WARPING DEIXIS: Distorting Gestures to Enhance Collaboration

Maurício Sousa
INESC-ID Lisboa, Técnico Lisboa, University of Lisbon
antonio.sousa@ist.utl.pt

Rafael Kuffner dos Anjos
INESC-ID Lisboa, Técnico Lisboa, University of Lisbon / FCSH-UNL
rkuffner@fcsh.unl.pt

Daniel Mendes
INESC-ID Lisboa, Técnico Lisboa, University of Lisbon
danielmendes@ist.utl.pt

Mark Billinghurst
School of Information and Mathematical Sciences, University of South Australia
mark.billinghurst@unisa.edu.au

Joaquim Jorge
INESC-ID Lisboa, Técnico Lisboa, University of Lisbon
jorgej@acm.org

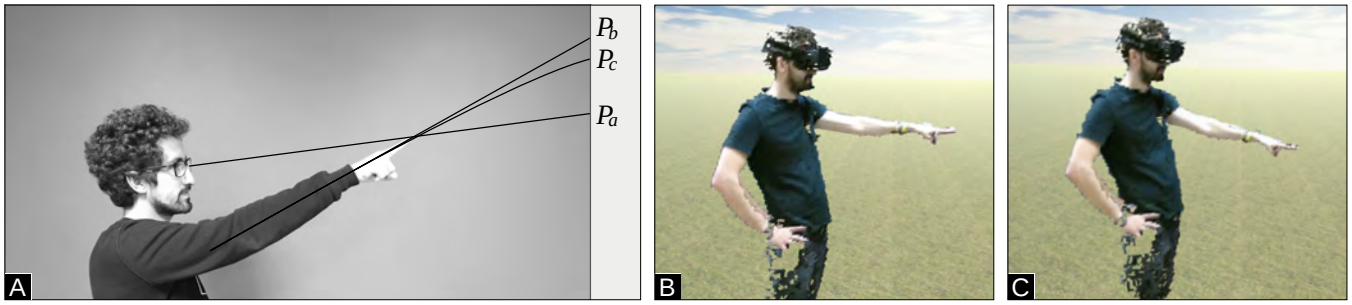

Figure 1: (A) When pointing to a distal referent (P_a), people usually put their index finger between their eyes and target. Yet, observers rely on a extrapolation of the arm-finger line to find the target of the gesture (perceiving something between P_b and P_c). We present Warping Deixis, an approach to reducing misinterpretation of deixis using body warping. Given a B) body representation of a pointing person, our approach C) changes rendering of the avatar's arm to reduce the gesture's ambiguity.

ABSTRACT

When engaged in communication, people often rely on pointing gestures to refer to out-of-reach content. However, observers frequently misinterpret the target of a pointing gesture. Previous research suggests that to perform a pointing gesture, people place the index finger on or close to a line connecting the eye to the referent, while observers interpret pointing gestures by extrapolating the referent using a vector defined by the arm and index finger. In this paper we present Warping Deixis, a novel approach to improving the perception of pointing gestures and facilitate communication in collaborative Extended Reality environments. By warping the virtual representation of the pointing individual, we are able to match the pointing expression to the observer's perception. We evaluated our approach in a co-located side by side virtual reality scenario. Results suggest that our approach is effective in improving the interpretation of pointing gestures in shared virtual environments.

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300838

CCS CONCEPTS

• **Human-centered computing → Computer supported cooperative work**.

KEYWORDS

Deixis, Pointing Gestures, Body Warping, Collaboration

ACM Reference Format:
Maurício Sousa, Rafael Kuffner dos Anjos, Daniel Mendes, Mark Billinghurst, and Joaquim Jorge. 2019. WARPING DEIXIS: Distorting Gestures to Enhance Collaboration. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300838

1 INTRODUCTION

We present a novel technique to improve communication via pointing gestures in Extended Reality (XR). When communicating, people often use deictic references (*Deixis*) – designating the referent by pointing at it [10, 26, 51]. Pointing gestures are widely used to indicate distal artifacts and locations to others forgoing lengthy verbal descriptions [37]. *Deixis* is key to facilitating collaboration, since it simplifies information sharing [17, 19].

In XR collaborative environments, showing full or upper-body representations of people is known to improve awareness [6, 13], since natural body language combined with speech can be used. Current XR technologies allow both local or remote users to be immersed in collaborative virtual environments (CVEs), making it possible for people to see each other either through realistic virtual avatars [40] or 3D-scanned representations [36]. Deixis fosters collaboration via natural gestures that indicate virtual objects in a 3D environment, since both pointing and task objects are visible.

Collaboration improves when people are able to accurately perceive others' pointing gestures. Indeed, these have a considerable impact both on efficiency and task performance when referencing objects or locations that are in close proximity [19]. For this reason, current computer-supported cooperative work (CSCW) approaches resort to simple proxies of deictic pointing – telepointers [18], virtual rays and highlighted targets [55] – to reference workspace artifacts. However, these proxies afford limited control, create visual clutter, and exhibit unclear ownership. Furthermore, these methods fall short when communicating areas, paths, and directions [55].

While it is desirable to improve people's ability to execute and perceive natural gestures in CVEs similarly to what they do in the real world, people observing pointing gestures are often unable to precisely determine where another person is pointing to [4, 11, 22, 43], causing people to engage in lengthy verbal descriptions to single out the location of interest [19]. There are other ways to resolve referent ambiguity besides natural language. However, our approach handles this in a natural and transparent manner.

Figure 1A demonstrates a typical pointing gesture where a person is pointing at a specific target (P_a). To this end, they typically align the tip of their index finger with the referent appearing in their field of view [4, 50, 53]. That is, the target location is intercepted by the vector from the eye to the tip of the index finger [22]. Yet, in contrast to referrer's gesture, observers use the direction of the pointer's arm and index finger to extrapolate its target [22, 53]. Thus, as exemplified in Figure 1A, a linear extrapolation following the arm to index finger vector leads to a perceived target (P_b) that is perceived as lying above the spot designate by the person performing the gesture. Furthermore, that extrapolation is nonlinear and most observers would judge that the person in Figure 1A is pointing at the vicinity of P_c.

In this work, we propose *Warping Deixis*, to improve the perception of deictic gestures in XR collaborative virtual environments. Our approach manipulates the pointer's avatar to rectify the pose of the pointing arm in real-time, for the representation to match the way people perceive the deictic gesture. We do this by dynamically relocating the arm on the pointer's virtual representation to create the illusion of gesturing towards another location, thus improving the perception by an observing collaborator, as demonstrated in Figure 1B (before warping) and Figure 1C (after warping).

