



Eye See What You See: Exploring How Bi-Directional Augmented Reality Gaze Visualisation Influences Co-Located Symmetric Collaboration

Allison Jing*, Kieran May, Gun Lee and Mark Billinghurst

Empathic Computing Lab, Australian Research Centre for Interactive and Virtual Environment, STEM, The University of South Australia, Mawson Lakes, SA, Australia

OPEN ACCESS

Edited by:

Parinya Punpongsanon,
Osaka University, Japan

Reviewed by:

Naoya Isoyama,
Nara Institute of Science and
Technology (NAIST), Japan
Thuong Hoang,
Deakin University, Australia

*Correspondence:

Allison Jing
allison.jing@mymail.unisa.edu.au

Specialty section:

This article was submitted to
Augmented Reality,
a section of the journal
Frontiers in Virtual Reality

Received: 19 April 2021

Accepted: 25 May 2021

Published: 14 June 2021

Citation:

Jing A, May K, Lee G and Billinghurst M
(2021) Eye See What You See:
Exploring How Bi-Directional
Augmented Reality Gaze Visualisation
Influences Co-Located
Symmetric Collaboration.
Front. Virtual Real. 2:697367.
doi: 10.3389/frvir.2021.697367

Gaze is one of the predominant communication cues and can provide valuable implicit information such as intention or focus when performing collaborative tasks. However, little research has been done on how virtual gaze cues combining spatial and temporal characteristics impact real-life physical tasks during face to face collaboration. In this study, we explore the effect of showing joint gaze interaction in an Augmented Reality (AR) interface by evaluating three bi-directional collaborative (BDC) gaze visualisations with three levels of gaze behaviours. Using three independent tasks, we found that all bi-directional collaborative BDC visualisations are rated significantly better at representing joint attention and user intention compared to a non-collaborative (NC) condition, and hence are considered more engaging. The Laser Eye condition, spatially embodied with gaze direction, is perceived significantly more effective as it encourages mutual gaze awareness with a relatively low mental effort in a less constrained workspace. In addition, by offering additional virtual representation that compensates for verbal descriptions and hand pointing, BDC gaze visualisations can encourage more conscious use of gaze cues coupled with deictic references during co-located symmetric collaboration. We provide a summary of the lessons learned, limitations of the study, and directions for future research.

Keywords: augmented reality collaboration, gaze visualisation, human-computer interaction, CSCW, design and evaluation methods

1 INTRODUCTION

This paper presents research on how visualisation of gaze cues in an Augmented Reality interface can impact face to face collaboration. As one of the most common interaction modalities, gaze communicates rich information during collaboration. We naturally gaze at objects to express our interests (Zhang et al., 2014) as our eyes move fast with less physical effort and can be used over a distance than other modalities which are often not as readily available. In both co-located and remote collaboration, we share gaze cues to improve the awareness of collaborator's focus (Lee et al., 2017) (Kuhn et al., 2009), minimise duplicate work (Zhang et al., 2017), predict another person's intention (Baron-Cohen et al., 1997), and share joint gaze to achieve common ground (Whittaker and O'Connell, 1997). For example, in a face-to-face operating room surgeons can use gaze as a referential pointer when their hands are occupied, or during online education training, gaze can be used to detect student behaviour patterns when other cues are not remotely accessible.

During a co-located collaboration, a dyad may intentionally add another explicit communication cue, such as verbal description or gesturing, together with gaze cues to further align their mutual

understanding, because gaze cues are implicit and are not always intentionally communicative. On the other hand, without the visual representation of gaze cues, collaborators may have to divide their attention between their partner's face and the task space, and may not be able to easily identify the object of interest without further verbal or gestural affirmation. To convey the rich context information that gaze cues produce, previous studies have confirmed the need for sharing gaze cues in face-to-face task collaboration over wall displays (Zhang et al., 2017) (Zhang et al., 2015). Situated displays have size limitations and are poor at conveying spatial cues which can cause confusion, disruption and distraction (Zhang et al., 2017).

In recent years, the rise of emerging reality-based technologies such as Augmented Reality (AR), Mixed Reality (MR), or Virtual Reality (VR) has enabled novel techniques to overcome the limitations of situated displays. However, the current co-located collaborative gaze visualisation studies conducted using reality-based technologies are often one-directional (single user gaze indicator) (Erickson et al., 2020) (Li et al., 2019), asynchronous (Rahman et al., 2020) with different knowledge level towards the task (Erickson et al., 2020), in a virtual task space (Li et al., 2019), or between a human and virtual collaborator (Erickson et al., 2020) (Li et al., 2019) (Rahman et al., 2020). It is common to represent all gaze behaviours (eye fixation, saccades, and blink etc) using the same virtual cue (e.g., a virtual gaze ray) while richer visualisation of different gaze behaviours combining both spatial and temporal information are neglected.

As a result, our research aims to examine how collaborative AR gaze visualisations can represent different dynamic gaze behaviours to facilitate pair communication in a range of co-located physical tasks. We postulate that sharing AR joint gaze visualisations between pairs may encourage more conscious use of gaze pointing coupled with deictic gesture references when multiple communication resources (Wang et al., 2012) such as hand pointing and verbal description are made available during co-located symmetric (Ens et al., 2019) tasks. This should result in a more effective and collaborative real-life task experience.

This paper first presents a prototype system **Figure 1** that offers four different styles of AR gaze visualisations with three levels of gaze behaviours. Defined as bi-directional collaborative (BDC) visualisation, both user's own and the partner's gaze indicators are made available to achieve explicit joint gaze interaction. We then evaluate our design in a lab-controlled user study consisting of three tasks (visual searching, matching, and puzzle-solving), spanning three task space setups ranging from a constrained desk to a relatively larger freely navigable environment. The results indicate that by combining both spatial and temporal factors in physical task space, AR gaze visualisations, especially Laser Eye (LE), improve pair communication by enabling virtual representation of joint attention and user intention using gaze behaviours in the relatively free-formed task space. Embodying gaze direction, LE maximises the view of gaze visualisations through peripheral vision and significantly encourages collaborator's mutual reaction with a relatively low mental effort. In contrast to the baseline of non-collaborative (NC) gaze cue, the overall experience with all three BDC gaze visualisations is considered significantly more engaging. Finally,

we provide insights from discussion, and a conclusion coupled with the direction for future work.

The main contributions of this paper are:

- The first user study that investigates the impact of AR enhanced gaze visualisations that combine both spatial and temporal factors during co-located symmetric collaboration.
- The first study that compares the effect of visualising joint gaze interactions (i.e. bi-directional collaborative gaze) to non-collaborative gaze over different task scales.
- The first study that explores the allocation of AR gaze cues compared to other available communication resources in physical collaboration.

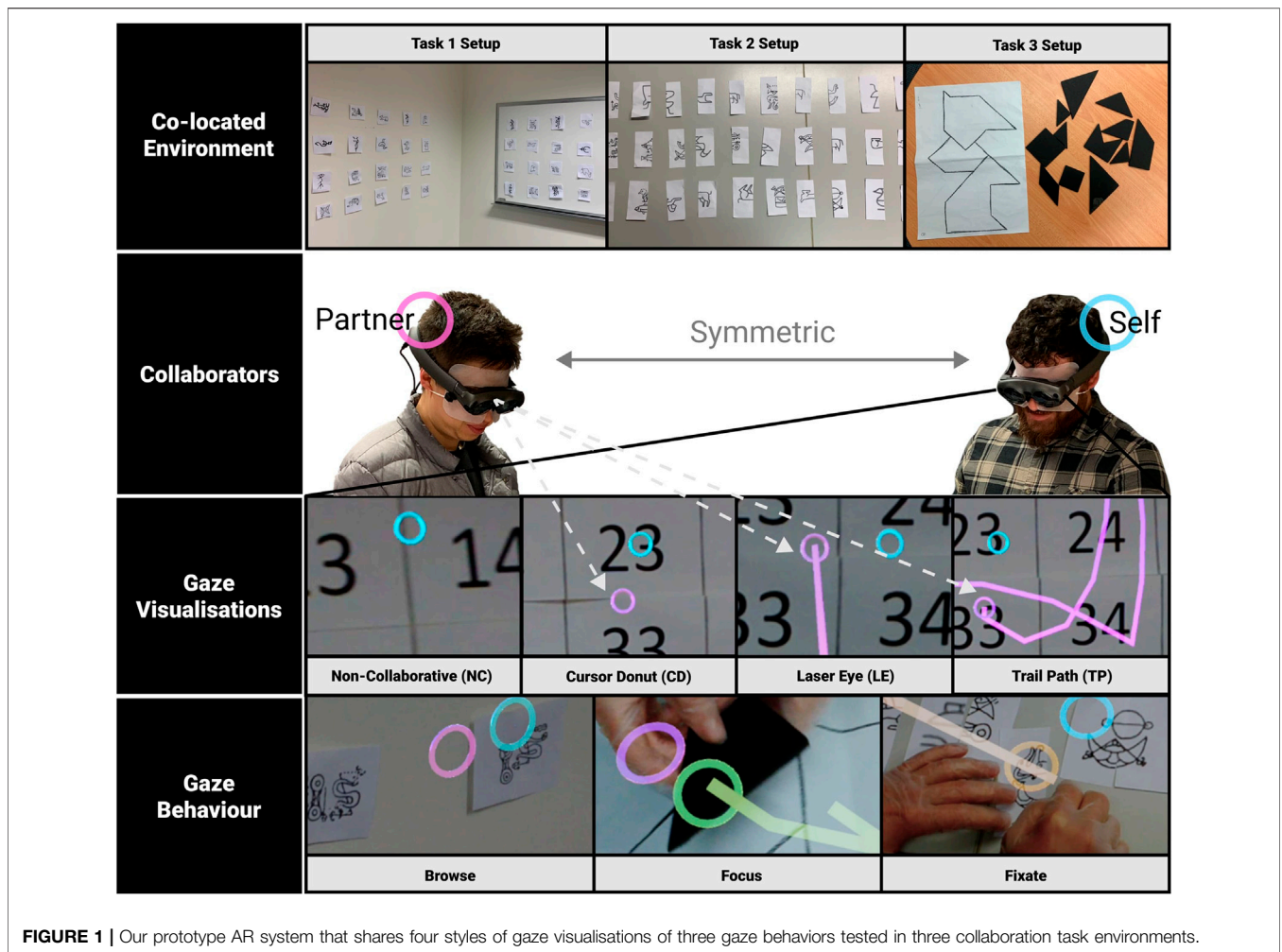
2 RELATED WORK

2.1 Gaze Behaviours in Collaboration

Gaze is considered one of the most valuable implicit visual communication cues in collaboration (Kraut et al., 2003) and can play multiple roles, including facilitating turn-taking, conveying cognitive activity, and expressing involvement, etc (Argyle and Cook, 1976). Eye movements such as fixations, saccades, and blinks are common gaze behaviours (Vickers, 1995) and for successful communication people typically exchange a range of different gaze signals. For instance, eye fixation can serve as a fast and precise pointer (Higuch et al., 2016), a method to confirm and clarify the object of interest (D'Angelo and Gergle, 2018), and a reference that simplifies linguistically complex objects with deictic references (e.g., "this" or "here") (Li et al., 2016) between collaborators. During joint gaze interaction, knowing where one's collaborator is looking helps and supports effortless mutual references which leads to easier collaboration (Tang and Fakourfar, 2017). Exchanging gaze can be more efficient than speech for the rapid communication of spatial information (Brennan et al., 2008) (Neider et al., 2010).

