Riccardo Bovo, Daniele Giunchi, Enrico Costanza, Anthony Steed, Thomas Heinis (2022): Shall I describe it or shall I move closer? Verbal references and locomotion in VR collaborative search tasks.. In: Proceedings of the 20th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies - Exploratory Papers, Reports of the European Society for Socially Embedded Technologies (ISSN 2510-2591), DOI: 10.48340/ecscw2022_ep02

Shall I describe it or shall I move closer? Verbal references and locomotion in VR collaborative search tasks.

Riccardo Bovo, Daniele Giunchi, Enrico Costanza, Anthony Steed, Thomas Heinis
Imperial College London, University College London
rb1619@ic.ac.uk, d.giunchi@ucl.ac.uk

Abstract. Research in pointing-based communication within immersive collaborative virtual environments (ICVE) remains a compelling area of study. Previous studies explored techniques to improve accuracy and reduce errors when hand-pointing from a distance. In this study, we explore how users adapt their behaviour to cope with the lack of accuracy during pointing. In an ICVE where users can move (i.e., locomotion) when faced with a lack of laser pointers, pointing inaccuracy can be avoided by getting closer to the object of interest. Alternatively, collaborators can enrich the utterances with details to compensate for the lack of pointing precision. Inspired by previous CSCW remote desktop collaboration, we measure visual coordination, the implicitness of deixis' utterances and the amount of locomotion. We design an experiment that compares the effects of the presence/absence of laser pointers across hard/easy-to-describe referents. Results show that when users face pointing inaccuracy, they prefer to move closer to the referent rather than enrich the verbal reference.

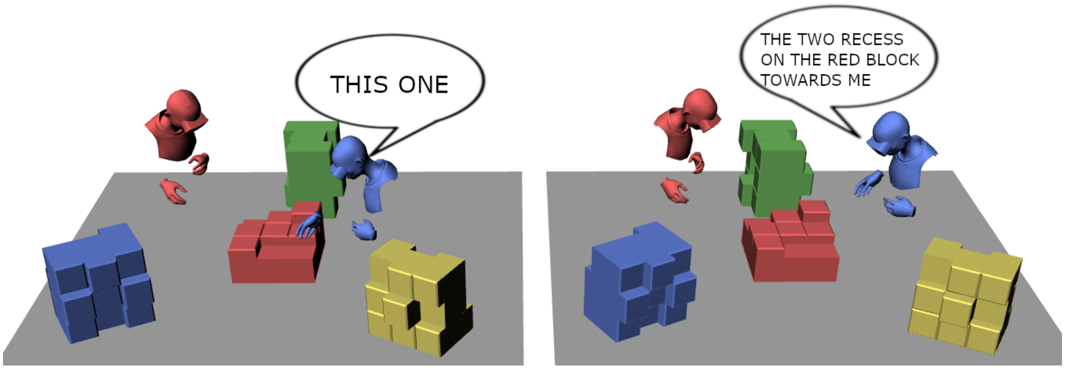


Figure 1: On the left an example of implicit verbal reference aided by a hand pointing action at a close distance from the referent. On the right, the equivalent reference is aided by a more detailed verbal description of the referent but lacks the hand pointing action from a close distance.

1 Introduction

Immersive collaborative virtual environments (ICVE) with user embodiment (i.e., avatars) support collaboration by providing a shared setting where collaborators have a sense of each other's presence (Benford et al., 1995). In ICVEs, the user's embodied hands behaviour is a non-verbal cue that integrates verbal communication during collaboration (Hindmarsh et al., 1998). For example, users can point to a referent during an utterance to trigger mutual orientation and visual coordination Moore et al. (2007). Hand pointing in conjunction with verbal, spatial references is called deictic pointing. Previous studies explore deictic pointing with distant targets (from fixed distances), measuring the accuracy of different hand pointing supports (Mayer et al., 2018, 2020; Wong and Gutwin, 2014, 2010). Outcomes from previous studies highlight how the degree of precision needed for the pointing gestures depends on how complex it is to describe the referent using utterances (Wong and Gutwin, 2014, 2010). However, in modern ICVE, users might get as close as needed to the referent and adapt to the accuracy required to perform the pointing gesture. Therefore, when faced with a lack of accuracy, will users spend time adjusting their distance from the target or overcome the difficulties of describing the target?

While in a physical environment, it is not always possible to move closer to an object of interest, in an immersive ICVE, this is not a problem as there are no physical barriers. In such scenarios, users can avoid inaccurate distance pointing by moving closer to the referent. However, the movement has a temporal cost: the time required to move closer to the point of interest. As Wong and Gutwin (2014) highlight, another approach consists of users enriching their verbal references with enough details to compensate for the pointing gesture's lack of precision. On the other hand, such a verbal supplement comes with a higher temporal and cognitive cost for both the performer and the reference recipient Wong and Gutwin (2014); D'Angelo and Begel (2017). Previous studies define pointing accuracy as a function of distance from the referent (Mayer et al., 2020; Wong and Gutwin, 2010). Accurate pointing can be performed from a far distance with a laser pointer for support or performed without a close distance to the referent. Inaccurate pointing consists of users who do not use/have laser pointers from a distance and choose to compensate with explicit verbal references.

The research community established the importance of laser pointers to achieve accurate pointing, but like any other tool or metaphor of interaction, laser pointers can be included or not in an ICVE. Some reasons for not including pointers can be the following: data visualisation issues such as a hidden or occluding cursor (especially if the informative area is dense), many users with cursors, and noise-induced by body jittering in high-density information areas Batmaz and Stuerzlinger (2019).

This study explores the trade-off between using locomotion to approach the referent or the alternative use of explicit verbal references to deal with the lack of pointing accuracy. We look at how this trade-off varies across conditions of lack/availability of laser pointers and conditions related to how complex/easy it is to describe the various referent in the scene (Figure 1). We explore such trade-offs in the context of visual search tasks, which are recognised as a proxy for many other tasks performed synchronously by pairs of participants in ICVE's Prilla (2019).

We run an experiment with 20 participants quantifying implicit/explicit references, locomotion and, in addition, visual coordination, which is highly correlated to the quality of pointing-based communication (Schneider and Pea, 2013). We use two datasets of different complexity representing two levels of difficulty in describing the referent: a simple puzzle and a very detailed 3D satellite map. In the simple 3D puzzle, pieces can be described by colours or labels, while on the map, places need to be referenced via 2D coordinates, which requires a greater cognitive effort. Inspired by previous CSCW work D'Angelo and Begel (2017) we measure the number of implicit/explicit deixis and the number of successful/unsuccessful deixis. Moreover, we measure users' movement in the space and task performance (task score and completion time).