The main contributions of this research include: 1) Warping Deixis, a novel body warping technique to improve how deictic gestures are interpreted in collaborative XR; 2) techniques to redirect arm poses applicable to different representations of virtual humans; 3) a user study, evaluating the impact of our approach in referent identification tasks; 4) and design considerations for future collaborative scenarios. A user study validated the assumption that warping the pointer's arm can significantly reduce misunderstandings of the referent and that people were not aware of the avatar distortion, showing that our technique does not impair communication in shared virtual environments as compared to the non-distorted representation.

2 RELATED WORK

Our research builds on prior work in three areas: 1) virtual representations of people, 2) collaboration in XR environments, and 3) the fundamentals of human production and interpretation of deictic gestures. Thus, in this section we discuss previous related work and provide a discussion of the state-of-the-art.

Virtual Representations of People

To enable collaboration, XR environments should rely on complete portrayals of people to allow for the understanding of nonverbal cues in addition to normal speech communication. Nonverbal communicative cues include facial expressions, gaze, body posture, *deixis* to indicate objects referred to in speech [33], and how people utilize the space and position themselves when communicating (*Proxemics* [21]). Being able to perceive such nonverbal cues is beneficial for the sense of co-presence and helps people to communicate naturally [12]. For this reason, such environments rely on virtual representations of people to provide the necessary awareness [20] of the collaborator's activities.

Early groupware approaches employed telepointers and cursors [18] as a mean to provide awareness of people's actions on a shared workspace. However, telepointers and

cursors provide limited knowledge about people's gestures, making it impossible to anticipate their actions. Hence, dynamic representations of arms [48, 49] and hands [45, 56] have been studied to convey such nonverbal cues, yet yielding limited awareness of other people's presence. Indeed, the more realistic fully 3D rigged virtual avatars are, the better they convey the feeling of co-presence [24]. Recent developments in commodity depth cameras enabled lifelike 3D full-body reconstructions of people. For example, Maimone et al. [32] presented a telepresence approach that employed full-body reconstructions of people in a 3D display for correct visualization. In the case of Beck et al. [5], 3D reconstructions are applied to virtual worlds where distributed groups meet. There is also previous research in utilizing Mixed and Augmented Reality to bring remote people to the local environment. Pejsa et al. [38] employed commodity projectors to render life-size representations of people creating the illusion of co-presence. Furthermore, Orts-Escolano et al. [36] demonstrated high-quality reconstruction through head-mounted displays (HMDs), creating an experience akin to physical presence.

The state-of-the-art suggests that full body virtual avatars can convey natural nonverbal communication cues essential for collaborating in a shared environment. Our research builds on this previous work to improve the perception of pointing gestures by manipulating the way virtual embodiments are presented to people in XR environments.

Deixis in XR Collaborative Environments

Pointing gestures are an important nonverbal communication tool to coordinate and maintain a up-to-date understanding of the context when collaborating, yet there are situations where people experience difficulties describing verbally distal referents with hard-to-describe shapes or locations [29]. In these cases, the ability to observe pointing gestures facilitates collaboration by making the communication task more natural.

Fussel et al. [16] suggested that, in collaborative distributed settings, perceiving gestures improves task performance. Indeed, deictic references increase workspace awareness by allowing people to qualify verbal references to artifacts in a shared workspace [19]. However, Wong and Gutwin [55] suggested that using deictic referencing in XR environments is more demanding than in the real world due to narrow fields of view (FOV) and poor resolution of current display technology.

Previous works showed that awareness cues can effectively support communication, such as using virtual pointers [14, 18, 35] or enhancing the collaboration through highlighting visual and audio cues [7, 41]. Yet, virtual pointers can provide inadequate or conflicting augmentations of pointing and produce a direction different from the pointers arm, and highlighted objects may not match the pointing gesture. Also, target highlighting is limited to predefined objects and its discrete movement makes it harder to control. When used in a collaborative environment these can contribute to clutter [55]. Furthermore, Piumsomboon et al. [39] concluded that such enhancements could obfuscate important social cues (facial expression or body gestures). Despite that, Piumsomboon et al. [40] introduced Mini-Me, an adaptive avatar that uses redirected gaze and gestures, and found that their approach was successful in improving user's awareness of their partner in a collaborative XR interface.

Our work focuses on improving the perception of deictic references, and consequently, expedite collaboration. Thus, our approach exploits the concept of gesture redirecting by warping virtual representations of people in an imperceptible way, without losing other important social cues.

Production and Interpretation of Deictic Gestures

Wong and Gutwin [54] suggested that people are "experts at (…) interpreting deictic gestures". Yet, people often fail to determine the exact location to which another person is pointing to. The perceptual accuracy depends on whether the pointing gesture is proximal or distal [43]. When indicating proximal referents, pointers are able to touch the target and observers can identify targets with confidence [4]. In contrast, when pointing at distal referents, people usually align the tip of their pointing finger with their dominant eye [53]. Indeed, ray pointing techniques exploiting the eye-index vector have been used to detect deictic gestures for object selection and manipulation in XR [30, 52] and large scale displays [1, 25], since they offer high accuracy [25]. However, previous research suggests that humans interpret distal pointing by extrapolating the vector defined by the pointer's posture [4, 53].

Herbort and Kunde [22] proposed that this difference between production and interpretation accounts for the systematic spatial misunderstanding of pointing to distant referents. Salomon [42] suggested that human attempts at vector extrapolation deviates from a geometric linear extrapolation. Inasmuch as people observing the arm to index finger vector in Figure 1A would interpret a target position between (P_b) and (P_c). Herbort and Kunde [22] asserted that people interpret pointing gestures by using a nonlinear extrapolation of the pointer's arm-finger vector. This non-linearity aspect of perceiving pointing gestures, can be described as a Bayesian-optimal integration of a linear extrapolation of the arm-finger vector and the observer's prior assumptions. Following this insight, Herbort and Kunde [22] introduced a predictive Bayesian model to estimate the position of referents. Their model is based on the assumptions that participants engage in geometric extrapolation of the arm–finger

line or eye–finger line and participants integrate the geometric extrapolation and a priori information according to Bayesian theory [27, 28]. The proposed Bayesian model can be expressed as follows:

$$\hat{y}_{Bayesian} = \frac{d^{-2}(1-w)y_{geo} + wy_0}{d^{-2}(1-w) + w} \quad (1)$$

Equation 1 considers d as the horizontal distance between the plane containing referents and the pointer's shoulder, y_{geo} as the result of geometric extrapolation, and y_0 as the a priori assumed average referent position, which is set to the pointer's shoulder height. The Bayesian model also considers the free parameter w, which relates to the variability associated with the linear extrapolations to the variability associated with the observer's unknown prior assumptions. The parameter w can assume values between 0 (participants rely exclusively on geometric extrapolation) and 1 (participants rely exclusively on the a priori assumption). The authors determined values of w individually for each participant and provided the average values for different gesture interpretation conditions. A study with participants revealed that the that the nonlinear extrapolation of pointing interpretation can successfully be described by the Bayesian model and referent estimates changed nonlinearly as a function of distance.