2.2 Visual Interfaces of Gaze Cues in Collaboration

Prior research has proposed multiple ways to visually convey gaze cues. Examples include a desktop gaze window of a remote partner's face (Kim et al., 2020b), a gaze cursor converted to a desktop display (Lee et al., 2016) (Lee et al., 2017), a trajectory gaze path on desktop monitors (D'Angelo and Gergle, 2016), a gaze spotlight on a wall display (Zhang et al., 2017), and a shared gaze through computer screen (D'Angelo and Begel, 2017). Those studies suggest that the design of gaze visualisations affects performance, coordination, searching behaviour, and perceived utility, although certain visualisations can overwhelm users by the excessive amount of visual information shown (D'Angelo and Gergle, 2018) (Zhang et al., 2017) (Li et al., 2016). For instance, a continuous gaze stream (D'Angelo and Gergle, 2016) or cumulative gaze heatmap (D'Angelo and Gergle, 2018) are both considered distracting, while with a gaze indicator on a 2D display users often need to constantly identify the gaze owner (Zhang et al., 2017).



Using reality-based virtual gaze interfaces may potentially solve the trade-off between the visibility of gaze indicators and the demand of cognition, by naturally simulating realistic gaze interaction in the collaborator’s immediate task space with an extended field of view. Besides, visual clusters of disruptive gaze information displayed during collaboration may also become less prominent. To investigate collaborative gaze systems that are ergonomic and immersive, researchers have explored the use of emerging technologies such as AR (Lee et al., 2017) (Li et al., 2019), VR (Piumsomboon et al., 2017) (Rahman et al., 2020), and MR (Kim et al., 2020a) (Bai et al., 2020).

However, very few studies have been conducted to investigate gaze visualisations and representations using AR/VR/MR compared to the traditional media. In AR applications, researchers have not yet considered using different forms of gaze behavioural feedback to signal the collaborator’s current status. Instead, gaze is represented as a simple real-time marker (cursor or crosshair) (Lee et al., 2017) (Gupta et al., 2016) or a ray (Li et al., 2019) (Bai et al., 2020). Gaze cues are also often depicted in a one-directional form that misses the immediate feedback for joint gaze interaction between collaborators (Wang et al., 2019) (Kim et al., 2020a). Designing joint gaze behavioural

visualisations in the appropriate form with different interfaces may have a significant impact on collaboration.

2.3 Visualised Gaze Cues Vs. Other Communication Cues in Co-Located Collaboration

A pair of collaborators who share the same basic roles and task knowledge [symmetric (Ens et al., 2019)] while accessing multiple communication resources (verbal and visual communication cues) at the same time (concurrent) in a co-located task space is defined by us as a Concurrent Symmetric Collaboration (CSC). For example, two people carrying out the same visual pattern-matching task while talking with each other and using the same shared interface is a CSC. The human cognitive system can execute multiple concurrent tasks while Threaded Cognition suggests that humans can perform more than one thing at the same time by maintaining a series of goal threads, although performance may suffer or improve from how resource allocation is optimised (Salvucci and Taatgen, 2008). For example, Brennan et al. (Brennan et al., 2008) have suggested that gaze-plus-voice performed slower than gaze alone in visual search tasks. This is because speaking incurred substantial coordination

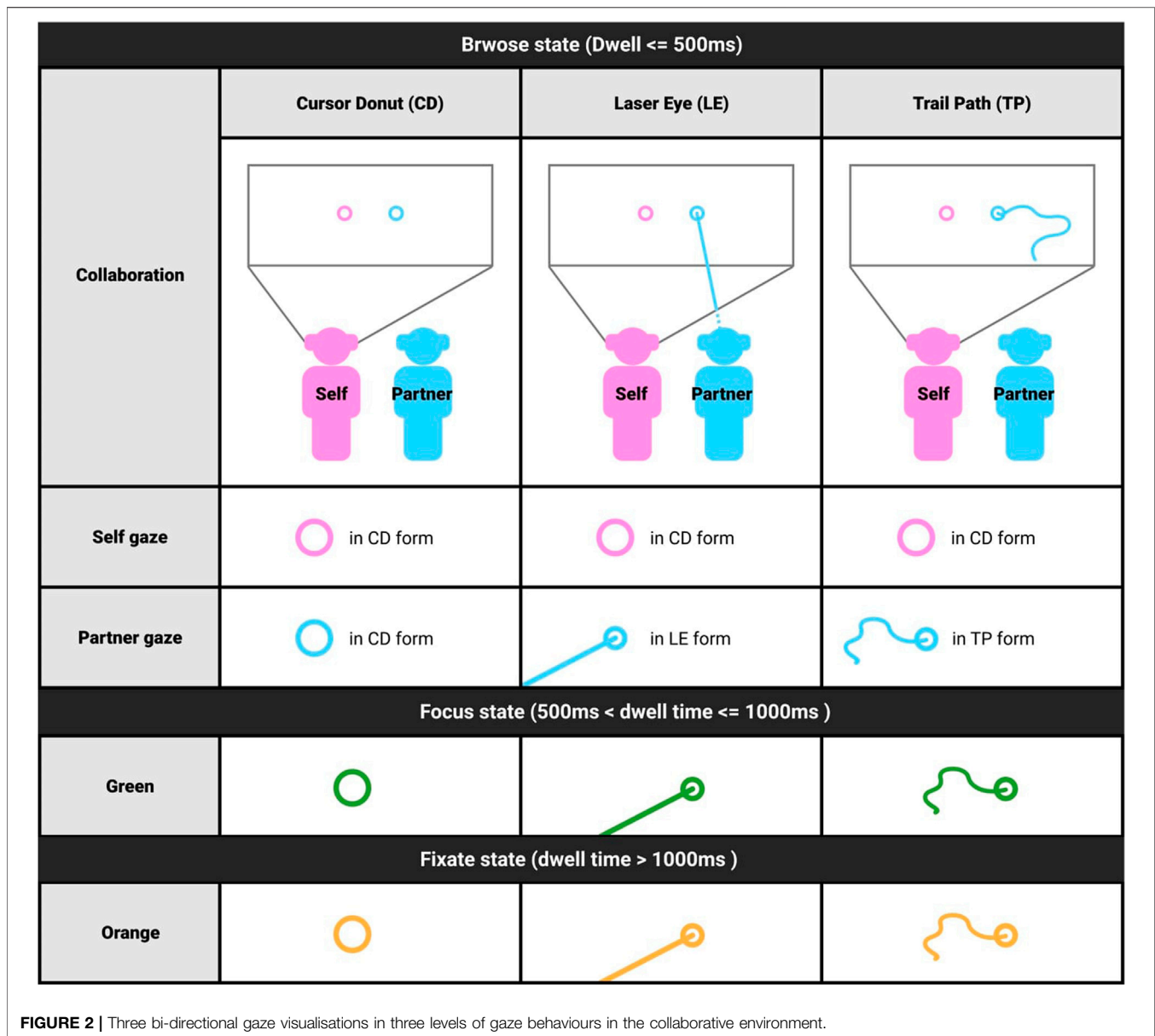


FIGURE 2 | Three bi-directional gaze visualisations in three levels of gaze behaviours in the collaborative environment.

costs and sharing gaze is more efficient than speech for the rapid communication of spatial information. Similarly, there is also a gaze-speech relationship between the gaze direction of two conversation partners (Vickers, 1995) and knowing what they are referring to is critical for task coordination during communication. As a result, in many face-to-face tasks in CSC where collaborators are able to access a number of communication resources (including speech, pointing, and gesturing), implicit communication cues such as gaze have yet to be fully optimised through virtual representations.

Additionally, in a face to face task, the orientation and direction of the collaborator’s body can affect the information that their partner can see. To achieve natural and consistent communication, understanding the perspective (Yang and Olson, 2002) (Tang and Fakourfar, 2017) of another user in a co-located environment is essential. Visual cues may be aligned less

accurately when viewing from a single perspective that warps spatial characteristics (Pan and Steed, 2016) (Kim et al., 2020b). Moreover, spatial elements can provide depth cues in the shared task space, enabling collaborators to gain a better understanding of the same visualised information from different perspectives (Jing et al., 2019) and improving communication.

The above listed challenges could possibly be overcome by conveying gaze cues as AR virtual representations. These could be dynamic representations that exhibit not only spatial elements of the complex physical task space but also bi-directional gaze behaviours among other available communication cues. Our research compares different AR gaze visualisations through co-located collaborative tasks across varied physical task space, task scale, and collaborator’s view perspective, attempting to fill an important gap in this research landscape.

TABLE 1 | Questions for the preliminary interview.

Q#	Questions
Q1	Self Introduction and the role you usually play during face-to-face collaboration?
Q2	What communication cues do you consider important during face-to-face collaboration?
Q3	If consider gaze cues important, what behaviour(s) should be visualised? How and why?
Q4	What form do you consider intuitive when used for your preferred visualisation(s)?
Q5	Would you be open to use AR visualisations (wearing HMD) to assist collaboration?

3 DESIGN AND IMPLEMENTATION

We investigate how different AR representations of gaze cues combining spatial and temporal factors affect face-to-face collaborative tasks in physical task space. Contrary to the prior user studies where one-directional gaze sharing is commonly used between collaborators (Li et al., 2019) (Erickson et al., 2020), our study explores mutual collaboration with equal visual communication cues shared between users. We design and implement three bi-directional collaborative (BDC) gaze visualisations (showing both participant's own gaze indicator and partner's simultaneously): Cursor Donut (CD), Laser Eye (LE), and Trail Path (TP). These BDC gaze visualisations can represent three levels of gaze behaviours: "browse," "focus" and "fixate" states (Figure 2). In the rest of this section we describe further details on the design and implementation of these gaze visualisations.

3.1 Preliminary Ideation, Prototyping and Usability Evaluation

Following the Star life cycle for human-computer interface development (Hartson and Hix, 1989), we first extracted user requirements and needs by conducting six preliminary interviews (questions in Table 1) *via* video call with six potential users ranging from academia (3 AR/VR engineering students) to industry (1 UX designer and 2 project managers). Our goal was to understand how gaze in general as a communication cue could be incorporated to help improve face-to-face collaboration experience from a real-life practical perspective. All of the potential users believed it would be helpful to use gaze visualisation to represent their partner's current focus as well as indicating joint awareness during collaboration. They felt that this could potentially "save time to search for attention and is more straight forward," make focus "easier to signal" and have the benefit that "feedback is instant." Participants also suggested several possible gaze representations they would like to experience for design ideation, such as a real-time spotlight of the current visit (the most common method), a bubble notification to alert previous visits with the name of the visitor (gaze identity), or a heatmap to show how many times an area has been visited (temporal factor) etc.

