Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality

Sven Mayer[1,2], **Jens Reinhardt**[3], **Robin Schweigert**[2], **Brighten Jelke**[4], **Valentin Schwind**[2,5], **Katrin Wolf**[3], **Niels Henze**[5]
[1] Carnegie Mellon University, Pittsburgh, Pennsylvania, United States, info@sven-mayer.com
[2] University of Stuttgart, Stuttgart, Germany, {firstname.lastname}@vis.uni-stuttgart.de
[3] Hamburg University of Applied Sciences, Hamburg, Germany, {firstname.lastname}@haw-hamburg.de
[4] Macalester College, Saint Paul, Minnesota, United States, bjelke@macalester.edu
[5] University of Regensburg, Regensburg, Germany, {niels.henze, valentin.schwind}@ur.de

ABSTRACT
Collaborative Virtual Environments (CVEs) offer unique opportunities for human communication. Humans can interact with each other over a distance in any environment and visual embodiment they want. Although deictic gestures are especially important as they can guide other humans' attention, humans make systematic errors when using and interpreting them. Recent work suggests that the interpretation of vertical deictic gestures can be significantly improved by warping the pointing arm. In this paper, we extend previous work by showing that models enable to also improve the interpretation of deictic gestures at targets all around the user. Through a study with 28 participants in a CVE, we analyzed the errors users make when interpreting deictic gestures. We derived a model that rotates the arm of a pointing user's avatar to improve the observing users' accuracy. A second study with 24 participants shows that we can improve observers' accuracy by 22.9%. As our approach is not noticeable for users, it improves their accuracy without requiring them to learn a new interaction technique or distracting from the experience.

Author Keywords
Deictic; ray tracing; virtual reality; correction model.

CCS Concepts
•Human-centered computing → Human computer interaction (HCI); Computer supported cooperative work;

INTRODUCTION
Collaborative Virtual Environments (CVEs) are virtual environments where multiple users connected by a network can meet, collaborate, and work [30]. Already proposed in the early 1990s, CVEs have been one of the earliest use cases for virtual reality (VR) [15, 17, 60] as the result of a convergence of research on VR and computer-supported cooperative work (CSCW) [6]. CVEs enable multiple users to work, play, and learn together while being in different locations. Due to the virtual environment, users can choose any environment that suits the task and the visual embodiment they prefer.

VR enables to go beyond the physical constraints of the real world. Both the avatar the virtual representation of the user and the virtual world are rendered in 3D and do not necessarily need to follow physical laws of the real world. Avatars, for example, can have a different virtual appearance than the own body and users can change their avatar as they have different preferences [53]. Due to the dominance of vision, it is even possible to alter the avatar pose without being noticed by the user [57]. That enables redirecting the user's movement, for example, to repurpose passive haptics of a physical object to simulate multiple virtual objects [2, 16].

A body of previous works aimed to optimize task performance in VR by changing the user's appearance but did not fully embrace the possibilities of VR. In VR, users could overcome limitations they face in the physical world. We are especially interested in improving users' ability to interpret deictic gestures as they are one of the most fundamental forms of non-verbal communication [14, 29, 33]. They are particularly important for CSCW applications in VR because, for example, they allow users to indicate objects they are talking about. Ultimately, such experience could be even more efficient and more enjoyable than the interaction in the physical world. As Mayer et al. [43, 44] showed that pointing in not accurate to overcome this, we envision that in VR pointing can be naturally improved. Here, Greenberg et al. [23] investigated displaying a reference cursor to enhance collaboration and accuracy. Moreover, Wong and Gutwin [67] investigated a wide range of visualization techniques to foster a better collaboration such as long arm, highlighting, and ray casting. On the other hand, Mayer et al. [43] showed that presenting feedback increases the selection time and will, therefore, break the conversation flow in a cooperative task [42]. Finally, Sousa et al. [57] recently presented a first step toward enhancing humans' ability to interpret deictic pointing gestures without adding feedback. They altered the pointing user's body posture by naturally rotating the arm of the pointer. Because the model proposed by

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376340

Sousa et al. [57] only accounted for vertical errors, they could only show a higher accuracy in one dimension. However, the models for 2D pointing tasks proposed by Mayer et al. [43, 44] suggest that there are pitch (up-down) and yaw (left-right) errors.

In this paper, we show how to improve the interaction between users in CVEs by increasing the accuracy while interpreting deictic gestures. In the first study, we confirm previous findings [57] and extend these by showing that users make systematic two-dimensional errors when interpreting pointing gestures. Our results suggest that, both the limited accuracy when performing deictic gestures as well as the limited accuracy when interpreting them contribute to the total error. We use the collected data to develop a model that rotates the arm of a pointing user's avatar to counteract this error. We show that while controlling the distance between the two users, the 2D pointing tasks are indeed subject to error in, both pitch and yaw rotation. Based on our first findings, we present a new general pointing correction model for deictic gestures. As the model is only applied to the perspective of the user interpreting the deictic gesture, it cannot affect the immersion of the pointer. Through a second study, we show that the developed approach significantly improves a user's accuracy when interpreting deictic gestures. Thus, our model provides improved capabilities to interpret deictic gestures in collaborative VR settings.

RELATED WORK

CVEs have been one of the early drivers for VR. They can enable natural communication between remote users. One of the most fundamental forms of non-verbal communication is deictic gestures. In the following, we first discuss the opportunities provided by VR and CVEs. Afterward, we provide an overview of work on deictic pointing and how the systematic errors that humans make can be compensated. Finally, we outline previous work on manipulating avatars' pose to provide new VR experiences.

Collaborative Virtual Environments

A VR is a simulated environment in which a user experiences presence [36, 50, 66] using a communication medium [59]. This occurs when a person's perception ignores the existence of the mediating technology and experiences the feeling of *being* or *acting* in a place, even when one is physically situated in another location [4, 48]. Users are immersed in virtual environments that can be similar to the real world but can also go beyond what is possible in the real world. Groom et al. [26], for example, proposed strategies to reduce racial prejudice by changing the user's appearance. Due to the underlying virtual environment, users or developers can choose the environment that suits the tasks and the virtual embodiment they prefer.

CVEs are virtual environments where multiple users are present at the same time [30]. Already proposed in the early 1990s, CVEs have been one of the early use cases for VR [15, 17, 60]. They have been seen as the result of a convergence of research on VR and CSCW [6]. CVEs enable multiple users to work, play, and learn in the same (virtual) environment while being physically distributed around the globe. CVEs have been even proposed as a collaborative space for researchers [8].

CVEs have been used to study various situations such as classroom situations [19] or to train first responders [7]. Greenwald et al. use CVEs for creating and manipulating virtual objects in multi-user scenarios [24] and collaborative learning [25]. The possibilities to recreate scenarios are endless. Consequently, their design has extensively been studied [5, 62, 49]. However, CVEs can also go beyond just mimicking the real world and have been used to enable users or developers to alter humans' perception [26, 41].

In summary, VR can enable immersive interactive systems. Particularly CVEs have the potential to improve human collaboration. Previous work mainly focused on replicating the real world with the highest fidelity. CVEs provide not only control over the virtual environment but also over the users' virtual bodies. Thus, CVEs have the potential to enable human collaboration beyond what is possible in the real world.

Deictic Pointing Gestures

After a child's first year, they develop the ability to point at distant objects using their hands and fingers, a behavior that is called "deictic gestures" [11, 13, 33]. Deictic gestures are fundamental to direct others' attention to objects and other beings and help develop a joint understanding of objects in space [1, 12]. It has been shown that the ability to perform deictic gestures is linked to developing an understanding of others' intentions [14]. Deictic gestures are typically performed by extending the arm and the index finger [3, 43, 61].

A large body of work from psychology investigated how humans point at objects in a distance. This research showed that humans' accuracy when pointing at remote objects with their hands or with tools is limited [21, 38]. This limited accuracy is a fundamental challenge for human-computer interaction (HCI) as, starting with the seminal work by Bolt [9], mid-air pointing has been frequently proposed as a natural way to interact with computing systems. Consequently, work in the CVE domain investigated how to design visual feedback to support users' ability to interpret the deictic gesture. Here, Wong and Gutwin [67] proposed visual indicators, such as an extended arm, a laser beam, or highlighting the targeted object to support deictic gestures in CVEs. However, Wachs et at. [64] showed that visual feedback can have negative effects on immersion and on cognitive workload. Previous work from HCI showed that pointing in VR, just as in the real world, has limited accuracy [40, 43, 44] when interacting with a system. Moreover, for collaborative tasks, the judgment of a deictic direction by an observer is fundamental for a positive outcome [22, 28, 58]. However, humans lack the ability to interpret deictic gestures precisely [34, 51]. Herbort and Kunde [34] model humans' limitation and present a model that can explain 91% of the variance. Based on the work by Herbort and Kunde, Sousa et al. [57] proposed an approach to improve humans' ability to interpret deictic gestures by rotating the arm to counteract the interpretation error. In contrast to work presented by Mayer et al. [43, 44], however, Herbort and Kunde [34, 35] as well as Sousa et al. [57] only analyzed the vertical errors by using targets within one defined column. Yet,

deictic gestures are 2D pointing tasks. Indeed, Bangerter and Oppenheimer [3], for example, showed that the accuracy for center targets and targets to the horizontal extremes differs.