The data collected shows statistically significant differences in locomotion performed when distance-pointing support is unavailable. Both data and observations confirm that when users lack support for distance pointing, they prefer to move closer to the referent to perform accurate pointing gestures rather than formulate a more complex verbal reference. We can see this change no matter the complexity of the task. The data collected also shows a statistically significant increase in visual coordination when laser pointers are available, which confirms previous work Moore et al. (2007).

Our results enable designers to understand how different elements (embodiment, locomotion, laser pointers) available in immersive ICVE impact pointing-based communication during a generic collaborative visual search task. Thus, our work can contribute to a more proficient interaction by outlining design implications. Presence of locomotion and the freedom for the user to move throughout the whole environment remove the need for distance pointing support.

In this way, an efficient locomotion system increases the rates of proximal pointing instead of promoting a cursor for distal pointing. However, laser pointers support may need to be considered if the collaborative task requires high visual coordination. Our study thus helps to make informed choices when designing an ICVE.

2 Related work

Pointing-based communication is ubiquitous in collaborative work. Within physically co-located scenarios, a pair of collaborators may use their hands and voice to engage in pointing-based communication. For example, indicating an object of interest by pointing hands towards it during an utterance is a common interaction called deictic pointing or deixis. During deixis, the interlocutor (i.e., recipient of the deixis) has to mentally project the collaborator's hand directly onto the observed scene to understand the referent of the deixis (i.e., understand the target object) (Higuch et al., 2016; Pfeiffer et al., 2008; Wong and Gutwin, 2014).

Pointing-based communication, however, can also be supported by laser pointers. A pointer's spotlight projected onto the observed scene allows identifying the referent unambiguously (Hindmarsh et al., 1998). Additionally, it facilitates the interpretation of the pointing gesture by removing the cognitive effort of projecting the hand/head directly onto the observed scene. Using a laser pointer might avoid any incorrect mental projection or ambiguous unclear projection results. Essentially pointers increase the awareness, during deixis, of a collaborator's visual focus (Piumsomboon et al., 2017).

Pointing-based communication is possible in co-located scenarios and remote scenarios thanks to either embodiment (i.e., avatars) and enhanced behaviour (i.e., pointers). There are several examples of remote collaboration scenarios in which pointing-based communication is possible, to mention a few: remote pair programming (D'Angelo and Begel, 2017), support of local workers by remote experts (Bai et al., 2020), remote collaboration in immersive VR environments (Moore et al., 2007).

2.1 Pointing-based communication in remote desktop collaborations

Pointing based communication can occur as long as collaborators have the means to point towards an object of interest while also communicating verbally. For example, several studies investigate pointing-based communication using gaze pointers (i.e., enhanced behaviour of eyes) in the context of 2D desktop remote collaboration (Villamor and Rodrigo, 2018; Jermann et al., 2011; Nüssli, 2011; Pietinen et al., 2008).

These studies show how visual aids based on the eye-tracked behaviour of collaborators (i.e., gaze-pointers) increase mutual awareness of visual focus, higher visual coordination and better collaboration quality. Schneider and Pea (2013) explore how depicting gaze in a remote desktop collaboration of two users performing a visual task increases visual coordination and enhances visual collaboration quality. When visual aids, such as pointers, are used, collaborators look at the same objects at the same time more often than without visual aids. Additionally, such increased visual coordination seems to aid communication about the visual context. For example D'Angelo and Begel (2017) explore visual aids (based on real-time eye-tracked behaviour) and prove that such visual aids improve communication by reducing the number of explicit utterances during deixis.

However, findings from the 2D desktop environment remote programming and visual analysis do not generalize to the immersive VR environments because the reviewed scenarios lack embodiment and locomotion (both elements present in state-of-the-art immersive VR collaboration environments). Embodiment, especially hand representation and hand real-time tracking behaviour, is the natural behaviour used in deictic pointing. However, in 2D desktop environments, the gaze is used as an input for pointing. While gaze can be thought of as coinciding with visual attention, it is a behaviour that is less deliberate and thus less controllable than the behaviour of hands. A second significant difference is related to fragmentation (Wong and Gutwin, 2014; Hindmarsh et al., 1998), or in other words, the fact that large parts of the environment in VR are not visible to the users, unlike the 2D desktop screen is. Fragmentation impacts pointing-based communication because the pair of collaborators may not be seeing the same subset of the 3D environment during deixis. They may thus not be able to see the collaborators' embodiment or the pointing visual aid. Moore et al. (2007) highlights how the observability of embodied activity and the projectability of gestures are essential aspects of pointing-based communication. While 2D desktop remote programming work may inspire metrics such as visual coordination and implicitness/explicitness of deixis utterances, their results are not necessarily generalizable to immersive VR collaboration.

2.2 Pointing-based communication in ICVE

Finally, ICVE offers the same degree of embodiment of mixed reality scenarios. Real-time tracked behaviour of hands/head allows natural pointing behaviour and natural exploration of the scene via head movements and locomotion. Several immersive VR studies explore the accuracy of hand pointing gestures. Mayer et al. (2018) propose adaptations to hand pointing in immersive VR that enhance the precision and accuracy of the pointers representations through spatial distortion. Mayer et al. (2020), in a similar way to Sousa et al. (2019) explores the approaches to improve precision by warping gestures to adjust pointing to the target.

However, while these recent studies aim to improve hand pointing accuracy, they do not evaluate the effect that pointers have on collaboration focusing only on the quantification of the pointing accuracy. All these works measure the accuracy of pointing from fixed distances, avoiding any form of locomotion within the scene. Our work aims to fill this gap, introducing specific tasks where we require the participants to move freely in the scene. An additional study from Bai et al. (2020) proposes a remote collaboration system that introduces an asymmetric interaction between a VR user and an AR user sharing a live 3D panorama of their surroundings. Differently from this study, our VR system provides both symmetric interaction and interface, and we focus on measuring the impact of locomotion on pointing-based communication.

2.3 How users compensate for inaccuracies during distance pointing

Previous studies explore techniques to improve accuracy and reduce errors when hand-pointing during pointing-based communication in immersive collaborative virtual environments (CVE). However, in a CVE in which users can move (i.e., locomotion), distance pointing (and its negative consequences) can be easily avoided by users' choice of increasing proximity to the referent. Additionally, a user could choose to compensate for imprecise distance pointing by enriching (adding details) to a verbal reference during a pointing gesture.

In an immersive CVE with embodiment and locomotion, we compare the presence and absence of pointers to understand if and how users compensate to avoid pointing errors and lack of precision. We also use several quantitative measures to understand how behaviour changes impact the quality of pointing-based communication. Inspired by previous CSCW remote desktop collaboration, we identify three easily quantifiable metrics: visual coordination, the implicitness of deixis' utterances, and references' success. Such metrics represent the quality of pointing-based communication during a collaborative task.