Based on the suggested gaze representations and previous gaze studies, an early concept prototype was developed and tested

between two pairs of collaborators in our lab. Each pair was given a symbol searching, memorising and matching task of three sessions using Magic Leap AR Head-Mounted Display (HMD) to test against no gaze visualisation (without AR HMD). For the first session of the task, participants were asked to match five duplicate symbols of Pictish and ancient Chinese characters from the pool of fifty scattered on an office desk and the wall above it. For the second section, participants were given a paper of ten symbols and were asked to memorise them in 2 min. They were then asked to collaboratively find all of the symbols in correct order as shown in the paper reference from the same pool. For the last session, they were instructed to complete a puzzle of twelve pieces together. Each session was timed and it averagely took around 10 min for the first session, 7 min for the second session, but over 15 min for the last session. The AR gaze representation was bi-directional in real-time cursor form. There were only two gaze behavioural visualisations demonstrated: a) "Browse state" (gaze dwell time <500 ms (Parisay et al., 2020)); b) "Fixate state" (gaze dwell time ≥500 ms) by changing the shape of respective single-ring cursor to a double-ring cursor with a larger radius. In a usability testing feedback, users mentioned that the constant change of "fixate" shape was distracting while Magic Leap HMD was uncomfortable to wear. Additionally, we also believe that the time to complete each session of the task should be around the same length (5 min ideally).

3.2 Gaze Visualisation Design

Based on the lessons learned from the preliminary feedback and current literature, we designed three bi-directional collaborative (BDC) visualisation styles (Figure 2) combining both spatial and temporal characteristics that we would further investigate in a user study. The designs were intended to explore the AR representation of joint gaze cues associated with other communication cues as opposed to simply optimising novel gaze visualisation designs in general.

As previously explained, to keep the visualised gaze cues symmetrical while representing joint gaze interaction, we decided to design gaze visualisations as a bi-directional collaborative (BDC). This means by visualising both user's own gaze and their partner's gaze, collaborators can access equal visual information as well as viewing joint gaze interaction *via* pair indicators. To visualise the user's own gaze, we decided to simply use a real-time cursor to reduce unnecessary information displayed to the user. We believe that this will minimise the distraction while still providing minimal cues to achieve symmetric gaze visualisation.

Subsequently, we also incorporated three levels of behavioural states into our design: 1) "Browse": active eye movements, shown in pink (for collaborator A) and blue (for collaborator B); 2) "Focus": when the gaze dwells on the same point for between 500 and 1000 ms, the gaze indicator will turn green; 3) "Fixate": where the gaze dwells on the same location for over 1000 ms, the gaze indicator will become orange. The idea of adding different duration of dwell time to represent three gaze behaviours aims at gaining understandings of how gaze focus exerts varied task intentions between collaborators. The dwell time threshold was taken from prior literature (Špakov and Miniotas, 2004) (Parisay et al., 2020) while using colours to signal the change of state instead

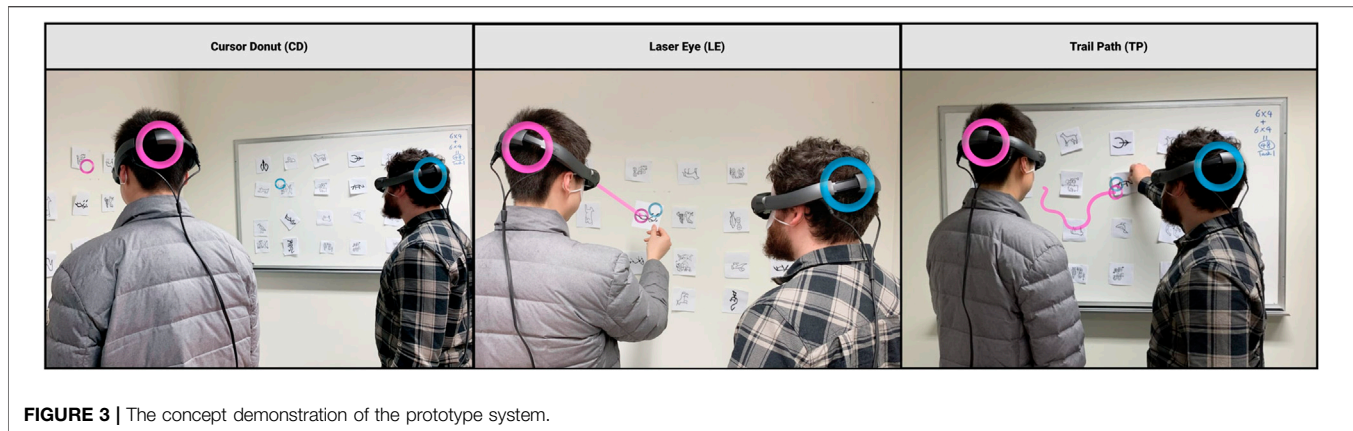


FIGURE 3 | The concept demonstration of the prototype system.

of shape was based on the pilot testing results to avoid distraction and mental load during collaboration. Lastly, we set the transparency (α) to 80% for all of the visualisations to avoid full occlusion of the target object.

To represent the partner's gaze we used three gaze visualisation techniques (**Figure 3** shows the concept): Cursor Donut (CD), Laser Eye (LE), and Trail Path (TP).

Cursor Donut (CD) represents the real-time gaze point that falls on the surface of target objects in the real world task space. The real-time marker is the most common representation of gaze cues for it contains fewer but critical gaze information in a less distracting visualisation form. This design is similar to an onscreen cursor pointer and is consistent with traditional gaze studies (Zhang et al., 2017).

Laser Eye (LE) extends the basic CD form with additional cues to improve spatial understandings of a user's gaze direction (Vickers, 1995). LE adds a virtual ray that starts from the collaborator's HMD (center of the lenses) and ends where the ray and a physical surface intersect. The design of a gaze ray has the benefit of identifying visual target (Erickson et al., 2020) yet symmetric physical tasks between real-life collaborators need to be evaluated.

Trail Path (TP) appears as a continuous trajectory path in motion incorporated with both spatial and temporal elements that shows the collaborator the current (in CD form) and past history (1500 ms; in line form) of the gaze point. The trajectory path of various visited locations is automatically turned off and only CD form is shown if users gaze dwell at one location for more than 1500 ms. Contrary to the previous study where the display threshold was set to 3000 ms (Zhang et al., 2017)(D'Angelo and Gergle, 2018), we decided to further reduce the threshold to half because behavioural colour states made available together with continuous path history have been considered as "too much" when presented with 1000, 2000, and 3000 ms during iterative testing.

3.3 Software and Hardware

The system prototype (concept demo in **Figure 3**) was built using the Magic Leap One¹, a see-through AR HMD with 30×40 field of view and a resolution of 1,440 by 1,600 pixels per eye, as well as

a built-in eye tracker². The software was developed using the Unity 3D game engine (v2019.2.15f1)³. We utilised the Magic Leap Toolkit (MLTK)⁴ to collect gaze data and build visualisations. The MLEyes API allows access to the 3D position of where the user is looking in world space and good eye calibration will lead to comfortable and accurate eye gaze suggested by Magic Leap. The offset can happen but is within 3–5 cm. The raw gaze fixation data was filtered and smoothed before converting to visualisations. ML Spatial Mapper⁵ was used to systematically generate the mesh of the task space. The collaborators' position and gaze point anchors were tracked and aligned using the Persistent Coordinate Frame⁶ and Spatial Alignment⁷ was used to ensure both co-located HMDs' tracking coordinate frames align accurately. All of the above synchronisation was achieved by using MLTK Transmission⁸ and a Wi-Fi network connection was provided *via* a mobile cellular Hotspot. All of the shared information was supported and maintained by Magic Leap's private cloud processing and storing center called "Shared World"⁹.

4 USER STUDY

The main goal of the user study was to explore how bi-directional AR gaze visualisations influence face-to-face concurrent symmetric collaboration (CSC). The research questions are listed below:

- Do AR bi-directional collaborative (BDC) gaze visualisations enhance face-to-face CSC?

²<https://developer.magicleap.com/en-us/learn/guides/design-eye-gaze>

³<https://unity.com/>

⁴<https://developer.magicleap.com/en-us/learn/guides/magic-leap-toolkit-overview>

⁵<https://developer.magicleap.com/en-us/learn/guides/meshing-in-unity>

⁶<https://developer.magicleap.com/en-us/learn/guides/persistent-coordinate-frames>

⁷<https://developer.magicleap.com/en-us/learn/guides/spatialalignment-mltk>

⁸<https://developer.magicleap.com/en-us/learn/guides/transmission-mltk>

⁹<https://developer.magicleap.com/en-us/learn/guides/shared-world-faq>

¹<https://www.magicleap.com/en-us/magic-leap-1>

- Does virtual gaze representation affect the utilisation of existing communication cues during collaboration?

4.1 Experimental Design

The experiment used a within-subject design with four conditions of visualisation styles: Cursor Donut (CD), Laser Eye (LE), Trail Path (TP), and Non-Collaborative (NC) as the baseline. NC showed only the user's own gaze using CD form without the partner's gaze interaction hence it was non-collaborative. We kept the user's own gaze in the baseline condition to balance out the distraction that the user's own gaze indicator may cause identified in the previous study (Lee et al., 2017), so the comparison between the conditions would be only affected by the visualisation of the collaborator's gaze. While an alternative baseline condition could be not wearing an AR HMD at all, we decided not to use such a condition as the baseline in order to avoid any bias generated based on the physical weight or discomfort of the HMD model as mentioned by the pilot users. As a result, we ensured the symmetric design was applied across all conditions except the BDC gaze visualisation techniques used for the partner's gaze in each condition. The order of conditions for each pair of participants was counter-balanced using a Balanced Latin Square method among all groups.

In each condition, participants were asked to collaboratively complete three tasks of completely independent setups, so each pair of participants needed to complete twelve (3 tasks \times 4 conditions) collaborative task sessions. The task knowledge levels for both collaborators were symmetric, meaning that they had the same amount of knowledge while performing all tasks under each condition, in contrast to asymmetric roles (Ens et al., 2019) where expert collaborators usually dominate the task coordination and communication.

Both qualitative and quantitative methods such as questionnaires, interviews, video analysis, completion time and error were included as the main measurements for this study.

4.2 Task Setup

To evaluate the research questions, we designed a set of concurrent collaborative tasks to reflect various aspects, including task property, the size of the task space, body orientation, body direction, general distance between collaborators and the target (Table 2). Figure 4 shows the Co-located Collaborative Task Definition Matrix (CCTDM) which is used to better explain how the tasks were designed to reflect various scenarios. In CCTDM, the task design space is summarised with two dimensions, the level of freedom allowed and the level of being systematic. We chose three tasks that fall between "free form" to "constrained" and "abstract" to "systematic". The first two tasks were visual searching and matching tasks that used three sets of pictographic symbols as target objects. The last task was a puzzle solving task where participants had to arrange 12 black-coloured Tangram pieces into a puzzle picture. Additionally, to encourage mutual communication between collaborators of different personalities and familiarity,

participants were asked to collaborate on the tasks together and to reach a unanimous agreement on each target symbol before proceeding to the next one.

