanation with solely eye movements ($p = 0.028$) or neutral facial expressions ($p = 0.0014$).

4.2.2 Gaze Frequency. We did not observe a statistically significant difference in the gaze frequency among neutral facial expressions (M: 8.46, SD: 9.75), eye movements (M: 8.81, SD: 9.81), mouth movements (M: 9.42, SD: 10.35), and their combination (M: 8.99, SD: 9.95) ($F(3,329) = 0.76$, $p > 0.05$). However, for the types of expressions, we found that the verbal explanation (M: 15.97, SD: 9.69) entailed an increased gaze frequency compared to the graphical task (M: 1.87, SD: 1.99). This finding is supported by a statistically significant main effect ($F(1,329) = 1471.05$, $p < 0.0001$, $\eta_g^2 = 0.817$). Lastly, an interaction effect for facial expression type*type of explanation was not statistically significant ($F(3,329) = 0.499$, $p > 0.05$).

4.2.3 Mouth Weight Changes. For the sum of mouth weight changes, we found no significant main effect for facial expressions ($F(3,329) = 0.35$, $p > 0.05$). No facial rendering (M: 61.39, SD: 31.99), rendering of eye (M: 60.05, SD: 9.81) and mouth movements (M: 61.24, SD: 32.23), and the combination of both (M: 60.27, SD: 33.31) were comparable to mouth movements. As for the type of explanation, we observed a difference in mouth weight changes. Performing the verbal explanation (M: 65.56, SD: 33.87) entailed more mouth movements than the graphical explanation (M: 55.92, SD: 30.07). This deviation was found to be significant by both a statistically significant main effect ($F(1,329) = 57.07$, $p < 0.0001$, $\eta_g^2 = 0.148$) and post-hoc pairwise

Table 3: Summary of ranking question results: For verbal explanation, graphical explanation, and Social VR applications in general the number of times that a rank was assigned per facial tracking method is displayed.

	Verbal Explanation				Graphical Explanation				In General			
	# Rank 1	# Rank 2	# Rank 3	# Rank 4	# Rank 1	# Rank 2	# Rank 3	# Rank 4	# Rank 1	# Rank 2	# Rank 3	# Rank 4
Neutral	7	4	7	30	18	5	6	19	4	4	11	29
Eyes	7	14	16	11	6	18	15	9	6	17	17	8
Mouth	4	20	17	7	2	17	16	13	3	20	16	9
Eyes&Mouth	30	10	8	0	22	8	11	7	35	7	4	2

comparison among the groups ($p < 0.0001$). An interaction effect for facial expression method and type of explanation is not present ($F(3,329) = 0.499$, $p > 0.05$).

4.2.4 Head Position. There was no significant main effect for changes in head position in terms of facial expressions ($F(3,329) = 0.07$, $p > 0.05$): No facial rendering (*M*:10.07, *SD*: 2.92), eye movements (M: 10.078, SD: 2.74), mouth movements (*M*: 10.023, *SD*: 2.51), and the combination of them (*M*: 9.96, *SD*: 2.56). However, verbal explanation (*M*: 10.74, *SD*: 3.17) with graphical explanation (*M*: 9.32, *SD*: 1.8) showed differences in a statistically significant main effect for type of explanation ($F(1,329) = 54.24$, $p < 0.0001$, $\eta_g^2 = 0.142$). This was further supported by a statistically significant deviation found with post-hoc analysis ($p < 0.0001$). There was no interaction effect for facial expression method*type of explanation identifiable ($F(3,329) = 1.66$, $p > 0.05$).

4.2.5 Head Rotations. Analysis of the head rotations indicated no deviations for neutral facial expression (*M*: 17.94, *SD*: 3.99), eye movement (*M*: 17.99, *SD*: 3.96), mouth movement (*M*: 17.83, *SD*: 3.51), and their combination (*M*: 17.64, *SD*: 3.54). The absence of a statistically significant main effect underpinned this notion ($F(3,329) = 0.07$, $p > 0.05$). The effect of head rotations on the type of explanation, on the other hand, revealed that the verbal explanation (*M*: 9.96, *SD*: 2.56) resulted in more head rotations than the graphical one. This was further supported by a statistically significant main effect for type of explanation ($F(1,329) = 302.68$, $p < 0.0001$, $\eta_g^2 = 0.48$), and the detection of a significant difference doing pairwise comparison ($p < 0.0001$). We did not detect any interaction effects for facial expression method*type of explanation with regards to head rotations ($F(3,329) = 1.37$, $p > 0.05$).

4.3 Qualitative Feedback

Overall, we found that less than half of the participants noticed the changes in the animation of their partner's avatar. Still, when presented with the choice, participants most often preferred the presence of both eye and mouth movement independent of the performed type of explanation but for different reasons. Participants looked more frequently at each other for verbal explanations to connect with them visually, i.e., look each other in the eyes. Participants focused more on the task than their partners for the graphical explanation. However, even here, they underlined the importance of having a fully animated avatar. According to the participants, this would increase the realism of the VR simulation, allow them to observe and recognize their partner's effort to solve the task through their facial expressions, and avoid life-less awkward staring in situations of looking at each other. An overview of the assigned ranks per facial expression method and per type of explanation is depicted in Table 3.