Previous literature allows us to define accurate pointing (both from the points of view of the producer and observer) and inaccurate pointing. Pointing gestures can be either proximal or distal Schmidt (1999). When indicating proximal referents, the producer of a pointing gesture can touch the target, and observers can identify targets with confidence Bangerter and Oppenheimer (2006). Therefore, consider proximal pointing is considered accurate as there is no room for misinterpretation.

With distal pointing, the observer needs instead to extrapolate the vector direction defined by the pointer's posture Bangerter and Oppenheimer (2006); Batmaz and Stuerzlinger (2019). However, previous studies have found that using a cursor improves mid-air pointing precision thanks to visual feedback and removes the need to extrapolate the direction of the pointing gesture again thanks to the visual depiction of the cursor Mayer et al. (2018). Therefore, we consider distal pointing with the cursor accurate as there is no room for misinterpretation, while we define distal pointing without the cursor as inaccurate.

While previous works offer several methods to improve the accuracy of pointing via machine learning models in our study, we explore how users deal with the lack of accuracy in an ecological context, in particular, related to visual analysis tasks.

3 Study Design

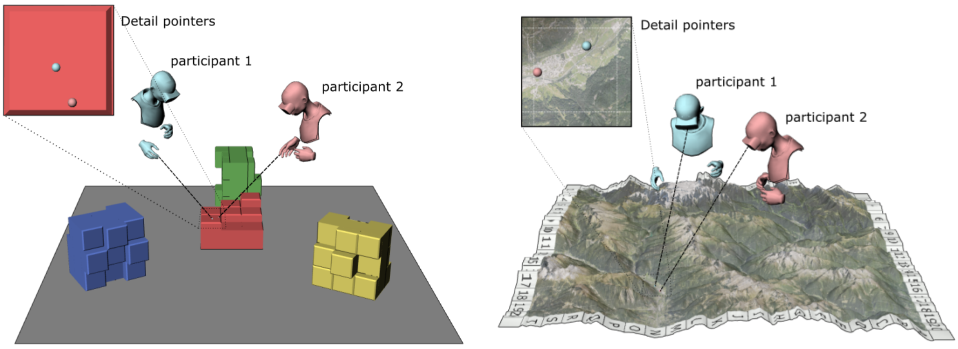
In the following subsections, we detail different aspects of the experiment. This study has been approved by the UCL Interaction Centre (UCLIC) Research Department's Ethics Chair.

3.1 Participants

Twenty-four participants (twelve pairs) volunteered to take part in the remote study. The data of two pairs of participants was used to pilot the study and test the application, while the remaining ten pairs were used for the data analysis. One condition during recruitment was for participants to own or have access to the specific VR HMD: Oculus Quest. This condition was because the experimental session was conducted remotely via teleconference software and then via the VR application. Participants were recruited online via forums and social networks groups dedicated to the Oculus Quest headset and Slack channels dedicated to HCI VR research participant pooling. Participants were recruited individually and then matched up in pairs based on their time availability to conduct the experiment. All participants provided informed consent and received £15 compensation for participating. For the study, pairs of participants were asked to work together in a remote collaborative visual analysis task. Participants were familiar with VR devices as they owned or had access to a HMD's headset. All participants had at least a university grade (6 PhD Candidates, 8 PhDs, 6 MScs, 4 BAs). The mean age was 33 years old with a standard deviation of 8.3. The 88% of the participants was male, and the 12% female.

3.2 Setup

To keep the application development simple and to avoid noise due to differences across VR HMDs, we decided to target a single device for the experiment. The selected headset (Oculus Quest) is 6 degrees of freedom (DoF) untethered VR HMD, with a 60 Hz refresh rate. We chose this headset because of its popularity and low retail price. We developed an application for collaborative visual analysis of 3D data using Unity (version 2018.4.14f1) and the Oculus unity SDK. The application enables the visualisation of different types of 3D data sets (i.e., terrains, 3D networks, CAD files). The application enables each participant to join a real-time session in which other participants' presence is represented by avatars (i.e., Oculus Avatar SDK) as shown in Figure 2. Each participant in the VR space is free to move in any direction using a thumbstick controller or physically move using the 6 DoF of the VR HMD. Avatar movements are streamed via the network, so their behaviour (head and hand movements) and position in the virtual space is reproduced with low latency. The application also enables participants to talk to each other using the embedded microphone and speakers of the VR HMD. Additionally, the setup supports an observer/moderator to be present in the VR session and environment.



(a) Puzzle Task

(b) Terrain Task

Figure 2: A pair of participants collaborate on the visual analysis tasks in the 3D environment. a) Participants are using a hand pointer while performing a four-part 3D puzzle. b) Participants identify the four largest settlements in a terrain dataset using a hand pointer. The hand direction visualised as a series of dotted lines is displayed in the image only to illustrate the difference between head and hand pointers. Both task environments have a size of 3x3 meters.

3.3 Pointers

The pointer consists of a small (1cm) sphere depicted at the intersection between the direction of the hand and the visualised data. Hands are tracked via controllers, and the hand pointer is associated with the dominant hand via the Oculus Unity SDK. The VR HMD tracks the head direction and position. The hand direction, or in other words, the ray departing from the hand, is not visualised; instead, the little sphere is visualised, depicting a small spotlight and therefore displaying the same effect of a laser pointer. The pointers can be seen in the "Detail pointers" window in Figure 2. When the pointer is not present, participants can still point using the hand embodiment as if they were in a physical co-located collaborative scene. The controller triggers approximate the posture of hands, so if a trigger is pressed/released, the correspondent finger is depicted fully contracted or in a straight position. Users can, therefore, intuitively use the index finger to point to referents (Figure 2).

3.4 Experiment Design

We design a 2(pointer)x2(reference difficulty) factors (Table 3a), within-subjects experiment. Participants collaborate on two visual search tasks consisting of identifying visual features in two data sets. The reference difficulty factor consists of two levels: a 3D terrain with hard-to-describe features and a 3D puzzle with an easy-to-describe feature. On the hard level, verbal references can be done using map coordinates or describing features in detail. On the easy level, verbal references can refer to the colour of puzzle blocks or a unique label number. We argue that the complexity of the features in the satellite map is higher than the simple puzzle geometric shapes to describe and disambiguate. Moreover, map coordinates are more complex to reference than a single puzzle label, as they require users to compose the coordinate by reading both longitudinal and latitudinal labels. Therefore, we argue that the cognitive effort required to describe the map's referent is higher than the puzzle. We validated such a hypothesis by pilots of the experiment. Moreover, experiment results of the number of implicit references further validate this level classification. The pointer factor consisted of two conditions: a condition without any pointer and a hand laser pointer, as previous work validates pointers as successfully supporting pointing based communication Moore et al. (2007).