Manipulating Avatars' Pose

VR theoretically offers the opportunity to create a virtual world "better" than the real one. While the laws of physics restrict human's actions in the real world, virtual worlds do not necessarily impose the same restrictions, such as displaying different arm visualizations for each person. Avatars do not have to be digital twins of the physical person they are representing. Avatars can look differently [52, 53], which can change how users type [39], perform target selection [54], or perceive surfaces [55]. They can also have skills the real person does not. Avatars can fly [63], shrink and grow [46], be teleported [10], or enable to look through the eyes of others [27]. They might even be able to perfectly point at objects so that their communication partner perfectly understands at which target they are pointing.

Previous work manipulated avatars' pose to overcome the restrictions imposed by the real world. Kasahara et al. [37] experimented with slightly deformed avatars. They found that the spatial-temporal deformation of an avatar can change how users perceive their own bodies. Predicting the user's pose in the near future and using the prediction when rendering the avatar instead of the actual pose makes the user feel lighter. Feuchtner and Müller [20] propose to control distant objects with a long virtual arm. They found that the virtual arm can be stretched to more than twice its real length without breaking the user's sense of ownership for the virtual limb. Azmandian et al. [2] exploit the dominance of vision when a user's senses conflict to realize what the authors call "haptic retargeting." When a user grabs an object, the position of the hand is dynamically altered which enables repurposing passive haptics from the same physical object for multiple virtual objects. Cheng et al. [16] extended this work by generalizing haptic retargeting. They analyzed a user's gaze and hand motions, and redirected their hand to a matching part of a sparse haptic proxy. Finally, Sousa et al. [57] proposed a body warping algorithm to improve pointing for 1D pointing tasks.

Overall, VR enables to overcome the restrictions of the real world. Avatars can look widely different than the user and not even necessarily like a human being. VR also enables spatial-temporal deformation of a user's avatar. Previous work focused on changing how users perceive their own avatar. To improve the interpretation of deictic gestures in CVE it is, however, necessary to change how other users perceive other users' avatars.

Summary

In summary, VR enables new communication pathways between users. While deictic gestures are one of the most fundamental forms of non-verbal communication, humans' precision when performing them is limited. To improve the precision when interacting with computing systems, a number of approaches have been developed. However, they cannot improve users' ability to interpret deictic gestures or distract from the gesture itself through additional visualizations. A potential approach to improve the interpretation of deictic gestures is manipulating the pose of the users' avatar. In contrast to previous work, however, it is necessary to manipulate the pose in two dimensions.

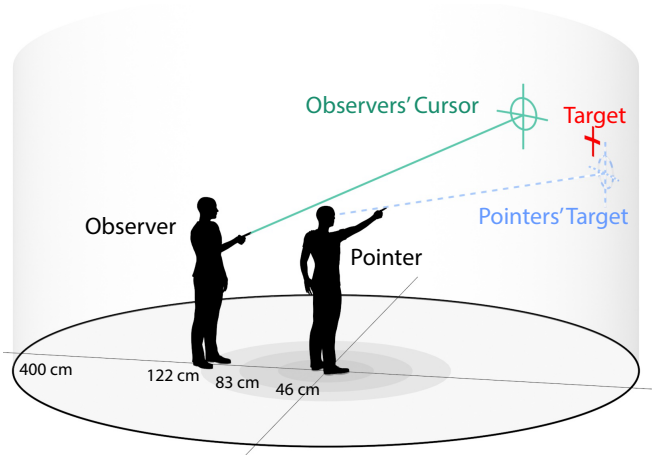

Figure 1. Concept illustration shows a cylinder with the pointer in the center and the observer in a DISTANCE of 122*cm*. The red target is only visible to the pointer. The blue dashed line represents the index finger ray cast (IFRC) of the pointer with the intersection of the cylinder at the end, which both are not visible to the participants. The green line represents the laser beam with its crosshair at the end, which is used for the observer to indicate the estimated pointing position of the pointer; this is only visible to the observer.

ERRORS WHEN INTERPRETING DEICTIC GESTURES

Previous work studying the errors made when interpreting deictic gestures only considered vertical errors. We, however, hypothesize that pointing is a task involving two rotation axes. Therefore, we conduct a study with two participants situated in a CVE. We asked one participant to estimate the position of targets another participant pointed at, see Figure 2. In the following, we refer to the participant who pointed at the targets as *Pointer* and the participant which estimated the target's position as the *Observer*.

In contrast to work presented by Sousa et al. [57], we hypothesize that pointing is a two-dimensional task and, therefore, the angle between target, *Pointer*, and *Observer* needs to be taken into account, see Figure 4. Moreover, Bangerter and Oppenheimer [3] showed that the accuracy is different for targets close to the horizontal extreme. Thus, the goal is to cover the full 360° around the participants while covering a large range in height. Thus, we are extending the work by Sousa et al. [57] by the rotation dimension.

Sousa et al. [57] showed that the distance between the *Pointer* has an effect on the *Observer*'s precision when estimating the target's position; thus, we systematically varied the distance between them. We choose the distances based on the proxemics zones by Hall [31]. According to Hall, the crossover from the intimate zone to the personal zone is 45*cm* from a person and the crossover from the personal zone to the social zone is 122*cm* from a person. In addition to these two distances, we use their center at 83*cm*. Therefore, we are spanning the range of distances where normal conversation

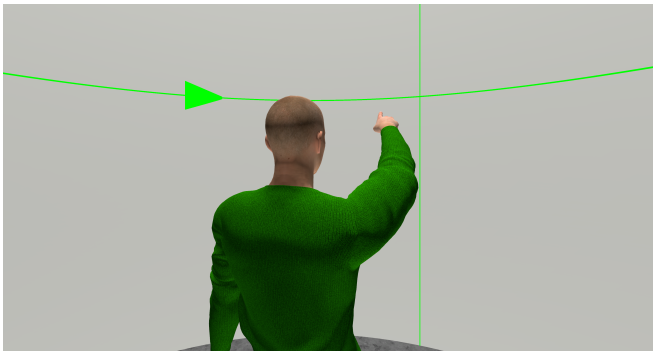

Figure 2. Third-person view of the pointer from the observer's perspective. The green arrows on the target indicate the rotation direction for the pointer to find the target's center faster. However, the target crosshair was never visible for the observer.

might occur while also investigating the extreme boundaries. Thus, we used a within-subject design with the independent variable DISTANCE. DISTANCE had the three levels, 46, 83, and 122cm, between pointer and observer.

For each of the three distances, a total of 80 targets were presented to the pointer. The targets were arranged in a 16×5 (ROTATION $\times$ HEIGHT) grid spanning the whole 360° rotation of the room, where the ROTATION variation results are projected every 22.5° and the variation of HEIGHT result in rows every .78m starting from one meter above the ground. The targets were only visible to the pointer.

Apparatus

We built a system in which two participants equipped with HTC Vives can interact with each other in the same VR scene. Each Vive was connected to a dedicated PC while participants' were shared via a network connection. We used a marker-based six degrees of freedom (6DOF) OptiTrack motion capture system to track the participants. The skeleton tracking provided by OptiTrack was used to track the upper body of the pointer. As the skeleton tracking delivers neither precise position nor orientation of the index finger, we followed the approach by Mayer et al. [43] to determine the different ray casts and attached an additional rigid body to the pointer's index finger. OptiTrack delivers the head rotation only with 240fps and has a higher latency than the rotation tracking of the Vive. Therefore, we determined participants' head rotation from the sensors of the Vive to counteract motion sickness. Using the OptiTrack system, we tracked both head-mounted displays (HMDs) using the Vive mount[1]. For the observer, we only tracked the head position using OptiTrack and the head rotation using the Vive. However, the observer was not rendered in VR. Additionally, we tracked a hand-held stick that the observer uses to describe the position the pointer indicated. In VR, the trajectory of the hand-held stick was shown as a laser beam with a crosshair at the end [67]. Both beam and crosshair were only visible to the observer.