4.2.1 Task 1: Memorisation and Searching

As shown in the left of Figure 5, this task was a combination of memorisation and visual searching. At the beginning of this task, both collaborators were given 1-minute to memorise six target symbols (two each from three categories of Pictographic symbols) followed by a 30-second cool-down period. The time and number of symbols were decided based on pilot testings. Subsequently, participants were guided to a corner of two adjacent walls, each about 2 m \times 3 m in size and with a 6 \times 4 grid of symbols, to search and locate the six target symbols that they memorised from a pool of forty-eight symbols on the walls. The pool was randomised for both orientation and position of the symbols in each trial. During the task, except for the 30-second cool-down where no communication was allowed, participants could freely choose their body orientations and directions while being allowed to use all communication cues including gaze, verbal, gesture, or hand pointing with each other.

4.2.2 Task 2: Searching and Matching

The second task (middle of Figure 5) consisted of visual searching and matching. Participants were exposed to a grid of symbols cut in half and were instructed to match four complete ones (eight halves) from a pool of forty-eight candidates (16 \times 3). The pool was also randomised for both orientation and position of the symbols in each trial. The participants stood side-by-side facing a 2 m \times 1 m desk of half symbols. Each participant was only able to hold one half at a time to match it with the other half which their collaborator picked. If the symbols each participant picked were not a match, they had to put them back in the original spots. Participants could rotate their body orientations and directions towards each other if needed, while standing on the same side of the table. As in the first task, participants had full access to all communication cues during the task.

4.2.3 Task 3: Puzzle Solving and Negotiating

The third task, as shown in the right of Figure 5, involved collaboratively negotiating and solving a Tangram-like puzzle. The participants stood facing each other across a 1 m \times 1 m table and were asked to place the puzzle pieces to fill up a shape outlined on a sheet of paper. There were twelve puzzle pieces to be placed with a time limit of 5 min. Same as the first two tasks, participants also had full access to all communication cues.

4.3 Participants

We recruited 24 participants with ages ranging from 21 to 39 years old ($M = 27.04$, $SD = 4.94$). Among the pairs of participants, 25% of the pairs had never met before while the rest were colleagues (42%) and friends (33%). The majority of the participants (71%) stated that they were familiar with AR in general while the rest had barely used one. Regarding the usage of gaze interfaces (HMD), 29% claimed to be experienced while 54% had very limited exposure (slight to none).

TABLE 2 | Task setup based on aspects including task property (the nature of the task), the size of the task space, participants' body orientations, participants' facing directions, and the general distance between collaborators and the target.

	Task 1	Task 2	Task 3
Task Property	Memorising + searching	Searching + matching	Puzzle-solving + negotiating
Task Space	Two adjacent wall surfaces (2 m × 3 m)	Desk surface (2 m × 1 m)	Desk surface (1 m × 1 m)
Body Orientation	Free form	Side by side	Face to face
Body Direction	Face straight	Face down	Face down
Distance from target	1.5 m	1 m	0.5 m

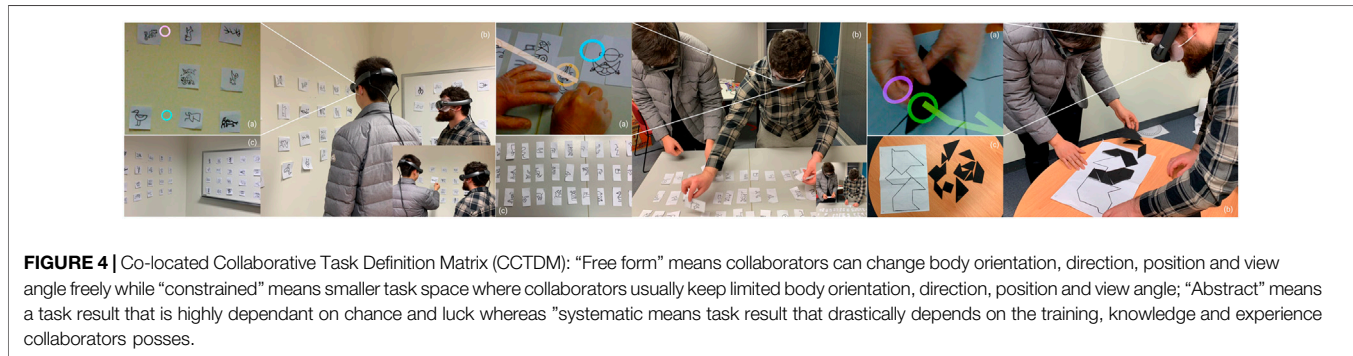


FIGURE 4 | Co-located Collaborative Task Definition Matrix (CCTDM): “Free form” means collaborators can change body orientation, direction, position and view angle freely while “constrained” means smaller task space where collaborators usually keep limited body orientation, direction, position and view angle; “Abstract” means a task result that is highly dependant on chance and luck whereas “systematic means task result that drastically depends on the training, knowledge and experience collaborators posses.

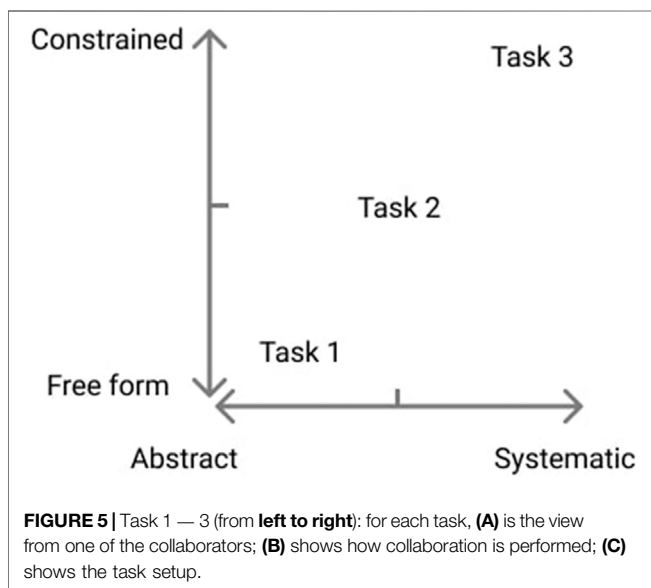


FIGURE 5 | Task 1 — 3 (from left to right): for each task, (A) is the view from one of the collaborators; (B) shows how collaboration is performed; (C) shows the task setup.

4.4 Experimental Procedure

A pair of participants were recruited for each experimental session. Prior to the study, participants were given an information sheet and consent form to sign, before being asked to fill up a demographic questionnaire. Then they were shown how to wear the Magic Leap One AR HMD and to calibrate the built-in eye-tracking system. A 5-min training session followed after explaining task descriptions. Participants used this time to get familiar with each visualisation and were asked if they were experiencing any sickness or visual discomfort.

During the experiment, participants finished Tasks 1 — 3 sequentially in each condition as tasks 1, 2, and 3 were completely

independent with different setups. However the order of conditions was counter-balanced between the groups. The tasks were designed to provide and include various configurations of use cases. While not treated as an independent variable, the tasks were used as a context during the interview as they were used solely for comparing between conditions (i.e., the visualisation styles). Since there was symmetric collaboration across all visualisations, the participants were required to wear the same HMD throughout the study to avoid any bias. Each session started with a given condition of visualisation style used for completing all three tasks. After finishing each condition, participants were given questionnaires including 7-point Likert scale rating items and open questions to answer and give feedback on their experience of collaboration in the condition. The experiment continued with the next condition until all four conditions were completed.

After finishing the experiment, we asked participants to provide rankings of all of the visualisation conditions based on their preference over five categories, followed by a semi-structured interview on the task collaboration experience. Every participant was compensated with a \$10 gift card to thank them for their participation.

The experiment took place in a lab room divided into three sections for the three tasks respectively and took 2 h per pair of participants on average to complete. The study was also video recorded on the participants’ HMDs during pair collaboration and was audio recorded during the post-experiment interviews.

4.5 Measures

Post-condition questionnaires presented at the end of each condition included ten 7-point Likert ratings on collaborative experiences three and the Subjective Mental Effort Question (SMEQ) to record participant’s mental effort.

TABLE 3 | Subjective rating questions (answered on a scale of 1: Strongly Disagree to 7: Strongly Agree).

Q#	Rating Statement
01	My intention is accurately represented
02	My partner's intention is accurately represented to me
03	It is better to understand my partner's focus
04	It is better for my partner to understand my focus
05	It is easy to observe my partner's attention
06	It is easy for my partner to observe my attention
07	I React to my partner frequently
08	My partner react to me frequently
09	This form of visualisation is effective
10	This form of visualisation is engaging

Post-study preference rankings regarding the sense of connect task assistance prediction of intention, “notice of focus” and overall preferred four of all conditions were collected upon finishing all conditions.

Semi-constructed post-study interviews intended to understand users experiences using each condition on each task. The interviews were audio recorded and transcribed.

Completion time was recorded as the time (in milliseconds) from when the collaborators were instructed to start the task until when they unanimously signal to finish. If technical issue happened or any participant called to stop, the time were subtracted.

Error referred to the incorrectly selected symbols or symbols with incorrect positions from the reference for Task 1; For Task 2, if two pieces of the half symbols did not come as a match, it was counted as an error; For Task 3, we did not include error calculation as there were more than one correct answer.

5 RESULTS

For non-parametric data (ratings and rankings) as well as parametric data that were not normally distributed (completion time and error), we applied a Friedman test ($\alpha = 0.05$, unless noted otherwise) followed by Wilcoxon Signed Rank (WSR) tests using the Bonferroni correction ($\alpha = 0.0083$) for post-hoc pairwise comparisons.

5.1 Subjective Rating Questionnaires

We collected ninety-six (4 conditions \times 24 participants) valid questionnaires that included rating items of collaborative experience (Table 3) and the Subjective Mental Effort Question (SMEQ) (Sauro and Dumas, 2009).

5.1.1 Collaborative Experiences

The results (Figure 6) drawn from the Friedman tests followed by pairwise WSR tests indicated that compared to the baseline NC (Q1: $M = 2.79$ $SD = 1.86$, Q2: $M = 2.62$ $SD = 1.86$), regardless of forms participants felt that with bi-directional collaborative (BDC) visualisations (Q1: $M = 4.12, 4.71, 4.46$, Q2: $M = 4.08, 4.62, 4.50$ for CD, LE, TP) their task intentions were significantly more accurately represented both to their partner [Q1:

$\chi^2(3) = 21.81, p < 0.0001$] and from their partner [Q2: $\chi^2(3) = 27.90, p < 0.0001$] during communication.