3.5 Task

The two tasks are collaborative visual search tasks. Visual search task is considered a proxy for many other tasks to be done together in VR synchronously, which include finding virtual objects or information together, jointly referencing the same referent Schmalstieg and Höllerer (2016); Prilla (2019).

For the hard task, we used a scenario common in HCI studies that consist in identifying features on 3D terrain maps. We took inspiration from previous works Šašinka et al. (2019); Liu et al. (2017). 3D terrain data is rich in details. Therefore it is complicated to describe it verbally. In the 3D terrain visual analysis task (i.e. hard verbal reference task), participants must identify the four largest settlements (i.e. cities) and the four largest lakes. The terrain consists of satellite images and elevation extracted from Mapbox, and the coordinates corners in the first dataset are for the top-left latitude 46.56, longitude 11.53 and bottom-right latitude 46.17, longitude 11.92; in the second dataset, the coordinates are top-left latitude 46.62, longitude 10.53 and bottom-right latitude 46.23, longitude 11.92.

For the easy task, we selected a scenario that is very common in collaborative VR tasks: puzzle. For example, many studies can be found in the literature using puzzle quiz Slater et al. (2000); Steptoe et al. (2009); Schroeder et al. (2001); Widestrom et al. (2000); Kim et al. (2014). Such tasks contain a visual analysis component which requires participants to identify compatible blocks by comparing them. In our specific case, we avoided any manipulation to focus on visual analysis and related pointing-base communication. In the 3D puzzle task, users must identify the four puzzle blocks that fit together (2 puzzles were present for each experiment condition). Each block measures 50x50x25 cm, and each of the two sides of the block contains 3x3 puzzle joints. Both puzzle conditions are available to be downloaded from ANON-REPOSITORY. At the beginning of each trial, participants were asked to collaboratively identify and report the four correct features to the experiment moderator. If there was a leading effect (i.e., one participant being the only one active), the experiment moderator would remind the pairs to discuss and agree upon features before reporting them. Both task search spaces are equal in size and correspond to 3x3 m. The time given to participants is displayed as a countdown on the VR scene and consists of 5 min max for each scene.

3.6 Procedure

At the beginning of each experimental block, participants are given a chance to practise the task and familiarise themselves with sample datasets. The practice time consists of a maximum of 5 min, but participants can interrupt it earlier if needed. The sample dataset used in practice was not used for the task. Users were allowed to train on both the easy (blocks) and hard (map) tasks. During familiarisation, participants can ask questions; this phase ends once both participants confirm understanding the task. Following the familiarisation, participants are asked to perform the task across the two conditions: hand pointer and no pointer. For each of the two conditions, an equivalent variation of each data set is used (two terrains and two puzzles) for four data sets (Table 3b). Trial order and experimental block order were randomised to counterbalance learning effects.

Once participants agree on a feature, they are asked to communicate it to the observer verbally. The observer only acknowledges the communicated data features as recorded if both participants explicitly agree on it; otherwise, the observer prompts a reminder that both participants have to agree. Such constraint forces pairs to work collaboratively. To incentivise engagement with the task, participants are told that if they score above a specific threshold value, they will receive a £15 voucher instead of a £10 voucher (in the end, every participant receives £15 regardless of their score). We recorded audio and video in VR and log position for all the experiment sessions.

		Factor1: Pointer	
		Level 1 No Pointer	Level 2 Hand Pointer
Factor2: Difficulty	Level1 Terrain	No Pointer Terrain	Hand Pointer Terrain
	Level2 Puzzle	No Pointer Puzzle	Hand Pointer Puzzle

(a) Experiment Design

Experimental Session Participants Diad					
Experimental Block1 Factor 2 Level 2: Terrain			**Experimental Block2** Factor 2 Level 2: Puzzle		
eg	**Trial1** F1 L1 No Pointer	**Trial2** F1 L2 Hand Pointer	eg	**Trial1** F1 L1 No Pointer	**Trial2** F1 L2 Hand Pointer

(b) Experiment procedure

Figure 3: (a) Experiment Design: the experiment has two factors: dataset and pointer. The dataset factor has two levels: 3D surface (terrain), 3D volumes (puzzle). The pointer factor has two levels: No Pointer, Hand Pointer. (b) The experimental procedure is divided in experimental blocks one for each level of the independent variable difficulty, and experimental trials one for each level of the independent variable pointer, plus one trial for task familiarization at the start of each experimental block. Trial order and experimental blocks order were randomised to counterbalance learning effects.

4 Measures

This section gives an overview of the measures collected during the experiment and how we post-process them. We record the head behaviour of both participants. Head gaze is the intersection between the ray starting from the Head position with the direction of Head rotation and the visualised data, which is used to calculate head concurrent pointing behaviour (i.e., visual coordination, section 4.1). Additionally, we record a video/audio stream of the virtual environment for every experimental session of the participants' avatars, containing verbal communication between participants. We use this data to perform implicit/explicit reference analysis (Section 4.3). To understand if the experimental conditions impact temporal and accuracy performances, we also record the task time and task score. Task time is capped to 300 seconds, 5 min to keep the duration of the whole experiment to 20 min max. The maximum number of correct answers for each task is four.

4.1 Visual Coordination

Visual coordination consists of participants' visual focus coupling, or in other words, how well synchronised their visual attention is. As previous work suggests, when users point to a referent during an utterance, this triggers mutual orientation, an essential part of visual coordination. Pointing-based communication is, in this sense, an effort aimed at negotiating shared visual attention during collaborative work Moore et al. (2007). Previous work also explores how visual coordination is highly correlated to the quality of collaboration Schneider and Pea (2013). Therefore visual coordination is a crucial dimension of collaboration in visual search tasks.

The ideal measure of visual coordination would require to use eye-gaze behaviour. However, our study did not use eye-trackers as most low-cost VR HMDs do not have them and running a remote user study during pandemics requires us to target popular low-cost headsets such as oculus quest. Instead, we use head-gaze behaviour, which several studies have reported as a good proxy of eye movements (Biguer et al., 1982; Pelz et al., 2001; Wang et al., 2019). Concurrent head pointing measures the time two participants concurrently point their heads towards the same target simultaneously. For example, when collaborators discuss a visual feature, they are likely to point their heads towards such a feature concurrently. This effect is also described as mutual orientation, identified by Moore et al. (2007) as the first stage of deixis in pointing-based communication.