A majority of participants (N = 28) did not notice the animation manipulation. The remaining participants recognized changes in "mouth" movements (16x), "eye" movements (6x), and general "mimics" (2x) of their partner's avatar.

4.3.1 Eye and Mouth. Although so few participants even noticed the animations, they preferred the combination of eye and mouth movements. As shown by the comments about the individual cues and their combination, facial expressions contributed content about the other's emotional state and attention, making the interaction feel more life-like. In effect, conditions without any facial expressions were ranked the lowest for both explaining types and the general VR experience (see Table 3). Participants stated "it felt most realistic when eyes and mouth were moving accordingly" (P98). In particular, eye and mouth movements helped recognize feelings and emotions during the interaction (P16, P94, P97, P23, P100). Accordingly, one participant stated, "eye and mouth movements clearly increased my awareness of the feelings of my counterpart" (P16). Even in the non-verbal context of graphical explanation, providing both eye and mouth movements was described as helpful for understanding the other's emotions (P4, P16, P64, P83). For example, P98 mentioned that "just to maintain realism, I would prefer a fully animated face (eye + mouth)".

4.3.2 Mouth. Mouth animations were ranked second for the verbal task and the general experience (see Table 3). Participants described mouth movements as helpful "to see what the other person says" (P48) and to notice whether the other person was happy (P55, P69, 95). This did not just occur for the verbal explanation task. For instance, regarding the graphical explanation, P61 stated: "when painting, the mouth movements show a little more effort of the counterpart".

Being able to detect their opponent even led some participants to mirror them. P95 stated, for instance: "The mouth grinned, that made me grin too" and P69 mentioned that "shared laughter unites". Interestingly, one participant disagreed, stating that "only mouth movements do not help, since no feelings are transported through them" (P82). Similarly, other participants also criticized the depicted mouth animations. Several participants stated that they were not yet satisfied with the quality of mouth animations (P73, P42, P35, P60) and that mouth movements looked "unnatural" (P35, P72) and "uncanny" (P1, P90). One participant said for instance that "mouth animations could have been more expressive" (P82). Others stated that they found mouth movements to be "distracting" (P60, P61).

4.3.3 Eyes. The eye animations were ranked second for the graphical task and third for the verbal explanation and general experience (see Table 3). Several participants noted that eye movements allowed them to notice where their opponent directed their attention (e.g. P76, P89). Eyes aided in determining whether they were currently addressed (e.g. P49, P76). Further, eye motions exposed thought processes, for instance by symbolizing "that one looks upwards while thinking" (P34). Moreover, participants stated that eye movements made the social interaction feel more natural and realistic (e.g. P61, P72). For instance, even when asked regarding the graphical explanation, P92 stated: *"Eyes are important. Otherwise, people stare at each other lifelessly."*. Yet, individuals criticized graphic quality as low (P58), making eye movements "not noticeable" enough (P94). Also, one participant felt the depicted eye movements to be "irritating" (P50). Lastly, for several participants, eye movements did not facilitate the completion of the explanation tasks (P32, P94, P96).

4.3.4 (Social) Interaction. The drawing task shifted the focus onto the tablet. As a result, facial animations were less important for graphical explanations for several participants (e.g., P2, P39, P12, P3, P49, P48). One participant stated: "I only looked at the drawing sheet. I probably would not even have noticed a crying counterpart" (P55). Participants acknowledged several downsides of interacting within the Social VR application. Several comments described facial animations as unrealistic and unsuitable for social interaction (P82, P55, P97). Some considered body tracking error-prone (P49, P42). This resulted for instance, in "hands constantly sliding through one's own body or the table" (P49). Additionally, some participants noted *"latency being present in the acoustic transmission"* (P76). Moreover, several participants described the questionnaire in some parts as cumbersome to operate (e.g. P94, P100) and that displaying it repeatedly throughout the study broke their presence in the VR. Also, the compulsory interaction with a stranger felt uncomfortable for some participants. P46 stated, for instance, that it was *"quite strange to talk to (and see) a person you don't know at all"*.

Nevertheless, most participants were fond of the interaction in the Social VR environment, feeling "much closer [to the counterpart] than during a video call" (P48). In effect, participants felt as if "their counterpart was with [them] in the room" (P72) and "standing at the other side" (P61). This closeness even resulted in communication partners sensing "sympathy" (P46) and sharing emotions: "we laughed together even though we were strangers." (P60). Further, several participants characterized the interaction as realistic and natural (P98, P32, P54, P1), mainly due to visible body movements such as "waving" (P95, P13). Overall, participants described the experiment as an immersive experience, stating that "one could almost forget that one is in the real world" (P82) and that "one thought one was talking to a real person" (P4).