In collaborative settings, misinterpreted deixis can thus undermine intentional communication. More specifically, failing to understand the exact target of a pointing gesture prevents other people from correctly understanding the context of collaborative tasks. In this research, we propose to distort the representation of the pointing arm to improve how others perceive the location of distal referents that the pointer wants to communicate.

3 WARPING DEIXIS

We propose Warping Deixis, an approach to reshape people's pointing poses in real-time, to improve human perception of deictic gestures in collaborative settings. We define Warping Deixis as any adjustment to the avatar of a person performing a pointing gesture in order to make distal referents both more explicit and easier to be identified. These adjustments should be plausible in order not to shift other people's attention away from the collaboration proper due to abrupt arm movements. We followed an approach analogous to the body warping technique by Azmandian et al. [2]. We also target pointing gestures towards distal referents, commonly executed with an almost fully extended arm [53], as depicted in Figure 1A.

In this work, we focus on MR environments where people collaborate with each other through virtual representations that can be manipulated whenever someone performs a pointing gesture. Therefore, our virtual representation manipulation approach incorporates two separate stages; 1) applying a Bayesian model to determine where people should be pointing when performing a gesture, 2) contributing a warping technique to suitably change virtual representations of people.

Bayesian-based Pointing Correction Model

As previously mentioned, when interpreting a pointing gesture, observers try to identify distal referents using a linear extrapolation of the vector that follows the pointer's arm (resulting in P_b when the pointer's intended target was P_a, in Figure 2A). However, experimental results from Herbort and Kunde [22], suggest that human attempts at linear extrapolation systematically deviate from a perfect geometric linear extrapolation and the observer's perceived position for the referent is usually located slightly further up, between P_b and P_c depending on the pointer's distance to the referent. Still, the observer's interpreted distal target location is disparate from the location intended by the person pointing (P_a). Accordingly, our approach follows these pointing production and gesture interpretation fundamentals to determine the optimal pointer's arm pose that would cause the deictic arm vector to appear to be pointing exactly above the intended target (P_d), as depicted in Figure 2. This enables the natural nonlinear human attempts of linear extrapolation to induce the observer to perceive the correct intended distal referent. Next we detail the steps necessary to calculate what the pointer's arm position that will induce the desired effect.

First, to realize the intended target (P_a), it is necessary to calculate the pointer's deictic vector and examine its direction. So $\vec{a}$ can be located by following the vector that starts from the pointer's eyes toward the index finger, $\vec{a} = P_{Index} - P_{Eyes}$. We define the vector representing the linear extrapolation of the pointer's arm as $\vec{b} = P_{Index} - P_{Elbow}$. We established the elbow as the vector's starting point considering that pointing gestures towards distal referents are not always executed with a fully extended arm [53] and, therefore, the arm segment between shoulder and elbow are usually not considered by observers when extrapolating the arm's pointing direction. However, any transformation of the pointer's arm should use the shoulder as a rotation pivot point to exclude awkward and unnatural arm postures. Moreover, when in a pointing stance, the shoulder offers more rotation freedom in contrast to the elbow.

Our approach relies on the Bayesian extrapolation model defined by Equation 1 to predict the referent's position estimated by the observer (P_c), thus $P_c = \hat{P}_b$. Given the pointer's shoulder as a rotation pivot, it is possible to determine the angular transformation needed to position the arm in the location where it should be. So, as depicted in Figure 2B,

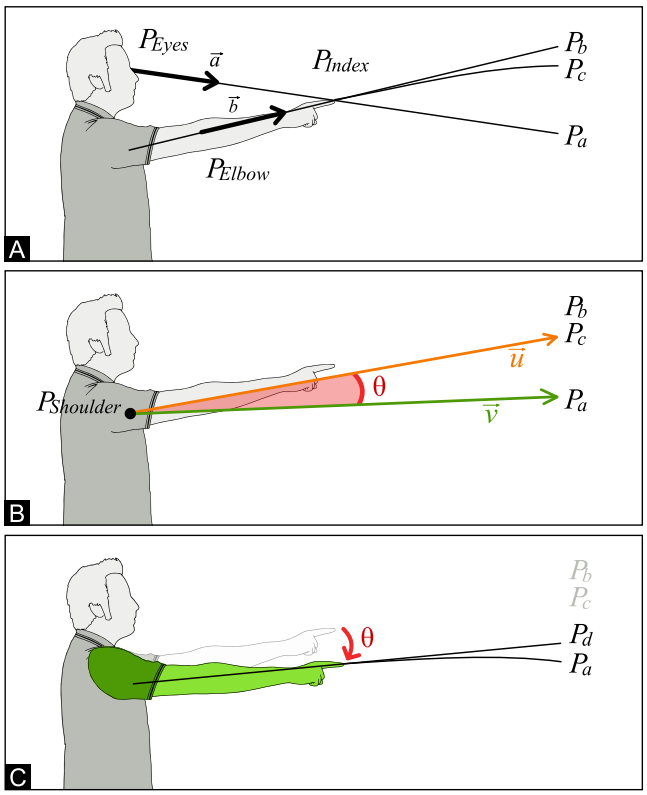

Figure 2: Warping Deixis uses A) a Bayesian model to predict the referent's location interpreted by the observer,P_c, and B) calculates the necessary arm displacement to a C) distal location P_d such that extrapolations would result in their correctly interpreting the pointer's intended target P_a.

the rotation required is the angular distance between (P_c) and (P_a). Given the vectors from shoulder ($P_{Shoulder}$) to the perceived target, $\vec{u} = P_c - P_{Shoulder}$, and to the pointer's intended target, $\vec{v} = P_a - P_{Shoulder}$, it is possible do determine the rotation axis $\vec{r} = \vec{u} \times \vec{v}$ and the rotation angle $\theta = \angle(\vec{u}, \vec{v})$.

The value of θ can be applied to the pointing arm of any body representation of a pointing person, since θ is the angular distance necessary for the arm to be pointing to a distal location (P_d) that should result in an interpretation of (P_a) through a nonlinear human extrapolation of the pointing gesture, as demonstrated in Figure 2C.