Both participants were located in the same VR room, a cylinder 8m in diameter and 7m in height. The pointer was positioned in the center of the room. During the study, we positioned the observer relative to the pointer in a DISTANCE according to the respective condition. They were separately positioned on 1m high cylinders with a diameter of 1m. We designed the room such that the floor would not limit the observer in selecting a position. Participants were asked to stay in the center of their cylinder; however, the experimenter had the option to recenter the participants if they leave their initial position. The observer was represented through the human-like avatar used by Schwind et al. [56] as they have shown this avatar leads to better pointing performance than less human-like characters. Moreover, today's VR technologies often use avatars that lead to a good immersion [45] and full body representation leads to an improved awareness [5].

While Sousa et al. [57] showed that distant targets have a higher absolute error, the angular error remains the same and the absolute error is proportional [44]. Indeed, Mayer et al. [44] modeled the angular error and derived distance-invariant models. Thus, in our study, we use a fixed target distance.

Mayer et al. [44] showed that sitting and standing result in the same pointing error. Sousa et al. [57] showed that arm- and head-pose are the dominant means when interpreting a deictic gesture. Additionally, arm and finger relationship is the same while sitting or even walking. Thus, we only investigated participants standing next to each other. As the observer's behavior could influence the pointer's behavior, the observer's avatar was not rendered in VR.

As aim of this study was to investigate and later improve the aiming performance of the observer, the pointer has to make the same errors in the study as which wound naturally occur to enable us to model the error. This, then allowed the model to combine the error made by pointer and observer. We, therefore, cannot provide a visible feedback to the pointer as it would change the pointer's behavior and thus error rendering the model only valid in cases where feedback is provided.

Procedure

The experiment was conducted with two participants at the same time. One participant was always the pointer and the other always the observer. After welcoming the participants, we asked them to sign the consent form and to fill in a demographic questionnaire. We then explained the procedure of the study and brought both participants into the VR.

After a familiarization phase, we guided the participants through the study. Targets were shown to the pointer on the wall using a green crosshair while not being visible for the observer. We added arrows pointing towards the target (see Figure 2) so that the pointer quickly finds the target. The pointer was asked to point at the target with the index finger of the right hand. No further instructors where given on how to point. When the pointer was sure to point into the right direction, they confirmed their posture using a remote control in their left hand. The confirmation changed the color of the target to red indicating that no further action is required. However, the pointer had to hold the posture until the observer's turn was over. In that moment, the crosshair at the end of the laser beam coming out of the observer's stick turned from red to green, indicating that action of the observer is required.

[1] https://github.com/interactionlab/htc-vive-marker-mount

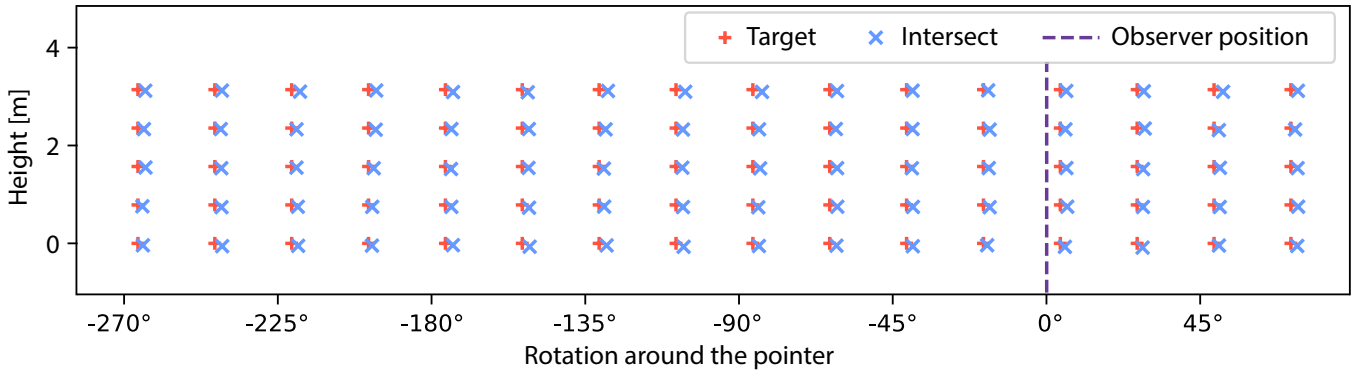

Figure 3. Offsets of the *Pointer* when applying the eye-finger ray cast (EFRC) direction estimation.

Afterwards, the observer had to estimate the target's position and indicate this using the laser beam. After positioning their crosshair, the observer had to confirm the position on a remote control with the left hand. The confirmation of the observer turned the observer's crosshair red again and showed the next target to the pointer.

This procedure was repeated 80 times until the pointer and observer had pointed at all 80 targets. After completing 80 trials for one DISTANCE, participants were asked to leave the VR and to fill a raw NASA-Task Load Index (raw TLX) [32] to observe possible fatigue effects and to give them a break to relax. When reentering the VR environment, they continued with the next DISTANCE.

We counter-balanced DISTANCE using a Latin Square design which accounts for first-order effects [65]. Further, we randomized the order of the targets. In total, each pair of participants had to point at 3 conditions $\times$ 80 targets = 240 targets. The time and effort of the participants was compensated with € 10.

Participants

We recruited 28 participants (19 male, and 9 female) via our institutions' mailing lists. Participants were between 19 and 41 years old ($M = 24.5$, $SD = 5.4$). As Plaumann et al. [47] showed a strong influence of handedness on the pointing performance, we only recruited right-handed participants who had no locomotor coordination problems and wore neither glasses nor contact lenses. We used the Porta test [18] to screen participants for eye-dominance: 23 participants had right-eye dominance, and 5 had left-eye dominance.

Results

We collected a total of 3,360 pointing postures of the pointer and additional 3,360 positions estimated by the observer. In line with previous work [3, 43, 61], participants performed the deictic gesture by extending the arm and the index finger to address the target in the distant. In the following, whenever Mauchly's test showed that the sphericity assumption was violated in the repeated measures analysis of variance (RM-ANOVA), we reported Greenhouse-Geisser (GG) or Huynh-Feldt (HF) corrected p-values.

Fatigue effects

First, we analyzed the raw TLX score (scale: 0-20) to determine if potential workload or fatigue effects had to be considered in the analysis. The mean raw TLX score was $M = 5.8$ ($SD = 2.3$) after the first, $M = 5.6$ ($SD = 2.7$) after the second, and $M = 5.4$ ($SD = 2.8$) after the last distance. We conducted a two-way RM-ANOVA to reveal significant differences between the pointer and the observer or if the round significantly influenced the work load. The analysis did not reveal a significant difference between the pointer and the observer ($F_{1,26} = .595$, $p = .447$, $\eta^2 = .019$) and no effect of round on the work load ($F_{1.604,41.699} = 1.001$, $p = .364$, $\eta^2 = .005$). The analysis also did not reveal a significant interaction effect ($F_{1.604,41.699} = .257$, $p = .725$, $\eta^2 = .001$). The mean results are similar to the ones reported by Mayer et al. [44, 43] for a similar pointing task. We assume that effects caused by fatigue or workload are negligible.

Accuracy of the Pointer

Previous work [43] showed a systematic effect for the commonly used ray casting methods when systems estimate the direction of a user. In the following, we investigate if the systematic error is also present in a cylinder, which has not been studied before. Therefore, we used the same four ray

Pointer	M	SD
Eye-finger ray cast (EFRC)	40.0cm	8.5
Forearm ray cast (FRC)	498.4cm	83.1
Head ray cast (HRC)	172.7cm	53.0
Index finger ray cast (IFRC)	284.1cm	175.0

Table 1. Average error for the four ray casting methods for the targets at a distance of $4m$ from the pointer.

Observer	M	SD
Distance 46cm	96.8cm	24.8
Distance 83cm	103.1cm	27.3
Distance 122cm	102.9cm	25.4
Average	100.9cm	25.8

Table 2. Observers' average error when estimating the position of the target at which the pointer points. The targets had a distance of $4m$ from the pointer.