5 DISCUSSION

Our findings show that facial expressions positively impact Social Presence levels. Rendering mouth movements significantly increases co-presence in comparison to neutral facial expressions. When performing verbal explanations, rendering eye movements elicits highest co-presence. Lastly, verbal explanations generally entail higher Social Presence than graphical ones. Moreover, an animated avatar with eyes and mouth rendering increases its realism, reflects the partner's effort of solving a task through facial expressions, and avoids life-less awkward staring. In the following, we discuss two main deductions from our findings: First, facial expressions influence Social Presence in an - at least partially - subconscious manner. Second, this effect depends on the type of explanation. As a result, we derive several design recommendations to increase Social Presence as well as limitations.

5.1 Let's Face It: Facial Expressions Increase Social Presence

Our results suggest that perceived Social Presence is influenced by facial expressions in Social VR settings (RQ1). The scores of the NMSPI demonstrate an increase in co-presence relative to neutral expressions when providing facial expressions, particularly mouth movements. Interestingly, when subdividing co-presence scores in perception of self and perception of other, we see this effect only with regards to perception of other. In other words, facial expressions do not affect a user's own feeling of co-presence. However, users assume that their counterpart perceives higher co-presence in a system with mouth movements. It is also notable that, unlike mouth movements alone, the combination of eyes and mouth movements is not significantly different from neutral facial expressions regarding co-presence. This may be due to an attention split. When both eye and mouth movements are displayed, two nonadjacent visual features compete for the user's limited attention resource. Focusing on both facial movements could limit the effect that mouth movements exert.

When interpreting the remaining scores of the NMSPI, there are no indications that facial expressions support the comprehension of the counterpart's emotions. The qualitative feedback, however, paints a different picture. Several participants stated that combining eye and mouth movements made them more aware of their partner's feelings. For some, this awareness altered their own affective state and perceived realism of the interaction. This observation aligns with insights from communication psychology, which regard facial expressions as significantly important for conveying and interpreting emotions [2]. In our evaluation, more participants associated this higher emotional salience with mouth movements than eye movements. Overall, most preferred the combination of mouth and eye animations. Moreover, we found that integrating facial expressions influences the time spent looking at the face of the interaction partner and the interpretation about their task-solving effort. Displaying conjunction of eye and mouth movements entails the longest gaze duration, followed by the sole rendering of mouth movements, then, the rendering of eye movements. With neutral facial expressions, participants spend the least time looking at their opponent's face. We interpret these reactions towards facial expressions as a sign of directed attention and, in line with related work [3, 30], more intimacy among users.

Our evaluation provides powerful implications on how facial expressions affect perceived Social Presence. From these, we conclude that rendering a combination of eye and mouth movements entails the largest positive impact on Social Presence. Considering the facial expressions independently of the types of explanation, the influence of mouth movements seems to be more pronounced than of eye movements. We argue that these outcomes are rooted in a change in mutual awareness when presented with facial expressions. We base this assumption on two findings: The time spent looking at the counterpart's face increased, which reflects more attention being devoted to the counterpart. Moreover, we detected a change in co-presence - a factor highly dependent on "the degree of mutual salience" [5]. We additionally argue that this impact is, at least partly, of implicit nature: Only less than half of the participants noticed a change in facial animations throughout the experiment. Yet, when facial expressions are present, users tend to direct their gaze longer at their opponent's face but not more frequently. This results in increased gaze duration, co-presence levels, and consequently, higher Social Presence. In combination with the lack of awareness, these findings point to underlying subconscious processes. This finding aligns with the results of other research projects, such as Oh et al. [54]. Both of these presumptions reveal diverse research potentials and should be explored in further research projects. Furthermore, future work should illuminate why the above-depicted discrepancy between the quantitative and qualitative feedback is found.

5.2 Explain Yourself... Verbally

For our evaluation, we employed explaining as a catalyst for social interaction. In effect, both types of explanation were generally attributed with high Social Presence scores, indicating a successful selection. Nevertheless, our results suggest that verbal explanation is better at generating a sense of Social Presence than a graphical one. While perceived affective understanding and interdependence scores were comparable among both types of explanation, co-presence scores are higher for verbal explanation. Looking at the co-presence score subdivided into the perception of self and perception of others, it becomes apparent that this result is mainly attributable to self-perception. In other words, the type of explanation primarily influences how one assesses their own co-presence but not how the co-presence of their counterpart is estimated. This effect can be ascribed to the differing focus on the interaction partner, seen by large deviations in gaze duration and gaze frequency between verbal and graphical explanation types. We acknowledge that within the scope of this experiment, we exclusively analyzed visual focus directed at faces. However, analyzing gazes at other body parts below the face, e.g., hand movements, is a promising future research direction, especially for graphical tasks. The observation further reflects that verbal and graphical explanation types differ in the number of body movements they elicit. We conclude this was due to verbal explanation encouraging more gestural behavior than graphical explanation, which required precise movements. Furthermore, as a graphical explanation requires few spoken words, verbal explanation also induces more mouth movements. Further research is needed to better understand these influences of verbal and graphic tasks on Social Presence and their underlying causes. For instance, accompanying graphical tasks with more present verbal instructions could alter the effect that graphical tasks exert on Social Presence.

