Computers in Human Behavior 167 (2025) 108598

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Avatar-mediated communication in collaborative virtual environments: A study on users' attention allocation and perception of social interactions ☆

Chen Li [a], Yixin Dai [b], Guang Chen [b],*, Jing Liu [c], Ping Li [d], Horace Ho-shing Ip [e]

[a] Department of Applied Social Sciences & Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[b] Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[c] Department of Accountancy, Economics and Finance, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong
[d] Department of Computing & School of Design, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[e] Department of Computer Science, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

Keywords:
Avatar-mediated communication
Collaborative virtual environment
Avatar anthropomorphism
Nonverbal social cues
Socio-emotional interaction
Task interaction

ABSTRACT

Collaborative virtual environments (CVEs) facilitate avatar-mediated communication (AMC), where users interact using human-like virtual characters in shared virtual worlds, enhancing the attractiveness, attentiveness, and connectedness of remote social experiences and thus becoming extremely popular nowadays in various application domains such as education and healthcare. Understanding how different aspects of avatar behaviours influence various types of social interactions is crucial for improving the design of CVEs. Grounded in a theoretical framework based on avatar anthropomorphic realism, nonverbal social cues, eye-mind hypothesis, and interaction process analysis, this study investigates the impact of avatars' gaze behaviours on users' attention allocation and perceptions during AMC in CVEs. A two-arm randomised controlled trial (RCT) with 60 participants (29 males and 31 females) compared static gaze and natural gaze avatars during socio-emotional and task interactions. Three-dimensional eye-tracking data revealed distinct attention patterns across three primary nonverbal social cues: eye gaze, head orientation, and pointing gesture. Furthermore, avatars' gaze type and interaction type were both found to significantly affect participants' attention allocation; natural gaze behaviour and task interactions mitigated the general gaze-avoidance pattern observed in previous studies. However, avatars' gaze type did not impact participants' perceptions of social presence and anxiety. This research provides a nuanced understanding of attention allocation across nonverbal social cues during AMC and underscores the importance of avatars' gaze and interaction types, highlighting important implications for the future design of CVE to enhance attention coordination and communication. Additionally, it calls for more comprehensive studies to explore avatars' anthropomorphic realism and its effects on user perceptions and overall experience during AMC.

1. Introduction

Collaborative virtual environments (CVEs) are online virtual worlds where multiple users share the same digital space to engage in interactions and communications (Benford, Greenhalgh, Rodden, & Pycock, 2001). They have recently gained significant interest and have been applied across various professional fields, including education, healthcare, and industry, due to their ability to overcome geographical barriers and enhance remote collaboration experience (Derouech, Hrimech, Lachgar, & Hanine, 2024; Li & Liu, 2022; Qiao, Xu, Li, & Ouyang, 2021; Ververidis, Nikolopoulos, & Kompatsiaris, 2022). In addition to their professional applications, CVEs are rapidly expanding in recreational contexts, such as gaming and social networking, where they offer engaging experiences and foster global communities by enabling users to interact and collaborate in online settings (Brown & Bell, 2004; Churchill, Snowdon, & Munro, 2012; Freeman, Acena, McNeese, & Schulenberg, 2022). Avatars, the digital characters that are controlled by human users to represent themselves in online virtual worlds, are a key component of CVEs (Schäfer, Reis, & Stricker, 2022). These avatars enhance the experience of interactions and communications in online and remote settings to a level that is very close to that in the real world (Benford et al., 2001; Nowak & Fox, 2018; Schäfer et al., 2022).

☆ The authors want to thank Mr Honglin Li, Ms Wujie Gao, Ms Qi Lyu, and Ms Qinyang Wu for their comments and suggestions on the virtual environment design and the study design.

* Correspondence to: HJ427,The Hong Kong Polytechnic University, 11, Yuk Choi Road, Hung Hom, Kowloon, Hong Kong.
E-mail addresses: richard-chen.li@polyu.edu.hk (C. Li), yixin.dai@connect.polyu.hk (Y. Dai), guang.chen@connect.polyu.hk (G. Chen), jingliu@hkbu.edu.hk (J. Liu), p.li@polyu.edu.hk (P. Li), horace.ip@cityu.edu.hk (H.H.-s. Ip).

https://doi.org/10.1016/j.chb.2025.108598
Received 16 August 2024; Received in revised form 25 November 2024; Accepted 6 February 2025
Available online 14 February 2025

This specific type of computer-mediated communication (CMC) centred around avatars is known as avatar-mediated communication (AMC) (Nowak & Fox, 2018). The best way to access AMC is by using virtual reality (VR) head-mounted displays (HMDs). Modern VR HMDs consist of not only stereoscopic displays but also multiple sensors to track the displays' position and orientation, as well as user behaviours in the physical world, such as voice, hand movements and gestures, facial expressions, and eye movements and openness. The captured user behaviours can be mirrored to avatars' behaviours and transmitted through CVEs to support higher levels of anthropomorphic realism of avatars, thus providing a more attractive, attentive, and connected social experience by conveying both verbal and nonverbal social cues (Garau, 2003; Herrera, Oh, & Bailenson, 2020; Nowak & Biocca, 2003; Nowak & Fox, 2018). The ability to better convey nonverbal social cues is what AMC in CVEs has been long appraised for when compared with other types of CMC (e.g., video conferencing) (Nowak & Fox, 2018). However, in reality, CVEs may have various abilities to convey nonverbal social cues. For example, although eye gaze is considered extremely important during social interactions (Cañigueral & Hamilton, 2019; Langton, Watt, & Bruce, 2000), not many commercially available CVEs support the natural gaze of avatars (i.e., directly mirroring human users' eye movements and openness to their avatars' in virtual worlds). It is largely unknown whether the lack of certain nonverbal social cues, such as eye gaze, during AMC would affect the user experience of AMC.

In this study, drawing mainly on theories of avatar anthropomorphic realism, nonverbal social cues, eye-mind hypothesis, and interaction process analysis (as shown in Fig. 1), we conducted a two-arm randomised controlled trial (RCT) to investigate how avatars' gaze behaviour affects human users' attention allocation and perceptions of social interactions during AMC in CVEs. Sixty participants were randomly assigned to either the static gaze group or the natural gaze group. They completed a social interaction task using avatars in a customised CVE. The task involved two stages: socio-emotional interactions and task interactions. We recorded participants' eye movements during the task to analyse their attention allocated to three primary nonverbal social cues (i.e., head orientation, eye gaze, and pointing gesture) (Langton et al., 2000). Additionally, we used well-validated questionnaires to study participants' perceptions of social interactions in the CVE.

The results suggested that (1) during both socio-emotional interactions and task interactions, participants' attention allocation was significantly different across the three nonverbal social cues, following a distinct pattern: the greatest amount of attention was directed towards the head orientation of the interactant's avatar, succeeded by the pointing gesture and then eye gaze, (2) avatars' gaze type (i.e., static gaze and natural gaze) and interaction type (i.e., socio-emotional interaction and task interaction) affected participants' attention during AMC: avatars' natural eye gaze and task interactions mitigated the general gaze-avoidance pattern during AMC found by prior studies; avatars with a higher level of anthropomorphic realism (i.e., with natural gaze) attracted greater attention on their eyes, while task interactions were associated with greater attention on less prominent cues (i.e., pointing gesture and eye gaze) in compared to socio-emotional interactions, and (3) avatars' gaze type showed no impact on participants' perceptions during AMC with respect to social presence and anxiety.

This study has two significant contributions. Firstly, the study offers a more nuanced understanding of how users allocate their attention across the three primary nonverbal social cues during AMC. Secondly, it highlights the crucial role of avatars' gaze type and interaction type in affecting participants' attention allocation during AMC. These findings enhance our understanding of user interaction dynamics in CVEs and offer valuable recommendations for designing CVEs, with the goal of improving the effectiveness and overall user experience in remote collaboration. It also calls for more comprehensive study designs to capture avatars' anthropomorphic realism and study its effect on users' perceptions and overall user experience during AMC.

2. Related work and research questions

2.1. Nonverbal social cues and attention allocation

Nonverbal behaviours are crucial in real-world social interactions. A prior study demonstrated that more than half of the information transmitted during face-to-face communication was through nonverbal behaviours (Argyle, 1988). That is, people's attention is often directed by interactants' nonverbal behaviours, with the eye gaze, head orientation, and pointing gesture as three primary nonverbal social cues that direct attention during social interactions and, therefore, also known as attentional cues (Langton et al., 2000).

According to the eye-mind hypothesis, which suggests that people's "visual attention is the proxy for mental attention", the importance of a nonverbal social cue can be inferred from the amount of attention people allocate to it (Webb & Renshaw, 2008). Among the three primary nonverbal social cues, eye gaze is considered extremely important; it can serve as a conduit for the expression of a wide range of meanings, including attractiveness (Argyle & Dean, 1965), intention (Baron-Cohen, Wheelwright, & Jolliffe, 1997), desire to communicate (Ho, Foulsham, & Kingstone, 2015), seeking approval (Efran, 1968), and notably, conveying directional messages (Kuhn, Tatler, & Cole, 2009). Through eye gaze, people can capture the focal points and follow the attention of their interactants; meanwhile, they can also use eye gaze to direct and guide their interactants' attention (Cañigueral & Hamilton, 2019). Therefore, eye gaze is crucial to the direction of attention during social interactions (Emery, 2000; Langton et al., 2000). Prior studies also noted the different importance of nonverbal social cues in directing people's attention, indicating a hierarchical order among them. For example, Perrett and Emery (1994) reported that the direction of attention was first signalled by the eyes, while the information from the head orientation was used if eye gaze was invisible or too far to be noticed. In turn, when the information from the eyes and head was unavailable, body orientation should signal the direction of attention (Perrett & Emery, 1994). Langton (2000) and Langton and Bruce (2000) further modified this model, suggesting that instead of completely overriding information from other nonverbal social cues, the information sent by eye gaze weakened that sent by head orientation when conflicts occurred between the two cues.

Despite this, how people allocate attention to different nonverbal social cues in complex scenarios remains controversial. For example, Birmingham, Bischof, and Kingstone (2008) suggested that people preferentially allocated their attention to others' faces, while Laidlaw, Foulsham, Kuhn, and Kingstone (2011) and Rubo and Gamer (2018) suggested that people tended to avoid rather than seek to look at others' faces. Additionally, few studies were conducted to investigate this difference in the attention allocated to different cues during AMC in CVEs. Thus, whether the importance of these cues differs in this process remains underexplored. While some nonverbal social cues, such as facial expressions and body postures, can attract attention and convey specific information, the primary focus of attention and information is predominantly guided by the three nonverbal social cues (i.e., eye gaze, head orientation, and pointing gesture) (Langton et al., 2000). Furthermore, at the application level, there are limited commercial and affordable VR devices that offer real-time and precise tracking of facial and body movements, and very few commercial and professional CVEs support these intricate forms of expression (Schäfer et al., 2022). Therefore, we want to explore how users' attention allocation varies across the three cues during AMCs in CVEs, with the research question as below (**RQ1**).