Focus-wise, LE ($M = 4.83$) was considered significantly better for understanding the partner's focus (Q3: $\chi^2(3) = 9.92, p < 0.05$) than NC while all three BDC visualisations ($M = 4.17, 4.71, 4.54$) were better (Q4: $\chi^2(3) = 18.90, p < 0.0001$) for the partner to understand participant's focus than NC ($M = 3.25$). Participants also believed that all three BDC visualisations (Q5: $M = 4.42, 5.00, 4.46$, Q6: $M = 4.17, 4.79, 4.38$ for CD, LE, TP) made it significantly easier to observe their partner's attention cue [Q5: $\chi^2(3) = 37.37, p < 0.0001$] as well as letting their partner observe the participant's attention cue [Q6: $\chi^2(3) = 36.00, p < 0.0001$] more easily, compared to the baseline NC (Q5: $M = 2.33$, Q6: $M = 2.42$).

In terms of reaction, compared to NC (Q7: $M = 2.96$, Q8: $M = 2.83$) participants felt that LE (Q7: $M = 4.33$) helped them react to their partners significantly more frequently [Q7: $\chi^2(3) = 17.14, p < 0.0001$] whereas they believed TP (Q8: $M = 3.83$) allowed their partners react to them more frequently [Q8: $\chi^2(3) = 14.95, p < 0.01$]. Moreover, participants agreed that the experience was significantly more effective [Q9: $\chi^2(3) = 9.06, p < 0.05$] with the aid of LE ($M = 4.58$) compared to NC ($M = 2.54$). In the meantime all three BDC visualisations ($M = 3.79, 4.58, 4.21$ for CD, LE, TP) were significantly more engaging [Q10: $\chi^2(3) = 24.73, p < 0.0001$] than NC ($M = 2.54$).

5.1.2 Subjective Mental Effort Question

There was no significant difference found between conditions in SMEQ although LE had the lowest mean value among all visualisations (Figure 7). However, One Sample Wilcoxon tests suggested that only LE had a significantly lower ($V = 24, p < 0.001$) rating below Rather hard to do (55) on a scale from 0 — 150, indicating that LE was considered less mentally demanding in assisting the task.

5.2 User Preferences

After performing all four conditions, participants ranked LE as the most preferred visualisation style for all five categories (Figure 8). A Chi-Square Goodness of Fit tests yielded significant differences against random choice for all five ranked items (Table 4). The results aligned with the descriptive statistics where LE showed strong preference (mean value) across all rated items among four proposed visualisation styles in the questionnaire analysis.

5.3 Interview Feedback

At the end of the study, we collected qualitative feedback on the usability of the preferred visualisation(s) for each task participants experienced. A total of 88% of the participants preferred LE for more free-formed task space (Task 1 and 2). They believed the 3D spatial characteristic improved peripheral visibility of the partner's gaze direction regardless of their current field of view hence it is easier to notice their partner's focus: “The addition of the laser helped me remain aware of my partner's focus when their focal point was out of view. I could follow the laser to the relevant point which was much more effective than the gaze cursor on its own.” (P22) and “I think laser helps more of the first two

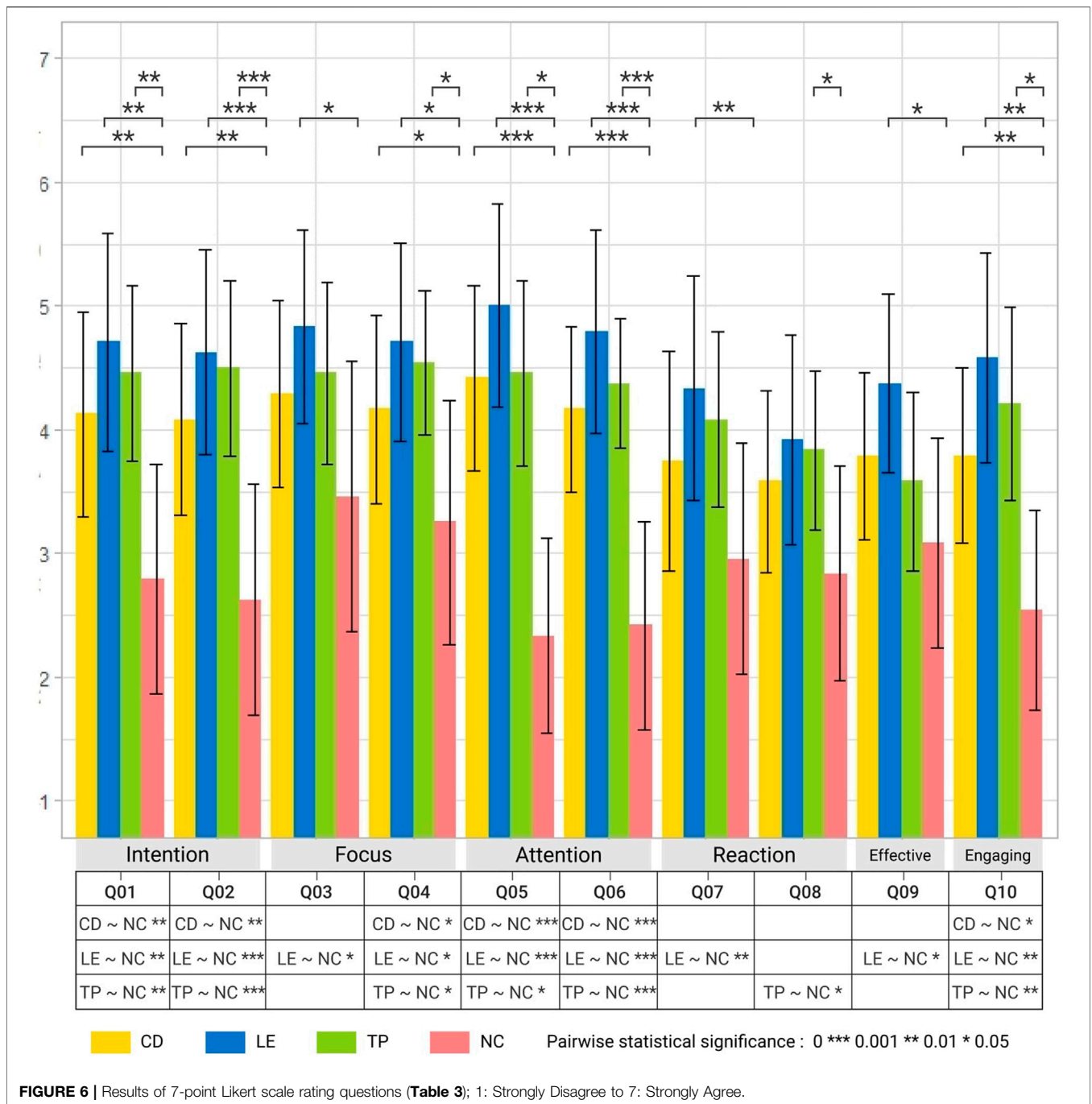


FIGURE 6 | Results of 7-point Likert scale rating questions (Table 3); 1: Strongly Disagree to 7: Strongly Agree.

tasks, because it is easier to follow my partner’s gaze direction without looking around for the partner.” (P18).

20% of the participants preferred TP for Task 2 because the temporal feature of this visualisation provides the history of the partner’s gaze focuses that did not require the collaborator’s real-time full attention: “I come back to check his gaze trail every now and then so I still see what he’s doing” (P6) and “the trail was useful for outlining things and indicating areas among many similar symbols” (P8). However for more constrained task space (Task 3) participants expressed negative opinions on

TP “The gaze trail is overlapping/interrupting my core focus and field of view from the puzzle desk.” (P24) and “it is more distracting than helping to communicate with my partner. Since we were allowed to speak and grab things physically from the desk” (P10).

On the contrary, participants disliked NC because of the amount of extra mental and physical effort they had to put into to achieve the same level of work in contrast to the BDC visualisations: “it felt like something was missing when I was unable to confirm what my partner was focusing on. This often

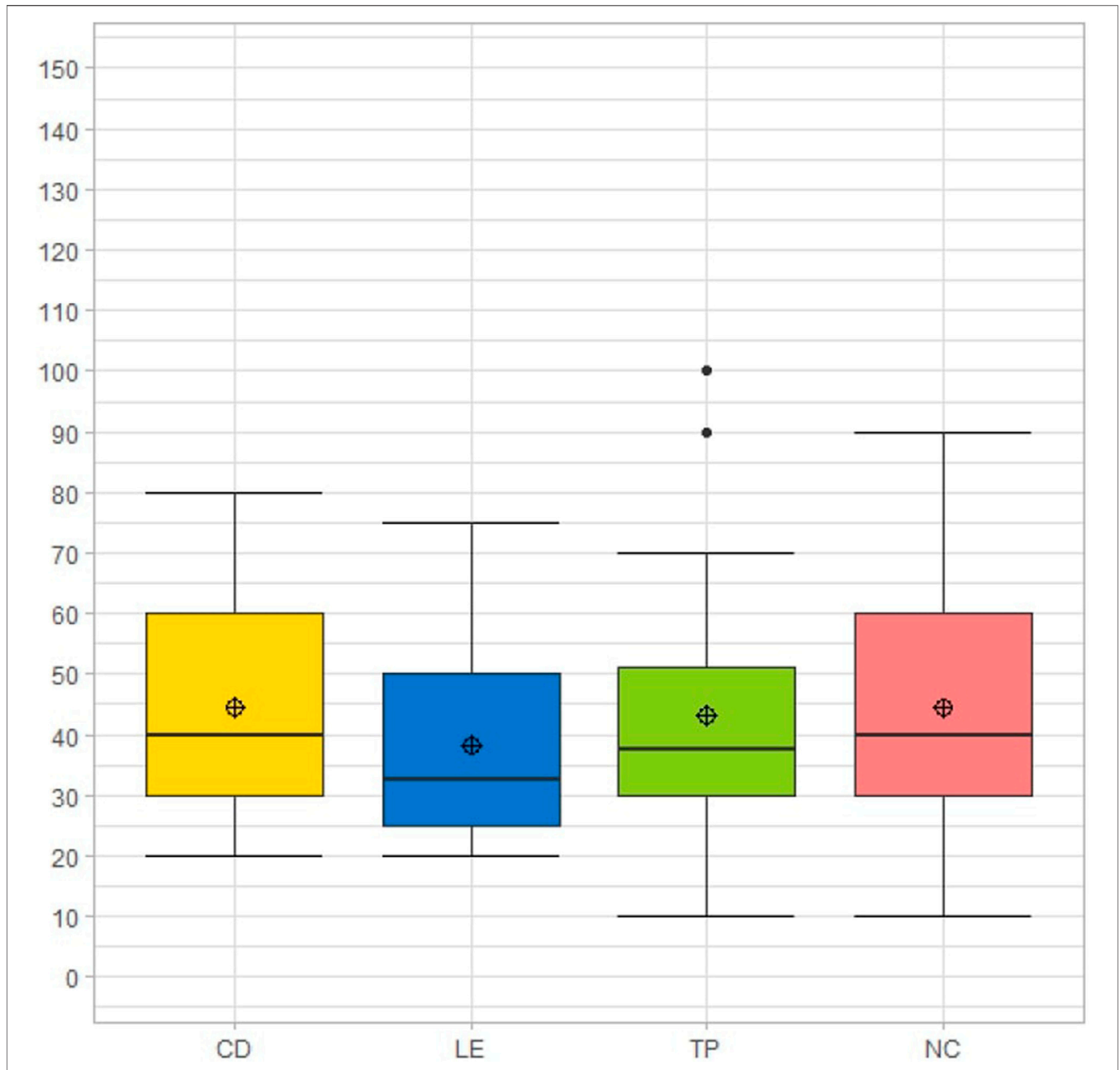


FIGURE 7 | Rating result for SMEQ (Sauro and Dumas, 2009) (from 0–150: 0 = not at all hard, 100 = very very hard to do, 115 and above = more than tremendously hard to do that no word can describe).

required usage of normal hand gestures or explanation to signify focus locations”(P24) and “I got used to not pointing to show my partner what I was looking at, but now I had to go back to pointing”(P10).