Types of explanation determine whether and to what extent facial expressions entail Social Presence. The verbal explanation enhances the effect of facial expressions on Social Presence, while the graphical one reduces it (RQ2). Moreover, the types of explanation even affect which of the facial expressions are more prominent. While in general, mouth movements were more significant for increasing co-presence, in combination with the verbal task type eye movements elicited highest co-presence. We argue that this is due to the fact that in a task that shifts a user's focus directly onto their counterpart, such as in the verbal explanation task, eye movements can be perceived better and thus gain greater importance for non-verbal communication. These presented interaction effects demonstrate the importance of the explanation type for shaping the experimental outcome. In more general terms, when evaluating how to increase Social Presence, other system characteristics (e.g. type of social interaction, technological properties, etc.) need to be considered. This finding is particularly noteworthy, regarding the diverse outcomes of related work in the field of facial animation for Social VR. Some authors produced significant increases in Social Presence levels across different smile conditions [54]. Yet, other approaches that rendered user facial expressions, reported purely qualitative tendencies of increasing Social Presence, lacking statistical significance [25, 55]. Our results propose a probable cause for the diverse outcomes and inconsistencies among many research efforts in the Social Presence domain, that did not ensure such comparable system characteristics. Accordingly, future research would benefit from greater consistency. This argumentation aligns with Oh et al. for whom "it would be beneficial to consider boundary conditions" for future research in the field of Social Presence [53]. Moreover, as our work illustrates how the impact of facial expressions is shaped by the type of collaboration, other forms of collaboration should be included in future work, i.e., usage of interactive tools and purely gesture-based collaboration.

5.3 Designing for Social Connectedness

The motivation of this work was to enhance connectedness among people by increasing Social Presence in Social VR applications. Grounded in our empirical analysis previously discussed, we propose the following design recommendations applicable for (future) Social VR developers to increase Social Presence:

(1) Integrate real-time tracked facial expressions. When struggling with restricted networked capacities, prioritize mouth movements over eye movements. Following this recommendation will change how the user estimates their counterpart's feeling of Social Presence.
(2) Use collaboration tasks that entail an increased amount of visual focus directed at the interaction partner, such as our verbal explanation. Doing so will enhance a user's own perception of Social Presence.
(3) Combine real-time tracked facial expressions with such collaboration tasks (i.e. verbal explanation). It will boost the impact that facial expressions - particularly eye movements - entail on Social Presence.

The following fictional application scenario exemplifies how these design recommendations may be implemented.

> *Example - Developing a Social VR application for joint development sessions for architects and contractors:* Tom is a developer of VR applications. An architecture firm

> hired him to develop a Social VR application for collaborating with contractors. His task is to connect employed architects remotely with hired contractors for joint development sessions. A core functionality of Tom's application should be to sketch and create designs collaboratively using a sketching desk. For the users, completing this design requires little focus on their counterparts. Knowing it would benefit the collaboration, Tom aims to make users feel as socially present as possible. He deduces from our design recommendations that additional collaboration phases should be implemented. Here, the application directs the collaborators' focus away from the sketching desk and toward their counterparts. In practice, Tom integrates a feature that disguises the sketch desk after a fixed time frame and encourages users to discuss their developments verbally. Afterward, the sketching continues. During these phases, Tom concludes that the tool should integrate real-time tracked facial expressions to increase the feeling of Social Presence further. Because the hired contractors often live in rural areas without access to high-end VR systems, Tom assumes limited system or network resources. Thus, if his application detects scarce resources, it is programmed to prefer mouth movements over other facial expressions.

As emphasized by this scenario, our findings are highly transferable and relevant for applications such as (business) communication technologies, (social) games, virtual collaboration settings, or education. Developers will better understand when to render and avoid facial expressions based on the context. This can facilitate a better system performance, similar to not rendering occluded objects. By implementing these recommendations, we deduce that future VR systems will need to incorporate sensor technologies to capture users' faces. Recently introduced VR HMDs, e.g. the HP Reverb G2 Omnicept [11] and the Meta Quest Pro [47], already integrate eye and mouth sensors as embedded in our prototype. Yet, these HMDs require a connection to a high-performance computer or are still costly. This limits their mobility and/or large-scale applicability. In contrast, our system presents a solution to animate VR avatars with facial expressions without needing high-performance computers and high costs. As a drawback, it required a large number of customization efforts. To make such a system more accessible, future affordable, and mobile off-the-shelves VR HMDs should integrate eye and mouth trackers. Once this is the case, Social VR applications integrating facial expressions can bring people "closer" together in work and private life. Moreover, future research should invest in more accurate and subtle representations of facial expressions, such as eyebrow movements, frowning, or wrinkles around one's mouth and nose. Fine-grained expressions have the potential to affect Social Presence even further. How and in which combination they map onto individual aspects of Social Presence needs to be explored in the future.

Furthermore, in terms of increasing perceived Social Presence, future systems could also benefit from modifying avatars' facial expressions based on users' emotions. Future VR system could analyze users' biosignals, such as heart rate variability, electroencephalography, galvanic skin conductance, or saliva, and derive emotions from them in real-time. Thus, using affective computing approaches might be highly applicable to further increase expressiveness of avatars' facial expressions in the future.

