



UNIVERSITY OF LEEDS

This is a repository copy of *MAGIC: Manipulating Avatars and Gestures to Improve Remote Collaboration*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/201242/>

Version: Accepted Version

Proceedings Paper:

Fidalgo, CG, Sousa, M, Mendes, D et al. (4 more authors) (2023) *MAGIC: Manipulating Avatars and Gestures to Improve Remote Collaboration*. In: *Proceeding of 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR), 25-29 Mar 2023, Shanghai, China. IEEE , pp. 438-448.

<https://doi.org/10.1109/vr55154.2023.00059>

© 2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

MAGIC: Manipulating Avatars and Gestures to Improve Remote Collaboration

Catarina G. Fidalgo*
Carnegie Mellon University
INESC-ID, Instituto Superior
Técnico, University of Lisbon

Maurício Sousa†
University of Toronto

Daniel Mendes‡
INESC TEC
Faculdade de Engenharia,
University of Porto

Rafael Kuffner dos Anjos§
University of Leeds

Daniel Medeiros¶
University of Glasgow

Karan Singh||
University of Toronto

Joaquim Jorge**
INESC-ID / Instituto Superior
Técnico, University of Lisbon

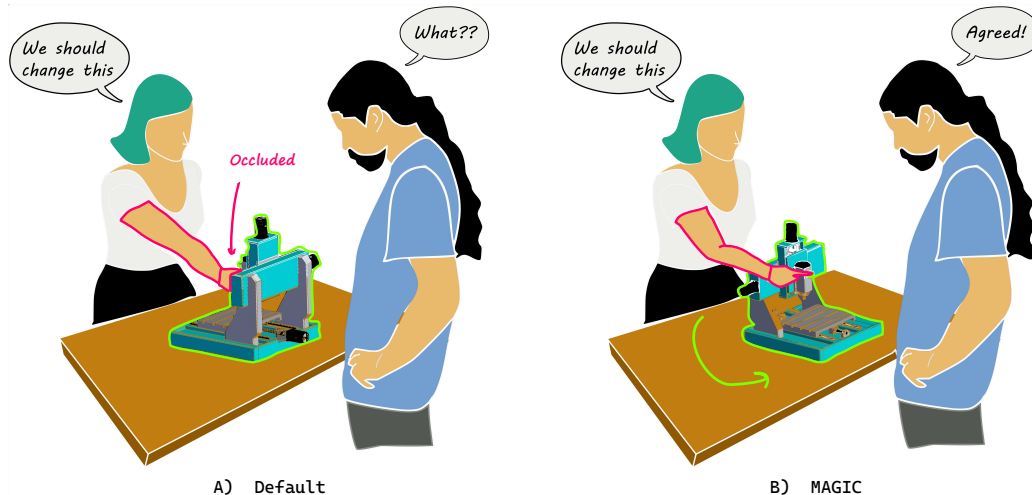


Figure 1: Collaborating remotely in a shared 3D workspace: A) *Veridical Face-to-face Remote Meeting*: Kate and George can see each other but have opposing points-of-view of the workspace. Kate indicates the optimal position for the component she is designing using proximal pointing gestures. George does not understand the location Kate is referring to as it is occluded. Kate needs to explain it through words or ask George to come by her side, which brings added complexity to the collaboration task. B) *MAGIC Remote Meeting*: Kate and George both share the “illusion” of standing on opposite sides of the workspace but are in fact sharing the same point-of-view. Kate’s representation is manipulated so that George sees her arm pointing to the location she intends to convey. George understands where she is referring to, without any added effort.

ABSTRACT

Remote collaborative work has become pervasive in many settings, ranging from engineering to medical professions. Users are immersed in virtual environments and communicate through life-sized avatars that enable face-to-face collaboration. Within this context, users often collaboratively view and interact with virtual 3D models, for example to assist in the design of new devices such as customized prosthetics, vehicles or buildings. Discussing such shared 3D content face-to-face, however, has a variety of challenges such as ambiguities, occlusions, and different viewpoints that all decrease mutual awareness, which in turn leads to decreased task perfor-

mance and increased errors. To address this challenge, we introduce MAGIC, a novel approach for understanding pointing gestures in a face-to-face shared 3D space, improving mutual understanding and awareness. Our approach distorts the remote user’s gestures to correctly reflect them in the local user’s reference space when face-to-face. To measure what two users perceive in common when using pointing gestures in a shared 3D space, we introduce a novel metric called pointing agreement. Results from a user study suggest that MAGIC significantly improves pointing agreement in face-to-face collaboration settings, improving co-presence and awareness of interactions performed in the shared space. We believe that MAGIC improves remote collaboration by enabling simpler communication mechanisms and better mutual awareness.