RQ1. Are there any differences in participants' attention allocation across the three primary nonverbal social cues (i.e., eye gaze, head orientation, and pointing gesture) during AMC?

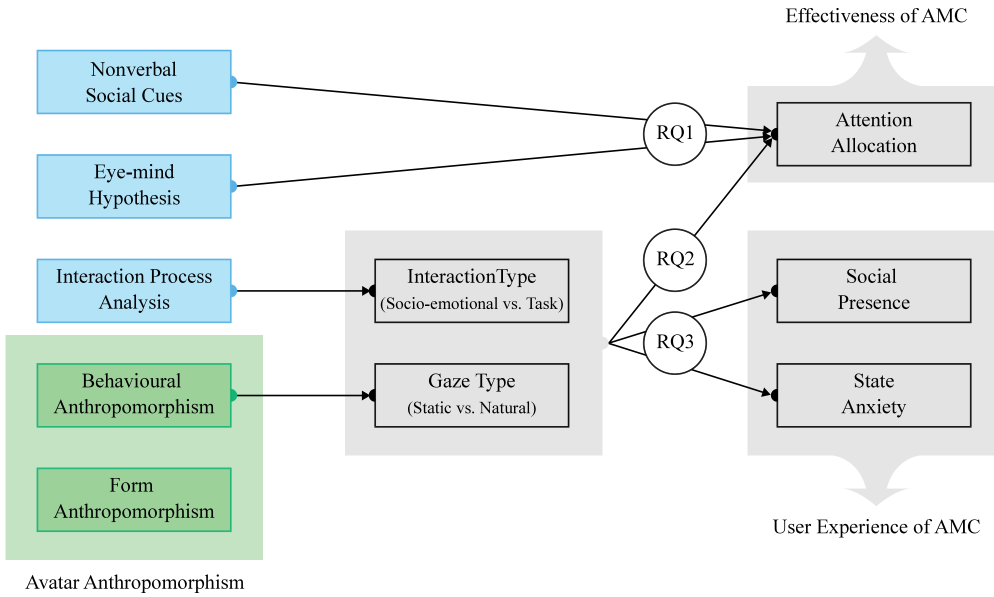


Fig. 1. The theoretical framework underpinning this study. It integrates theories related to face-to-face social interactions, including nonverbal social cues (Langton et al., 2000), eye-mind hypothesis (Webb & Renshaw, 2008), and interaction process analysis (Bales, 1950), which are highlighted in blue, and theories related to AMC under the umbrella of avatar anthropomorphism (Nowak & Fox, 2018), marked in green. A comprehensive discussion of these theories and their roles in supporting the three research questions are provided in Section 2.

2.2. Avatars' gaze type, social interaction type, and attention allocation

In the context of AMC in CVEs, avatars' eye gaze is the most interesting nonverbal social cue to study among the three primary cues. Despite its importance in social interactions and communication, avatars' eye gaze is often rendered more "unnaturally" than head orientation and pointing gestures. To naturally render avatars' gaze shifts, enabling devices (e.g., VR HMDs) need to capture human users' eye behaviours in real-time and convey the natural gaze shifts of avatars to interactants in CVEs, which is referred to as *natural gaze* in this paper. Enabling natural gaze of avatars requires relatively costly hardware and can be computationally expensive to implement (Plopski et al., 2022). Consequently, most commercially available CVEs choose to simulate avatars' eye behaviours unnaturally using simple algorithms. The roughest but most common simulation keeps avatars' eyes centred in the eye sockets and static relative to the head, regardless of how human users' eyes move, eliminating the natural function of eye gaze as an information signal, thus potentially affecting users' attention allocation (Cañigueral & Hamilton, 2019). This simulation is referred to as *static gaze* in this paper.

Besides avatars' gaze type, another factor that may contribute to users' attention allocation is the type of social interactions that they were engaged with during AMC. According to Bales' interaction process analysis (Bales, 1950), interactions can be categorised into two types according to their intended purpose: socio-emotional interactions and task interactions. The former refers to interactions that express social information and emotions, while the latter refers to inquiries about the current task or procedure (Peña & Hancock, 2006).

While earlier studies on social attention relied on simple scene-viewing tasks (Birmingham et al., 2008), recent studies focused on the effect of social interactions brought by a live person on participants' attention allocation (Jing, May, Matthews, Lee, & Billinghurst, 2022; Laidlaw et al., 2011; Rubo & Gamer, 2018). For example, Birmingham et al. (2008) reported that when viewing a static picture containing gaze and other visual information, participants preferentially allocated their attention to people's faces in the picture; specifically, their eye gaze fixated most frequently on the eyes, followed by heads, other parts of the human body, and then other objects in the picture. However, Laidlaw et al. (2011) reported that in complex situations that afforded social interactions, people tend to avoid rather than seek to look at others' faces, which is different from completing simple scene-viewing tasks where eye gaze in pictures or videos, as pre-recorded stimuli, is passive and non-interactive. The gaze-avoidance phenomenon has also been observed in virtual worlds where users also looked more frequently at objects than at the virtual characters facing them (Rubo & Gamer, 2018).

While some recent studies have examined the impact of various tasks (e.g., freeview, search, memory, and navigation) on participants' attention in virtual worlds (Hadnett-Hunter, Nicolaou, O'Neill, & Proulx, 2019; Mills, Hollingworth, Van der Stigchel, Hoffman, & Dodd, 2011), there has not been any study to our knowledge that explores the effects of interaction type (i.e., socio-emotional interaction and task interaction) combined with avatars' gaze type (i.e., static gaze and natural gaze) on participants' attention allocation in CVEs. Therefore, we aim to address the following research question (**RQ2**).

RQ2. How would avatars' gaze type (i.e., static gaze and natural gaze) and interaction type (i.e., socio-emotional interaction and task interaction) affect participants' attention allocation during AMC?

2.3. Avatars' anthropomorphism, social presence, and anxiety

Whether the avatars' eye gaze is naturally controlled by and therefore conveys users' real eye gaze can greatly affect avatar anthropomorphism, which is a critical factor affecting users' perceptions of the verbal and nonverbal interpersonal interactions during AMC (Burden & Savin-Baden, 2019; Nowak & Fox, 2018). For example, Nowak and Biocca (2003) found that users perceived a higher level of social presence when interacting with avatars with high anthropomorphism. There are two types of anthropomorphism: form anthropomorphism and behavioural anthropomorphism (Nowak & Fox, 2018). Form anthropomorphism describes how human-like an avatar looks; more human-like avatars are often considered as more attractive and credible (Gong, 2008; Nowak & Rauh, 2005). Behavioural anthropomorphism describes how human-like avatars behave (Nowak & Fox, 2018).

Prior studies have investigated the impact of avatars' gaze type, as a factor of behavioural anthropomorphism, on users' perceptions of social interactions during AMC. For example, Jing et al. (2022) reported that visualising eye gaze of the remote collaborator amplified meaningful joint attention and improved the participants' sense of co-presence. Steptoe et al. (2009) also found that in comparison to avatars with static gaze or gaze simulated via a simple head-orientation-based algorithm, avatars with natural gaze were associated with the highest engagement level and collaboration performance in AMC. In this study, we want to explore the impact of avatars' gaze type on users' perceptions of social interactions during AMC, with a focus on social presence and anxiety.

Social presence refers to human users' subjective experience of being present with a "real" person and "having access to his or her thoughts and emotions" (Oh, Bailenson, & Welch, 2018), which is a crucial component that enables effective social interactions and leads to natural behaviours during AMC (Herrera et al., 2020). Unlike the technology-driven definitions of social presence, the definition of social presence adopted in this study emphasises intimacy and immediacy during CMC (Short, Williams, & Christie, 1976; Walther, 1992). Specifically, intimacy refers to the sense of connectedness perceived by communicators, whereas immediacy refers to the psychological distance between them (Oh et al., 2018). Based on this definition, several studies have investigated avatar-related approaches to improve social presence during AMC (Kyrlitsias & Michael-Grigoriou, 2022); some approaches focused on improving avatars' form anthropomorphism (Oh et al., 2018; Zibrek & McDonnell, 2019), while others studied the effects of behavioural anthropomorphism on social presence, with avatars' gaze type being a critical yet underexplored variable. For example, Bente, Eschenburg, and Aelker (2007) investigated the effect of the duration of eye gaze on participants' sense of social presence via AMC using computer monitors, suggesting that the longer gaze duration caused higher levels of social presence. However, those studies used algorithm-simulated gaze rather than real human users' natural gaze, and the computer monitors are less immersive than VR HMDs. Therefore, whether interacting with avatars with natural gaze affects users' sense of social presence during AMC remains unclear.

Anxiety is a normal emotional response characterised by feelings of worry, fear, or apprehension (Spielberger, Gonzalez-Reigosa, Martinez-Urrutia, Natalicio, & Natalicio, 1971). It is important to distinguish anxiety from anxiety disorders, which refer to diagnosed conditions where anxiety becomes chronic and disproportionate to the actual threat posed by a situation (Craske et al., 2011). Our study focuses specifically on participants' feelings of anxiety during AMC as a state rather than the anxiety disorder. Previous research has shown that individuals with high anxiety are more likely to avoid social interactions and tend to exhibit shorter durations and frequencies of eye contact during conversations (Howell, Zibulsky, Srivastav, & Weeks, 2016; Turner, 1988). However, these studies were not carried out in CVEs involving human-controlled avatars. Thus, besides the sense of social presence, we are interested in exploring whether avatars' gaze type affects users' state anxiety during AMC.

RQ3. Does the avatar's gaze type (i.e., static gaze and natural gaze) affect participants' sense of social presence and state anxiety during AMC?

3. Method

3.1. Study design and participants

The study was designed as a two-arm RCT with participants either in the natural gaze group or the static gaze group. Seventy-six individuals registered for participation through posted flyers on the campus of The Hong Kong Polytechnic University. Thirteen participants withdrew from the study due to time constraints, and three participants were excluded from the data analyses due to technical errors during data collection. Hence, a total of 60 participants (29 males and 31 females) were included in our analyses. The participants were randomly assigned to either the natural or static gaze group via a random number generator,[1] with 30 participants in each group. All participants had normal or corrected-to-normal vision and were native Chinese speakers. The language restriction was implemented to facilitate effective interactions with the experimenters and eliminate language as a potential confounder of the study. The participants' demographics are shown in Table 1 and the flow chart of the study is shown in Fig. 2.

Table 1
Demographics of sample participants.

Group	Male n	Female n	Age (years) Mean (SD)
Static gaze (n = 30)	15	15	23.93 (3.17)
Natural gaze (n = 30)	14	16	26.33 (7.47)
Combined (n = 60)	29	31	25.13 (5.82)

Two experimenters, referred to as Experimenter A and Experimenter B for the rest of the paper, were involved in conducting this user study. Experimenter A, the third author, interacted and completed a social interaction task with participants in CVE using avatars. Experimenter B, the second author, provided instructions and assistance to participants throughout the study in the physical space. To maintain consistency across sessions, Experimenter A was required to strictly follow the study protocol when interacting with participants. Before the study began, the first author provided a comprehensive training session to Experimenter A. Following this training, Experimenter A completed three mock sessions in the CVE under the supervision of the first author. These mock sessions were video recorded and subsequently reviewed by the two experimenters and the first author together to ensure adherence to the protocol and to refine any necessary aspects of the interaction process.