Warping People's Virtual Representations

Different methods have been used to create a virtual representation of people in XR environments (e.g. avatar model, point clouds, virtual hands). In all these representations, arms are usually defined by a set of 3D points representing either joints or surface points. Therefore, any warping operation will consist of transforming a set of points according to an estimated matrix. For an avatar model, skeleton transformations would bring a rigged mesh to the right location, yet, for point cloud-based representations, the transformation must be applied to each individual point comprising the full arm. Given a virtual representation $\mathcal{V}$, consisting of a set of 3D points p, warping the pointing arm to another location can be achieved considering that we have the position of the pivot point (which in our case is $P_{Shoulder}$) and that the point set representing the arm, $\mathcal{A} \in \mathcal{V}$, can be estimated. Thereby, the rotation matrix R representing the angular rotation θ about the axis defined by $\vec{r}$, and can be calculated by:

$$R = (\cos\theta)I + (\sin\theta)[\vec{r}]_\times + (1 - \cos\theta)(\vec{r} \otimes \vec{r}) \quad (2)$$

Where $[\vec{r}]_\times$ is the cross product matrix of $\vec{r}$, $\otimes$ is the tensor product, and I is the identity matrix. Then, our warping matrix W is:

$$W = T_{P_{Shoulder}} R T_{-P_{Shoulder}} \quad (3)$$

Which represents a translation of the representation $\mathcal{A}$ to the origin so it is centered around the pivot point $P_{Shoulder}$, followed by the rotation R, and translating $\mathcal{A}$ back to its original position. Finally, we apply the warping matrix to each 3D point in the virtual representation of the arm:

$$\vec{p}_{warped} = W\vec{p}, \; \forall p \in \mathcal{A} \quad (4)$$

Figure 2C describes this process visually, highlighting in green the points that were affected by the warping transformation. In the next section, we introduce the user study, describe the evaluation prototype and discuss implementation details.

4 EVALUATION

To assess whether our approach improves the perception of pointing gestures in collaborative settings, we conducted a user study using pairs of participants. During the evaluation, participants were asked to alternate between the roles of pointer and observer. The main goal was to check how warping the pointer's arm would benefit the observer's attempts at extrapolating the target location. We also evaluated whether our warping technique was perceptible to the participants.

For this, we employed a fully-immersive virtual environment to accommodate participants in a side-by-side formation (at a distance of 2m from each other), facing the location were targets would appear. We followed the arrangement of participants and location of targets previously utilized by Herbort and Kunde [23]. However, in our evaluation, participants were immersed in a virtual environment, yet they could see each other's 3D avatars in real-time. Accordingly, we compared task performance and gathered user preferences in two conditions: (1) with Warping Deixis and (2) without Warping Deixis (baseline).

Procedure

Participants were asked to perform a set of three tasks for each of the two conditions. The order of conditions was counterbalanced between sessions to avoid biased results. All sessions followed the same structure: 1) an introductory briefing; 2) filling in a consent form and a profile questionnaire; 3) executing the tasks with the first condition; 4) filling a questionnaire for the first condition; 5) executing the tasks with the second condition; 6) filling the final questionnaire for the second condition. This took approximately 30 minutes in total.

We started by introducing the user study procedure to each pair of participants, followed by a description of the evaluation's main objective without revealing our body warping technique. Participants were only informed that the evaluation was a study on perception of pointing gestures. Each participant was then randomly assigned to their location, left or right in a side-by-side formation. Afterwards, participants jointly executed both sets of tasks.

Task execution for each condition was made up of two stages. In the first stage, the participant on the right initially assumed the role of observer, while the left participant was given referents to perform pointing gestures. Then, participants followed a set of three number identification tasks on a vertical pole at three different distances, similarly to Herbort and Kunde [23]. The second stage consisted of repeating the first stage, but with the roles of pointer and observer reversed. In the following, we detail the evaluation tasks.

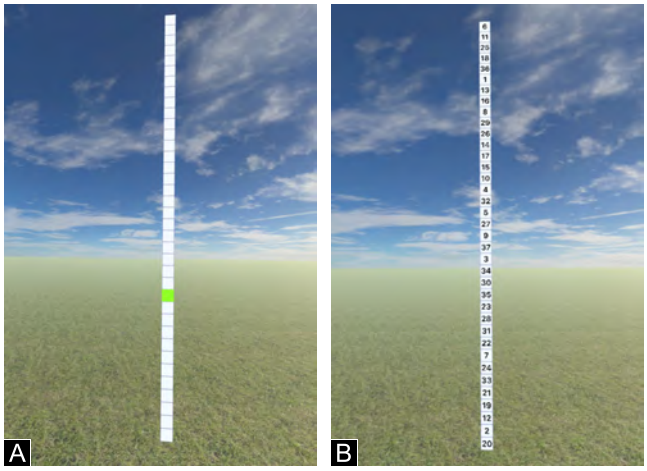

Figure 3: Pole task: A) from the point-of-view of the pointer, the pole featured blank squares and the target was highlighted in green; B) on the other hand, observers were unaware of the green target and the squares were numbered.

Tasks

For the purpose of this research, we reduced the need for supplementary verbal or contextual information as much as possible, since our objective relates to the accuracy of the information conveyed by the pointing gesture alone. Therefore, tasks were designed to not allow participants to use verbal descriptions to convey the location of the target referents, also, the participants were encouraged to not use speech communication and just perform a pointing gesture.

We replicated the numbered pole experiment from Herbort and Kunde [23] in a virtual environment. In all tasks, different information was presented individually to the pointer and the observer, as shown in Figure 3. When participants assume the roles of pointer and observer, they were asked to perform three tasks using a vertical numbered pole at different distances to the pointer: one, two and three meters (Figure 4). While the participant in a pointing role was presented with a highlighted target and no numbers, the observer was unaware of the target's location but could see the numbers. The observer was asked to report the referent's exact location based on how they interpret other participant's pointing gesture.

The pole consisted of a vertical numbered line with 37 white squares with black borders (8cm x 8cm), starting from the floor to 296cm of height. Thus, the vertical distance between the center of adjacent squares was 8cm. We doubled the square size used by [23] to improve the readability in Virtual Reality head-mounted displays, and the pole was positioned in front of the pointer. As shown in Figure 3A, the pointer's view of the pole consisted of blank squares with the referent highlighted in green. Pointers were instructed to point at the green square. On the other hand, the observer's view showed numbered white squares (Figure 3B). The numbers on the square labels were previously assigned to each square randomly and were used by the observer to report where the pointer was pointing to. Each pole task displayed numbered squares in a different order, and the top and bottom squares were excluded as referents.

Setup and Prototype

We configured the evaluation environment for both participants side by side in the same room. Each setup consisted of a desktop computer connected to an Oculus Rift headset as depicted in Figure 5. We used a non-intrusive open source toolkit [46] for body tracking, to combine skeleton information with the 3D representations of people in the same coordinate system. Our capture setup included two Microsoft Kinect v2 sensors mounted on tripods, 2m above the floor, facing down to ensure that pointing arms were always unobstructed during capture.