6 LIMITATIONS

In most of this paper, we consider Social Presence desirable in Social VR. Yet, Social Presence may also be unfavorable in particular application domains (e.g., when anonymity is of importance) or with particular user groups (e.g., those feeling discomfort during social interactions) [53]. In these cases, the provided design recommendations do not apply. For such circumstances, we consequently propose: Developments should refrain from employing facial expressions. Instead, a task that allocates the user's focus away from the interaction partner, such as our graphical explanation, may be favorable. Choosing verbal and graphical tasks for maintaining comparability provided an additional limitation. Particularly during graphic explanation, participants frequently leaned over the drawing table, and their focus was often not directed towards their opponent, but on the drawing tablet. Allowing a stronger focus on their opponent during this task, e.g. by positioning the drawing tablets in a higher, angled position, or by using a whiteboard for joint drawing, could therefore result in further conclusions regarding the effect of facial expressions in graphic tasks. Employing other explanation activities, such as pantomimic ones, in which non-verbal cues are more explicit, might also reveal additional implications of facial expressions in Social Presence. During our evaluation, we found that the VIVE facial tracker used for tracking and measuring mouth movements is still in its early development and is primarily designed for entertainment purposes. Accordingly, it lacks a uniform calibration procedure. We implemented our own calibration sequence for the evaluation. Notably, this required instructors to calibrate the facial tracking mechanism by eye. As we employed a within-study design, this should not impact the data quality of the presented research. Still, future facial tracking systems would benefit from integrating a more standardized calibration procedure for their sensors. In addition, the range of facial expressions displayed was limited in our setup. Other hardware systems, such as the Meta Quest Pro [47], which can also display changes in eyebrows, facial wrinkles, and further additional facial animations, could have provided additional interesting findings. However, due to the high cost of this hardware, this would also limit the transferability of the results to today's average social VR user. Lastly, our apparatus only provided four pre-defined avatars as embodiment choices for participants. It is ambiguous whether a more individualized user representation would have impacted the study results. Nevertheless, several participants stated in the qualitative feedback that they felt highly embodied in our Social VR application. Thus, we argue that our findings still present valuable insights about the impact of facial expressions on Social Presence.

7 CONCLUSION

In this paper, we investigated four types of facial expressions for two collaborative types of explanation in Social VR. To assess the impact of the proposed facial expressions and types of explanation on Social Presence, we conducted a controlled lab evaluation with 48 participants. Our results revealed that verbal explanation entails higher Social Presence than graphical explanation in collaborative Social VR, independent of the facial expression type. Moreover, we discovered that the type of facial expressions influences perceived Social Presence in Social VR settings, particularly regarding co-presence. We found that rendering mouth movements elicited higher co-presence than neutral facial expressions regardless of explanation type. However, rendering eye movements elicited the highest co-presence when performing verbal explanations. Lastly, we found that users directed their gaze at an interaction partner's face within Social VR the longest when both eye and mouth movements were rendered.

ACKNOWLEDGMENTS

This project is funded under project number 16SV8712 by the German Federal Ministry for Education and Research. We would like to thank Eric Landwehr for his assistance in implementing the prototype. We further want to thank Ronja Krüger for the illustrations, and for the support in planning and conducting the experiment sessions.

REFERENCES

[1] Sahar Aseeri and Victoria Interrante. 2021. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE transactions on visualization and computer graphics* 27, 5 (May 2021), 2608–2617. https://doi.org/10.1109/TVCG.2021.3067783

[2] Avinash Awasthi and Manas K. Mandal. 2015. Facial Expressions of Emotions: Research Perspectives. In *Understanding Facial Expressions in Communication: Cross-cultural and Multidisciplinary Perspectives*. Springer Science + Business Media, New York, NY, US, 1–18. https://doi.org/10.1007/978-81-322-1934-7_1

[3] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2001. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 10, 6 (Dec. 2001), 583–598. https://doi.org/10.1162/105474601753272844

[4] Frank Biocca and Chad Harms. 2003. Guide to the Networked Minds Social Presence Inventory v. 1.2: Measures of Co-Presence, Social Presence, Subjective Symmetry, and Intersubjective Symmetry.

[5] Frank Biocca, Chad Harms, and Judee Burgoon. 2003. Towards A More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence* 12 (Oct. 2003), 456–480. https://doi.org/10.1162/105474603322761270

[6] Y Bondareva, L.M.J. Meesters, and D. Bouwhuis. 2006. Eye Contact as a Determinant of Social Presence in Video Communication. (Jan. 2006).

[7] Caio Brito and Kenny Mitchell. 2019. Repurposing Labeled Photographs for Facial Tracking with Alternative Camera Intrinsics. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 864–865. https://doi.org/10.1109/VR.2019.8798303

[8] Deutschland Bundeszentrale für Politische Bildung, Wissenschaftszentrum Berlin für Sozialforschung Statistisches Bundesamt, and Projektgruppe Das Sozio-ökonomische Panel Deutsches Institut für Wirtschaftsforschung. 2021. *Datenreport 2021 ein Sozialbericht für die Bundesrepublik Deutschland.*

[9] Judee K. Burgoon, Valerie Manusov, and Laura K. Guerrero. 2021. *Nonverbal Communication* (second ed.). Routledge, New York. https://doi.org/10.4324/9781003095552

[10] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in Communication. In *Perspectives on Socially Shared Cognition.*, Lauren B. Resnick, John M. Levine, and Stephanie D. Teasley (Eds.). American Psychological Association, Washington, 127–149. https://doi.org/10.1037/10096-006

[11] HP Development Company. 2022. HP Reverb G2 Omnicept Edition. https://www.hp.com/de-de/vr/reverb-g2-vr-headset-omnicept-edition.html.