In the natural gaze group, real-time eye movements of the participants and Experimenter A, captured by the eye-tracker in the HMDs, were used to drive the eye gaze and eye openness of their avatars. In the static gaze group, participants and Experimenter A were embodied in avatars, of which the virtual eyes remained centred in the eye sockets and static relative to the head, but the avatars' eye openness was controlled by a simple blinking computer programme (Bentivoglio et al., 1997), mimicking the most common implementation of avatars' eye behaviour simulation in commercially available CVEs.

Considering the two types of interactions (i.e., socio-emotional interactions and task interactions) (Peña & Hancock, 2006), we referred to Herrera et al. (2020) and designed the social interaction task with two stages, namely *Stage1* and *Stage2* for the rest of the paper. Specifically, *Stage1* involved casual interactions between Experimenter A and the participants in fostering a friendly atmosphere. During this stage, Experimenter A implemented a carefully curated set of ice-breaker questions designed to elicit participants' personal experiences, including their academic backgrounds and preferred locations in participants' hometowns. These questions were strategically selected to be non-intrusive and broadly applicable, fostering a positive and unbiased initial connection. *Stage1* lasted for around eight minutes. *Stage2* involved a discussion regarding two different designs of student dormitories (see Fig. 4). During this stage, Experimenter A briefly introduced the two dormitory layouts and allowed participants time to explore them. Then, participants were prompted to select their preferred design and articulate the reasons for their choice. It is noteworthy that neither dormitory design possessed inherent advantages or disadvantages. Experimenter A adhered to a consistent communication strategy and flow: after participants selected their preferred design, they were asked to explain their reasoning, their opinions were acknowledged, and they were encouraged to discuss any dislikes regarding the alternative

[1] https://www.randomizer.org/.

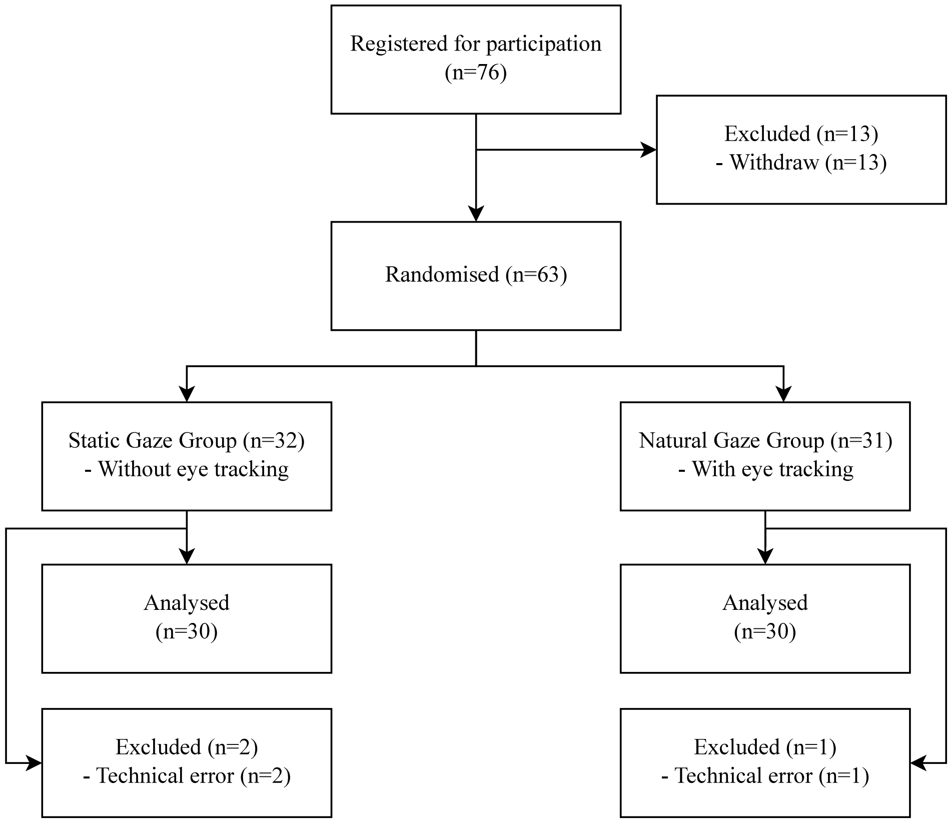


Fig. 2. The flow chart of the study.

design. Lastly, Experimenter A presented arguments in favour of the participants' less preferred design to stimulate further discussion on potential improvements, thereby concluding *Stage2*. *Stage2* lasted around ten minutes.

Additionally, instead of organising participants into dyads to interact with one another in CVE, they were instructed to engage solely with the same experimenter (i.e., Experimenter A) whose avatar's gaze type was set as the same as the participants in each group. This method ensured strict adherence to the study protocol, maximised consistency in avatar appearance and social behaviours, and minimised potential inter-session variations that might arise from employing multiple experimenters. Specifically, Experimenter A underwent three supervised mock sessions to familiarise himself with the study protocol and was directed to meticulously follow the two-stage design during the social interaction task with participants. His avatar was crafted to closely resemble his real-life appearance, providing a stable and neutral visual reference that remained constant across all experimental sessions. Furthermore, Experimenter A was unknown to all participants, and the interaction was intentionally brief to mitigate any potential effects of familiarity.

3.2. Apparatus and collaborative virtual environment

The participants and Experimenter A were provided with two sets of identical hardware to complete the study. Each set included one computer and one set of VR HMDs with the paired controllers and tracking devices. The computer was equipped with an Intel Core i7-12700H processor, 16 GB RAM, and an NVIDIA GeForce RTX 3060 graphics card. The VR HMDs used in this study were the HTC VIVE Pro Eye, which supported real-time eye tracking at up to 125 Hz. The HMDs' built-in microphone and speakers were used for the real-time bidirectional voice chat in the CVE. The paired controllers were the Valve Index controllers capable of tracking each finger's movements. The two SteamVR Base Station 2.0 tracking devices were diagonally positioned from each other in the physical space for interactions to allow for blind-spot-free tracking of the HMDs and controllers.

The participants and Experimenter A were physically separated into two rooms during the experiment. Each room had a set of prepared hardware mentioned above, with the CVE as the only medium to support the communication and interactions between participants and Experimenter A. Each of the two rooms where Experimenter A and the participants used the apparatus measured approximately 2.5 m by 2.5 m. However, to ensure a sufficient safety margin, the boundary for the safe use of the apparatus was set to 2 m by 2 m in the VR HMDs with visual hints. As the social interaction task required minimal movement, this arrangement did not restrict their movements or impact the quality of their interactions. The overall setup is illustrated in Fig. 3.

The CVE used in the study was designed and developed to support AMC, including real-time bidirectional voice chat, using the Unity[2] game engine. Participants were able to customise their avatars using a browser-based interface[3] while Experimenter A used the same pre-customised avatar throughout the study. To simplify network programming and facilitate reliable and low-latency voice chat, we utilised the Photon Unity Networking (PUN) framework.[4] The Vive SRanipal SDK[5] was used to collect the raw gaze data from the VR HMDs' built-in eye-tracker, which provides gaze directions (i.e., eye rotation), eye positions, and eye openness with timestamps in milliseconds. For the natural gaze group, the captured gaze data were transported directly via the PUN framework to drive the avatars' eye movements and blinks

2 https://unity.com/.
3 https://vr.readyplayer.me/avatar.
4 https://photonengine.com/pun.
5 https://developer.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/.

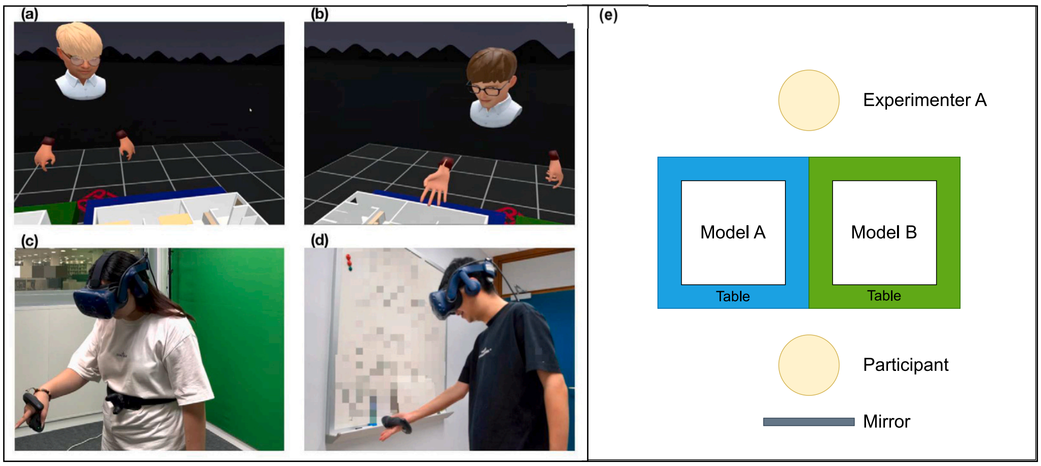


Fig. 3. The apparatus and the layout of the virtual scene. (a) The view of Experimenter A through VR HMDs; (b) The view of a participant through VR HMDs; (c) A participant in the physical space; (d) Experimenter A in a separated physical space; (e) The layout of the virtual scene.

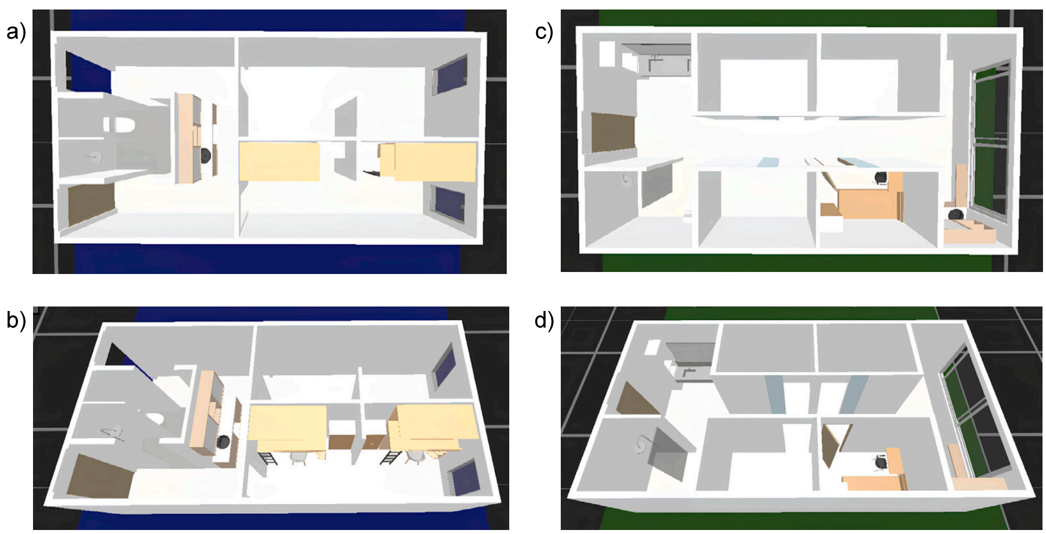


Fig. 4. Dormitory layouts and corresponding three-dimensional models utilised in the social interaction task within the CVE. (a) Top view of the first dormitory layout, characterised by an open design with excellent visibility and flexible space for various activities, though it offers less privacy and smaller windows that may affect ventilation and lighting. (b) Side view of the first dormitory layout. (c) Top view of the second dormitory layout, featuring high privacy with individual rooms and ample lighting in study areas, yet potentially feeling more enclosed due to limited socialising space. (d) Side view of the second dormitory layout.

in the CVE. SteamVR Unity Plugin[6] was used to capture the head, hand, and finger movements of our participants and Experimenter A and drive their avatars. OBS Studio[7] was utilised on both computers to record videos of the renderings viewed by participants and Experimenter A through the VR HMDs. The captured video footage provided a means to systematically review adherence to the established interaction protocols (see Section 3.1).