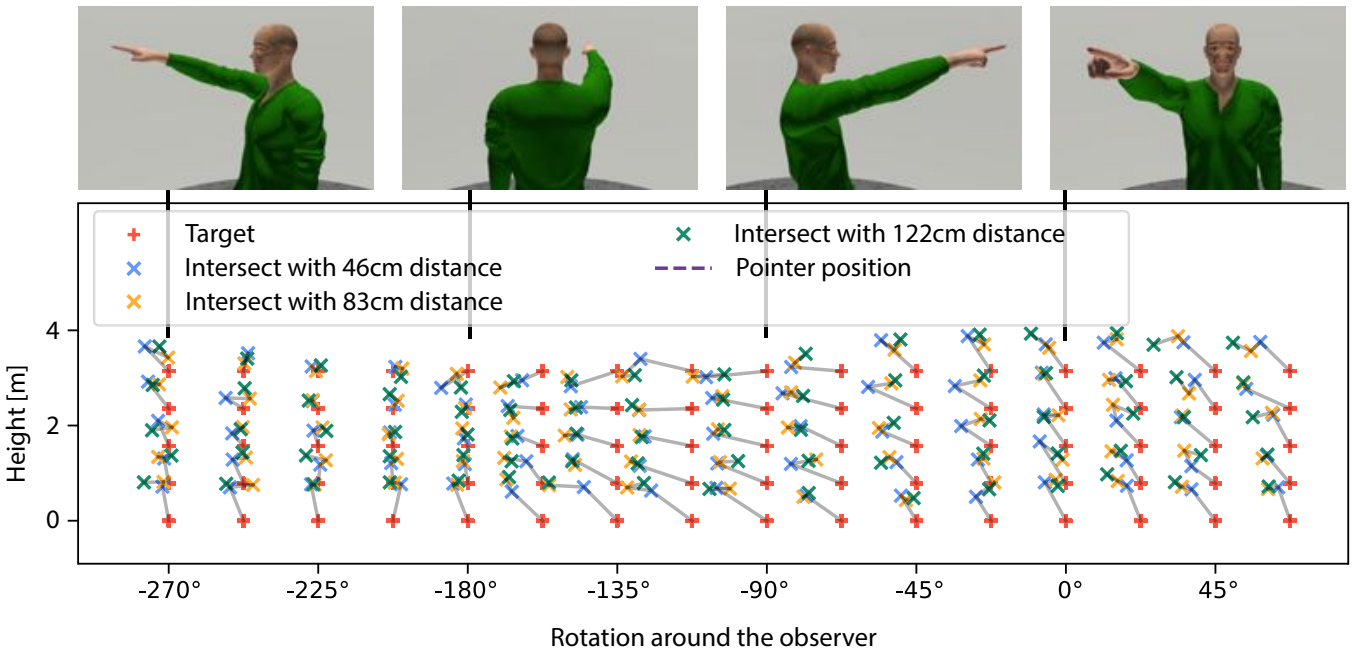

Figure 4. Observer's estimation offsets for the three DISTANCES with four respective postures of the pointer.

casting methods (METHOD) eye-finger ray cast (EFRC), index finger ray cast (IFRC), forearm ray cast (FRC), and head ray cast (HRC) to determine the pointers' accuracy. Before analyzing the pointing precision of the *Pointer*, we filtered the pointing gestures to remove outliers. We use two times the standard deviation as an upper bound. We choose this filter based on a visual inspection of the histogram to meet the normality assumption. We thereby removed 0.48% of the data. We conducted a one-way RM-ANOVA to determine if the distance between the intersect of the *Pointer* and the target is different for the four METHODS. The analysis revealed a significant effect of the METHOD on distance ($F_{1.636,21.267} = 60.005, p < .001, \eta^2 = .751$). The two-tailed t-test with Bonferroni correction revealed that IFRC vs. HRC did not significantly differ ($p = .223$), IFRC vs. FRC are significantly different ($p = .003$), and all other combinations also significantly differ (all $p < .001$), see Table 1. Therefore, EFRC performed significantly better than the other METHODS, see Figure 3.

Accuracy of the Observer

Before analyzing how accurate the *Observer* can estimate the positions of the targets, we filtered the *Observers'* estimations to remove outliers. We use three times the standard deviation as an upper bound. Again, this filter was selected based on a visual inspection of the histogram to meet the normality assumption. Logically this filter was not as strict as the spread for the observer is higher. We thereby removed .42% of the data. Afterward, we conducted a three-way RM-ANOVA to determine if DISTANCE × HEIGHT × ROTATION significantly influence the observers' accuracy. The analysis revealed a significant effect of DISTANCE, HEIGHT and ROTATION ($F_{2,26} = 4.995$, $p < .015$, $\eta^2 = .004$; $F_{4,52} = 4.958$, $p < .002$, $\eta^2 = .006$; $F_{15,195} = 5.842$, $p < .001$, $\eta^2 = .063$; respectively). We also found a significant two-interaction effect for HEIGHT × ROTATION, $F_{60,780} = 1.680$, $p < .002$, $\eta^2 = .032$. However, no two-way interaction effect for DISTANCE × HEIGHT and DISTANCE × ROTATION ($F_{8,104} = .768$, $p > .631$, $\eta^2 = .002$; $F_{30,390} = 1.450$, $p > .062$, $\eta^2 = .015$; respectively). We also found no three-way interaction effect for DISTANCE × HEIGHT × DISTANCE, $F_{120,1560} = .902$, $p > .763$, $\eta^2 = .026$. The observers' average error over all distances was 100.9cm, see Figure 4 and Table 2.

Discussion

We conducted a study to understand how a person interprets where another person is pointing. We first validated if the apparatus leads to results similar to previous work for the pointer. Any discrepancy would render invalid results for the observer. For the pointer, the results are in line with previous work that only focused on the precision of the pointer [43, 56]. As in previous work, using an eye-finger ray cast to determine where a person points at is more precise than related approaches. An error of 40.0cm in a distance of 4m remains the most precise approach. The results for the pointer are in line with previous works [43, 44]; therefore, we consider the apparatus as a valid tool to further study the observer's perspective.

Analyzing the precision with which an observer can estimate where another person points, we found an average error of around 1m in a distance of 4m. The error is, however, not evenly distributed, see Figure 4. We found significant effects for DISTANCE, HEIGHT, and ROTATION. Thus, the accuracy of an observer depends on the relative position of the pointer and the target. This is especially apparent for the angle between *Pointer* and *Observer* (ROTATION). Observers achieve the highest accuracy when the pointer stands between them

and the target (-180°). They remain fairly accurate when the target rotates to the left and the pointer starts facing the observer (-225°). This confirms intuition as pointing at an object while facing another person is common when discussing or highlighting distant objects. Rotating targets to the right (-135°) results in the lowest accuracy. At this point, the observer faces the back of the pointer, which seems to make it especially difficult to interpret the deictic gesture. Here, we found a significant effects on the observers' behavior, see Figure 4. Therefore, we conclude that the errors made, when observers estimate the position of the target, are not random but can partially explained by a systematic component. Furthermore, we showed that the 1D pointing task used by Sousa et al. [57] is not sufficient to model the full scope of deictic pointing. Our results confirm that deictic gestures are 2D pointing tasks.

It is widely accepted that humans' precision when performing deictic gestures is limited [21, 38, 43]. Our study shows that when interpreting deictic gestures, human observers are less precise than an IFRC. Consequently, we assume that the total error when deictic gestures are interpreted is caused by a combination of the pointer's limited accuracy when performing gestures and the observer's limited accuracy when interpreting them. With this study, we lay the foundation for improving the interaction between users in CVEs by increasing the accuracy when interpreting deictic gestures. Therefore, we conducted the study in CVE. As previous work showed that VR affects how users point at targets, we cannot make assumptions about deictic gestures in general or why they occur. We, however, assume that further investigations outside a VR could reveal interesting insights.

IMPROVING HUMANS' ABILITIES TO INTERPRET DEICTIC GESTURES

In our first study, we showed how a person interprets the pointing gesture of another person. As we showed a systematic misinterpretation of a pointer's gesture by an observer, we aim to systematically compensate the error by changing the pose of the pointing user's avatar. By changing the pose, we aim to manipulate where another user believes the avatar truly points at. The changed pose only needs to be visible to the observer. Therefore, it remains unnoticeable for the pointer but can improve the observer's accuracy. In the following, we first develop a model that describes the systematic component of the error when observers interpret deictic gestures. Afterward, we describe how the model can be applied to pointing avatars to achieve a better deictic gesture interpretation.

Model to Improve Estimation Accuracy

Sousa et al. [57] proposed an initial model to improve deictic pointing by manipulating the avatar of the pointing person. However, they treated deictic gestures as a 1D task. While Sousa et al. [57] incorporated the height of the target in relation to the pointer, they ignored the angle between target, pointer, and observer. Our results show that this angle is an important factor to be considered for a general model of deictic gestures; however, this is missing in a 1D task. In the following, we further extend the idea of manipulating the avatar to derive a general model for all deictic pointing gestures.