[12] Elena Dzardanova, Vlasios Kasapakis, and Damianos Gavalas. 2018. Social Virtual Reality. In *Encyclopedia of Computer Graphics and Games*, Newton Lee (Ed.). Springer International Publishing, Cham, 1–3. https://doi.org/10.1007/978-3-319-08234-9_204-1

[13] Paul Ekman and W. Friesen. 1978. Facial Action Coding System. https://doi.org/10.1037/t27734-000.

[14] Paul Ekman and Erika Rosenberg (Eds.). 2005. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (second ed.). Oxford University Press.

[15] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. arXiv:2102.11824 [cs, stat]

[16] Danilo Fanes. 2022. Pico Neo 3 Pro Eye - VR Headset with Native Eye Tracking. https://vr.tobii.com/integrations/pico-neo-3-pro-eye/.

[17] Martin Feick, Niko Kleer, Anthony Tang, and Antonio Krüger. 2020. The Virtual Reality Questionnaire Toolkit. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*. ACM, Virtual Event USA, 68–69. https://doi.org/10.1145/3379350.3416188

[18] Maxwell Foxman, Alex P. Leith, David Beyea, Brian Klebig, Vivian Hsueh Hua Chen, and Rabindra Ratan. 2020. Virtual Reality Genres: Comparing Preferences in Immersive Experiences and Games. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*. ACM, Virtual Event Canada, 237–241. https://doi.org/10.1145/3383668.3419881

[19] C.d. Frith, D.m. Wolpert, and R. J. R. Blair. 2003. Facial Expressions, Their Communicatory Functions and Neuro–Cognitive Substrates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358, 1431 (March 2003), 561–572. https://doi.org/10.1098/rstb.2002.1220

[20] Elias Giannopoulos, Victor Eslava, María Oyarzabal, Teresa Hierro, Laura González, Manuel Ferre, and Mel Slater. 2008. The Effect of Haptic Feedback on Basic Social Interaction within Shared Virtual Environments. In *Proceedings of the 6th International Conference on Haptics: Perception, Devices and Scenarios (EuroHaptics '08)*. Springer-Verlag, Berlin, Heidelberg, 301–307. https://doi.org/10.1007/978-3-540-69057-3_36

[21] Jihun Hamm, Christian G. Kohler, Ruben C. Gur, and Ragini Verma. 2011. Automated Facial Action Coding System for Dynamic Analysis of Facial Expressions in Neuropsychiatric Disorders. *Journal of Neuroscience Methods* 200, 2 (Sept. 2011), 237–256. https://doi.org/10.1016/j.jneumeth.2011.06.023

[22] Ryan Handley, Bert Guerra, Rukkmini Goli, and Douglas Zytko. 2022. Designing Social VR: A Collection of Design Choices Across Commercial and Research Applications. https://doi.org/10.48550/arXiv.2201.02253 arXiv:2201.02253 [cs]

[23] Chad Harms and Frank Biocca. 2004. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. In *Seventh Annual International Workshop: Presence*, M. Alcaniz and B. Rey (Eds.). Valencia: Universidad Politecnica de Valencia., 7.

[24] Jonathon D. Hart, Thammathip Piumsomboon, Louise Lawrence, Gun A. Lee, Ross T. Smith, and Mark Billinghurst. 2018. Emotion Sharing and Augmentation in Cooperative Virtual Reality Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. ACM, Melbourne VIC Australia, 453–460. https://doi.org/10.1145/3270316.3271543

[25] Jonathon D. Hart, Thammathip Piumsomboon, Gun A. Lee, Ross T. Smith, and Mark Billinghurst. 2021. Manipulating Avatars for Enhanced Communication in Extended Reality. In *2021 IEEE International Conference on Intelligent Reality (ICIR)*. 9–16. https://doi.org/10.1109/ICIR51845.2021.00011

[26] Hasbro. 2022. Taboo Game. https://shop.hasbro.com/en-us/product/taboo-game:304C0329-5056-9047-F5D1-8C8A886E0D35.

[27] Zhenyi He, Ruofei Du, and Ken Perlin. 2020. CollaboVR: A Reconfigurable Framework for Creative Collaboration in Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 542–554. https://doi.org/10.1109/ISMAR50242.2020.00082

[28] Fernanda Herrera, Soo Youn Oh, and Jeremy N. Bailenson. 2020. Effect of Behavioral Realism on Social Interactions Inside Collaborative Virtual Environments. *PRESENCE: Virtual and Augmented Reality* 27, 2 (Feb. 2020), 163–182. https://doi.org/10.1162/pres_a_00324

[29] htc. 2022. VIVE Facial Tracker. https://www.vive.com/de/accessory/facial-tracker/.