To facilitate the social interaction task, two virtual tables were placed in the middle of the CVE. On top of these tables were two different dormitory layout models, which were the focus of interactions during *Stage2*. The tables and models were not interactable and were set to stationary, which helped minimise potential distractions and encouraged a more focused AMC. Participants and Experimenter A were spawned on opposite sides of the tables and remained standing throughout the interaction. The layout of the virtual scene is shown in Fig. 3(e) and the two dormitory layouts are shown in Fig. 4.

[6] https://valvesoftware.github.io/steamvr_unity_plugin/.
[7] https://obsproject.com/.

3.3. Procedure

Successfully registered participants were randomly assigned to either the static gaze or natural gaze group. Reminder emails were sent to eligible participants one day before their participation in the study. The emails included a link to the questionnaire for collecting their demographic information.

On the experiment day, Experimenter B checked the participants' identity upon arrival and briefed them about the study procedure. An information sheet with detailed procedures, potential risks, data accessibility, and contact information of the project team was provided to the participants. Participants were then instructed to read and sign the consent form carefully. Afterwards, participants were directed to complete the pre-assessment on a computer, which included the State-Trait Anxiety Inventory-Trait (STAI-T), State-Trait Anxiety Inventory-State (STAI-S), and Simulator Sickness Questionnaire (SSQ) (see Section 3.4).

After completing the pre-assessment, participants were invited to customise their avatars on the computer using the browser-based interface. They were instructed to avoid adding face covers to their avatars, such as sunglasses and masks, because of the nature of this study on

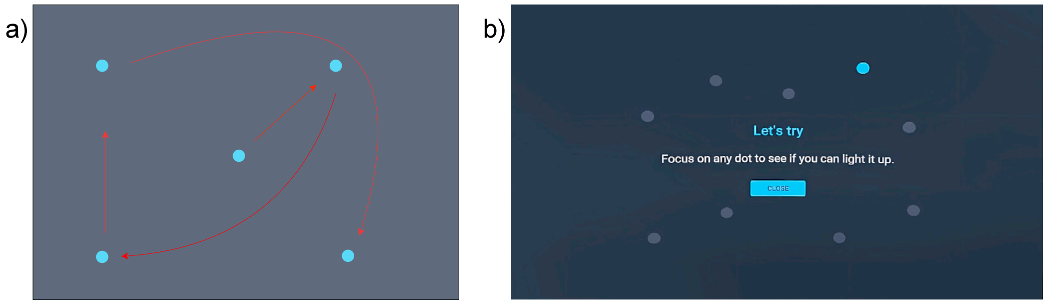


Fig. 5. The eye tracker's built-in calibration and testing tools. (a) The five-dot pattern used for calibration; (b) Testing targets that would illuminate when being looked at.

nonverbal social cues, including eye gaze. After that, the customised avatars were imported into the CVE, and participants would wear and adjust the VR HMDs to ensure a clear and comfortable view.

Next, participants were guided through the eye tracker calibration and testing processes. First, they adjusted the fit and interpupillary distance of the VR HMDs by following the in-VR instructions and the guidance provided by Experimenter B. After making these adjustments, participants were instructed not to change the fit of the HMDs any further. The eye tracker was then calibrated using a five-dot pattern. Following the calibration process, the accuracy of the eye tracker was tested with its built-in tool. Participants were asked to look at each testing target to ensure it could be illuminated correctly. Fig. 5 shows the eye tracker's built-in calibration and testing tools.

Once the eye tracker calibration and testing were completed, Experimenter B started the CVE and enabled the virtual mirror. Participants were asked to follow Experimenter B's instructions to perform a series of actions — a commonly used method to enhance their sense of embodiment (Oliva, Beacco, Navarro, & Slater, 2022). These actions began with participants looking at themselves in the virtual mirror with their heads slowly turning from left to right and up to down. Then, participants focused on their hands and gripped the controllers properly. Next, participants moved each finger away from the controller and then returned it to its original position on the controller, starting from the thumb to the pinkie. This sequence was performed for both hands. After this, Experimenter B disabled the virtual mirror and informed Experimenter A to enter the CVE to start the two-stage social interaction task. With the completion of the task, participants removed the HMDs and were asked to complete the post-assessment on the computer, which included the SSQ, Networked Minds Measure of Social Presence (NMMSP), and STAI-S (see Section 3.4). After the post-assessment, each participant received a supermarket coupon worth 50 Hong Kong dollars as a token of appreciation.

3.4. Measures

3.4.1. Attention allocation

Participants' gaze data were analysed to investigate their attention allocation as people's visual attention reflects their mental attention (Goldberg & Wichansky, 2003; Webb & Renshaw, 2008). Participants' attention allocation was calculated using the percentage of eye fixation time on each of the three nonverbal social cues (i.e., eye gaze, head orientation, and pointing gesture) of their interactant's (i.e., Experimenter A's) avatar.

It is noteworthy that while participants' attention on Experimenter A's eye gaze and head orientation was measured by recording their eye fixations on the eyes and heads of Experimenter A's avatar respectively, participants' attention on the third cue (i.e., pointing gesture) was measured by their eye fixations on the hands and index fingers of the interactant's avatar. It is because (1) participants' hands were occupied during AMC as they needed to hold a pair of controllers (see Section 3.2) to track their hand and finger movements, which might limit their capability of pointing; (2) the accuracy of finger tracking could also introduce an unwanted factor that might affect the effectiveness of pointing gestures; and (3) while it is common to use index fingers to express directional information, other fingers can also serve the same purpose (Langton & Bruce, 2000). Therefore, the hands and index fingers of Experimenter A's avatar were both considered the nonverbal social cue of the pointing gesture. Details about eye gaze data processing can be seen in Section 3.5.

3.4.2. Social presence

The Networked Minds Measure of Social Presence (NMMSP) (Harms & Biocca, 2004) was adopted to assess participants' perceived sense of social presence in the CVE. With 36 items, it measures six subscales of social presence: co-presence, attentional allocation, perceived message understanding, perceived affective understanding, perceived emotional interdependence, and perceived behavioural interdependence. Responses are given on a seven-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (7). The measure was taken during the post-assessment.

3.4.3. Anxiety

The State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1971) was selected to assess participants' anxiety levels before and after completing the social interaction task in the CVE. Two types of anxiety can be measured by this questionnaire: (1) the state anxiety (STAI-S) captures participants' temporary feeling of anxiety at a specific time and situation, while (2) the trait anxiety (STAI-T) captures the individual differences in the frequency and intensity of anxiety experienced across different situations (Guillén-Riquelme & Buela-Casal, 2014), with each type measured by 20 items. Each item was rated on a four-point scale, from "almost never" (1) to "almost always" (4), with a higher score indicating a higher anxiety level. Note that the trait anxiety was measured during the pre-assessment only; the state anxiety was measured during both the pre-assessment and post-assessment so that we could attribute the change of the state anxiety to the AMC in CVE.

3.4.4. Cybersickness

As a common negative side effect of VR exposure, cybersickness is associated with participants' perceptions, such as anxiety (Pot-Kolder, Veling, Counotte, & Van Der Gaag, 2018) and sense of presence (Weech, Kenny, & Barnett-Cowan, 2019). Therefore, we also measured the severity of cybersickness our participants experienced during the pre-assessment and post-assessment using the Simulator Sickness Questionnaire (SSQ) (Kennedy, Lane, Berbaum, & Lilienthal, 1993). This self-reported questionnaire comprises 16 items that measure three sickness symptom clusters labelled Oculomotor (eyestrain, difficulty focusing, blurred vision, headache, etc.), Disorientation (dizziness, vertigo, etc.), and Nausea (nausea, stomach awareness, increased salivation, burping, etc.), respectively. Each item is rated by participants based on a four-point scale ranging from "no symptoms" (0) to "severe symptoms" (3) to indicate the severity of the corresponding symptom they experienced.

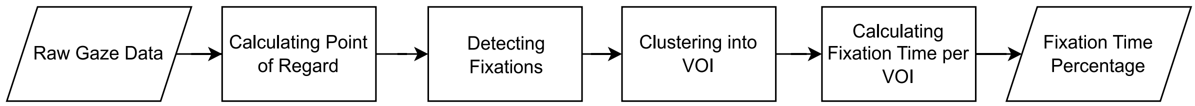


Fig. 6. The pipeline of gaze data processing. The input is the gaze data obtained from VIVE Pro Eye. The output represents the percentage of eye fixation time allocated to each attentional cue (i.e., head orientation, eye gaze, and pointing gesture) as a percentage of the total fixation time on all attentional cues.

3.5. Gaze data analysis

We designed a five-stage pipeline to process the gaze data and analyse participants' attention allocation (see Fig. 6). The five stages are explained as follows.

Calculating point of regard. In this stage, we converted pairs of left and right gaze rays into specific points in the virtual world where users were looking, known as points of regard. These points of regard at each timestamp were calculated by finding the intersection point of the left and right gaze rays. However, due to the precision of the eye tracker, the two gaze rays might not always intersect. In this case, the midpoint of the line segment representing the shortest distance between the pair of gaze rays was calculated as the point of regard, following the method described by Duchowski et al. (2022).

Detecting fixations. In this stage, we detected fixations from points of regard by implementing the Velocity-Threshold Algorithm (I-VT) described by Salvucci and Goldberg (2000), which is commonly used for fixation detection (Gao et al., 2021; Olsen, 2012). I-VT identifies fixations based on the low point-to-point velocity property, with a threshold of less than 20°/s (Salvucci & Goldberg, 2000). The duration of each fixation was limited to the range of 50 to 600 milliseconds (ms) following the hardware manufacturer's recommendations.[8]

Clustering into VOIs. The volume of interest (VOI) refers to the specific target regions in the three-dimensional virtual world that we are concerned about, which are the three nonverbal social cues. Since these regions in the virtual world are three-dimensional, we use the term VOI instead of the traditional area of interest (AOI) to describe them more accurately. Fig. 7 shows the VOIs of the three cues. Because these VOIs were constantly moving during AMC and thus could not be pre-defined using three-dimensional coordinates, we first attached colliders, components for enabling physical interactions in VR, to the three cues to define the VOIs. Then, we used Unity's built-in physics engine to cast a ray from the participant's eye position in the gaze direction and detect collisions with the pre-attached colliders. Lastly, we clustered each fixation into a VOI based on the collision detection result.