We post-process head gaze recorded data to measure the time head-gaze overlap during one experimental trial. We define a distance of 20 cm as the threshold for the euclidean distance calculation. Below such threshold, the two head-gaze are considered to point at the same location and above. They are considered to be pointing at different data features. The distance between the two head gaze points

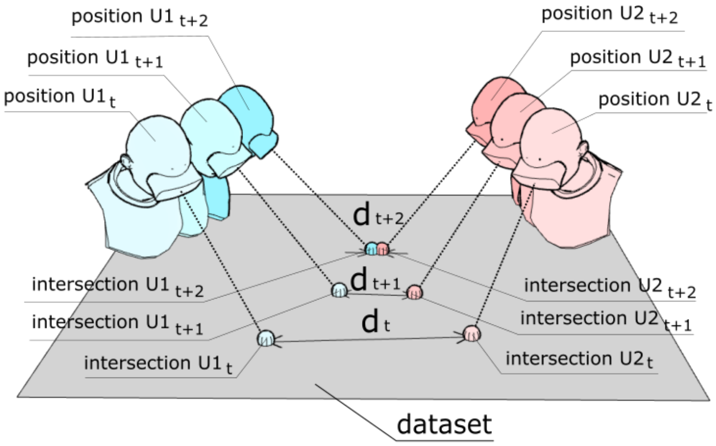


Figure 4: A view of the collected measure of head position, head direction and head signal intersection at a specific moment in time. We post-process head gaze recorded data (intersection) to measure the time head-gaze overlap during one experimental trial by computing the euclidean distance for each time frame. We post-process head position data to measure the time (seconds) participants stay still/move and compare it throughout the different experimental conditions.

is calculated for every sample at time *t*; then, we multiply the number of samples by the sampling frequency to obtain the cumulative time of concurrent head pointing (Figure 4).

4.2 Locomotion

When performing an implicit spatial reference (i.e. pointing/utterance) during pointing-base communication, the referent can be misunderstood by the collaborator (i.e., recipient). Such misunderstanding happens because the gesture performer might point imprecisely. Alternatively, the recipient may fail to correctly project the direction of the hand/arm onto the observed scene. A way to improve the accuracy of a pointing action during pointing-based communication consists in moving closer to the referent, so to make sure that the observer/listener won't miss-interpret the direction of the pointing action Wong and Gutwin (2010). Laser Pointers instead allow participants to perform precise pointing. Using a laser pointer, the performer of the pointing action can adjust the cursor position until the cursor lays on the referent, removing ambiguities. Pointers, therefore, allow to perform accurate pointing gestures from a distance (i.e., without having to travel towards the referent) Wong and Gutwin (2014). However, during collaborative visual tasks, participants might be interested in reducing the distance to a referent for other reasons, such as observing it in greater detail or simply increasing its presence by joining a collaborator's working area.

To investigate the impact of locomotion on pointing based communication, we measure how much time each participant spends moving in ICVE during each trial. As part of the experiment guideline, we expressly asked participants to explore the space only via a thumb-stick controller rather than moving physically for safety reasons. Therefore we used the locomotion speed set in the unity environment of 1.6 m/s to determine the ideal threshold to classify intended movement and noise.

We post-process head position data to measure the cumulative time of locomotion and compare it throughout the different experimental conditions. To calculate the locomotion time, we considered only the samples where the velocity is above the threshold of 0.8 m/s, calculating the distance using sampling frequency and velocity and removing small movements and noise.

4.3 Implicit references

Deixis consists of verbal references supported by a pointing gesture. Within a visual search task, deixes are common occurrences as they allow negotiating the collaborative shared visual context.

Deixes can be implicit or explicit: the first requires less information uttered and are also cognitively less demanding D'Angelo and Begel (2017); Wong and Gutwin (2014). Implicit deixis tends to rely more on the accuracy of the pointing action as the utterance does not carry sufficient information to disambiguate the referent. We consider an implicit spatial reference occurring whenever a participant refereed to a data feature without explicitly naming any unique property of the object (i.e., name, location, colour). Instead, explicit deixis contains information to disambiguate the referent from the rest of the data set. Such explicit information can consist of: position relative to the user (e.g., on my left/right etc.), object characteristics (e.g., the red block etc.), labels (i.e., a unique textual description) or its absolute position expressed in coordinates (i.e., the data feature in B5).

Understanding how pointing based communication changes when hard-to-describe referents are present, or a lack of distance pointing support means classifying each deixis as implicit/explicit. Such a classification gives us an understanding of how smooth/fast verbal communication is. Additionally allows us to understand the balance with behavioural alternatives, such as getting close to the referent to pinpoint it more accurately.

Inspired by previous CSCW work proposed by D'Angelo and Begel (2017), we transcribed audio of the collected videos and carried out a double-blinded video/text classification of the spatial, verbal references. Two analysts performed the analysis to countereffect the subjectivity of the classification process. If the two interpreters were unclear if an instance was implicit or explicit, they conducted a collaborative post-analysis to reach convergence.

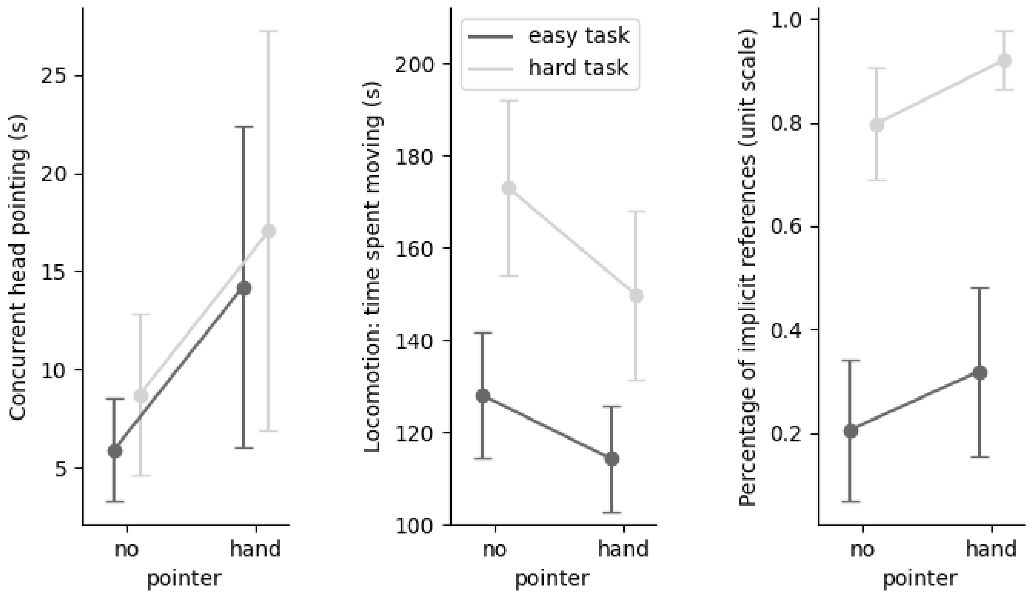
We also classify each reference as successful/unsuccessful. Such classification allows us to understand if and how locomotion impacts the effectiveness of point-based communication when there is a lack of support for distant pointing. A reference is considered unsuccessful when the recipient misinterprets the correct referent or if the recipient ignores the deixis.