Previous work suggests that humans, when performing pointing gestures, heavily rely on the index finger in relation to the head. Therefore, considering the relation between the tip of the index finger and the position of the head is crucial to improve estimation accuracy of the pointing arm. To do so, all joints between the index finger tip and the head can be manipulated. However, the smaller the manipulation the smaller the potentially negative effects from artificial postures arise. Thus, we only manipulate the orientation of the avatar's shoulder. On the one hand, a small rotation at the shoulder will have a large impact on the finger tips in the distance. On the other hand, the shoulder is a ball and socket joint and, thus, can naturally lead arm and hand rotations to perform rotations in all directions in front of the human body. Hence, we use the shoulder joint to rotate the index finger to compensate the systematic error when humans interpret deictic gestures.

The following model is inspired by previous work, such as Mayer et al. [43, 44] and Schwind et al. [56], who presented regression models to improve systems' ability to better interpret the pointer's deictic gesture. Their models use an angular representation of the ray cast representation with respect to the head of the user. This allows the model to be distance-invariant. We define Δ_ω with $\omega \in pitch, yaw$ to be the correction rotations index finger tip in respect to the pointers also known as EFRC. To predict Δ_ω, we used ordinary least squares regression to fit the collected data to the parameters (amplitude: *amp*; frequency: *freq*; phase shift: *pshift*; and *offset* for pitch and yaw) of the function $f_\omega(\alpha_p, \alpha_y)$, see Equation (1). Here, α_p and α_y are the actual orientation of the EFRC of the pointer in relation to the observer. This allows the observer to freely move around and still get the right correction applied.

$$\begin{aligned}\Delta_\omega = f_\omega(\alpha_p, \alpha_y) = \\ (sin(freq_p(\alpha_p - pshift_p))amp_p + offset_p + 1)* \\ (sin(freq_y(\alpha_y - pshift_y))amp_y + offset_y + 1) \quad (1)\end{aligned}$$

The function $f_\omega(\alpha_p, \alpha_y)$ returns the rotation of the index finger tip around the origin of the EFRC. A simple transformation needs to be applied to rotate the finger around the shoulder to align the index finger tip with the corrected EFRC.

The polynomial correction model proposed by Mayer et al. [43, 44] poses the issue of motion jumps when applied to the full 360° scenario. To overcome the limitation of their model, we used a sine model (Equation (1)), to correct the *Pointers'* arm. A sine wave has the unique property that we can apply any rotation to the function without encountering gaps, or jumps neither in the arm's position nor in its orientation.

Arm Rotation of the Pointers' Avatars

To correct the rotation of the pointers' arm, we used the EFRC and entered the angles of yaw α_y and pitch α_p into the model in Equation (1). The corrected ray cast $EFRC_c$ was used to apply the new rotation for the joint between the right shoulder and right arm of the avatar's skeleton. As the origin of the $EFRC_c$ is at the center between both eyes, the correction of the shoulder-arm joint can be transformed for their combined ray cast.

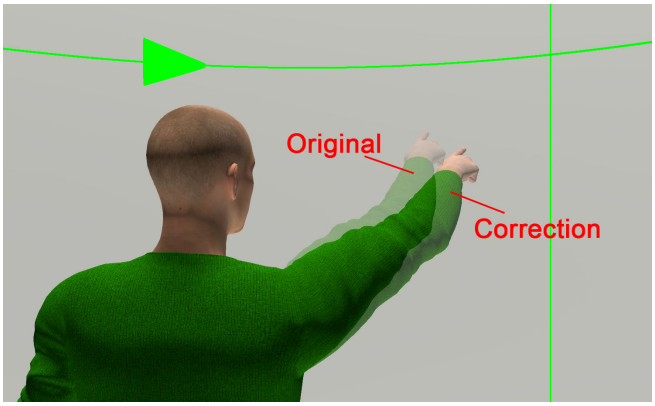

Figure 5. Observer's view with and without the arm correction of the pointing avatar.

We used an imaginary sphere at the arm joint with a dynamic radius determined by the distance between the position of the index finger tip and its center at the arm joint. The intersection of $EFRC_c$ with this sphere was used as the target for joint orientation of the arm. The new joint orientation was shifted from the vector between arm joint and index finger tip to the vector between arm joint and intersection with the $EFRC_c$. As the radius of the intersecting sphere and the $EFRC_c$ were determined for each frame, the correction of the right arm orientation was dynamically applied according to the view and pose of the user. Original and corrected models of the avatar's arm are shown in Figure 5. The correction was only visible to the observer.

EVALUATION STUDY

We conducted a second study to evaluate if changing the pose of the pointing user's avatar increases the accuracy of an observer when determining where the pointer points. We used an additional independent variable CORRECTION with the two levels, *NoCorrection* and *WithCorrection* to understand the impact of the model. We used the same task as in the first study to validate the results and added a second task to understand how the proposed approach performs other tasks. We used an apple-picking task, where the pointer instructs the observer to pick specific apples from trees surrounding them. The correction was applied to the *Pointer* as seen by the *Observer*. Thus the *Observer* sees a corrected arm while the *Pointer* does not perceive the correction. Again, we used DISTANCE with three levels (46*cm*, 83*cm*, and 122*cm*) as the independent variable. Furthermore, we used participants' ROLE with the two levels, *Pointer* and *Observer*, as the third independent variable.

We counterbalanced the order of CORRECTION within each task. However, participants started always in the validation task to avoid influencing the results of the apple-picking task, were participants may adapt their pointing behavior.

Apparatus

We used the same hardware setup as in the first study. We modified the original task to apply the correction directly to the shoulder of the *Pointer* during the *WithCorrection* condition. Moreover, we reduced the number of targets to provide room for the second task without adding more pointing gestures that could potentially lead to fatigue effects. Therefore, we used a 8×3 (ROTATION $\times$ HEIGHT) grid spanning the whole 360° rotation of the room, where the ROTATION variations are projected every 45° and the HEIGHT varies over rows every $1.56m$ starting on the elevated position of $1m$ the avatars stands on. 8 ROTATIONS $\times$ 3 HEIGHTS $\times$ 3 DISTANCES $\times$ 2 repetitions $=$ 144 deictic gestures.

In the second task, we placed 80 apples in a $4m$ radius from the pointer, covering the same area as during the validation task without an even distribution of the targets. Only one apple at a time was visible to the *Pointer* while the *Observer* saw all 80 apples. We presented eight apples one after another to the *Pointer*. The *Observer* had to pick which of the 80 apples the *Pointer* was pointing at. After picking eight apples, the system showed the number of correctly picked apples. In total, participants had to pick 96 apples $=$ 3 DISTANCES $\times$ 8 apples to pick per round $\times$ 4 repetitions.

Procedure

As in the first study, one participant was always the pointer and the other the observer. We welcomed both participants, explained the procedure of the study, and asked them to sign the consent form as well as to fill in the demographic data from. We then brought both participants into the VR. After a familiarization phase, we guided the participants through the study. We monitored their actions on a separate screen. When the participants switched between the *NoCorrection* and the *WithCorrection* condition as well as between the tasks, we asked them to fill in a raw TLX [32] and an iGroup Presence Questionnaire (IPQ) [48].

Participants

We recruited 24 participants (19 male, 5 female) via our institutions' mailing lists. Participants were between 17 and 35 years old ($M = 23.9$, $SD = 4.1$). We again only recruited right-handed participants who had no locomotor coordination problems. We used the Porta test [18] to screen participants for eye-dominance: 13 participants had right-eye dominance, 9 had left-eye dominance, and 2 were undecided.

Results

For the validation task, observers estimated a total of 1,728 positions for each of the two conditions. For the apple-picking task, observers selected a total of 1,152 apples for each of the two conditions. As in the analysis of the first study, whenever Mauchly's test showed that the sphericity assumption was violated in the RM-ANOVA, we reported Greenhouse-Geisser (GG) or Huynh-Feldt (HF) corrected p-values.