Calculating fixation time per VOI. After clustering participants' eye fixations into VOIs in the previous step, we calculated their total eye fixation time on three nonverbal social cues with the below formula:

$$T_{s_i} = \sum_j t_{s_{ij}}, \quad i \in \{\text{head, eye, gesture}\}, s \in \{Stage1,\ Stage2\}$$

where $t_{s_{ij}}$ represents the time of a participant's jth eye fixation on the cue i in Stage s with $i \in \{\text{head, eye, gesture}\}$, $s \in \{Stage1,\ Stage2\}$, and T_{s_i} refers to the sum of the participant's eye fixation time on the cue i in Stage s.

Calculating attention allocation. In the last step, to quantify a participant's attention allocation on a specific cue out of the three, we calculated the ratio of the participant's eye fixation time on this cue to his eye fixation time on all three cues. The calculation is illustrated in the below formula:

$$R_{s_i} = \frac{T_{s_i}}{T_{s_{head}} + T_{s_{eye}} + T_{s_{gesture}}},$$

$$i \in \{\text{head, eye, gesture}\}, s \in \{Stage1,\ Stage2\}$$

where R_{s_i} refers to the ratio of the participant's eye fixation time on the nonverbal social cue i in Stage s, T_{s_i} refers to the sum of the participant's eye fixation time on the cue i in Stage s, and $T_{s_{head}}$, $T_{s_{eye}}$, and $T_{s_{gesture}}$ represent the sum of participant's eye fixation time on eye gaze, head orientation, and pointing gesture, respectively, in Stage s.

Table 2
Descriptive statistics of participants' attention across cues by Group (i.e., gaze type) and Stage (i.e., interaction type).

	Group	Cue	Mean (SD)	Median (Q1,Q3)
Stage 1	Static gaze (n = 30)	Head orientation	0.78(0.13)	0.81(0.71,0.88)
		Eye gaze	0.04(0.03)	0.04(0.02,0.06)
		Pointing gesture	0.18(0.13)	0.15(0.08,0.24)
	Natural gaze (n = 30)	Head orientation	0.81(0.13)	0.85(0.75,0.89)
		Eye gaze	0.06(0.05)	0.05(0.02,0.08)
		Pointing gesture	0.13(0.14)	0.08(0.04,0.15)
Stage 2	Static gaze (n = 30)	Head orientation	0.56(0.15)	0.55(0.50,0.66)
		Eye gaze	0.09(0.07)	0.09(0.03,0.13)
		Pointing gesture	0.35(0.17)	0.35(0.19,0.49)
	Natural gaze (n = 30)	Head orientation	0.55(0.21)	0.58(0.43,0.71)
		Eye gaze	0.12(0.07)	0.12(0.09,0.16)
		Pointing gesture	0.33(0.22)	0.29(0.13,0.47)

Table 3
Descriptive statistics of participants' NMMSP and STAI scores.

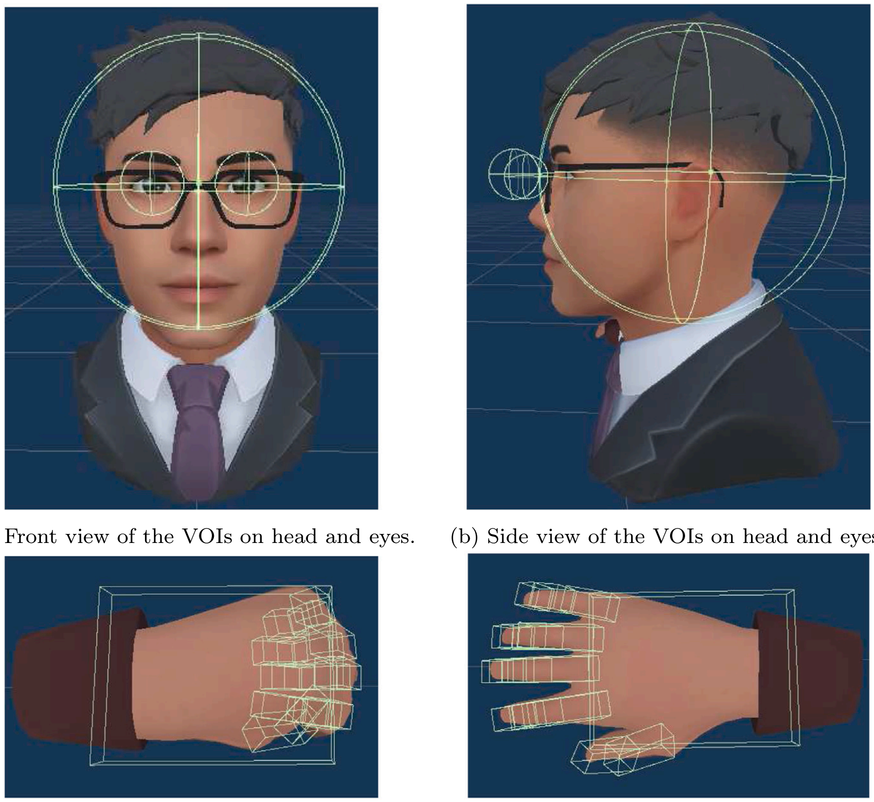
		Static gaze (n = 30)		Natural gaze (n = 30)	
		Mean	SD	Mean	SD
Social presence (NMMSP)	CP	6.54	0.662	6.38	0.909
	AA	5.63	1.201	5.88	1.064
	PMU	6.69	0.682	6.67	0.533
	PAU	6.03	1.234	5.83	1.067
	PEI	4.99	1.648	4.96	1.621
	PBI	6.00	0.897	6.09	0.920
	Total	5.98	0.675	5.97	0.656
Anxiety (STAI)	STAI-T	41.97	9.619	42.50	9.515
	STAI-S (Δ)	−6.63	8.779	−5.20	6.429

CP: Co-presence.
AA: Attentional Allocation.
PMU: Perceived Message Understanding.
PAU: Perceived Affective Understanding.
PEI: Perceived Emotional Interdependence.
PBI: Perceived Behavioural Interdependence.

4. Results

The descriptive statistics of our participants' attention across three nonverbal social cues (i.e., eye gaze, head orientation, and pointing gesture) in the two groups during *Stage1* and *Stage2* are shown in Table 2. The descriptive statistics of the participants' NMMSP and STAI scores are shown in Table 3. The STAI-T scores suggested the participants in both the static gaze group ($M = 41.97$, $SD = 9.619$) and the natural gaze group ($M = 42.50$, $SD = 9.515$) exhibited low to moderate stress levels (Kayikcioglu, Bilgin, Seymenoglu, & Deveci, 2017; Spielberger, Goruch, Lushene, Vagg, & Jacobs, 1983). These scores are consistent with those reported in a prior study involving participants of similar age demographics in Hong Kong (Jones, Dean, & Lo, 2002).

[8] https://connect.tobii.com/s/article/types-of-eye-movements.



(a) Front view of the VOIs on head and eyes. (b) Side view of the VOIs on head and eyes.

(c) Top view of the VOIs on the left hand. (d) Top view of the VOIs on the right hand.

Fig. 7. The VOIs we defined for the three nonverbal social cues using colliders in Unity (regions covered in the green wireframes).

Table 4
Pairwise comparisons of attention for participants in the natural gaze group.

	Pair	Mean Diff.	t value	$Pr(>\|t\|)$	95% CI
Stage 1	$Head - Eye$	0.74	28.77	<0.001***	[0.69, 0.80]
	$Head - Gesture$	0.67	13.48	<0.001***	[0.57,0.77]
	$Eye - Gesture$	−0.07	−2.41	0.023*	[−0.14,−0.01]
Stage 2	$Head - Eye$	0.43	10.67	<0.001***	[0.35,0.51]
	$Head - Gesture$	0.23	2.88	0.007**	[0.07,0.39]
	$Eye - Gesture$	−0.21	−4.43	<0.001***	[−0.30,−0.11]

Note: $p < 0.1(\cdot)$ $p < 0.05(^{*})$ $p < 0.01(^{**})$ $p < 0.001(^{***})$.

4.1. Attention allocation across cues

To address **RQ1**, three pairwise comparisons of participants' attention on different cues (i.e., head orientation versus eye gaze, head orientation versus pointing gesture, and eye gaze versus pointing gesture) were conducted for the natural gaze group with paired sample t-test.

As shown in Table 4, the result indicated that participants' attention differed across the three cues during both socio-emotional interactions ($Head - Eye$: $t(29) = 28.77, p < 0.001$, $Head - Gesture$: $t(29) = 13.48, p < 0.001$, $Eye - Gesture$: $t(29) = -2.41, p = 0.023$) and task interactions ($Head - Eye$: $t(29) = 10.67, p < 0.001$, $Head - Gesture$: $t(29) = 2.88, p = 0.007$, $Eye - Gesture$: $t(29) = -4.43, p < 0.001$). The mean amount of attention participants in the natural gaze group allocated on the three cues is plotted in Fig. 8.

In addition, the results showed that in both socio-emotional interactions (i.e., *Stage1*) and task interactions (i.e., *Stage2*), participants' attention allocation across the three cues followed a specific order: the prominence of the head orientation superseded that of pointing gesture, which, in turn, was greater than the focus on the eyes.

4.2. Effects of avatars' gaze type and interaction type on attention allocation

To investigate how avatars' gaze type and interaction type that participants were engaged with affected their attention allocation during AMC (**RQ2**), a linear regression model was used to explore the effect of Group (i.e., gaze type) and Stage (i.e., interaction type) on participants' attention on each nonverbal social cue separately.

As displayed in Table 5, the results suggested avatars' gaze type significantly affected participants' attention allocation on the interactant's eye gaze ($B = 0.02, SE = 0.01, p = 0.029$): i.e., participants interacting with avatars with natural gaze allocated significantly greater attention on the interactant's eyes in comparison to those interacting with avatars with static gaze. Meanwhile, it is noteworthy that avatars' gaze type did not affect participants' attention on the other two cues (i.e., head orientation and pointing gesture).

Also, the results showed that interaction type made a significant difference in participants' attention on all three cues; in comparison to socio-emotional interactions (i.e., *Stage 1*), participants in task interactions (i.e., *Stage 2*) allocated more attention on their interactant's eye gaze ($B = 0.05, SE = 0.01, p < 0.001$) and pointing gesture ($B = 0.18, SE = 0.03, p < 0.001$), but less attention on the interactant's head orientation ($B = -0.24, SE = 0.03, p < 0.001$).