Additionally, when asked to rate the how much participants react to the change of behaviour states (“browse, focus, and fixate”) on a scale of 1 (Not at all) to 7 (Completely), the average rating for CD, LE and TP were 3.04, 3.2, and 2.86. The explanation given by the participants including “changes

are too subtle “too occupied on the task to pay attention to” and focused for a moment but did not actively react to it.”

5.4 Task Completion Time and Error Rate

As pointed out in “Experimental Design” section, due to the nature of Task 3 having fixed duration of 5 min, we only recorded the completion time and error for Task 1 and 2. However there is no significant result shown in **Figure 9** mainly because a) the collected samples are too small (12

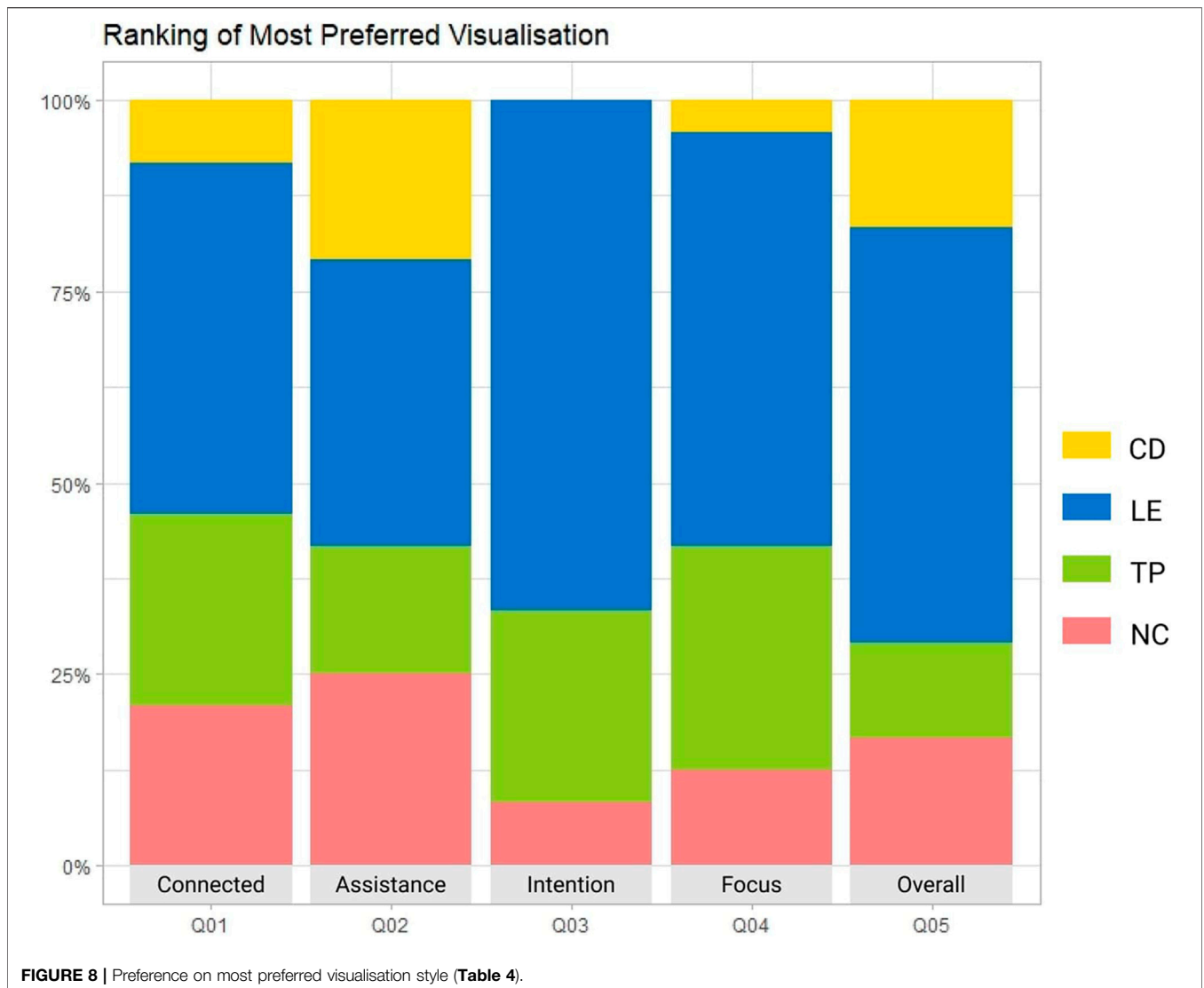


TABLE 4 | Ranking results of the most preferred visualisation style.

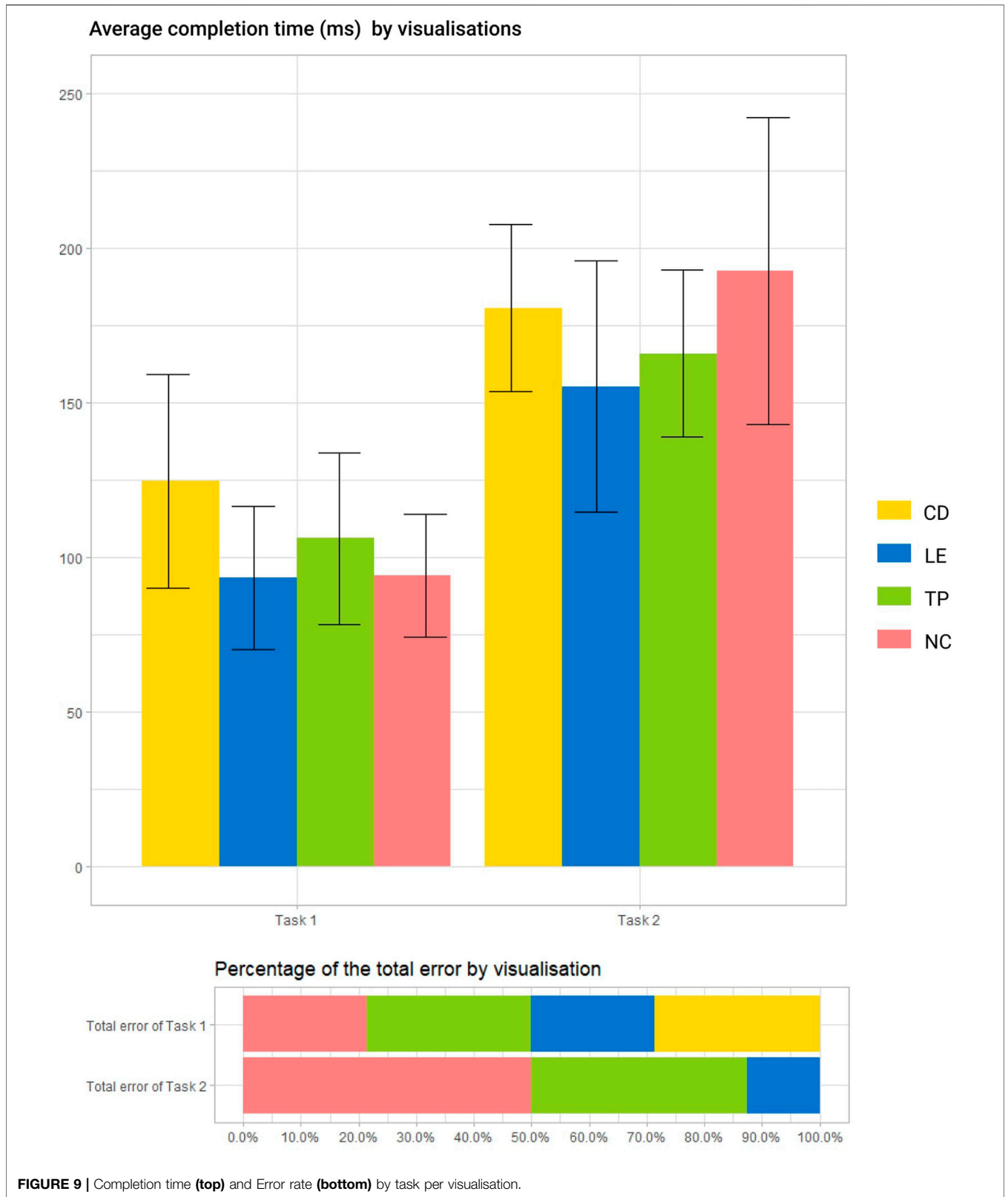
Questions	χ^2	df	p Value
Q01: Most connected	22.67	9	<0.01
Q02: Best task assistant	16.67	9	<0.05
Q03: Best predicts partner's intention	67.33	9	<0.001
Q04: Best notice partner's focus	88.33	9	<0.001
Q05: Overall most preferred	32.67	9	<0.001

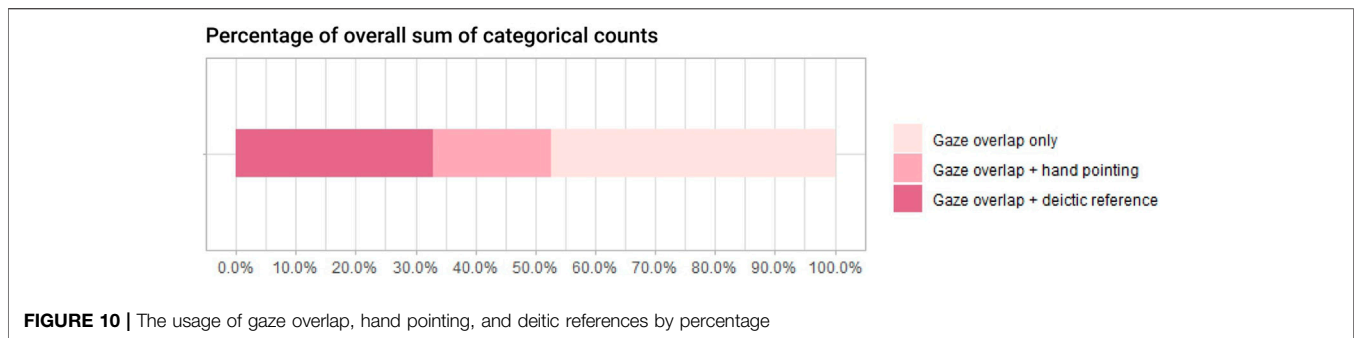
pairs of results in each task) and b) the relative “abstract” task nature might have lead to luck-driven results. As the error numbers are relatively low, we decided to use percent stacked bar to represent the contribution of each visualisation to Task 1 and 2. LE had the lowest average completion time (T1: 93 ms, T2: 155 ms) and second lowest error percentage (T1: 20%, T2: 12%), which may show its potential to outperform other visualisation styles in general.