Index Terms: Human-centered computing—Collaborative and social computing—Collaborative and social computing theory, concepts and paradigms—Computer supported cooperative work

1 INTRODUCTION

In our globalized world, remote collaboration is increasingly important. Air travel cost and time overheads, global warming, and more recently, the pandemics made the ability to work remotely very popular, if not essential.

*e-mail: cfidalgo@andrew.cmu.edu

†e-mail: mauricio@dgp.toronto.edu

‡e-mail: danielmendes@fe.up.pt

§e-mail: r.kuffnerdosanjos@leeds.ac.uk

¶e-mail: daniel.piresdesamedeiros@glasgow.ac.uk

||e-mail: karan@dgp.toronto.edu

**e-mail: jorgej@tecnico.ulisboa.pt

Traditional approaches to remote collaboration, using voice and video, have been used for decades. However, these technologies neglect essential aspects of interpersonal nonverbal communication such as gaze, body posture, gestures to indicate objects referred to in speech (*deictic gestures* [51]), or even how people use the space and position themselves when communicating (i.e. *proxemics* [27]). These limitations negatively affect how remote people interact with a local user, especially for operations that require 3D spatial referencing and action demonstration [18].

Mixed Reality (MR) is a key technology for enabling effective remote collaboration. Fields such as engineering, architecture, and medicine already take advantage of MR approaches for 3D models' analysis. Remote collaboration through MR allows teams to meet in life-sized representations and discuss topics as if sharing the same office. This enables a high level of co-presence and allows a more rigorous spatial comprehension of the 3D models they are working on [11, 42]. Indeed, previous research found that for a remote meeting to be closer to a co-located experience, it should rely on a real size portrayal of remote people to maintain the sense of "being there" [23, 54, 55]. Furthermore, people perform tasks better when communicating via full-body gestures [17].

A *face-to-face f-formation*, i.e., two people position themselves facing each other [71], was found to have potential advantages over *side-by-side f-formations*, i.e., two people position themselves side to side facing the same direction [71], including gaze awareness, shared attention and the ability to see the other person's visual cues [36, 76]. In addition to hindering the capacity to recognize sequential and intentional communicative cues, in a side-by-side f-formation participants share the same personal space, which makes it hard to implement mechanisms that can coordinate the use of this shared space [36]. However, when virtually collaborating in a "veridical face-to-face setting", people have to deal with the problem of sharing opposing points of view (POVs) of the workspace. In these situations, people do not share the same *forward-backward* orientation, and there can be occlusions of parts of the workspace. This constrains the ability for people to use descriptions of relative positions and affects the understanding of where or what the remote person is pointing at, leading to communication missteps and causing tasks to be more laborious [64, 76]. This problem of disjoint visual perspectives when referring the workspace has been identified in previous work [36, 48, 72, 75, 76], especially when using pointing gestures [33, 68]. Previous work found that for 3D-object-centered collaboration, a shared POV is more effective and preferred when compared to an opposing POV [19, 44]. Although shared POV is a well-established approach for this purpose, users lose the communicational advantages of seeing their partner in front of them [37], being deprived of non-verbal cues such as gaze, facial expressions, and body posture which are essential in terms of awareness of all remote users involved in the collaborative task. Indeed, Gutwin and Greenberg [26] suggest that to ensure workspace awareness people should gather information in ways familiar to what happens in real environments, with mechanisms such as consequential communication, feedthrough, and intentional communication. This first source of information, *consequential communication*, relies on the other person's body: posture, movement of head, eyes, arms, and hands, or facial expressions provide an abundance of information about what is happening in the workspace.

We conclude that shared viewpoints and face-to-face dialogues are both beneficial but incompatible. To address this challenge, we explore integrating both by manipulating user's perception of each other and the environment, to provide a novel share of perspective where the local user is also aware of their partner's consequential communication.