4.3. Effects of avatars' gaze type on social presence and anxiety

To explore the effect of avatars' gaze type on participants' perceptions of AMC in CVEs (**RQ3**), we compared the static gaze and natural gaze groups in terms of participants' social presence and anxiety with two independent two-sample t-tests. The results are reported in Table 6.

The between-group comparisons of participants' sense of social presence showed no significant difference between the two groups ($t = 0.09, p = 0.932$). Further comparisons of the six sub-scales of

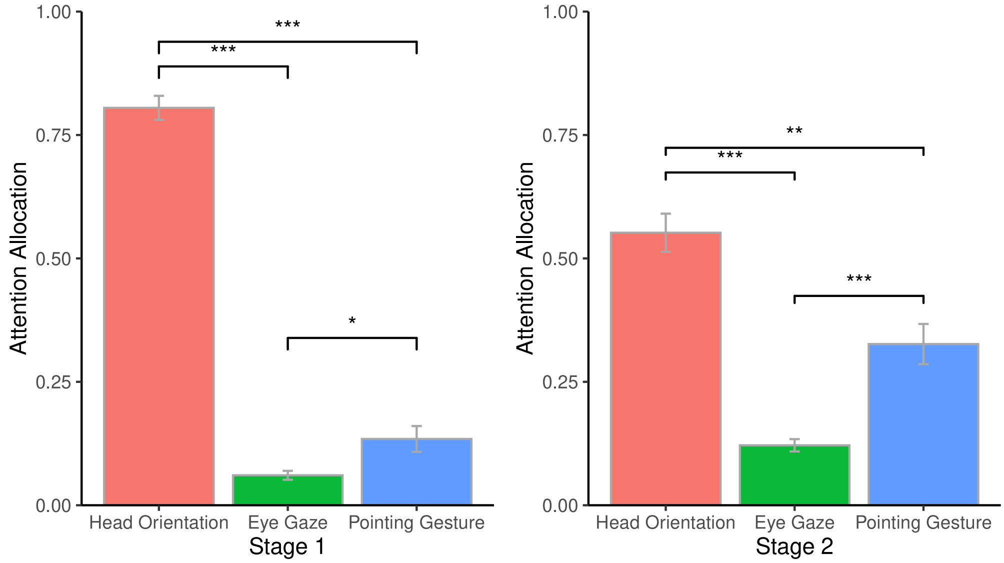


Fig. 8. Average attention allocation across cues for participants in the natural gaze group during *Stage1* (left) and *Stage2* (right). The data processing pipeline for calculating attention allocation is outlined in Section 3.5.

Table 5
Effects of avatar's gaze type (Group) and interaction type (Stage) on attention allocation across three nonverbal social cues.

		B (SE)	t value	Pr (> |t|)	Adj. R^2
Head orientation	(Intercept)	0.79(0.03)	31.463	<0.001***	
	Gaze type	0.01(0.03)	0.276	0.783	0.35***
	Interaction type	−0.24(0.03)	−8.143	<0.001***	
Eye gaze	(Intercept)	0.04(0.01)	4.482	<0.001***	
	Gaze type	0.02(0.01)	2.212	0.029*	0.19***
	Interaction type	0.05(0.01)	5.084	<0.001***	
Pointing gesture	(Intercept)	0.17(0.03)	6.356	<0.001***	
	Gaze type	−0.03(0.03)	−1.010	0.315	0.22***
	Interaction type	0.18(0.03)	5.868	<0.001***	

Note: $p < 0.1(\cdot)$ $p < 0.05(^{*})$ $p < 0.01(^{**})$ $p < 0.001(^{***})$.

Table 6
Between-group comparisons of participants' sense of social presence and anxiety.

	Scales	Mean Diff.	t value	Pr (> |t|)	95% CI
Social presence (NMMSP)	CP	0.17	0.81	0.420	[−0.25, 0.58]
	AA	−0.24	−0.83	0.408	[−0.83, 0.34]
	PMU	0.02	0.11	0.916	[−0.30, 0.33]
	PAU	0.20	0.67	0.505	[−0.40, 0.80]
	PEI	0.04	0.09	0.927	[−0.81, 0.88]
	PBI	−0.09	−0.38	0.706	[−0.56, 0.38]
	Total	0.01	0.09	0.932	[−0.33, 0.36]
Anxiety (STAI)	STAI-T	−0.53	−0.22	0.830	[−5.48, 4.41]
	STAI-S (Δ)	−1.43	−0.72	0.474	[−5.42, 2.55]

Note: $p < 0.1(\cdot)$ $p < 0.05(^{*})$ $p < 0.01(^{**})$ $p < 0.001(^{***})$.
CP: Co-presence.
AA: Attentional Allocation.
PMU: Perceived Message Understanding.
PAU: Perceived Affective Understanding.
PEI: Perceived Emotional Interdependence.
PBI: Perceived Behavioural Interdependence.

social presence also showed no significant difference between the two groups (co-presence: $t = 0.81, p = 0.420$; attentional allocation: $t = -0.83, p = 0.408$; perceived message understanding: $t = 0.11, p = 0.916$; perceived affective understanding: $t = 0.67, p = 0.505$; perceived emotional interdependence: $t = 0.09, p = 0.927$; perceived behavioural interdependence: $t = -0.38, p = 0.706$), indicating avatars' gaze type did not affect participants' sense of social presence.

Similarly, the between-group comparisons of participants' trait anxiety suggested no significant difference between the static gaze and natural gaze groups (STAI-T: $t = -0.22, p = 0.830$), further addressing the risk of biases introduced during grouping. In addition, a comparison of the change in participants' state anxiety before and after AMC still showed no difference between the two groups (STAI-S (Δ): $t = -0.72, p = 0.474$), indicating that avatars' gaze type did not affect participants' state anxiety.

To ensure that the above measures of social presence and anxiety were not affected by participants' experienced cybersickness during AMC in the CVE, we also compared the difference in the severity of cybersickness experienced by our participants with paired sample t-tests. The results suggested that there was no significant difference in the severity of cybersickness before and after AMC, either in the three sub-scales (Nausea: $t = 0.93, p = 0.356$, Oculomotor: $t = 1.09, p = 0.282$, Disorientation: $t = -0.44, p = 0.664$) or the total scores (TS: $t = 0.64, p = 0.527$), indicating participants' did not experience significant cybersickness induced by AMC in the CVE. Hence, it is safe to conclude that the above measures of social presence and anxiety were not affected by the severity of cybersickness experienced by our participants.

5. Discussion

With the two-arm RCT, we explored the importance of three nonverbal social cues in directing people's attention, the effect of avatars' gaze type and interaction type on users' attentional allocation, as well as the effect of avatars' gaze type on participants' perceptions of social interactions (i.e., social presence and anxiety) during AMC in the CVE.

5.1. Importance of nonverbal social cues during AMC

The pairwise comparison of the amount of attention that participants of the natural gaze group allocated to the three nonverbal social cues (i.e., head orientation, eye gaze, and pointing gesture) suggested that the cues were not equally important in directing users' attention when completing the social interaction tasks in the CVE (**RQ1**). In addition, it revealed a very interesting pattern regarding the importance of nonverbal social cues in AMC: during both socio-emotional and task interactions, the participants tended to allocate most of their attention to the head orientation of the interactant's avatar, followed by pointing gesture and then eye gaze.

This finding contradicts the sequence of cues (i.e., first the eyes, then the head, and then other parts of the body) that users' attention allocation follows when viewing static images (Birmingham et al., 2008), but corroborates the gaze avoidance phenomenon in both real-world (Laidlaw et al., 2011) and virtual environments (Rubo & Gamer, 2018) — people tended to avoid rather than seek to look at others' faces in complex situations that afforded social interaction; the phenomenon can be influenced by factors such as culture (Haensel, Smith, & Senju, 2022), task demands (Abeles & Yuval-Greenberg, 2017), and personality traits (Larsen & Shackelford, 1996), but not by anxiety during social interactions (Rösler, Göhring, Strunz, & Gamer, 2021; Tönsing et al., 2022). This finding substantiated the claim that communications with participants embodied in computer-generated avatars could provide social interactions similar to those in real-world situations. In our study, the chance of social interaction provided by AMC elicited the gaze-avoidance phenomenon as participants tend to avoid unwanted attention or interaction from the interactant. In addition, as previous studies (Laidlaw et al., 2011; Rubo & Gamer, 2018) compared users' attention on the interactant and other objects only, our study provided a nuanced understanding regarding users' attention allocation across the nonverbal social cues signalled by the interactant in AMC.

Another possible explanation would be that VR provided a richer modality communication form than images (Steuer, 1992), which have been used in previous studies on gaze patterns. With images, the only way that participants could obtain information was by observing static images; it was reasonable for participants to attend more to the eyes of characters in images because of the rich information eyes could convey (Cañigueral & Hamilton, 2019). Meanwhile, in our study, with live avatars and immersive content in the CVE, participants could obtain information through multiple modalities, including non-verbal (e.g., movements of the head and hands) and verbal communications with the interactant. These modalities enhanced the richness of the media (Daft & Lengel, 1986; Steuer, 1992) but might also weaken the importance of eye gaze and consequently reduce the amount of attention participants allocate to it. In addition, object size could explain our findings. According to Proulx (2010), large objects tend to capture attention. Because the size of the head is the largest among the three nonverbal social cues, it was more likely to capture participants' attention and occupied most of the participants' attention during AMC for social interactions.

In summary, the prioritisation of head orientation over eye gaze and pointing gestures indicates that AMC in CVEs may more accurately replicate real-world social interactions compared to other forms of CMC that utilise media with lower levels of richness. This underscores the potential of CVEs to enhance the naturalness and enjoyment of social experiences in remote contexts.

5.2. Impact of avatars' gaze type and interaction type on attention allocation

The results of multiple linear regressions implied that both avatars' natural gaze and task interaction affected attention allocation and potentially mitigated the general gaze-avoidance pattern during AMC (**RQ2**).

The results suggested that the interaction type did affect participants' attention allocation. Specifically, in comparison to socio-emotional interactions (i.e., *Stage1*), participants in task interactions (i.e., *Stage2*) tended to attend more to the interactant's eye gaze and pointing gesture but less to the most prominent cue - head orientation. While socio-emotional interactions might not have specific agendas or goals for participants to achieve, and therefore, participants should have less motivation or urgency to interact with others, task interactions required participants to gather information about their interactant's behaviours and intentions quickly for effective attention coordination and task completion. Moreover, in comparison to socio-emotional interactions in which participants' gaze avoidance behaviour might not have any immediate consequence, in task interactions participants might be forced to obtain additional information and confirmation from the interactant's eyes to complete the task collaboratively and effectively. Consequently, participants' attention in task interactions was more spread out on different cues, including eye gaze and pointing gesture, instead of focusing on the most prominent one - head orientation. Therefore, although the eye-gaze avoidance phenomenon (Laidlaw et al., 2011) may still exist in AMC for task interactions, participants' increased attention to the interactant's less prominent cues (i.e., pointing gesture and eye gaze) contributed to the decrease in their attention to the head orientation.