Fatigue effect

First, we analyzed the raw TLX score (scale: 0 - 20) to determine if potential workload or fatigue effects had to be considered in the analysis. Therefore, we conducted two univariate two-way analysis of variances (ANOVAs): one for each task to understand if ROUND or ROLE (with two levels: *Pointer* and *Observer*) influenced the perceived task load. The analysis revealed no significant effect for ROUND and ROLE ($F_{1,22} = 1.541$, $p > .227$, $\eta^2 = .004$; $F_{1,22} = .231$, $p > .636$, $\eta^2 = .010$, respectively) and also no significant interaction

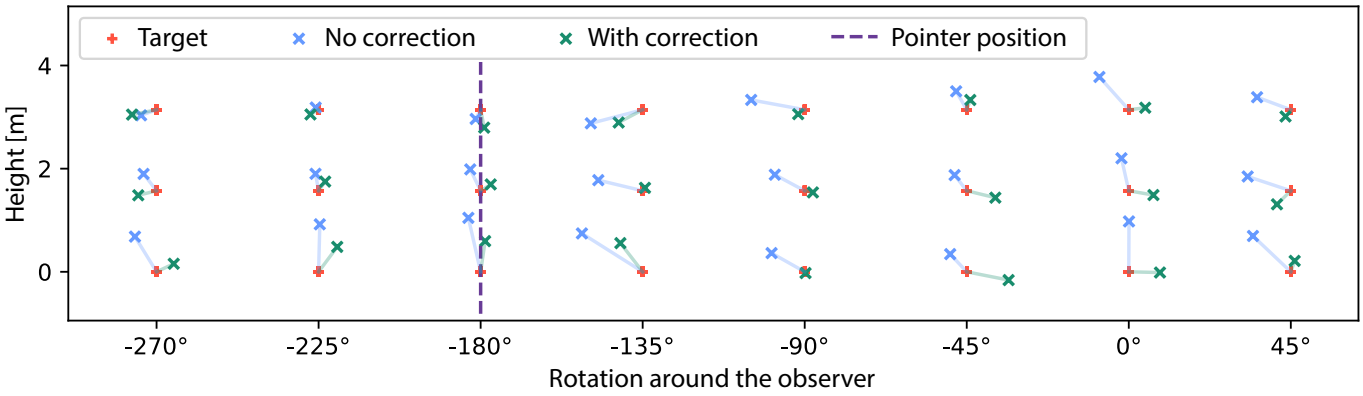

Figure 6. Observers' estimated positions of the targets in the validation task split by CORRECTION.

effect ($F_{1,22} = 3.781$, $p > .064$, $\eta^2 = .010$) in the validation task. The mean raw TLX score was $M = 6.0$ ($SD = 3.1$) after the first and $M = 5.6$ ($SD = 3.8$) after the second task during the validation. The analysis revealed no significant effect for ROUND and ROLE ($F_{1,22} = .059$, $p > .809$, $\eta^2 = .011$; $F_{1,22} = .268$, $p > .610$, $\eta^2 < .001$, respectively) and also no significant interaction effect ($F_{1,22} = .179$, $p > .676$, $\eta^2 < .001$) in the apple task. The mean raw TLX score was $M = 7.6$ ($SD = 4.1$) after the first, $M = 7.5$ ($SD = 4.4$) after the second rounds in the apple task. Thus, we assume that the effect of participants' fatigue or task load was negligible.

Validation Task

We analyzed how accurate the *Observer* can estimate the target with and without correction. Therefore, we conducted a two-way RM-ANOVA to determine whether DISTANCE or CORRECTION had a significant effect on the *Observer's* accuracy, see Table 3. The analysis revealed a significant effect of CORRECTION and DISTANCE ($F_{1,11} = 8.240$, $p = .015$, $\eta^2 = .110$; $F_{2,22} = 3.708$, $p = .041$, $\eta^2 = .045$; respectively), see Figure 7. The analysis did not reveal a significant interaction effect ($F_{1.311,14.425} = 2.402$, $p = .456$, $\eta^2 = .009$).

Apple-Picking Task

We conducted a two-way RM-ANOVA to determine whether CORRECTION or DISTANCE significantly influenced the amount of correctly picked apples. Our analysis revealed no significant effects for CORRECTION and DISTANCE on the correctly picked apples ($F_{1,11} = .147$, $p > .709$, $\eta^2 = .005$; $F_{2,22} = 2.275$, $p > .126$, $\eta^2 = .022$; respectively). The analysis did not reveal a significant interaction effect ($F_{2,22} = .011$, $p > .989$, $\eta^2 < .001$). With correction the participants picked on average 4.6 ($SD = 1.6$) correct apples out of 8, whereas in the condition without correction they picked only 4.4 ($SD = 1.9$).

Observer	*NoCorrection*		*WithCorrection*	
	M	**SD**	**M**	**SD**
Distance 46*cm*	89.4	32.4	70.	27.
Distance 83*cm*	102.4	36.3	84.3	27.1
Distance 122*cm*	112.2	48.2	80.2	30.9
Average	101.4	39.6	78.2	28.2

Table 3. Overall offsets (in cm) between interact and target in the validation task at $4m$ distance from the pointer.

Presence

First, we analyzed the IPQ questionnaire [48] (scale: from -3 to 3) to determine whether ROLE or CORRECTION significantly influenced the presence in VR in one of the two tasks. Therefore, we conducted two univariate two-way ANOVAs. In the validation task, the analysis revealed a significant effect for ROLE ($F_{1,22} = 12.79$, $p < .001$, $\eta^2 = .334$). However, we found no significant effect for CORRECTION ($F_{1,22} = 1.614$, $p > .217$, $\eta^2 = .010$) and also no significant interaction effect ($F_{1,22} = .078$, $p > .782$, $\eta^2 < .001$). The mean IPQ score was $M = .28$ ($SD = .62$) for the *Pointer* and $M = -.89$ ($SD = 1.02$) for the *Observer* in the validation task. In the apple-picking task, the analysis revealed a significant effect for ROLE ($F_{1,22} = 6.441$, $p < .019$, $\eta^2 = .212$). However, we found no significant effect for CORRECTION ($F_{1,22} = .534$, $p > .472$, $\eta^2 = .002$) and also no significant interaction effect ($F_{1,22} = .382$, $p > .542$, $\eta^2 = .001$). The mean IPQ score was $M = .76$ ($SD = .68$) for the *Pointer* and $M = -.11$ ($SD = 1.$) for the *Observer* in the apple task.

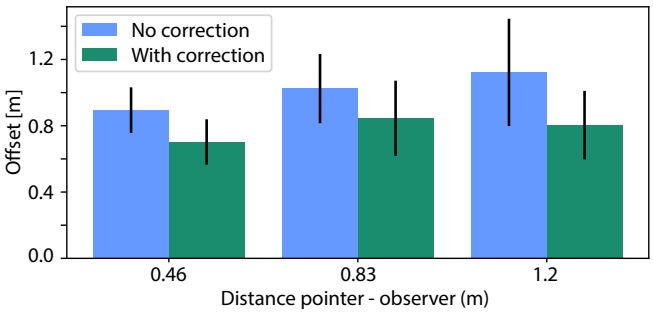

Figure 7. Average estimation error of the *Observer* in the validation task split by CORRECTION and DISTANCE.

Discussion

In our second study, we determined if manipulating a person's arm can help an observer to estimate the target's position. Results of the validation task that replicated the procedure of the first study were in line with the first study when estimating the position of the target to which the pointer points. The average error was 101.4cm, which is similar to the average error of 100.9*cm* found in the first study. Moreover, the validation task revealed that correcting the pointer's pose increases the observer's accuracy by 22.9%. While the error was lower when participants stood close to each other, we observed that the accuracy increases for all conditions.

While participants' performance improved in the apple-picking task when the correction was applied, we could not reveal a significant difference. We assume that this is caused by the sparse distribution of apples on the trees. Thus, the visual clue of seeing only one apple in a certain direction is enough to pick the right apple. In the future, a study with a higher object density could determine the relationship between density and the potential to improve the accuracy.

Applying the correction had no significant effect on the workload or the participants' immersion. In fact, the corrected pose is only visible to the observer; therefore, it cannot affect the pointer. However, unsurprisingly, the pointer who had an avatar had significantly higher immersion than the observer who had no avatar. Our results show that applying the correction also remains unnoticeable to the observer.

CONCLUSIONS AND FUTURE WORK

We investigated how to improve the interaction between users in CVEs by increasing the accuracy when interpreting deictic gestures. In the first study, we showed that humans' accuracy when interpreting deictic gestures is indeed limited. We found systematic effects that depend on the relative position of the pointer, the observer, and the target. As the errors are at least partially systematic, we developed a model that describes the error and used it to rotate a pointing person's arm to improve an observer's interpretation accuracy. Through a second study, we showed that by changing the pose of a pointing user's avatar as seen by an observer, we can significantly improve the observer's accuracy by 22.9%. For a second task we found no statistically significant difference. The descriptive results are, however, encouraging. Even for tasks with a low object density, changing the pointer's pose might be able to improve the accuracy. As the model is only applied to the perspective of the user interpreting the deictic gesture, it is not noticeable for the pointing user. We show that the correction does also not affect the observer. Moreover, it neither requires users to learn a new interaction technique nor to adapt their behavior. We conclude that changing the pose of a pointing user's avatar can improve an observer's accuracy without costs.