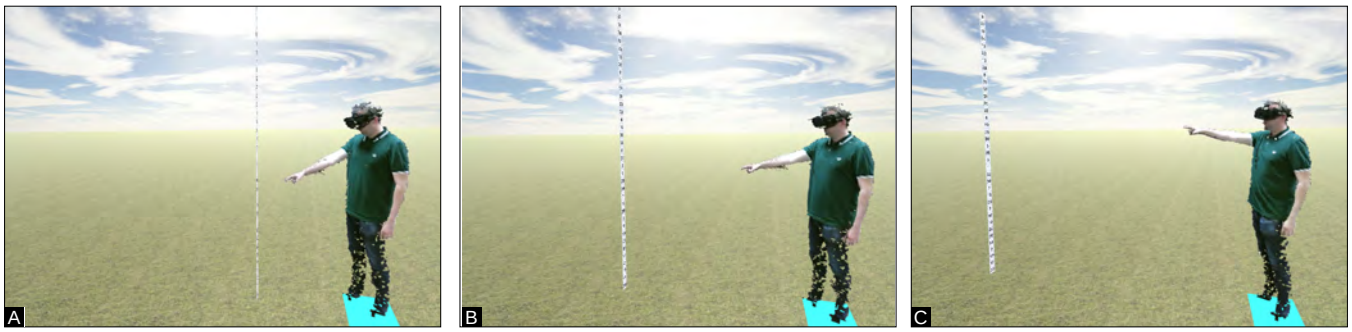

Figure 4: Participants were asked to perform three referent identification tasks in a vertical pole positioned at A) one meter, B) two meters and C) three meters.

We developed a prototype in Unity3D, and both setups were connected through a LAN evaluation server using TCP connections for both user's representation and the synchronization of the evaluation environment, as depicted in Figure 5. In the virtual environment, participants could see their own body representations and their partners standing to their side. The virtual representations in the virtual environment matched the real world position of the participants. The virtual environment also included visual indicators of the participants assigned positions, matching physical floor mats providing passive haptic feedback. A separate controlling application was used by the evaluation moderator to advance the tasks in both environments, instantiating targets and indicators accordingly, and setting the given answers during the pole tasks.

The participants' virtual representations were drawn as a 3D polygon mesh using color and depth values obtained from the depth cameras. Body warping was implemented in a vertex shader, applying Equation 4 to each point belonging to the arm. To predict the observer's interpreted location of referents, we employed the average values of w for Equation 1 provided by Herbort and Kunde [22] for side view gesture interpretation for each referent distance.

Warping can be triggered when someone is pointing to a target location or virtual object. In this case, smooth transitions can be applied to avoid gross discontinuities in arm movements. For the purpose of this evaluation, and since the only target consisted of one pole, we employed a collider much larger than the pole (four meters height and a width of two meters), which triggered the warping as soon as the participant raised an arm. This triggering approach allowed for the warping to start earlier and gradually. However, this strategy would need to be refined for virtual environments with multiple targets.

In regards to warping virtual representations, whenever a participant pointed to the target area, the shader would be updated with the relevant skeleton joint positions. To determine what point-cloud elements would be warped, our approach selected the points that were contained within a bounding volume, representing the person's arm. To determine that volume, we considered all space at the distance of

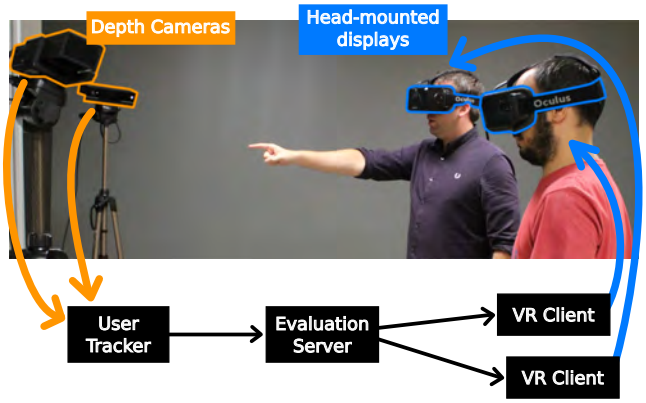

Figure 5: Evaluation setup with two user study participants and the prototype's architecture design.

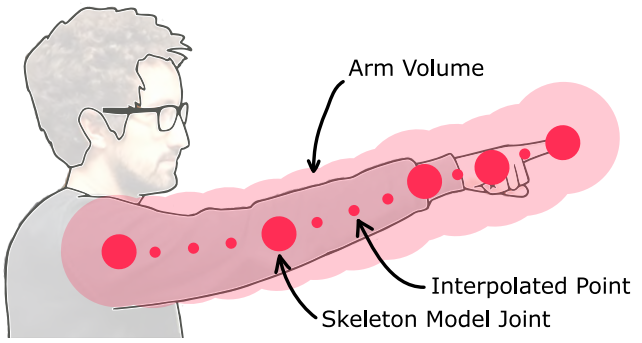

Figure 6: To identify the pointing arm 3D points to warp, we consider all points within a volume defined by a set of spheres centered across the arm skeleton model joints and other interpolated points between those joints.

15cm from the center of each Kinect skeleton model joints and interpolated bone positions calculated from increments of 5cm, as demonstrated in Figure 6.

Apparatus

The evaluation trials were performed in a controlled laboratory environment (Figure 5). All trials featured two moderators. One managed the evaluation server and guided the experiment, while another took notes and observed whether participants experienced any difficulty or discomfort. The server fired each trial and collected targets perceived by the observer (manually introduced by the first moderator). To collect targets, the second moderator used a scripted dialogue that required the observer to report and confirm the perceived target's number.

Each participant was instructed to stand on top of the floor mats positioned to match the positional indicators in the virtual environment. Participants were also instructed not to move around freely and keep to their assigned positions during each session.

Participants

Our subject group included 18 people (11 male, 7 female), organized in pairs. While participants' ages ranged from 18 to 44 years, most (14) were between 18 and 25 years old. All reported having previous usage experience in Virtual Environments.

5 RESULTS AND DISCUSSION

During the evaluation sessions we collected *Task Performance* data through logging, and *User Preferences* from questionnaires completed after finishing each set of tasks under both conditions.

Task Performance

We measured participants' task performance using the distance between the task's target, as indicated by the pointer, and the perceived target reported by the observing participant. Similarly to Herbort and Kunde [23], we measured distances between the centers of task targets and perceived target squares on the virtual pole, and then converted these to meters. Figure 7 shows the logged mean error distances of the observers for each task under both evaluation conditions.