Avatars' natural gaze was positively associated with participants' attention on the interactant's eye gaze implying that despite the gaze avoidance phenomenon (Laidlaw et al., 2011; Rubo & Gamer, 2018) in AMC, participants tended to look at the eyes of the interactant's avatar if it could mirror the interactant's real eye behaviours and therefore provide more useful information for attention coordination. In the case of avatars with static gaze, directional information conveyed through gaze direction was the same as through the head orientation of avatars because the virtual eyeballs remained centred in the eye sockets throughout the AMC. It is interesting that avatars' gaze type did not affect participants' attention on the other two nonverbal social cues (i.e., head orientation and pointing gesture) sent by the interactant. A possible explanation is that the change in avatars' gaze type was insufficient to dilute the relative importance of other cues in attention coordination. Future studies may explore this by manipulating the forms and functionalities of all nonverbal social cues in AMC.

In summary, the findings here suggested that incorporating natural gaze behaviours in avatars could enhance attention coordination by prompting users to infer the interactant's intentions based on eye gaze, even when the gaze-avoidance phenomenon occurs. This could foster improved communication and remote collaboration, particularly during task interactions. Given that task interactions are more prevalent than socio-emotional interactions in professional settings, such as education and healthcare, we advocate for the integration of natural gaze in CVEs tailored for professional use.

5.3. Impact of avatars' gaze type on social presence and anxiety

The results suggested that avatars' gaze type did not affect participants' sense of social presence or anxiety during AMC (**RQ3**). While the participants' anxiety levels showed no difference when interacting with avatars with natural or static gaze is consistent with previous studies (Reichenberger, Wechsler, Diemer, Mühlberger, & Notzon, 2022), there are some potential explanations for the observed indifference between the two groups regarding participants' sense of social presence.

First, the two gaze types of avatars used in our study (i.e., static gaze and natural gaze) may only partially contribute to the avatars' behavioural anthropomorphism (i.e., how human-like an avatar behaves), because behavioural anthropomorphism involves more behavioural dimensions beyond gaze behaviours. Therefore, although behavioural anthropomorphism is a powerful predictor for users' sense of social presence during CMCs (Oh et al., 2018), the mere difference in avatars' gaze type might not be able to capture the variance of behavioural anthropomorphism. Avatars with static gaze in our study already had a high level of behavioural anthropomorphism and therefore participants in the static gaze group might have already perceived a relatively high level of social presence, which consequently contributed to the indifference of participants' sense of social presence between the two groups (see Table 3). Future studies may consider capturing avatar anthropomorphism with more discriminative measures to study its effect on users' perceptions during AMC.

Second, two sub-scales of social presence (i.e., perceived affective understanding and perceived emotional interdependence) focused on the participants' feelings and emotions with the interactant. However, due to the nature of the social interaction task in this study, we did not expect many emotional exchanges between participants and the interactant. Future studies may consider designing social interaction tasks involving more emotional exchanges to ensure two sub-scales are sensitive to measure the potential effects of avatars' gaze types on social presence.

6. Limitations and future work

The first limitation is that we had participants interacting with the same experimenter. Although it enabled us to control the interaction process and duration, it might limit the gender of the interactant (i.e., male) and consequently affect the generalisability of the findings. Although it is out of the scope of this study as it might require significantly more participants to test the effect of the interactant's gender on attention distribution and participants' sense of social presence and anxiety, it is worth analysing in future studies.

The second limitation is the weight of the HMDs and the fixed height of the table. Several participants reported that the HMDs were heavy, and they felt fatigue and discomfort during the study. In addition, a few participants claimed that the models were placed at a relatively low height, which required them to bend over to explore the models closely. This might result in increased fatigue and discomfort, particularly in the neck. Although the analysis of participants' severity of cybersickness showed no significant change before and after the experiment, future studies might consider improving the design and using lighter VR HMDs to avoid any potential influence on participants' overall perception and performance during the study.

The third limitation is using Valve Index controllers to track participants' finger movements restricted their ability to move their fingers and perform hand gestures freely. Although they were informed that the controllers were well tied to their hands and they were free to move their fingers and utilise gestures, most participants rarely detached all their fingers from the controllers throughout the entire interaction. Thus, we suggest future studies use other methods, such as depth cameras and computer vision based approaches, for tracking finger movements and hand gestures to provide higher flexibility and freedom to participants' hand movements.

The fourth limitation of our study is the restricted duration of social interactions, which were confined to relatively short time frames. This constraint may not adequately capture the long-term effects of different gaze types on attention allocation, anxiety, and social presence in CVEs. This is a common limitation inherent to laboratory experiments, where the need for controlled conditions often necessitates shorter interaction periods. Consequently, our findings may not fully represent the dynamics that develop with prolonged use of CVEs. To address this limitation, future studies should consider extending the duration of interactions and investigating the long-term use of CVEs. Such studies could lead to valuable insights into the sustained impact of avatar anthropomorphism on the effectiveness and overall user experience of AMC.

Lastly, this study does not delve into the reasons behind the observed gaze avoidance behaviours in the CVE, but gaining a deeper understanding of its implications for social interactions, both in the real world and in CVEs, is important for future research. Exploring this phenomenon from both quantitative and qualitative perspectives through future studies could provide valuable insights that inform the development of targeted interventions and strategies aimed at promoting more effective and comfortable social interactions in CVEs.

7. Conclusion

Grounded in a framework integrating theories related to face-to-face social interactions, including nonverbal social cues, the eye-mind hypothesis, and interaction process analysis, with theories related to AMC under the umbrella of avatar anthropomorphism and motivated by the importance of nonverbal social behaviours during social interaction, we conducted this two-arm RCT to examine users' attention distribution across nonverbal social cues, the impact of avatars' gaze type and interaction type on users' attention allocation, and the effects of avatars' gaze type on participants' perceptions during AMC. Sixty participants (29 males and 31 females) joined the study and were randomly assigned to either the natural or the static gaze group to complete social interaction with an experimenter in our custom-designed CVE for less than 20 minutes.

The findings of the study revealed several important insights. Firstly, the three primary nonverbal social cues were not equally important in directing participants' attention during AMC: the most important cue was head orientation, followed by pointing gestures and then the eye gaze. That is, human users tended to avoid the unwanted eye gaze from their interactant. Secondly, the type of social interactions that participants were engaged with during AMC played an important role in their attention allocation: task interaction is associated with increased attention on the two less prominent cues (i.e., pointing gestures and eye gaze) in comparison with socio-emotional interactions. Meanwhile, avatars' gaze type, as an important proxy for avatar anthropomorphism, mitigated the gaze-avoidance effect in social interactions as it was associated with participants' increased attention on eye gaze, but showed no effect on users' attention on other cues. Lastly, avatars' gaze type showed no effect on users' perceptions of social presence and anxiety during AMC for social interactions.

The study provided a more nuanced understanding regarding human users' attention allocation across the three primary nonverbal social cues during AMC; that is, despite the general existence of gaze avoidance, the avatars' gaze type and interaction type played an important role in affecting users' attention allocation. Additionally, the study highlighted important implications for the future design of CVEs, which are increasingly being adopted across various professional fields such as education, healthcare, and industry. The broader implementation of natural gaze is anticipated to enhance attention coordination and communication, which are essential for effective task interactions — the primary form of social interactions in the professional use of CVEs. As CVEs continue to evolve and become integral to more professional and recreational contexts, these insights will be instrumental in shaping future innovations that enhance collaboration and communication in digital spaces.

CRediT authorship contribution statement

Chen Li: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Yixin Dai:** Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Guang Chen:** Writing – original draft, Software, Resources, Methodology, Formal analysis, Data curation. **Jing Liu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Investigation, Formal analysis. **Ping Li:** Writing – review & editing, Validation. **Horace Ho-shing Ip:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Formal analysis.

Funding and Ethical review

This work was supported by The Hong Kong Polytechnic University (project number: P0035264). The study was approved by the Institutional Review Board of The Hong Kong Polytechnic University (application number: HSEARS20230210016).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors want to thank Mr Honglin Li, Ms Wujie Gao, Ms Qi Lyu, and Ms Qinyang Wu for their comments and suggestions on the virtual environment design and the study design.

Data availability

Data will be made available on request.

References

Abeles, D., & Yuval-Greenberg, S. (2017). Just look away: Gaze aversions as an overt attentional disengagement mechanism. *Cognition*, *168*, 99–109. http://dx.doi.org/10.1016/j.cognition.2017.06.021.

Argyle, M. (1988). *Bodily communication*. Methuen & Co Ltd, London.

Argyle, M., & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, 289–304. http://dx.doi.org/10.1111/j.2044-8295.1970.tb01257.x.

Bales, R. F. (1950). *Interaction process analysis; a method for the study of small groups*. Addison-Wesley.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a" language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, *4*(3), 311–331. http://dx.doi.org/10.1080/713756761.

Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. *Communications of the ACM*, *44*(7), 79–85. http://dx.doi.org/10.1145/379300.379322.

Bente, G., Eschenburg, F., & Aelker, L. (2007). Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. In *Presence 2007: Proceedings of the 10th annual international workshop on presence, October 25-27, 2007, Barcelona, Spain* (pp. 207–214).

Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., & Albanese, A. (1997). Analysis of blink rate patterns in normal subjects. *Movement Disorders*, *12*(6), 1028–1034. http://dx.doi.org/10.1002/mds.870120629.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition*, *16*(2–3), 341–355. http://dx.doi.org/10.1080/13506280701434532.

Brown, B., & Bell, M. (2004). Cscw at play: 'there'as a collaborative virtual environment. In *Proceedings of the 2004 ACM conference on computer supported cooperative work* (pp. 350–359). http://dx.doi.org/10.1145/1031607.1031666.

Burden, D., & Savin-Baden, M. (2019). *Virtual humans: Today and tomorrow*. Chapman and Hall/CRC, http://dx.doi.org/10.1201/9781315151199.

Cañigueral, R., & Hamilton, A. F. d. C. (2019). The role of eye gaze during natural social interactions in typical and autistic people. *Frontiers in Psychology*, *10*, 560. http://dx.doi.org/10.3389/fpsyg.2019.00560.

Churchill, E. F., Snowdon, D. N., & Munro, A. J. (2012). *Collaborative virtual environments: Digital places and spaces for interaction*. Springer Science & Business Media.

Craske, M. G., Rauch, S. L., Ursano, R., Prenoveau, J., Pine, D. S., & Zinbarg, R. E. (2011). What is an anxiety disorder? *Focus*, *9*(3), 369–388. http://dx.doi.org/10.1176/foc.9.3.foc369.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554–571. http://dx.doi.org/10.1287/mnsc.32.5.554.

Derouech, O., Hrimech, H., Lachgar, M., & Hanine, M. (2024). A literature review of collaborative virtual environments: Impacts, design principles, and challenges. *Journal of Information Technology Education: Research*, *23*, 11. http://dx.doi.org/10.28945/5283.