[30] Wijnand IJsselsteijn, Joy van Baren, and Froukje van Lanen. 2003. Staying in Touch Social Presence and Connectedness through Synchronous and Asynchronous Communication Media. (2003), 5.

[31] Seung-A Annie Jin. 2010. Parasocial Interaction with an Avatar in Second Life: A Typology of the Self and an Empirical Test of the Mediating Role of Social Presence. *Presence: Teleoperators and Virtual Environments* 19, 4 (Aug. 2010), 331–340. https://doi.org/10.1162/PRES_a_00001

[32] Richard D. Johnson. 2011. Gender Differences in E-Learning: Communication, Social Presence, and Learning Outcomes. *Journal of Organizational and End User Computing (JOEUC)* 23, 1 (2011), 79–94. https://doi.org/10.4018/joeuc.2011010105

[33] Marcel Jonas, Steven Said, Daniel Yu, Chris Aiello, Nick Furlo, and Doug Zytko. 2019. Towards a Taxonomy of Social VR Application Design. 437–444. https://doi.org/10.1145/3341215.3356271

[34] Chiara Jongerius, Roy S. Hessels, Johannes A. Romijn, Ellen M. A. Smets, and Marij A. Hillen. 2020. The Measurement of Eye Contact in Human Interactions: A Scoping Review. *Journal of Nonverbal Behavior* 44, 3 (Sept. 2020), 363–389. https://doi.org/10.1007/s10919-020-00333-3

[35] Oliwier Krawczyk. 2021. Universal HTC Vive Facial Tracker Mount V2.

[36] Karel Kreijns, Kate Xu, and Joshua Weidlich. 2022. Social Presence: Conceptualization and Measurement. *Educational Psychology Review* 34, 1 (March 2022), 139–170. https://doi.org/10.1007/s10648-021-09623-8