By embracing the possibilities of VR, we showed that users can be provided with individualized views that improve their performance. This is not only true for pairs of users but also for multiple persons because everybody can realize their own optimized perspective on the virtual world. Future work should investigate additional ways to optimize communication in CVEs. Potential approaches include further subtle manipulations of the avatars' pose as well as changing the audio signal to improve verbal communication. Additionally, as VR has the potential to display pointing feedback, e.g., Wong and Gutwin [67], a next step should investigate how our method compares to feedback in respect to accuracy and task completion time as well as user experience.

Throughout this paper, pointing at targets and estimating where a person points were the only tasks. In most CVE applications, this is different. We did not investigate the effects of changing the avatar's pose when a user is not pointing. It might be necessary to determine when a user points to only apply the correction in such situations. We assume that monitoring users' pose and using simple thresholds to identify when a user starts pointing is sufficient. However, more advanced machine learning-based approaches might also be useful. In both studies, we observed that in VR, humans' accuracy when interpreting deictic gestures is limited. We, however, assume that this limited accuracy does not only manifest in VR. Therefore, future work should investigate the interpretation of deictic gestures outside of VR.

With this paper, we also contribute the source code and the scenes[2] of the Unity project used in the studies. We provide the code to correct the pose of the pointer's avatar under the MIT license. This enables other researchers to apply the model in their own work and to further investigate our findings.

REFERENCES

[1] Martha W. Alibali. 2005. Gesture in Spatial Cognition: Expressing, Communicating, and Thinking About Spatial Information. *Spatial Cognition and Computation* 5, 4 (2005), 307–331. DOI: http://dx.doi.org/10.1207/s15427633scc0504_2

[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 (CHI '16)*. ACM, New York, NY, USA, 1968–1979. DOI: http://dx.doi.org/10.1145/2858036.2858226

[3] Adrian Bangerter and Daniel M. Oppenheimer. 2006. Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture* 6, 1 (2006), 85–102. DOI: http://dx.doi.org/10.1075/gest.6.1.05ban

[4] R.M. Baños, C. Botella, M. Alcañiz, V. Liaño, B. Guerrero, and B. Rey. 2004. Immersion and Emotion: Their Impact on the Sense of Presence. *CyberPsychology & Behavior* 7, 6 (dec 2004), 734–741. DOI:http://dx.doi.org/10.1089/cpb.2004.7.734

[5] 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*

[2] https://github.com/interactionlab/Deictic-Pointing-in-VR

Factors in Computing Systems (CHI '95). ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 242–249. DOI: http://dx.doi.org/10.1145/223904.223935

[6] Steve Benford, Chris Greenhalgh, Tom Rodden, and James Pycock. 2001. Collaborative Virtual Environments. *Commun. ACM* 44, 7 (July 2001), 79–85. DOI:http://dx.doi.org/10.1145/379300.379322

[7] Johanna Bertram, Johannes Moskaliuk, and Ulrike Cress. 2015. Virtual training: Making reality work? *Computers in Human Behavior* 43 (2015), 284 – 292. DOI:http://dx.doi.org/https://doi.org/10.1016/j.chb.2014.10.032

[8] Frank Biocca. 1992. Communication Within Virtual Reality: Creating a Space for Research. *Journal of Communication* 42, 4 (1992), 5–22. DOI: http://dx.doi.org/10.1111/j.1460-2466.1992.tb00810.x

[9] Richard A. Bolt. 1980. Put-that-there: Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques (SIGGRAPH '80)*. ACM, New York, NY, USA, 262–270. DOI: http://dx.doi.org/10.1145/800250.807503

[10] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. 2016. Point & Teleport Locomotion Technique for Virtual Reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. ACM, New York, NY, USA, 205–216. DOI: http://dx.doi.org/10.1145/2967934.2968105

[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. DOI:http://dx.doi.org/10.1348/026151000165553

[12] George Butterworth and Nicholas Jarrett. 1991. What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology* 9, 1 (1991), 55–72. DOI: http://dx.doi.org/10.1111/j.2044-835X.1991.tb00862.x

[13] Luigia Camaioni. 1997. The Emergence of Intentional Communication in Ontogeny, Phylogeny, and Pathology. *European Psychologist* 2, 3 (1997), 216–225. DOI: http://dx.doi.org/10.1027/1016-9040.2.3.216

[14] Luigia Camaioni, Paola Perucchini, Francesca Bellagamba, and Cristina Colonnesi. 2004. The Role of Declarative Pointing in Developing a Theory of Mind. *Infancy* 5, 3 (2004), 291–308. DOI: http://dx.doi.org/10.1207/s15327078in0503_3

[15] Christer Carlsson and Lennart E. Fahlén. 1993. Integrated CSCW Tools Within a Shared 3D Virtual Environment (Abstract). In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 513–. DOI: http://dx.doi.org/10.1145/169059.169452

[16] Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D. Wilson. 2017. Sparse Haptic Proxy: Touch Feedback in Virtual Environments Using a General Passive Prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3718–3728. DOI:http://dx.doi.org/10.1145/3025453.3025753

[17] Christopher Codella, Reza Jalili, Lawrence Koved, J. Bryan Lewis, Daniel T. Ling, James S. Lipscomb, David A. Rabenhorst, Chu P. Wang, Alan Norton, Paula Sweeney, and Greg Turk. 1992. Interactive Simulation in a Multi-person Virtual World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, USA, 329–334. DOI: http://dx.doi.org/10.1145/142750.142825

[18] Giambattista Della Porta. 1593. *De refractione Optices Parte: Libri Novem...* Ex officina Horatii Salviani, apud Jo. Jacobum Carlinum, & Antonium Pacem.

[19] Daphne Economou, Ioannis Doumanis, Lemonia Argyriou, and Nektarios Georgalas. 2017. User Experience Evaluation of Human Representation in Collaborative Virtual Environments. *Personal Ubiquitous Comput.* 21, 6 (Dec. 2017), 989–1001. DOI: http://dx.doi.org/10.1007/s00779-017-1075-4

[20] Tiare Feuchtner and Jörg Müller. 2017. Extending the Body for Interaction with Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5145–5157. DOI: http://dx.doi.org/10.1145/3025453.3025689

[21] John M. Foley and Richard Held. 1972. Visually directed pointing as a function of target distance, direction, and available cues. *Perception & Psychophysics* 12, 3 (01 May 1972), 263–268. DOI: http://dx.doi.org/10.3758/BF03207201

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

[23] Saul Greenberg, Carl Gutwin, and Mark Roseman. 1996b. Semantic telepointers for groupware. In *Proceedings Sixth Australian Conference on Computer-Human Interaction*. 54–61. DOI: http://dx.doi.org/10.1109/OZCHI.1996.559988

[24] Scott W Greenwald, Wiley Corning, and Pattie Maes. 2017a. Multi-user framework for collaboration and co-creation in virtual reality *(CSCL '17)*. 12th International Conference on Computer Supported Collaborative Learning. http://hdl.handle.net/1721.1/108440

[25] Scott W Greenwald, Alexander Kulik, André Kunert, Stephan Beck, Bernd Fröhlich, Sue Cobb, Sarah Parsons, Nigel Newbutt, Christine Gouveia, Claire Cook, and others. 2017b. Technology and applications for collaborative learning in virtual reality. Philadelphia, PA. https://repository.isls.org/handle/1/210

[26] Victoria Groom, Jeremy N Bailenson, and Clifford Nass. 2009. The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence* 4, 3 (2009), 231–248.

[27] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017. ShareVR: Enabling Co-Located Experiences for Virtual Reality Between HMD and Non-HMD Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4021–4033. DOI: http://dx.doi.org/10.1145/3025453.3025683

[28] Carl Gutwin and Saul Greenberg. 2002. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work* 11, 3 (01 Sep 2002), 411–446. DOI: http://dx.doi.org/10.1023/A:1021271517844

[29] Anthony Guye-Vuillème, Tolga K. Capin, Igor Sunday Pandzic, Nadia Magnenat Thalmann, and Daniel Thalmann. 1999. Nonverbal communication interface for collaborative virtual environments. *Virtual Reality* 4, 1 (01 Mar 1999), 49–59. DOI: http://dx.doi.org/10.1007/BF01434994

[30] Olof Hagsand. 1996. Interactive multiuser VEs in the DIVE system. *IEEE MultiMedia* 3, 1 (Spring 1996), 30–39. DOI:http://dx.doi.org/10.1109/93.486702

[31] Edward Twitchell Hall. 1966. *The hidden dimension*. Doubleday & Co, New York, NY, USA.