Duchowski, A. T., Krejtz, K., Volonte, M., Hughes, C. J., Brescia-Zapata, M., & Orero, P. (2022). 3D gaze in virtual reality: vergence, calibration, event detection. *Procedia Computer Science*, *207*, 1641–1648. http://dx.doi.org/10.1016/j.procs.2022.09.221.

Efran, J. S. (1968). Looking for approval: effects on visual behavior of approbation from persons differing in importance. *Journal of Personality and Social Psychology*, *10*(1), 21. http://dx.doi.org/10.1037/h0026383.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581–604. http://dx.doi.org/10.1016/S0149-7634(00)00025-7.

Freeman, G., Acena, D., McNeese, N. J., & Schulenberg, K. (2022). Working together apart through embodiment: Engaging in everyday collaborative activities in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, *6*(GROUP), 1–25. http://dx.doi.org/10.1145/3492836.

Gao, H., Bozkir, E., Hasenbein, L., Hahn, J.-U., Göllner, R., & Kasneci, E. (2021). Digital transformations of classrooms in virtual reality. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–10). http://dx.doi.org/10.1145/3411764.3445596.

Garau, M. (2003). *The impact of avatar fidelity on social interaction in virtual environments*. University of London, University College London (United Kingdom).

Goldberg, J. H., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In *The mind's eye* (pp. 493–516). Elsevier, http://dx.doi.org/10.1016/B978-044451020-4/50027-X.

Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, *24*(4), 1494–1509. http://dx.doi.org/10.1016/j.chb.2007.05.007.

Guillén-Riquelme, A., & Buela-Casal, G. (2014). Meta-analysis of group comparison and meta-analysis of reliability generalization of the State-Trait Anxiety Inventory Questionnaire (STAI). *Revista Espanola de Salud Publica*, *88*(1), 101–112. http://dx.doi.org/10.4321/s1135-57272014000100007.

Hadnett-Hunter, J., Nicolaou, G., O'Neill, E., & Proulx, M. (2019). The effect of task on visual attention in interactive virtual environments. *ACM Transactions on Applied Perception (TAP)*, *16*(3), 1–17. http://dx.doi.org/10.1145/3352763.

Haensel, J. X., Smith, T. J., & Senju, A. (2022). Cultural differences in mutual gaze during face-to-face interactions: A dual head-mounted eye-tracking study. *Visual Cognition*, *30*(1–2), 100–115. http://dx.doi.org/10.1080/13506285.2021.1928354.

Harms, C., & Biocca, F. (2004). Internal consistency and reliability of the networked minds measure of social presence. *vol. 2004*, In *Seventh annual international workshop: Presence* (pp. 246–251). Universidad Politecnica de Valencia Valencia.

Herrera, F., Oh, S. Y., & Bailenson, J. N. (2020). Effect of behavioral realism on social interactions inside collaborative virtual environments. *Presence*, *27*(2), 163–182. http://dx.doi.org/10.1162/pres_a_00324.

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLoS One*, *10*(8), Article e0136905. http://dx.doi.org/10.1371/journal.pone.0136905.

Howell, A. N., Zibulsky, D. A., Srivastav, A., & Weeks, J. W. (2016). Relations among social anxiety, eye contact avoidance, state anxiety, and perception of interaction performance during a live conversation. *Cognitive Behaviour Therapy*, *45*(2), 111–122. http://dx.doi.org/10.1080/16506073.2015.1111932.

Jing, A., May, K., Matthews, B., Lee, G., & Billinghurst, M. (2022). The impact of sharing gaze behaviours in collaborative mixed reality. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–27. http://dx.doi.org/10.1145/3555564.

Jones, A. Y., Dean, E., & Lo, S. K. (2002). Interrelationships between anxiety, lifestyle self-reports and fitness in a sample of hong kong university students. *Stress*, *5*(1), 65–71. http://dx.doi.org/10.1080/102538902900012350.

Kayikcioglu, O., Bilgin, S., Seymenoglu, G., & Deveci, A. (2017). State and trait anxiety scores of patients receiving intravitreal injections. *Biomedicine Hub*, *2*(2), 1–5. http://dx.doi.org/10.1159/000478993.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. http://dx.doi.org/10.1207/s15327108ijap0303_3.

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition*, *17*(6–7), 925–944. http://dx.doi.org/10.1080/13506280902826775.

Kyrlitsias, C., & Michael-Grigoriou, D. (2022). Social interaction with agents and avatars in immersive virtual environments: A survey. *Frontiers in Virtual Reality*, *2*, Article 786665. http://dx.doi.org/10.3389/frvir.2021.786665.

Laidlaw, K. E., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, *108*(14), 5548–5553. http://dx.doi.org/10.1073/pnas.1017022108.

Langton, S. R. (2000). The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(3), 825–845. http://dx.doi.org/10.1080/713755908.

Langton, S. R., & Bruce, V. (2000). You must see the point: automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 747. http://dx.doi.org/10.1037/0096-1523.26.2.747.

Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59. http://dx.doi.org/10.1016/S1364-6613(99)01436-9.

Larsen, R. J., & Shackelford, T. K. (1996). Gaze avoidance: Personality and social judgments of people who avoid direct face-to-face contact. *Personality and Individual Differences*, *21*(6), 907–917. http://dx.doi.org/10.1016/S0191-8869(96)00148-1.

Li, C., & Liu, J. (2022). Collaborative virtual environment for distant and blended learning in the higher education setting: A systematic review. In *International conference on blended learning* (pp. 135–146). Springer.

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision*, *11*(8), http://dx.doi.org/10.1167/11.8.17, 17–17.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *12*(5), 481–494. http://dx.doi.org/10.1162/105474603322761289.

Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, *6*, 30–53. http://dx.doi.org/10.12840/issn.2255-4165.2018.06.01.015.

Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, *11*(1), 153–178. http://dx.doi.org/10.1111/j.1083-6101.2006.tb00308.x.

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 114. http://dx.doi.org/10.3389/frobt.2018.00114.

Oliva, R., Beacco, A., Navarro, X., & Slater, M. (2022). QuickVR: A standard library for virtual embodiment in unity. *Frontiers in Virtual Reality*, *3*, Article 937191. http://dx.doi.org/10.3389/frvir.2022.937191.

Olsen, A. (2012). The Tobii I-VT fixation filter. *Tobii Technology*, *21*, 4–19.

Peña, J., & Hancock, J. T. (2006). An analysis of socioemotional and task communication in online multiplayer video games. *Communication Research*, *33*(1), 92–109. http://dx.doi.org/10.1177/0093650205283103.

Perrett, D. I., & Emery, N. J. (1994). Understanding the intentions of others from visual signals: neurophysiological evidence. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *13*(5), 683–694.

Plopski, A., Hirzle, T., Norouzi, N., Qian, L., Bruder, G., & Langlotz, T. (2022). The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Computing Surveys*, *55*(3), 1–39. http://dx.doi.org/10.1145/3491207.

Pot-Kolder, R., Veling, W., Counotte, J., & Van Der Gaag, M. (2018). Anxiety partially mediates cybersickness symptoms in immersive virtual reality environments. *Cyberpsychology, Behavior, and Social Networking*, *21*(3), 187–193. http://dx.doi.org/10.1089/cyber.2017.0082.

Proulx, M. J. (2010). Size matters: large objects capture attention in visual search. *PLoS One*, *5*(12), Article e15293. http://dx.doi.org/10.1371/journal.pone.0015293.

Qiao, J., Xu, J., Li, L., & Ouyang, Y.-Q. (2021). The integration of immersive virtual reality simulation in interprofessional education: A scoping review. *Nurse Education Today*, *98*, Article 104773. http://dx.doi.org/10.1016/j.nedt.2021.104773.

Reichenberger, J., Wechsler, T. F., Diemer, J., Mühlberger, A., & Notzon, S. (2022). Fear, psychophysiological arousal, and cognitions during a virtual social skills training in social anxiety disorder while manipulating gaze duration. *Biological Psychology*, *175*, Article 108432. http://dx.doi.org/10.1016/j.biopsycho.2022.108432.

Rösler, L., Göhring, S., Strunz, M., & Gamer, M. (2021). Social anxiety is associated with heart rate but not gaze behavior in a real social interaction. *Journal of Behavior Therapy and Experimental Psychiatry*, *70*, Article 101600. http://dx.doi.org/10.1016/j.jbtep.2020.101600.

Rubo, M., & Gamer, M. (2018). Virtual reality as a proxy for real-life social attention? In *Proceedings of the 2018 ACM symposium on eye tracking research & applications* (pp. 1–2). http://dx.doi.org/10.1145/3204493.3207411.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 71–78). http://dx.doi.org/10.1145/355017.355028.

Schäfer, A., Reis, G., & Stricker, D. (2022). A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems. *ACM Computing Surveys*, *55*(6), 1–27. http://dx.doi.org/10.1145/3533376.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Toronto; London; New York: Wiley.

Spielberger, C. D., Gonzalez-Reigosa, F., Martinez-Urrutia, A., Natalicio, L. F., & Natalicio, D. S. (1971). The state-trait anxiety inventory. *Revista Interamericana de Psicologia/Interamerican Journal of Psychology*, *5*(3 & 4).

Spielberger, C. D., Goruch, R., Lushene, R., Vagg, P., & Jacobs, G. (1983). *Manual for the state-trait inventory STAI (form Y)*. Palo Alto, CA, USA: Mind Garden.

Steptoe, W., Oyekoya, O., Murgia, A., Wolff, R., Rae, J., Guimaraes, E., et al. (2009). Eye tracking for avatar eye gaze control during object-focused multiparty interaction in immersive collaborative virtual environments. In *2009 IEEE virtual reality conference* (pp. 83–90). IEEE, http://dx.doi.org/10.1109/VR.2009.4811003.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, *42*(4), 73–93. http://dx.doi.org/10.1111/j.1460-2466.1992.tb00812.x.

Tönsing, D., Schiller, B., Vehlen, A., Spenthof, I., Domes, G., & Heinrichs, M. (2022). No evidence that gaze anxiety predicts gaze avoidance behavior during face-to-face social interaction. *Scientific Reports*, *12*(1), 21332. http://dx.doi.org/10.1038/s41598-022-25189-z.

Turner, J. H. (1988). *A theory of social interaction*. Stanford University Press.

Ververidis, D., Nikolopoulos, S., & Kompatsiaris, I. (2022). A review of collaborative virtual reality systems for the architecture, engineering, and construction industry. *Architecture*, *2*(3), 476–496. http://dx.doi.org/10.3390/architecture2030027.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, *19*(1), 52–90. http://dx.doi.org/10.1177/009365092019001003.

Webb, N., & Renshaw, T. (2008). Eyetracking in HCI. In *Research methods for human-computer interaction* (pp. 35–69). Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511814570.004.

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: a review. *Frontiers in Psychology*, *10*, 158. http://dx.doi.org/10.3389/fpsyg.2019.00158.

Zibrek, K., & McDonnell, R. (2019). Social presence and place illusion are affected by photorealism in embodied VR. In *Proceedings of the 12th ACM SIGGRAPH conference on motion, interaction and games* (pp. 1–7). http://dx.doi.org/10.1145/3359566.3360064.