[37] Udo Kuckartz. 2010. *Einführung in die computergestützte Analyse qualitativer Daten* (3., aktualisierte aufl ed.). VS, Verl. für Sozialwiss, Wiesbaden.
[38] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. 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, 1–10. https://doi.org/10.1145/3139131.3139156
[39] Chiwon Lee, Hyunjong Joo, and Soojin Jun. 2021. Social VR as the New Normal? Understanding User Interactions for the Business Arena. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3411763.3451740
[40] Na Liu and Ruifeng Yu. 2017. Influence of Social Presence on Eye Movements in Visual Search Tasks. *Ergonomics* 60, 12 (Dec. 2017), 1667–1681. https://doi.org/10.1080/00140139.2017.1342870
[41] Paul Ekman Group LLC. 2022. Facial Action Coding System. https://www.paulekman.com/facial-action-coding-system/.
[42] Jean-Luc Lugrin, David Zilch, Daniel Roth, Gary Bente, and Marc Erich Latoschik. 2016. FaceBo: Real-time Face and Body Tracking for Faithful Avatar Synthesis. In *2016 IEEE Virtual Reality (VR)*. IEEE, Greenville, SC, USA, 225–226. https://doi.org/10.1109/VR.2016.7504735
[43] Divine Maloney, Guo Freeman, and Donghee Wohn. 2020. "Talking without a Voice": Understanding Non-verbal Communication in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 4 (Oct. 2020), 1–25. https://doi.org/10.1145/3415246
[44] Mattel. 2022. Pictionary Quick-Draw Guessing Game For Family, Kids, Teens And Adults, 8 Year Old & Up. https://shop.mattel.com/products/pictionary-dkd47.
[45] Gordon McIntyre, Roland Gocke, Matthew Hyett, Melissa Green, and Michael Breakspear. 2009. An Approach for Automatically Measuring Facial Activity in Depressed Subjects. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE, Amsterdam, 1–8. https://doi.org/10.1109/ACII.2009.5349593
[46] Meta. 2021. Oculus Lipsync Unity Integration.
[47] Meta. 2022. Meta Quest Pro: Unser neues und fortschrittlichstes VR-Headset | Meta Store. https://www.meta.com/de/quest/quest-pro/.
[48] Rachel Metz. 14.06.17. Virtual Reality's Missing Element: Other People. *MIT Technology Review* (14.06.17).
[49] Chris Miller and Mia Poston. 2020. Types of Nonverbal Communication. *Exploring Communication in the Real World* (Aug. 2020).
[50] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine* 19, 2 (June 2012), 98–100. https://doi.org/10.1109/MRA.2012.2192811
[51] Dooley Murphy. 2017. Building a Hybrid Virtual Agent for Testing User Empathy and Arousal in Response to Avatar (Micro-)Expressions. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3139131.3141217
[52] Aki Myllyneva and Jari K. Hietanen. 2015. There Is More to Eye Contact than Meets the Eye. *Cognition* 134 (Jan. 2015), 100–109. https://doi.org/10.1016/j.cognition.2014.09.011
[53] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (Oct. 2018), 114. https://doi.org/10.3389/frobt.2018.00114
[54] Soo Youn Oh, Jeremy Bailenson, Nicole Krämer, and Benjamin Li. 2016. Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PLOS ONE* 11, 9 (Sept. 2016), e0161794. https://doi.org/10.1371/journal.pone.0161794
[55] Catherine Oh Kruzic, David Kruzic, Fernanda Herrera, and Jeremy Bailenson. 2020. Facial Expressions Contribute More than Body Movements to Conversational Outcomes in Avatar-Mediated Virtual Environments. *Scientific Reports* 10, 1 (Dec. 2020), 20626. https://doi.org/10.1038/s41598-020-76672-4
[56] Xueni Pan, Marco Gillies, and Mel Slater. 2008. The Impact of Avatar Blushing on the Duration of Interaction between a Real and Virtual Person. (2008), 7.
[57] Markus Petrykowski, Philipp Berger, Patrick Hennig, and Christoph Meinel. 2019. Digital Collaboration with a Whiteboard in Virtual Reality. In *Proceedings of the Future Technologies Conference (FTC) 2018 (Advances in Intelligent Systems and Computing)*, Kohei Arai, Rahul Bhatia, and Supriya Kapoor (Eds.). Springer International Publishing, Cham, 962–981. https://doi.org/10.1007/978-3-030-02686-8_72
[58] Sandra Poeschl and Nicola Doering. 2015. Measuring Co-Presence and Social Presence in Virtual Environments - Psychometric Construction of a German Scale for a Fear of Public Speaking Scenario. *Studies in Health Technology and Informatics* 219 (2015), 58–63.
[59] Anna Samira Praetorius, Lara Krautmacher, Gabriela Tullius, and Cristóbal Curio. 2021. User-Avatar Relationships in Various Contexts: Does Context Influence a Users’ Perception and Choice of an Avatar?. In *Mensch Und Computer 2021 (MuC '21)*. Association for Computing Machinery, New York, NY, USA, 275–280. https://doi.org/10.1145/3473856.3474007
[60] Daniel Roth, Constantin Klelnbeck, Tobias Feigl, Christopher Mutschler, and Marc Erich Latoschik. 2018. Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Tuebingen/Reutlingen, Germany, 215–222. https://doi.org/10.1109/VR.2018.8447550
[61] Daniel Roth, Kristoffer Waldow, Marc Erich Latoschik, Arnulph Fuhrmann, and Gary Bente. 2017. Socially Immersive Avatar-Based Communication. In *2017 IEEE Virtual Reality (VR)*. IEEE, Los Angeles, CA, USA, 259–260. https://doi.org/10.1109/VR.2017.7892275
[62] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. *Using Presence Questionnaires in Virtual Reality*. https://doi.org/10.1145/3290605.3300590
[63] Harrison Jesse Smith and Michael Neff. 2018. Communication Behavior in Embodied Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173863
[64] Radosław Sterna and Katja Zibrek. 2021. Psychology in Virtual Reality: Toward a Validated Measure of Social Presence. *Frontiers in Psychology* 12 (Oct. 2021), 705448. https://doi.org/10.3389/fpsyg.2021.705448
[65] Tiffany Y. Tang and Yongfu Wang. 2015. Alone Together: Multiplayer Online Ball Passing Using Kinect - An Experimental Study. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW'15 Companion)*. Association for Computing Machinery, New York, NY, USA, 187–190. https://doi.org/10.1145/2685553.2699003
[66] Bavo Van Kerrebroeck, Giusy Caruso, and Pieter-Jan Maes. 2021. A Methodological Framework for Assessing Social Presence in Music Interactions in Virtual Reality. *Frontiers in Psychology* 12 (2021).
[67] VERBI. 2021. MAXQDA. VERBI.
[68] Astrid M. von der Pütten, Nicole C. Krämer, Jonathan Gratch, and Sin-Hwa Kang. 2010. "It Doesn't Matter What You Are!" Explaining Social Effects of Agents and Avatars. *Computers in Human Behavior* 26, 6 (Nov. 2010), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012
[69] Paul Watzlawick and Janet Beavin. 1967. Some Formal Aspects of Communication. *American Behavioral Scientist* 10, 8 (April 1967), 4–8. https://doi.org/10.1177/0002764201000802
[70] C. Wienrich, K. Schindler, N. Döllinqer, S. Kock, and O. Traupe. 2018. 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)*. 207–214. https://doi.org/10.1109/VR.2018.8446575
[71] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Vancouver BC Canada, 143–146. https://doi.org/10.1145/1978942.1978963
[72] Wolf3D. 2021. Ready Player Me.
[73] Wolf3D. 2022. Partner Apps & Games with 3D Avatars | Ready Player Me. https://readyplayer.me/apps.
[74] Amal Yassien, Passant ElAgroudy, Elhassan Makled, and Slim Abdennadher. 2020. A Design Space for Social Presence in VR. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. ACM, Tallinn Estonia, 1–12. https://doi.org/10.1145/3419249.3420112
[75] Jihun Yu and Jungwoon Park. 2016. Real-Time Facial Tracking in Virtual Reality. In *SIGGRAPH ASIA 2016 VR Showcase (SA '16)*. Association for Computing Machinery, New York, NY, USA, 1. https://doi.org/10.1145/2996376.2996390