[32] Sandra G. Hart. 2006. NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomic Society annual meeting*, Vol. 50. SAGE Publications, SAGE Publications, Los Angeles, CA, USA, 904–908. DOI: http://dx.doi.org/10.1177/154193120605000909

[33] John B. Haviland. 2003. *How to point in Zinacantán*. Psychology Press, Mahwah, New Jersey, USA, 139–169.

[34] 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), 78–89. DOI: http://dx.doi.org/10.1037/xhp0000126

[35] Oliver Herbort and Wilfried Kunde. 2018. How to point and to interpret pointing gestures? Instructions can reduce pointer–observer misunderstandings. *Psychological Research* 82, 2 (01 Mar 2018), 395–406. DOI:http://dx.doi.org/10.1007/s00426-016-0824-8

[36] Jason Jerald. 2016. *The VR Book: Human-Centered Design for Virtual Reality*. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA.

[37] Shunichi Kasahara, Keina Konno, Richi Owaki, Tsubasa Nishi, Akiko Takeshita, Takayuki Ito, Shoko Kasuga, and Junichi Ushiba. 2017. Malleable Embodiment: Changing Sense of Embodiment by Spatial-Temporal Deformation of Virtual Human Body. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6438–6448. DOI: http://dx.doi.org/10.1145/3025453.3025962

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

[39] Pascal Knierim, Valentin Schwind, Anna Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA. DOI: http://dx.doi.org/10.1145/3173574.3173919

[40] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 81, 14 pages. DOI:http://dx.doi.org/10.1145/3173574.3173655

[41] Sarah Lopez, Yi Yang, Kevin Beltran, Soo Jung Kim, Jennifer Cruz Hernandez, Chelsy Simran, Bingkun Yang, and Beste F. Yuksel. 2019. Investigating Implicit Gender Bias and Embodiment of White Males in Virtual Reality with Full Body Visuomotor Synchrony. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 557, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300787

[42] Kristina Lundholm Fors. 2015. *Production and Perception of Pauses in Speech*. Ph.D. Dissertation. University of Gothenburg.

[43] Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. 2018. The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 653, 13 pages. DOI:http://dx.doi.org/10.1145/3173574.3174227

[44] Sven Mayer, Katrin Wolf, Stefan Schneegass, and Niels Henze. 2015. Modeling Distant Pointing for Compensating Systematic Displacements. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 4165–4168. DOI: http://dx.doi.org/10.1145/2702123.2702332

[45] 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. DOI:http://dx.doi.org/10.1145/3173574.3173620

[46] 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 (SA '17)*. ACM, New York, NY, USA, Article 3, 2 pages. DOI: http://dx.doi.org/10.1145/3132818.3132822

[47] Katrin Plaumann, Matthias Weing, Christian Winkler, Michael Müller, and Enrico Rukzio. 2017. Towards accurate cursorless pointing: the effects of ocular dominance and handedness. *Personal and Ubiquitous Computing* (07 Dec 2017). DOI: http://dx.doi.org/10.1007/s00779-017-1100-7

[48] Holger T. Regenbrecht, Thomas W. Schubert, and Frank Friedmann. 1998. Measuring the Sense of Presence and its Relations to Fear of Heights in Virtual Environments. *International Journal of Human-Computer Interaction* 10, 3 (1998), 233–249. DOI: http://dx.doi.org/10.1207/s15327590ijhc1003_2

[49] Rufat Rzayev, Gürkan Karaman, Katrin Wolf, Niels Henze, and Valentin Schwind. 2019. The Effect of Presence and Appearance of Guides in Virtual Reality Exhibitions. In *Proceedings of Mensch Und Computer 2019 (MuC'19)*. ACM, New York, NY, USA, 11–20. DOI:http://dx.doi.org/10.1145/3340764.3340802

[50] Maria V. Sanchez-Vives and Mel Slater. 2005. From presence to consciousness through virtual reality. *Nature reviews. Neuroscience* 6, 4 (apr 2005), 332–339. DOI: http://dx.doi.org/10.1038/nrn1651

[51] Chris L. Schmidt. 1999. Adult understanding of spontaneous attention-directing events: What does gesture contribute? *Ecological Psychology* 11, 2 (1999), 139–174. DOI: http://dx.doi.org/10.1207/s15326969eco1102_2

[52] Valentin Schwind, Pascal Knierim, Lews Chuang, and Niels Henze. 2017a. "Where's Pinky?": The Effects of a Reduced Number of Fingers in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Computer-Human Interaction in Play (CHI PLAY'17)*. ACM, New York, NY, USA, 6. DOI: http://dx.doi.org/10.1145/3116595.3116596

[53] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017b. "These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1577–1582. DOI: http://dx.doi.org/10.1145/3025453.3025602

[54] Valentin Schwind, Jan Leusmann, and Niels Henze. 2019. Understanding Visual-Haptic Integration of Avatar Hands Using a Fitts' Law Task in Virtual Reality. In *Proceedings of Mensch Und Computer 2019 (MuC'19)*. ACM, New York, NY, USA, 211–222. DOI: http://dx.doi.org/10.1145/3340764.3340769

[55] Valentin Schwind, Lorraine Lin, Massimiliano Di Luca, Sophie Jörg, and James Hillis. 2018a. Touch with Foreign Hands: The Effect of Virtual Hand Appearance on Visual-haptic Integration. In *Proceedings of the 15th ACM Symposium on Applied Perception (SAP '18)*. ACM, New York, NY, USA, Article 9, 8 pages. DOI: http://dx.doi.org/10.1145/3225153.3225158

[56] Valentin Schwind, Sven Mayer, Alexandre Comeau-Vermeersch, Robin Schweigert, and Niels Henze. 2018b. Up to the Finger Tip: The Effect of Avatars on Mid-Air Pointing Accuracy in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Computer-Human Interaction in Play*. DOI: http://dx.doi.org/10.1145/3242671.3242675

[57] Maurício Sousa, Rafael Kufner dos Anjos, Daniel Mendes, Mark Billinghurst, and Joaquim Jorge. 2019. Warping Deixis: Distorting Gestures to Enhance Collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 608, 12 pages. DOI:http://dx.doi.org/10.1145/3290605.3300838

[58] Mark Jeffrey Stefik, Daniel G. Bobrow, Gregg Foster, Stan Lanning, and Deborah Tatar. 1987. WYSIWIS Revised: Early Experiences with Multiuser Interfaces. *ACM Transactions on Information Systems* 5, 2 (April 1987), 147–167. DOI: http://dx.doi.org/10.1145/27636.28056

[59] Jonathan Steuer. 1992. Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication* 42, 4 (1992), 73–93. DOI: http://dx.doi.org/10.1111/j.1460-2466.1992.tb00812.x

[60] Haruo Takemura and Fumio Kishino. 1992. Cooperative Work Environment Using Virtual Workspace. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work (CSCW '92)*. ACM, New York, NY, USA, 226–232. DOI: http://dx.doi.org/10.1145/143457.269747

[61] Janet L. Taylor and D. I. McCloskey. 1988. Pointing. *Behavioural Brain Research* 29, 1 (1988), 1 – 5. DOI: http://dx.doi.org/https://doi.org/10.1016/0166-4328(88)90046-0

[62] Jolanda G. Tromp, Anthony Steed, and John R. Wilson. 2003. Systematic Usability Evaluation and Design Issues for Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments* 12, 3 (2003), 241–267. DOI: http://dx.doi.org/10.1162/105474603765879512

[63] Martin Usoh, Kevin Arthur, Mary C Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P Brooks Jr. 1999. Walking> walking-in-place> flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 359–364.

[64] Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan. 2011. Vision-based Hand-gesture Applications. *Commun. ACM* 54, 2 (Feb. 2011), 60–71. DOI:http://dx.doi.org/10.1145/1897816.1897838

[65] E. J. Williams. 1949. Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. *Australian Journal of Chemistry* 2, 2 (01 Jun 1949), 149–168. https://doi.org/10.1071/CH9490149

[66] Bob G Witmer and Michael J Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (jun 1998), 225–240. DOI: http://dx.doi.org/10.1162/105474698565686

[67] 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 (CSCW '14)*. ACM, New York, NY, USA, 1377–1387. DOI: http://dx.doi.org/10.1145/2531602.2531691