Hence, we present MAGIC, a novel approach to collaborative remote work that enables two people to interact using pointing gestures, adopting a face-to-face f-formation while sharing a common

perspective of the workspace. MAGIC distorts gestures performed by a remote person to present a corrected virtual image to the local participant, ensuring accurate communication. Prior work suggests that the interpretation of pointing gestures can be significantly improved by warping the pointing arm [49, 63]. Sousa et al. [63] developed a model for correcting vertical errors using targets within one fixed column (1D), while Mayer et al. [49] manipulated pointing gestures for two dimensions. However, these works are focused on distal pointing for side-by-side f-formations with targets at a fixed distance. Our work focuses on proximal pointing for face-to-face f-formations where targets can be anywhere in the shared 3D space. Collaborators can use pointing gestures to refer to proximal objects in the workspace and interpret their partner's movements in a modified reference space. This creates the illusion that they performed a gesture as expected since it corresponds to the reference space of the local participant, as demonstrated by the usage scenario in Figure 1. We implemented a telepresence environment using virtual representations of people rendered life-size around a shared workspace to evaluate our approach. Additionally, we introduce *Pointing Agreement* to classify what participants perceive in common when using proximal pointing gestures to indicate targets located in a shared workspace, measured through the application of the *Jaccard Similarity Coefficient* [38, 67] in a 3D space. The user study results showed improved capabilities to understand pointing gestures using MAGIC as compared to a veridical face-to-face f-formation. Furthermore, participants did not detect our manipulations, and MAGIC distortions did not negatively affect the participants' sense of presence or the time they spent doing a task. Our experiments validated the assumption that sharing the same perspective while in a face-to-face f-formation improves *pointing agreement* and facilitates virtual object-centered remote collaboration.

The main contributions of this research include: 1) MAGIC, a novel interaction technique to integrate person-, task-, and reference spaces to improve the understanding of proximal pointing gestures in face-to-face remote collaboration in virtual environments; 2) Pointing Agreement, a concept that encapsulates referential consensus between two face-to-face collaborators; 3) A user study, demonstrating the effectiveness of our technique.

2 RELATED WORK

MAGIC builds on research in workspace awareness in shared spaces, interpretation of pointing gestures in Collaborative Virtual Environments (CVEs), and perception manipulation techniques.

2.1 Workspace Awareness in Shared Spaces

Workspace awareness relates to the immediate changes that occur in a shared workspace, and can be defined as the "*up-to-the-moment understanding of another person's interaction with a shared workspace*" [26]. Greenberg et al. [24] suggest that this real-time knowledge of another person's interactions and their effects on the workspace is essential to effective collaboration. As working together causes people to undertake the additional task of maintaining collaboration through communication and decision making, apart from their individual domain jobs [2], inadequate workspace awareness causes people to perform more challenging and awkward collaboration tasks, which, in turn, causes the domain tasks to be more laborious [64].

In addition to mechanisms such as *feedthrough*, *consequential communication* and *intentional communication* being challenging to recreate when collaborating remotely, most telepresence systems typically rely on a separation between the *person space* and the *task space*.

Buxton [12] suggested that "*effective telepresence depends on quality sharing of both person and task space*", with collaborators meeting in a shared space where they can refer the workspace using gaze or deictic gestures, the *reference space*.

An early example of this concept is the Clearboard [36], where two participants engage in collaborative drawing tasks while seeing each other face-to-face. To correct for the inaccurate reference system, the authors resorted to horizontally reverse the video streams to establish the same point-of-view for both participants. To enable users to collaborate in a common reference space, other approaches combined physical and virtual objects on arbitrary surfaces [39], rendered the remote user's hands on the local space [62], or mirrored remote users' representations rendered on top of a whiteboard content [76]. Leithinger et al. enabled users to manipulate physical shapes, with the person space in a face-to-face or corner-to-corner f-formation on the sides of the shared workspace [47]. These demonstrated that face-to-face telepresence approaches allow for improved effectiveness when compared to side-by-side settings [76], and suggested that to maintain awareness in face-to-face interactions, telepresence systems should resort to selective image reversal of text and graphics [48].

Speicher et al. [65] provided users with a third-person view by default but enabled one collaborator at a time to gain control over everyone's 360° video. This enabled all collaborators to view directions in a synchronized manner. Similarly, Teo et al. [69] allowed experts to interact with trainees in a third-person view by default, but with the option to immerse themselves into the same point of view as the local collaborator, live-streamed through a 360° camera attached to the trainee's head. Cai et al. [13] demonstrated a system that shows the remote user's virtual head and hands in a shared live 360 panorama captured from a backpack-mounted 360 camera. Lee et al. [45] created a system that allows sharing of hand gestures and awareness cues over live 360 panoramas captured by a head-worn camera, enabling two-way non-verbal communication between a pair of users wearing AR or VR displays. Similarly, Teo et al. [70] investigated adding ray pointing and drawing annotations to hand gestures as non-verbal communication cues in a live 360 panorama-based MR remote collaboration system, showing that participants acting as a local users were able to perform tasks faster and with less error with the help of visual annotation cues. On the other hand, Rhee et al. [59] created a system for enabling face-to-face collaboration using 360 panoramas that enabled a remote user to be teleported to a visually rich remote location with a visually-matched virtual object and share their interaction intentions with pointing gestures and voice.