5.5 Observation and User Experience

Through cross-checked video recordings, we analysed the percentage of collaborator hand pointing compared to deictic language references when their gaze visualisations overlapped. Regardless of visualisations, we collected the total counts across three tasks to provide a general understanding of how joint gaze was utilised compared to hand pointing and verbal communication when gaze visualisations were made explicitly available during face-to-face collaboration. As shown in **Figure 10**, the overall deictic reference counts were almost 35% more than hand pointing counts. Meanwhile, the use of a deictic reference combined with joint gaze overlap took around 30% of the total count while hand pointing under the same combination only 20%.

This also corresponds with the user feedback where 40% of the comments being negative in Task 3, participants still found that in Task 1 and 2 BDC gaze visualisations especially LE (54% as 1st preference) were helpful. Provided with extra spatially or





temporally incorporated visual awareness, participants could use gaze cues to coordinate task without distributing other communication resources (such as verbal or hand pointing) to achieve the same result in relatively relaxed space like Task 1 and 2. Whereas having both spatial and temporal factors in one visualisation such as TP in a constrained space like Task 3, the continuous information displayed may start to distract participants when they could distribute other communication cues easily and quickly.

The above findings favoured our postulation that gaze visualisations might optimise the allocation of communication cues but further investigation is required to validate the point. We think that in a relatively free-formed task space, reaching out a hand to point may be physically more demanding compared to gaze pointing. In addition, different viewing perspectives between collaborators may affect correct understating of pointing directions (Kim et al., 2020b), resulting in participants gradually learning that hand pointing might not be as effective and accurate.

Both quantitative and qualitative results have so far answered our first research question indicating that AR BDC gaze visualisations have significantly enhanced face to face CSC compared to the baseline condition.

6 DISCUSSION

Besides using bi-directional collaborative (BDC) gaze visualisations as attention-driven (progress checking and focus checking) or task coordination-driven (work allocation or prediction) cues mentioned in the interview subsection, many participants noted that BDC visualisations also worked as a referential pointer or affirmation checker: “it was very useful for confirming that we were looking at the same object.” (P22) as opposed to NC where participant felt “I didn’t like the idea of moving my hands around so much to point to the piece I’m talking about” (P21). We think that BDC gaze visualisations may have possibly encouraged participants to subconsciously pay more attention to each other’s task status using their gaze cues. As a result, adding BDC gaze cues may visually help Concurrent Symmetric Collaboration (CSC) where participants with the same level of knowledge towards the same task have to leverage multiple cognition layers to memorise, search, match, and coordinate concurrently. It is possible that AR visualisations

have eliminated certain cognitive threads (Salvucci and Taatgen, 2008) and freed up resources by offering another visual representation that otherwise requires verbal communication or hand pointing to compensate. Commented on as a strategy by a pair of collaborators, temporal visualisations were used to partition tasks where participants could periodically cross check each other’s focus history to achieve more effective collaboration. Whereas spatial visualisations were utilised to minimise proactive seeking or searching for the partner’s current focus.

When asked if they found any visualisations to be distracting, participants expressed neutral feelings yet four participants mentioned wanting to have control over displaying or hiding visualisations based on tasks. While some argued that their own gaze indicator could be distracting, half of the people expressed that they mentally “forget” or “shut down” visualisations (either their own or partner’s) when not needing them. This aligned with both SMEQ results as well as our postulation where we believed that people optimise their communication resources during CSC based on what cue is deemed useful (speech or hand pointing) or not useful (gaze) from the traditional point of view. Additionally, participants suggested that TP sometimes became distracting, especially when their view focus became more constrained (Task 3) while their communication resources become more direct (face to face close within reach). It also explained our Co-located Collaborative Task Definition Matrix (CCTDM) that the more constrained the task space is, the less helpful co-located gaze cues may become as other communication cues (such as hand pointing, grabbing, or gesturing) may easily compensate for one another. The more systematic the task property is the less helpful symmetric gaze cues become as the expert may simply dominate the collaboration hence joint awareness might be largely undermined. This finding exerts the need for further evaluation on combining gaze visualisations with the appropriate physical tasks and the performing roles.

Additionally, the above discussion might also explain why LE is more preferred. As it adds more spatial awareness to the periphery of the other collaborator’s field of view (FOV) via symmetric AR visualisations, it would have resulted in activating more joint gaze interaction without seeking or signaling information directly from and to the partner’s current body orientation for both Task 1 and 2 (less systematic task in a more relaxed space). However traditional desktop visualisations were unable to achieve similar embodied symmetric gaze collaboration. Moreover, participants did not seem to actively

respond to colour differentiation between behaviour states (browse to focus to fixate) across all tasks. This could be caused by the symmetrical nature of the task setup. Rather, they tended to collaboratively work on the task with equal amount of focus on each other's work. As a result, we suspect that gaze behaviour will be more prominent if the collaboration is between an expert and a novice (asymmetrical). The above findings may partially answer our last research question however further studies are needed to fully validate the claim.

The study was conducted using Magic Leap One HMD which has a relatively narrow viewing angle. With a wider field of view, the experience can possibly be improved as it allows more space for eyes to augment gaze visualisations while less head rotation is required to extend the view, hence potentially decreasing physical effort during task collaboration. While this paper mainly discusses the use between two collaborators, multi-user scenarios haven't been explored. With more people made available in the collaborative task space, more visualisations are also made available within the limited field of view. For visualisations containing temporal elements such as TP, the multi-user experience will most likely distract collaborators. A refined design of gaze color or gaze control coupled with further studies may alleviate this situation. For example, adding fade out effect on TP to lighten the line color is one way to solve the problem; Having individual user to display or hide TP can be another solution.

Although defined on the task scale of CCTDM as abstract against systematic, task difficulty may be perceived differently based on participants spatial cognitive skills. In our study it was considered as a random factor when we selected participants, however we noticed that in Task 3 some pairs were able to complete the entire puzzle within 5 min for more than one condition as opposed to some which barely complete a puzzle in a single condition. We plan to conduct future studies to fully understand the effect to further build the CCTDM concept as well.

Last but not the least, we believe that other than optimising communication cues in a co-located task space, compensating communication cues can also make impact in a practical setup. For instance, in a operation room especially during COVID-19 when medical professionals are equipped with protective gears, seeing gaze visualisations may compensate the hard-to-access verbal or hand pointing cues thus making the communicate less physically or mentally demanding. For users who are temporarily or permanently disabled, using gaze visualisations as a representation to replace hand or speech to indicate user intention, focus, or attention may also turn task collaboration more effective. There is a vast space for our research to grow and iterate, therefore further studies would be needed to solve more practical real-life problems.

6.1 Limitations

The results of our user study were interesting, but there were a number of limitations that should be addressed in the future. Due to the accuracy of the eye-trackers, the gaze visualisations had at most 3 cm of depth difference between the gaze visualisation projection on the virtual mesh and the actual physical objects.

Our solution was to select physical objects that were around 10 by 10 cm in size (symbols and Tangram pieces) to maintain the rough accuracy while informing the participants about the slight gap they might be seeing. There was also a slight latency (around 50–100 ms) in synchronisation when participants changed targets that shared a large physical distance, however the gap is probably too slight for the participants to mention. Nevertheless it is also worth mentioning that errors in precision and lag were considered within acceptable range based on previous literature (Erickson et al., 2020), and no participants have mentioned any accuracy or latency problems.

Another issue was multi-session instability: during collaboration, the Magic Leap spatial alignment and transmission was sometimes lost when starting a new session, and several participants became impatient while we reset the device. This was caused by the Magic Leap Toolkit being in beta stage, but the participants were well informed on that and was taken into consideration during the user study.

This user study was also conducted in a controlled lab environment with tasks that were designed for symmetric face to face collaboration. This limits us from making claims on similar collaborations in real world environments where various confounding factors may happen from time to time. E.g., a real world environment with a busier setup may confuse users from locating the target of interest or partner's focus without experiencing distraction. Occlusion within the physical task space might limit the view of gaze visualisations. Since the study took place at a university, students became our main source of participants. Our selection on the participants did not include a mental rotation tests, therefore as noted in CCTDM some tasks may be considered as "systematic" for the participants who have better spatial cognitive skills. Those limitations all require further investigation and evaluation in our future studies.

Finally, to ensure that the results were not biased we acknowledge that the task performance could have been affected by factors that are out of control (e.g., participants personal abilities or prior experience with memorisation, visual searching, and Tangram puzzle solving). Our intention in the future is to explore how AR visualisations can be combined with various task setups in hope of finding an optimised method to best support face to face collaboration.

7 CONCLUSION AND FUTURE WORK

In this research, we compared the effect of sharing AR bi-directional collaborative (BDC) gaze visualisations to non-collaborative (NC) visualisation in three co-located symmetric task collaborations. Combining both spatial and temporal factors, we designed and implemented a prototype system of three BDC gaze representations, including Cursor Donut (CD), Laser Eye (LE), and Trail Path(TP), that also represented three levels of gaze behaviours including browse, focus and fixate states. We subsequently evaluated the prototype with twenty-four participants completing three independent pair tasks. The rating results have suggested that BDC gaze visualisations are

significantly better at representing joint attention and user intention especially in a relatively free formed physical task space. LE which superimposes gaze direction from the participant's eyes to the surface of the target maximises peripheral vision of BDC visualisation without sharp body movements or reorientation, thus it was ranked as the most preferred visualisation style. LE is also considered significantly more effective as it encourages participants to actively react to each other's focus with a relatively low mental effort to achieve. However, different gaze behaviours represented through the change of colours did not show much benefit.

In addition, all BDC visualisations have been perceived as significantly more engaging in the co-located symmetric collaboration in contrast to NC visualisation. By offering another virtual representation that otherwise requires verbal descriptions or hand pointing to compensate, BDC gaze visualisations have encouraged participants to consciously use a combination of gaze pointing and deictic references to communicate and coordinate during face-to-face collaboration.

In the future, we plan to further investigate the use of AR gaze visualisations in asymmetric pair tasks where one person plays the expert role while the other as a novice to evaluate how knowledge level affects co-located collaboration. Studying the AR representation of interaction between shared gaze (e.g., highlight of gaze overlap) in concurrent multitasking co-located collaboration can also help us understand the design of gaze interface against visual distraction. Additionally, how

different interaction modalities of gaze behaviours (shape, colour, sound, or pattern) affect co-located collaboration could also be a valuable realm to explore.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of South Australia Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AJ contributed to the conception, design, development, conduct and report of the analysis of the study. KM contributed to the conduct of the study. GL and MB supervised the study as well as part of the report of the study. AJ wrote the draft of the manuscript. AJ, GL, and MB contributed to manuscript revision, read, and approved the submitted version.