We performed a two-way repeated measures ANOVA to assess how the independent variables, pole distance and technique, affected the perceived distance to the target. Pole distance included three levels (1, 2 and 3 meters) and technique consisted of two levels (baseline and Warping Deixis). All effects were statistically significant at the 0.05 significance level. The main effect for distance yielded an F ratio of $F(2, 34) = 60.325, p < .0005, \eta_p^2 = .780$. Post-hoc Paired T-Tests revealed significant differences between 1m

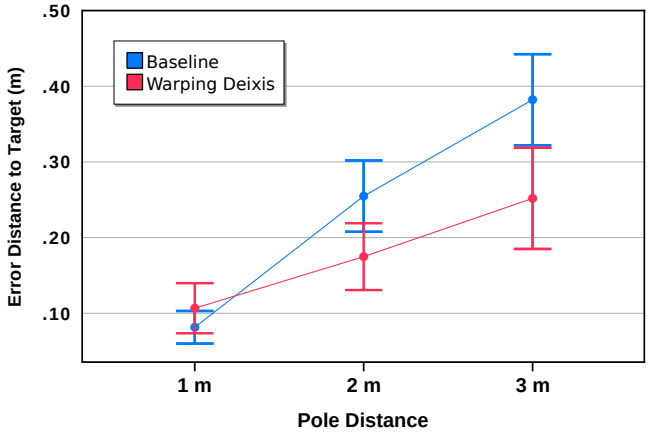

Figure 7: Task performance results for each condition: mean and 95% confidence interval error bars.

($M = .094, SD = .009$), and 2m ($M = .215, SD = .016, t(35) = -6.726, p < .0005, d = -1.121$), between 1 and 3m ($M = .317, SD = .023, t(35) = -8.972, p < .0005, d = -1.495$), and between 2m and 3m ($t(35) = -7.122, p < .0005, d = -1.187$).

The main effect for technique yielded an F ratio of $F(1, 17) = 5.753, p = .025, \eta_p^2 = .253$, indicating a significant difference between baseline ($M = .240, SD = .018$) and Warping Deixis ($M = .178, SD = .017$). The interaction effect was significant, $F(2, 34) = 16.747, \eta_p^2 = .496$.

Post-hoc tests, using a Paired T-Test with Holm-Bonferroni correction (Table 1), revealed no statistically significant difference between our approach and the baseline condition in the first task (1m to pole), whereas for the other tasks (2m and 3m to pole), Warping Deixis (1m: $M = .107m, SD = .067$; 2m: $M = .174m, SD = .089$; 3m: $M = .252m, SD = .135$) successfully improved the observers' perception in comparison to the baseline (1m: $M = .085m, SD = .043$; 2m:

Comparison	*t*	*df*	*p*	*d*	*α*
BL 1m - WD 1m	-1.327	17	0.202	-0.312	0.05
BL 2m - WD 2m	2.671	17	0.016 *	0.629	0.017
BL 3m - WD 3m	3.386	17	0.004 *	0.798	0.01
BL 1m - BL 2m	-9.331	17	<0.0005 *	-2.199	0.006
BL 1m - BL 3m	-11.212	17	<0.0005 *	-2.642	0.006
BL 2m - BL 3m	-8.57	17	<0.0005 *	-2.019	0.007
WD 1m - WD 2m	-2.66	17	0.016 *	-0.627	0.025
WD 1m - WD 3m	-4.354	17	<0.0005 *	-1.026	0.008
WD 2m - WD 3m	-3.277	17	0.004 *	-0.772	0.013

Table 1: Statistical tests reported at $p = .05$ significance levels (BL: baseline, WD: Warping Deixis). $*$ denotes statistical significance compared to the Holm-Bonferroni corrected α value.

Question	*Warping Deixis*	*Baseline*
Q1. I felt present in the Virtual Environment.	5 (1.5)	5 (1.5)
Q2. I felt that my colleague was present in the Virtual Environment.	5 (1.75)	5 (1)
Q3. I felt that I was pointing to were I wanted to point.	4.5 (1)	5 (0.75)
Q4. It was easy to understand where my colleague was pointing to.	4 (0)	4 (1.75)

Table 2: Results for the user preference questionnaires (Median, Inter-quartile Range).

$M = 0.255m, SD = .094$; 3m: $M = 0.382m, SD = .121$). At one meter, the pointer's index finger is so close to the referent that our approach yields no significant gain. This result agrees with the findings at one meter reported by Herbort and Kunde [23]. However, for either technique, longer distances to the pole significantly increase the error to the perceived target as shown on the last three lines of Table 1. For these, Warping Deixis shows significant advantage.

User Preferences

After completing each set of tasks, participants were asked to fill in a preferences questionnaire related to the condition they had just experienced. One of our key goals was to assess whether warping was perceived by the observers. The questionnaire included statements scored on a 6-point Likert Scale where a value of 1 meant that users did not agree at all with a statement and 6 meant that they fully agreed with it. Table 2 shows posed questions and corresponding results for both conditions.

For all questions, the Wilcoxon-Signed Ranks test revealed no statistically significant differences between the baseline and Warping Deixis conditions. This suggests that our approach warped the arm in a convincing manner, since participants did not seemingly distinguish any morphological changes in the body representations of their companions.

In addition, the questionnaire also featured the open question: "Did you find anything strange about your partner's body representation in the Virtual Environment? If so, please state what.". Seven participants reported the somewhat noisy representations caused by the depth sensor for both conditions. However, they did not report anything specifically relatable to the Warping Deixis condition, reinforcing that avatar distortion was not noticeable.

6 LIMITATIONS

From the findings, as revealed by the evaluation, we conclude that Warping Deixis demonstrates a significant improvement in the interpretation of deictic gestures to distal referents in a Virtual Reality environment. The tasks presented show that observers benefit from our warping technique when interpreting the referents located two and three meters in front of the person performing the pointing gesture. Still, our approach has some limitations.

The employed Bayesian model only considers the vertical axis to extrapolate the observers' interpretations of pointing gestures. In this research we focused on improving the accuracy of the vertical component, because misunderstandings occur consistently due to the elevation of the arm [4, 22, 53]. Furthermore, arm elevation is not only relevant to indicate referents in a vertical plane, but also is useful to refer to objects at different depths/distances. Yet, previous research suggests that human vector extrapolation is often biased toward both the vertical and horizontal axis [8]. Further research is necessary to assess the benefits of using a Bayesian correction approach to horizontally distributed referents.