5 Statistical Analysis

We performed a repeated measure ANOVA test (using JASP) on the data we collected and post-processed. For the measures of temporal and accuracy performances and the number of unsuccessful deixes, the analysis did not return any significant difference across conditions. For these measures, we won't report the analysis result for conciseness. Our results of visual coordination are achieved by a set of 10 samples (10 pairs of participants). While for locomotion and implicit references, all 20 participants are measured individually, thus equivalent to 20 samples.

5.1 Visual Coordination

The 2 way ANOVA analysis results show one main effect related to the factor: *Pointer* p-value $<.001$ (Table I and Figure 5a). When participants have a laser pointer, they spend approximately 8 seconds more pointing their head towards the same data subset. To contextualize this measure, the average duration of a task is 230 sec, representing approximately 3.4% of the time. However, from observations, we can see that the task time is split between independent work (scanning data visualization independently) and collaborative work (discussing the interpretation of data features). Considering that visual coordination only relates to collaborative work, we argue that the 3.4% of time represents a much higher value within the collaborative stages.



(a) Visual Coordination

(b) Locomotion

(c) Implicit references

Figure 5: Descriptive plots: on the horizontal axes the pointer conditions, on the separate lines the difficulty of explicit references (i.e., hard task and easy task), error bars display the confidence interval of 95%.

Table I: ANNOVA: Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
(a) Visual Coordination					
pointer	1001.618	1	1001.618	22.919	< .001*
difficulty	11.694	1	11.694	0.114	0.743
pointer * difficulty	29.941	1	29.941	0.695	0.426
(b) Locomotion					
pointer	6906.361	1	6906.361	15.816	< .001*
difficulty	32328.758	1	32328.758	19.590	< .001*
pointer * difficulty	469.447	1	469.447	0.887	0.358
(c) Implicit references					
pointer	0.140	1	0.140	7.031	0.026
difficulty	3.560	1	3.560	80.807	< .001*
pointer * difficulty	2.984e-4	1	2.984e-4	0.010	0.924

* p < .005

5.2 Locomotion

We statistically compare the measures of locomotion (i.e. time spent moving) by performing a two way repeated measure ANOVA (Table I and Figure 5a). The ANOVA analysis results show two main effect related to the factors *Pointer* (p-value <.001) and *Difficulty* (p-value <.001). While we see an effect of locomotion related to the differences in the task, the important result is the effect on the pointer level and the lack of interaction between the two levels. When participants do not have a laser pointer, they spend approximately 18s more moving. To give a contextual understanding of this measure, the average duration of a task is 230 seconds, therefore representing approximately 7% of the time. If we consider that the average locomotion speed for this experiment is set to 1.6 m/s. This means that participants without support for distance pointing travelled approximately 28 meters more (in a 3m x 3m visualization space).

5.3 Implicit References

We statistically compare the repeated measures of the dependent variable: *number of implicit Deixes* by performing a two way repeated measure ANOVA (Table I and Figure 5a). The ANOVA analysis results show two main effects related to the factor: *difficulty* (p-value <.001). This result validates the design level of difficulty: if the referent is simple to identify by an explicit reference, the user tends to verbally describe it. On the other hand, when the referent is difficult to identify by verbal description, the user will adopt the strategy of pointing it and adding implicit references.

6 Discussion

Previous studies based on distance pointing in ICVE and real-world scenarios show that collaborators pointing accuracy from a distance often depends on either having access to a laser pointer or on how hard to describe it the referent (Wong and Gutwin, 2010, 2014). However, ICVE allows participants to move in the environment and, therefore, get as close as they need to the referent to perform an accurate pointing gesture. Therefore, what would users do when faced with the option of moving closer to the referent or describing it in better detail? Such a question is worth answering to understand better the dynamics of pointing-based communication in ICVEs. A better understanding of such collaborative dynamics is fundamental to developing solutions that can better support collaboration in ICVEs. Therefore within this study, we introduce the ability for users to move in the ICVE to investigate the trade-off between moving close to a referent and the effort of composing a verbal reference when the referent is difficult to describe. We do so within the context of a collaborative visual search task which is recognised to be a proxy of many other collaborative tasks in VR Prilla (2019).

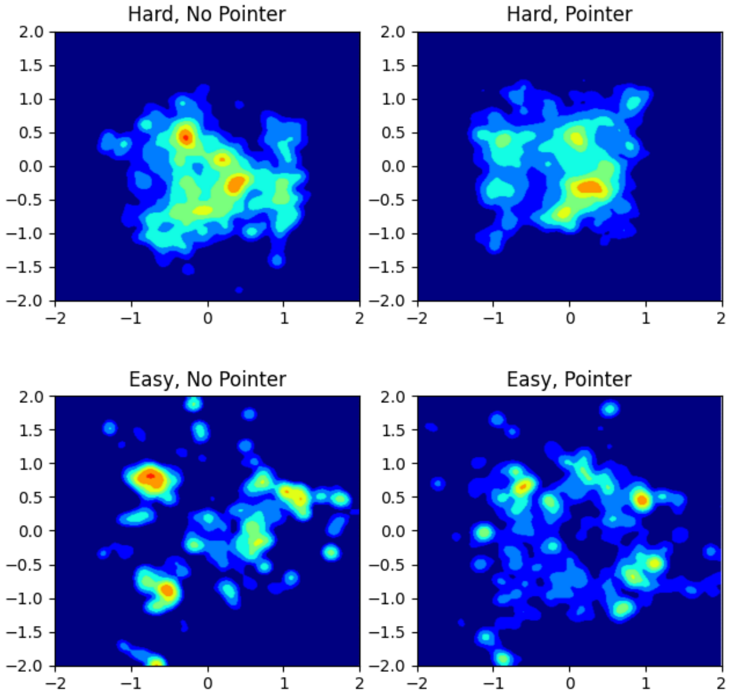


Figure 6: Heat-map of physical movement for the 4 experimental conditions.

6.1 Impact of locomotion on pointing-based communication

Our results extends the work of Wong and Gutwin (2010, 2014) by exploring a different dynamic of pointing based communication in the collaborative search task. While Wong measured accuracy in the context of fixed user distances from the referent, we explore a more ecologically valid scenario. Users are free to move in the ICVE and are instructed to perform a generalise search task. We extend his work by showing how users choose to locomote no matter how hard-to-describe is the referent in front of the choice of verbally describing a referent or moving closer to it. Such a statement is supported by the statistical analysis of locomotion, which shows a significant movement increment in hard and easy tasks when the pointer is absent.