REFERENCES

- Argyle, M., and Cook, M. (1976). *Gaze And Mutual Gaze*. New York: Cambridge University Press.
- Bai, H., Sasikumar, P., Yang, J., and Billinghamurst, M. (2020). A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 25 April, 2020-30 April, 2020, New York, NY, USA: Association for Computing Machinery), 1–13. doi:10.1145/3313831.3376550
- Baron-Cohen, S., Wheelwright, S., and Jolliffe, a. T. (1997). Is There a "Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome, *VISUAL COGNITION*, 4, 311–331. doi:10.1080/713756761
- Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., and Zelinsky, G. J. (2008). Coordinating Cognition: The Costs and Benefits of Shared Gaze during Collaborative Search. *Cognition* 106, 1465–1477. doi:10.1016/j.cognition.2007.05.012
- D'Angelo, S., and Begel, A. (2017). Improving Communication between Pair Programmers Using Shared Gaze Awareness. Conference on Human Factors in Computing Systems - Proceedings, 6 May, 2017-11 May, 2017, New York, NY, USA: Association for Computing Machinery, Vol. 2017-Janua, 6245–6255. doi:10.1145/3025453.3025573
- D'Angelo, S., and Gergle, D. (2018). An Eye for Design. Conference on Human Factors in Computing Systems - Proceedings, April, 2018- 26 April, 2018, New York, NY, USA: . Association for Computing Machinery), CHI '16, 1–12. doi:10.1145/3173574.3173923
- D'Angelo, S., and Gergle, D. (2016). Gazed and Confused. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 7 May, 2016-12 May, 2016, New York, NY, USA: Association for Computing Machinery), CHI '16, 2492–2496. doi:10.1145/2858036.2858499
- Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T., et al. (2019). Revisiting Collaboration through Mixed Reality: The Evolution of Groupware. *Int. J. Human-Computer Stud.* 131, 81–98. doi:10.1016/j.ijhcs.2019.05.011
- Erickson, A., Norouzi, N., Kim, K., Schubert, R., Jules, J., LaViola, J. J., et al. (2020). Sharing Gaze Rays for Visual Target Identification Tasks in Collaborative Augmented Reality. *J. Multimodal User Inter.* 14, 353–371. doi:10.1007/s12193-020-00330-2
- Gupta, K., Lee, G. A., and Billinghamurst, M. (2016). Do you See what I See? the Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE Trans. Vis. Comput. Graphics* 22, 2413–2422. doi:10.1109/TVCG.2016.2593778
- Hartson, H. R., and Hix, D. (1989). Human-Computer Interface Development: Concepts and Systems for its Management. *ACM Comput. Surv.* 21, 5–92. doi:10.1145/62029.62031
- Higuch, K., Yonetani, R., and Sato, Y. (2016). Can Eye Help You? Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 7 May, 2016- 12 May, 2016. New York, NY, USA: Association for Computing Machinery) CHI'16, 5180–5190. doi:10.1145/2858036.2858438
- Jing, A., Xiang, C., Kim, S., Billinghamurst, M., and Quigley, A. (2019). SnapChart: An Augmented Reality Analytics Toolkit to Enhance Interactivity in a Collaborative Environment. Proceedings - VRCAI 2019: 17th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry, 14 November, 2019- 16 November, 2019. New York, NY, USA: . Association for Computing Machinery) CHI'16, 0–1. doi:10.1145/3359997.3365725
- Kim, S., Jing, A., Park, H., Kim, S.-h., Lee, G., and Billinghamurst, M. (2020a). Use of Gaze and Hand Pointers in Mixed Reality Remote Collaboration. The 9th International Conference on Smart Media and Applications (Jeju, Republic of Korea: SMA), 1–6.
- Kim, S., Jing, A., Park, H., Lee, G. A., Huang, W., and Billinghamurst, M. (2020b). Hand-in-Air (HiA) and Hand-On-Target (HoT) Style Gesture Cues for Mixed Reality Collaboration. *IEEE Access* 8, 224145–224161. doi:10.1109/ACCESS.2020.3043783
- Kraut, R. E., Fussell, S. R., and Siegel, J. (2003). Visual Information as a Conversational Resource in Collaborative Physical Tasks. *Human-Computer Interaction* 18, 13–49. doi:10.1207/S15327051HCI1812(_)_210.1207/s15327051hci1812_2

- Kuhn, G., Tatler, B. W., and Cole, G. G. (2009). You Look where I Look! Effect of Gaze Cues on Overt and covert Attention in Misdirection. *Vis. Cogn.* 17, 925–944. doi:10.1080/13506280902826775
- Lee, G., Kim, S., Lee, Y., Dey, A., Piumsomboon, T., Norman, M., et al. (2017). Improving Collaboration in Augmented Video Conference Using Mutually Shared Gaze. *ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. Adelaide, Australia: The Eurographics Association), 197–204. doi:10.2312/egve.20171359
- Lee, Y., Masai, K., Kunze, K., Sugimoto, M., and Billinghamurst, M. (2016). A Remote Collaboration System with Empathy Glasses. Adjunct Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2016, (Yucatan, Mexico. Institute of Electrical and Electronics Engineers Inc.), 342–343. doi:10.1109/ISMAR-Adjunct.2016.0112
- Li, J., Manavalan, M. E., D'Angelo, S., and Gergle, D. (2016). Designing Shared Gaze Awareness for Remote Collaboration. Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW 26-Februar, 26 February, 2016- 2 March, 2016, 325–328. doi:10.1145/2818052.2869097
- Li, Y., Lu, F., Lages, W. S., and Bowman, D. (2019). Gaze Direction Visualization Techniques for Collaborative Wide-Area Model-free Augmented Reality. Symposium on Spatial User Interaction. 19 October, 2019- 20 October, 2019, New York, NY, USA: Association for Computing Machinery), 1–11. doi:10.1145/3357251.3357583
- Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., and Zelinsky, G. J. (2010). Coordinating Spatial Referencing Using Shared Gaze. *Psychon. Bull. Rev.* 17, 718–724. doi:10.3758/PBR.17.5.718
- Pan, Y., and Steed, A. (2016). Effects of 3D Perspective on Head Gaze Estimation with a Multiview Autostereoscopic Display. *Int. J. Human-Computer Stud.* 86, 138–148. doi:10.1016/j.ijhcs.2015.10.004
- Parisay, M., Poullis, C., and Kersten-Oertel, M. (2020). FELIX: Fixation-Based Eye Fatigue Load index a Multi-Factor Measure for Gaze-Based Interactions. International Conference on Human System Interaction. 6-8 June 2020, Tokyo, Japan, IEEE, 74–81. doi:10.1109/HSI49210.2020.9142677
- Piumsomboon, T., Dey, A., Ens, B., Lee, G., and Billinghamurst, M. (2017). [POSTER] CoVAR: Mixed-Platform Remote Collaborative Augmented and Virtual Realities System with Shared Collaboration Cues. Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, 9-13 Oct. 2017, Nantes, France, . IEEE, 218–219. doi:10.1109/ISMAR-Adjunct.2017.72
- Rahman, Y., Asish, S. M., Fisher, N. P., Bruce, E. C., Kulshreshtha, A. K., and Borst, C. W. (2020). Exploring Eye Gaze Visualization Techniques for Identifying Distracted Students in Educational VR. 22-26 March 2020, Atlanta, GA, USA, . Institute of Electrical and Electronics Engineers, 868–877. doi:10.1109/vr46266.2020.00009
- Salvucci, D. D., and Taatgen, N. A. (2008). Threaded Cognition: An Integrated Theory of Concurrent Multitasking. *Psychol. Rev.* 115, 101–130. doi:10.1037/0033-295X.115.1.101
- Sauro, J., and Dumas, J. S. (2009). Comparison of Three One-Question, Post-Task Usability Questionnaires. Proceedings of the 27th international conference on Human factors in computing systems - CHI 09. 4 April, 2009- 9 April, 2009, New York, USA: ACM Press), 1599–1608. doi:10.1145/1518701.1518946
- Špakov, O., and Miniotos, D. (2004). On-line Adjustment of Dwell Time for Target Selection by Gaze. *ACM Int. Conf. Proceeding Ser.* 82, 203–206. doi:10.1145/1028014.1028045
- Tang, A., and Fakourfar, O. (2017). Watching 360° Videos Together. Conference on Human Factors in Computing Systems - Proceedings, 6 May, 2017- 11 May, 2017, Vol. 2017. New York, NY, USA: Association for Computing Machinery), 4501–4506. doi:10.1145/3025453.3025519
- Vickers, J. N. (1995). Gaze Control in Basketball Foul Shooting. *Stud. Vis. Inf. Process.* 6, 527–541. doi:10.1016/S0926-907X(05)80044-3
- Wang, P., Zhang, S., Bai, X., Billinghamurst, M., He, W., Wang, S., et al. (2019). Head Pointer or Eye Gaze: Which Helps More in MR Remote Collaboration? 26th IEEE Conference on Virtual Reality and 3D User Interfaces, 23-27 March 2019, Osaka, Japan: IEEE, 1219–1220. doi:10.1109/VR.2019.8798024
- Wang, Z., David, P., Srivastava, J., Powers, S., Brady, C., D'Angelo, J., et al. (2012). Behavioral Performance and Visual Attention in Communication Multitasking: A Comparison between Instant Messaging and Online Voice Chat. *Comput. Hum. Behav.* 28, 968–975. doi:10.1016/j.chb.2011.12.018
- Whittaker, S., and O'Connell, B. (1997). *The Role of Vision in Face-To-Face and Mediated Communication*.
- Yang, H., and Olson, G. M. (2002). Exploring Collaborative Navigation: The Effect of Perspectives on Group Performance. Proceedings of the 4th International Conference on Collaborative Virtual Environments, 30 September, 2002- 2 October, 2002, New York, NY, USA: Association for Computing Machinery, 135–142. doi:10.1145/571878.571899
- Zhang, Y., Chong, M. K., Müller, J., Bulling, A., and Gellersen, H. (2015). Eye Tracking for Public Displays in the Wild. *Pers Ubiquit Comput.* 19, 967–981. doi:10.1007/s00779-015-0866-8
- Zhang, Y., Müller, J., Chong, M. K., Bulling, A., and Gellersen, H. (2014). GazeHorizon. UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 13 September, 2014- 17 September, 2014, New York, NY, USA: Association for Computing Machinery, 559–563. doi:10.1145/2632048.2636071
- Zhang, Y., Pfeuffer, K., Chong, M. K., Alexander, J., Bulling, A., and Gellersen, H. (2017). Look Together: Using Gaze for Assisting Co-located Collaborative Search. *Pers Ubiquit Comput.* 21, 173–186. doi:10.1007/s00779-016-0969-x

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

Copyright © 2021 Jing, May, Lee and Billinghamurst. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.