In our evaluation prototype, we employed a virtual representation of the participants based on point cloud data converted to a textured mesh, using data from commodity depth cameras. Our approach showed some noisy contours, especially in parts of the participants' body that were not facing the depth cameras. Some participants reported this, although none suggested that the issue affected the experience. In future research, more accurate representations of people should be used to assess body warping techniques. One might argue that camera noise had the positive effect of masking distortions induced by warping limbs during deictic gestures. A more accurate representation might require more work to make geometric distortions imperceptible.

Finally, our approach provides the means to reduce the ambiguity of deictic gestures but does not allow for precise identification of referents. Indeed, if the evaluation participants were able to use verbal communication to resolve target misunderstandings, tasks would require considerably longer periods of time to be accomplished and the identification of referents would be more exact. Yet, pointing gestures also function as a complement to speech, when verbal communication combined with deictic references is difficult [22]. Furthermore, pointing gestures to ambiguous referents require longer verbal descriptions than unambiguous ones [3], allowing people in collaborative environments to become more focused on domain tasks and less involved in the tasks of maintaining the collaboration [19].

7 CONCLUSIONS AND FUTURE WORK

In this paper we introduced Warping Deixis, a body distortion approach to improve the perception of pointing gestures in virtual collaborative environments. The effectiveness of the technique, is backed by an experimental evaluation as compared to a baseline condition. To this end, we compared our warping method with not applying it at all in a series of tasks to identify referents on a numbered pole. We devised a virtual environment where two participants alternately

assumed roles of pointer and observer. Results suggest that Warping Deixis is successful at reducing the ambiguity of pointing gestures. Furthermore, people failed to notice the effects of our body warping approach when interpreting pointing gestures and arm motions.

Our results, beyond suggesting that Warping Deixis can improve collaboration in XR scenarios, also indicate that retargeting pointing gestures could benefit future human-technology approaches. Environments that exploit avatar-like or real-time 3D reconstructions of people are not the only systems that would benefit from retargeting the direction of pointing gestures, since any setting relying on the interpretation of *deixis* to interact with humans, currently suffers from the misunderstandings and miscommunication previously described. Thus, improvements in retargeting pointing gestures should enhance the effectiveness of virtual humanoid companions and non-player characters (NPCs) [34] in virtual environments, as well as, physical robot instructors and guiding helpers [9, 15, 31, 44, 47].

As for future work, we intend to further our research on improving the perception of gestures on collaborative environments. Namely, we plan to study Warping Deixis to improve workspace awareness in 3D collaborative task spaces with more than two collaborators. For this, we propose exploring a broader set of collaborative tasks. Also, a general strategy to trigger body warping for virtual environments with multiple targets, should be the subject of future research. In contrast to real life settings, Extended Reality Environments support other means to resolve pointing inaccuracies, including virtual objects (halos, light rays, etc) to enhance deixis. However, we might argue that conflicting indications such as misunderstood gestures could decrease the effectiveness of these enhancers. In the future, we plan to evaluate such observations via additional user experiments. It might also be useful to examine additional contextual, gesture or pose recognition approaches to trigger avatar warping. Another promising direction is to study other forms of body warping, focusing on ensuring that manipulated actions do not force people to convey different meanings than they originally intended and further explore deixis warping in real use case scenarios. Furthermore, since the predictive Bayesian model from Herbort and Kunde [22] is limited to a narrow set of arrangements of people, it would be interesting to explore machine learning approaches to dynamically predict the tendency of people to rely exclusively on either geometric extrapolation or on a-priori assumptions (parameter w of Equation 1), for different group formations and distances in proxemic interactions.

ACKNOWLEDGEMENTS

This work was partially supported by FCT, through grants IT-MEDEX PTDC/EEISII/6038/2014 and UID/CEC/50021/2019, and the European Research Council under the project Ref. 336200.

REFERENCES

[1] Ferran Argelaguet and Carlos Andujar. 2009. Visual feedback techniques for virtual pointing on stereoscopic displays. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. ACM, 163–170.

[2] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. 2016. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1968–1979.

[3] Adrian Bangerter. 2004. Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science* 15, 6 (2004), 415–419.

[4] Adrian Bangerter and Daniel M Oppenheimer. 2006. Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture* 6, 1 (2006), 85–102.

[5] Stephan Beck, Andre Kunert, Alexander Kulik, and Bernd Froehlich. 2013. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (2013), 616–625.

[6] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowdon. 1995. User Embodiment in Collaborative Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 242–249. https://doi.org/10.1145/223904.223935

[7] Mark Billinghurst and Hirokazu Kato. 1999. Collaborative mixed reality. In *Proceedings of the First International Symposium on Mixed Reality*. 261–284.

[8] H Bouma and JJ Andriessen. 1968. Perceived orientation of isolated line segments. *Vision Research* 8, 5 (1968), 493–507.

[9] Paul Bremner and Ute Leonards. 2016. Iconic gestures for robot avatars, recognition and integration with speech. *Frontiers in psychology* 7 (2016), 183.

[10] George Butterworth. 2003. Pointing is the royal road to language for babies. In *Pointing*. Psychology Press, 17–42.

[11] George Butterworth and Shoji Itakura. 2000. How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology* 18, 1 (2000), 25–50.

[12] Bill Buxton. 2009. *Mediaspace – Meaningspace – Meetingspace*. Springer London, London, 217–231. https://doi.org/10.1007/978-1-84882-483-6_13

[13] William Buxton. 1992. Telepresence: Integrating shared task and person spaces. In *Proceedings of graphics interface*. 123–129.

[14] Thierry Duval, Thi Thuong Huyen Nguyen, Cédric Fleury, Alain Chauffaut, Georges Dumont, and Valérie Gouranton. 2014. Improving awareness for 3D virtual collaboration by embedding the features of users' physical environments and by augmenting interaction tools with cognitive feedback cues. *Journal on Multimodal User Interfaces* 8, 2 (2014), 187–197.

[15] Felix Faber, Maren Bennewitz, Clemens Eppner, Attila Gorog, Christoph Gonsior, Dominik Joho, Michael Schreiber, and Sven Behnke. 2009. The humanoid museum tour guide Robotinho. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*. IEEE, 891–896.

[16] Susan R Fussell, Leslie D Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam DI Kramer. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction* 19, 3 (2004), 273–309.

[17] Saul Greenberg, Carl Gutwin, and Andy Cockburn. 1996. Awareness through fisheye views in relaxed-WYSIWIS groupware. In *Graphics interface*, Vol. 96. 28–38.

[18] Saul Greenberg, Carl Gutwin, and Mark Roseman. 1996. Semantic telepointers for groupware. In *Computer-Human Interaction, 1996. Proceedings., Sixth Australian Conference on*. IEEE, 54–61.