Furthermore, we integrate the analysis of locomotion by generating cumulative head position heat maps for each experimental condition Figure 6. It is evident that the different datasets led to different exploration patterns and that the support for distance pointing did not impact how users explored the environment. If we cross the data from Figure 6 and Figure 5b and we notice that the locomotion last 20 seconds more in the absence of the pointer condition, we infer that such difference is not due to the exploration but to compensate lack of a laser pointer.

6.2 Impact of pointers on verbal communication

Previous CSCW studies in 2D desktop collaboration in remote programming show how pointers can increase the number of implicit references during deixis, making verbal communication faster and smoother (D'Angelo and Begel, 2017). Inspired by such a study, we counted and analysed the number of implicit references. In our ICVE experiment, results and observations suggest that when a pointer is not available, the number of implicit references (Fig 5c) during deixis stay the same. Our results differ from D'Angelo and Begel (2017) suggesting that when the embodiment is available, and users are free to move throughout the data pointers, visualisations do not influence verbal communication.

6.3 Impact of pointers on visual coordination

Previous research explored visual attention cues from head behaviour or eye gaze behaviour in ICVE during visual search tasks Piumsomboon et al. (2017) measuring how visual attention cues increase visual coordination. In general, hand pointing is recognized to trigger mutual orientation and visual coordination Wong and Gutwin (2010); Moore et al. (2007), however to the best of our knowledge, no study measure visual coordination with and without laser pointers in ICVEs. Our study fills this gap by showing that hand pointers availability increases the amount of time that collaborators spend concurrently pointing their heads towards the same subset of the data (section 5a).

7 Future work and Design Implications

In this study, we answered the following question: what will users do when faced with a lack of pointing accuracy: moving closer to the referent or describing it in better detail? While pointers in VR are proved extremely useful from previous studiesHindmarsh et al. (1998); Hoppe et al. (2018); Bai et al. (2020), we observe that visual pointers inclusion might depend on several factors: the complexity of the user interface, how crowded the ICVE is, and the confusion that multiple pointers may cause. Such considerations impact the design of ICVE, which needs to balance the advantages and disadvantages of pointers, compensating with alternative approaches that help to point accuracy. In addition, since there are benefits in moving closer to a referent, such as observing it in more detail or improving engagement with collaborators, we aim to identify methods that allow participants to semi-automatically move closer to an intended referent with or without pointing at it. A further approach can be identifying the intended referent by leveraging shared focus or adding semantic augmentation.

Our study does not consider distance perception as a crucial factor. This assumption is inherited from different works Mayer et al. (2020, 2018, 2015); Schweigert et al. (2019); Sousa et al. (2019); Wong and Gutwin (2014) that conversely consider distance with an active role in pointing accuracy. However, this possible implication of distance perception in deictic pointing could be a good topic for future studies, as the research community is not yet detailed; studies that explore the perception of distance in VR are Finnegan et al. (2016); Maruhn et al. (2019).

Another interesting aspect is the implication of different locomotion strategies in ICVEs. For example, teleportation is a locomotion method which requires pointing to translate a user's location in the ICVE. Such a technique depends on the individual and the environment. However, our study, which explores the relations between pointing and locomotion, could inspire the community to investigate a collaborative version of locomotion. For example, when someone is making a pointing reference, the system can offer a "privileged" position and orientation for the observer that can be instantly applied. In addition, such a mechanism can be used for different collaboration tasks.

Moreover, we hope that the research community could use our results to explore novel ways of referencing targets based on a different paradigm or input channels such as speech. Previous studies demonstrate that a natural language processing pipeline could be used to describe and possibly display visual cues on some specific object parts Giunchi et al. (2021). Our study entails that when the referent is easy-to-describe, such a speech-based system could be used to highlight referents, such as collaborators are doing this naturally during a collaboration task. On the other hand, if the referent is hard-to-describe, that system may not be effectively used.

8 Conclusions

This paper designed and carried out an experiment to test the participants' attitude in a pointing-based task in ICVE. We conclude that deictic referencing in ICVEs with embodiment and locomotion does not require pointers to be accurate and implicit, as long as the users are free to move as close as they need to the data they are observing. One main reason is that when users are facing the problem of inaccuracy during pointing, they instinctively move closer to the referent rather than using verbal references to improve the precision of their pointing. Moreover, this effect is independent of how hard-to-describe the referent is. Locomotion allows users to move closer to the referent while performing deixis, improving pointing accuracy. We outline some design implications by highlighting how designers and engineers should consider two essential elements in support of distance-pointing: first, if users are able to move within the environment, and second if the collaborative task requires high visual coordination.

References

Bai, H., P. Sasikumar, J. Yang, and M. Billinghurst (2020): 'A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing'. In: *Conference on Human Factors in Computing Systems - Proceedings*.

Bangerter, A. and D. M. Oppenheimer (2006): ‘Accuracy in detecting referents of pointing gestures unaccompanied by language’. *Gesture*, vol. 6, no. 1, pp. 85–102.

Batmaz, A. U. and W. Stuerzlinger (2019): ‘Effects of 3d rotational jitter and selection methods on 3d pointing tasks’. In: *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*. pp. 1687–1692.

Benford, S., J. Bowers, L. E. Fahlen, C. Greenhalgh, and D. Snowdon (1995): ‘User embodiment in collaborative virtual environments’. *Conference on Human Factors in Computing Systems - Proceedings*, vol. 1, pp. 242–249.

Biguer, B., M. Jeannerod, and C. Prablanc (1982): ‘The coordination of eye, head, and arm movements during reaching at a single visual target’. *Experimental Brain Research*, vol. 46, no. 2, pp. 301–304.

D’Angelo, S. and A. Begel (2017): ‘Improving communication between pair programmers using shared gaze awareness’. *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-Janua, pp. 6245–6255.

Finnegan, D. J., E. O’Neill, and M. J. Proulx (2016): ‘Compensating for distance compression in audiovisual virtual environments using incongruence’. In: *Conference on Human Factors in Computing Systems - Proceedings*. pp. 200–212.

Giunchi, D., A. Sztrajman, S. James, and A. Steed (2021): ‘Mixing modalities of 3D sketching and speech for interactive model retrieval in virtual reality’. In: *IMX 2021 - Proceedings of the 2021 ACM International Conference on Interactive Media Experiences*. pp. 144–155, Association for Computing Machinery.

Higuch, K., R. Yonetani, and Y. Sato (2016): ‘Can eye help you?: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks’. In: *Conference on Human Factors in Computing Systems - Proceedings*. New York, NY, USA, pp. 5180–5190, Association for Computing Machinery.

Hindmarsh, J., M. Fraser, C. Heath, S. Benford, and C. Greenhalgh (1998): ‘Fragmented interaction: Establishing mutual orientation in virtual environments’. In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. New York, New York, USA, pp. 217–226, ACM Press.

Hoppe, A. H., K. Westerkamp, S. Maier, F. van de Camp, and R. Stiefelhagen (2018): ‘Multi-user collaboration on complex data in virtual and augmented reality’. *Communications in Computer and Information Science*, vol. 851, pp. 258–265.

Jermann, P. ., D. . Mullins, and M.-A. . Nüssli (2011): ‘Collaborative Gaze Footprints Correlates of Interaction Quality’.

Kim, S., G. Lee, N. Sakata, and M. Billinghurst (2014): ‘Improving co-presence with augmented visual communication cues for sharing experience through video conference’. In: *ISMAR 2014 - IEEE International Symposium on Mixed and Augmented Reality - Science and Technology 2014, Proceedings*. pp. 83–92, IEEE.

Liu, C., O. Chapuis, M. Beaudouin-Lafon, and E. Lecolinet (2017): ‘CoReach: Cooperative gestures for data manipulation on wall-sized displays’. *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, pp. 6730–6741.

Maruhn, P., S. Schneider, and K. Bengler (2019): ‘Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains’. *PLoS ONE*, vol. 14, no. 10.

Mayer, S., J. Reinhardt, R. Schweigert, B. Jelke, V. Schwind, K. Wolf, and N. Henze (2020): 'Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality'. In: *Conference on Human Factors in Computing Systems - Proceedings*. pp. 1–14.

Mayer, S., V. Schwind, R. Schweigert, and N. Henze (2018): 'The effect of offset correction and cursor on mid-air Pointing in real and virtual environments'. *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2018-April, pp. 1–13.

Mayer, S., K. Wolf, S. Schneegass, and N. Henze (2015): 'Modeling distant pointing for compensating systematic displacements'. In: *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2015-April. pp. 4165–4168.

Moore, R. J., N. Ducheneaut, and E. Nickell (2007): 'Doing virtually nothing: Awareness and accountability in massively multiplayer online worlds'. *Computer Supported Cooperative Work*, vol. 16, no. 3, pp. 265–305.

Nüssli, M.-A. (2011): 'Dual Eye-Tracking Methods for the Study of Remote Collaborative Problem Solving'. *PhD Thesis, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE*, vol. 5232.

Pelz, J., M. Hayhoe, and R. Loeber (2001): 'The coordination of eye, head, and hand movements in a natural task'. *Experimental Brain Research*, vol. 139, no. 3, pp. 266–277.

Pfeiffer, T., M. E. Latoschik, and I. Wachsmuth (2008): 'Conversational pointing gestures for virtual reality interaction: Implications from an empirical study'. In: *Proceedings - IEEE Virtual Reality*. pp. 281–282.

Pietinen, S., R. Bednarik, T. Glotova, V. Tenhunen, and M. Tukiainen (2008): *A Method to Study Visual Attention Aspects of Collaboration: Eye-Tracking Pair Programmers Simultaneously*.

Piumsomboon, T., A. Dey, B. Ens, G. Lee, and M. Billinghurst (2017): 'CoVAR: Mixed-Platform Remote Collaborative Augmented and Virtual Realities System with Shared Collaboration Cues'. In: *Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2017*. pp. 218–219, Institute of Electrical and Electronics Engineers Inc.

Prilla, M. (2019): '"I simply watched where she was looking at": Coordination in short-term synchronous cooperative mixed reality'. *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. GROUP.

Šašinka, C., Z. Stachoň, M. Sedlák, J. Chmelík, L. Herman, P. Kubíček, A. Šašinková, M. Doležal, H. Tejkl, T. Urbánek, H. Svatoňová, P. Ugwitz, and V. Juřík (2019): 'Collaborative immersive virtual environments for education in geography'. *ISPRS International Journal of Geo-Information*, vol. 8, no. 1.

Schmalstieg, D. and T. Höllerer (2016): *AR Textbook Tobias*.

Schmidt, C. L. (1999): 'Adult Understanding of Spontaneous Attention-Directing Events: What Does Gesture Contribute?'. In: *Ecological Psychology*, Vol. 11. pp. 139–174.

Schneider, B. and R. Pea (2013): 'Real-time mutual gaze perception enhances collaborative learning and collaboration quality'. *International Journal of Computer-Supported Collaborative Learning*, vol. 8, no. 4, pp. 375–397.

Schroeder, R., A. Steed, A. S. Axelsson, I. Heldal, A. Abelin, J. Widestrom, A. Nilsson, and M. Slater (2001): 'Collaborating in networked immersive spaces: as good as being there together?'. *Computers & Graphics*, vol. 25, no. 5, pp. 781–788.

Schweigert, R., V. Schwind, and S. Mayer (2019): 'EyePointing: A Gaze-Based Selection Technique'. vol. 19.

Slater, M., A. Sadagic, M. Usoh, and R. Schroeder (2000): 'Small-group behavior in a virtual and real environment: A comparative study'. *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 1, pp. 37–51.

Sousa, M., R. K. Dos Anjos, D. Mendes, M. Billinghurst, and J. Jorge (2019): 'Warping deixis: Distorting gestures to enhance collaboration'. In: *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 12. New York, NY, USA, pp. 1–12, Association for Computing Machinery.

Steptoe, W., O. Oyekoya, A. Murgia, R. Wolff, J. Rae, E. Guimaraes, D. Roberts, and A. Steed (2009): 'Eye tracking for avatar eye gaze control during Object-Focused multiparty interaction in immersive collaborative virtual environments'. In: *Proceedings - IEEE Virtual Reality*. pp. 83–90, IEEE.

Villamor, M. and M. M. Rodrigo (2018): 'Predicting successful collaboration in a pair programming eye tracking experiment'. In: *UMAP 2018 - Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. pp. 263–268.

Wang, P., S. Zhang, X. Bai, M. Billinghurst, W. He, S. Wang, X. Zhang, J. Du, and Y. Chen (2019): 'Head pointer or eye gaze: Which helps more in MR remote collaboration'. In: *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*. pp. 1219–1220, Institute of Electrical and Electronics Engineers Inc.

Widestrom, J., A. S. Axelsson, R. Schroeder, A. Nilsson, I. Heldal, and A. Abelin (2000): 'The Collaborative Cube Puzzle: A Comparison of Virtual and Real Environments'.

Wong, N. and C. Gutwin (2010): 'Where are you pointing? The accuracy of deictic pointing in CVEs'. In: *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2. New York, New York, USA, pp. 1029–1038, ACM Press.

Wong, N. and C. Gutwin (2014): 'Support for deictic pointing in CVEs'. pp. 1377–1387.