[19] Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)* 11, 3-4 (2002), 411–446.

[20] Carl Gutwin, Saul Greenberg, and Mark Roseman. 1996. Workspace awareness in real-time distributed groupware: Framework, widgets, and evaluation. In *People and Computers XI*. Springer, 281–298.

[21] Edward Twitchell Hall. 1966. *The hidden dimension*. Doubleday & Co.

[22] Oliver Herbort and Wilfried Kunde. 2016. Spatial (mis-) interpretation of pointing gestures to distal referents. *Journal of Experimental Psychology: Human Perception and Performance* 42, 1 (2016), 78.

[23] Oliver Herbort and Wilfried Kunde. 2018. How to point and to interpret pointing gestures? Instructions can reduce pointer–observer misunderstandings. *Psychological research* 82, 2 (2018), 395–406.

[24] Dongsik Jo, Ki-Hong Kim, and Gerard Jounghyun Kim. 2016. Effects of avatar and background representation forms to co-presence in mixed reality (MR) tele-conference systems. In *SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*. ACM, 12.

[25] Ricardo Jota, Miguel A Nacenta, Joaquim A Jorge, Sheelagh Carpendale, and Saul Greenberg. 2010. A comparison of ray pointing techniques for very large displays. In *Proceedings of graphics interface 2010*. Canadian Information Processing Society, 269–276.

[26] Sotaro Kita. 2003. *Pointing: Where language, culture, and cognition meet*. Psychology Press.

[27] David C Knill and Alexandre Pouget. 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences* 27, 12 (2004), 712–719.

[28] Konrad P Körding and Daniel M Wolpert. 2004. Bayesian integration in sensorimotor learning. *Nature* 427, 6971 (2004), 244.

[29] Robert M Krauss and Susan R Fussell. 1990. Mutual knowledge and communicative effectiveness. *Intellectual teamwork: Social and technological foundations of cooperative work* (1990), 111–146.

[30] Sangyoon Lee, Jinseok Seo, Gerard Jounghyun Kim, and Chan-Mo Park. 2003. Evaluation of pointing techniques for ray casting selection in virtual environments. In *Third international conference on virtual reality and its application in industry*, Vol. 4756. International Society for Optics and Photonics, 38–45.

[31] Karina R. Liles, Clifton D. Perry, Scotty D. Craig, and Jenay M. Beer. 2017. Student Perceptions: The Test of Spatial Contiguity and Gestures for Robot Instructors. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. ACM, New York, NY, USA, 185–186. https://doi.org/10.1145/3029798.3038297

[32] Andrew Maimone and Henry Fuchs. 2011. Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*. IEEE, 137–146.

[33] David McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago press.

[34] Tsukasa Noma, Liwei Zhao, and Norman I. Badler. 2000. Design of a Virtual Human Presenter. *IEEE Comput. Graph. Appl.* 20, 4 (July 2000), 79–85. https://doi.org/10.1109/38.851755

[35] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 405–415.

[36] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 741–754. https://doi.org/10.1145/2984511.2984517

[37] Thomas Pechmann and Werner Deutsch. 1982. The development of verbal and nonverbal devices for reference. *Journal of experimental child psychology* 34, 2 (1982), 330–341.

[38] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1716–1725. https://doi.org/10.1145/2818048.2819965

[39] Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. ACM, 16.

[40] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 46, 13 pages. https://doi.org/10.1145/3173574.3173620

[41] Thammathip Piumsomboon, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. CoVAR: a collaborative virtual and augmented reality system for remote collaboration. In *SIGGRAPH Asia 2017 Emerging Technologies*. ACM, 3.

[42] Ann D Salomon. 1947. Visual field factors in the perception of direction. *The American journal of psychology* 60, 1 (1947), 68–88.

[43] Chris L Schmidt. 1999. Adult understanding of spontaneous attention-directing events: What does gesture contribute? *Ecological Psychology* 11, 2 (1999), 139–174.

[44] Zhuoyu Shen and Yan Wu. 2016. Investigation of Practical Use of Humanoid Robots in Elderly Care Centres. In *Proceedings of the Fourth International Conference on Human Agent Interaction (HAI '16)*. ACM, New York, NY, USA, 63–66. https://doi.org/10.1145/2974804.2980485

[45] Rajinder S Sodhi, Brett R Jones, David Forsyth, Brian P Bailey, and Giuliano Maciocci. 2013. BeThere: 3D mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 179–188.

[46] Maurício Sousa, Daniel Mendes, Rafael Kuffner Dos Anjos, Daniel Medeiros, Alfredo Ferreira, Alberto Raposo, João Madeiras Pereira, and Joaquim Jorge. 2017. Creepy Tracker Toolkit for Context-aware Interfaces. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17)*. ACM, New York, NY, USA, 191–200. https://doi.org/10.1145/3132272.3134113

[47] Osamu Sugiyama, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, Norihiro Hagita, and Yuichiro Anzai. 2006. Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection science* 18, 4 (2006), 379–402.

[48] Anthony Tang, Carman Neustaedter, and Saul Greenberg. 2007. Videoarms: embodiments for mixed presence groupware. In *People and Computers XX—Engage*. Springer, 85–102.

[49] Anthony Tang, Michel Pahud, Kori Inkpen, Hrvoje Benko, John C Tang, and Bill Buxton. 2010. Three's company: understanding communication channels in three-way distributed collaboration. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*.

ACM, 271–280.

[50] Janet L Taylor and DI McCloskey. 1988. Pointing. *Behavioural Brain Research* (1988).

[51] Michael Tomasello, Malinda Carpenter, and Ulf Liszkowski. 2007. A new look at infant pointing. *Child development* 78, 3 (2007), 705–722.

[52] Stephen Voida, Mark Podlaseck, Rick Kjeldsen, and Claudio Pinhanez. 2005. A study on the manipulation of 2D objects in a projector/camera-based augmented reality environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 611–620.

[53] Marta Wnuczko and John M Kennedy. 2011. Pivots for pointing: Visually-monitored pointing has higher arm elevations than pointing blindfolded. *Journal of Experimental Psychology: Human Perception and Performance* 37, 5 (2011), 1485.

[54] Nelson Wong and Carl Gutwin. 2010. Where are you pointing?: the accuracy of deictic pointing in CVEs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1029–1038.

[55] Nelson Wong and Carl Gutwin. 2014. Support for deictic pointing in CVEs: still fragmented after all these years'. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 1377–1387.

[56] Erroll Wood, Jonathan Taylor, John Fogarty, Andrew Fitzgibbon, and Jamie Shotton. 2016. Shadowhands: High-fidelity remote hand gesture visualization using a hand tracker. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*. ACM, 77–84.