Regarding 3D object manipulation, Benko et al. [10] introduced *MirageTable*, to bring people together face-to-face as if they were working at the same table. Participants share the same task space, interacting with physical objects, although this approach did not seemingly avoid occlusions. Additionally, Kim et al. [40] explored the combination of different visual cues (e.g., pointer, sketching, and hand gestures) showing a significantly higher level of usability when the sketch cue is added to a hand-pointing gesture cue alone.

We conclude that face-to-face encounters provide benefits to remote collaboration, although pointing at 3D objects from different perspectives still induces ambiguities since collaborator's gestures may be occluded. We tackle this challenge by implementing task- and personal spaces in combination. We also hypothesize that enabling the same perspective while correcting people's actions can improve intentional and consequential communication mechanisms as it can make users' gestures and other visual actions more understandable by both parts of a remote encounter.

2.2 Manipulating People's Perception

In virtual meetings, the sense of presence of remote users has an essential role in the meeting participant's capacity to communicate and collaborate. Previous research suggests that utilizing complete or upper-body representations improves people's awareness [9, 12] since a richer vocabulary combining body language with speech can be used. Furthermore, understanding the other person's gaze [53], communicative gestures [8, 41] and deictics [21, 66] are known to

improve remote collaboration. To improve how one perceives the remote participants' intentions and actions, different warping techniques has been used by taking advantage of manipulating human perception.

A classic example of this approach is the *Redirected Walking* [58], where the virtual environment is transformed so that the person can use a natural metaphor to cover larger distances in a limited physical space [6]. Besides the environment, retargeting techniques can be applied to the representation of people. This was first introduced with inverse kinematics to animate virtual characters when incomplete motion was available [22, 46], and for one single person to simultaneously animate multiple skeletons [29]. Feuchtnner et al. [20] extended the users' arm to allow for access to devices that are out of reach. Other approaches use motion remapping to modify real people's poses to match the motion of a person in a given video [15]. Recent work also suggests that the interpretation of deictic gestures can be significantly improved by using retargeting techniques that warp the pointing arm [49, 63, 76].

Piumsomboon et al. [56] found that manipulating the remote user's representation in order to guarantee that his gestures were always in the field of view of the local user enables a high sense of presence and increases task performance over an unmodified avatar. Sousa et al. [64] studied different manipulations of remote people and workspaces, mirroring one or the other in order to enable a shared point of view for the users. Results suggested remote collaboration benefits more from workspace consistency rather than people's representation fidelity. Hoppe et al. [34] also manipulated virtual people's representation to provide a shared point of view, finding that modifications of the workspace and the user's avatar to induce a shared perspective reduces mental load and increases task performance. And later, in [35], Hope et al. further explore shared perspective by comparing the baseline condition of two users standing in the same location and working with overlapping avatars, with their approach of shifting the remote collaborator's representation to the side and redirecting their gestures to match the reference space using laser pointing.

In *MAGIC*, we warp the remote collaborator's representation to guarantee that their gestures match the reference space when enabling a share of perspective. Similarly to [22, 46, 56], we apply inverse kinematics when redirecting the remote avatar's gestures.

2.3 Interpreting Pointing Gestures in CVEs

The problem of interpreting remote pointing gestures in collaborative tasks has been long considered. In the context of 2D gestures, Greenberg et al. [25] considered the problem of representational differences between two collaborators' views when using 2D telepointers, outlining potential solutions such as scaling or object-based pointing. Hindmarsh et al. [31, 32] also focused on the difficulties in identifying what other people are pointing at in a VR environments, identifying difficulties resolving pointing gestures when the pointing embodiments and target objects were not both in view. As a possible solution, they suggested to focus on exaggerating the representation of actions so that they can easily be seen by others. Wong et al. extended Hindmarsh et al's work to quantify the amount of error in people's interpretation of others' gestures [73, 74] when pointing in distance. Similarly, Sousa et al. [63], and later Mayer et al. [49], proposed solutions for correcting this problem by using retargeting techniques. However, these works are focused on *distal pointing* for *side-by-side f-formations* with targets at a fixed distance along 1D [63] or 2D [49]. In our work, we address the problem of *proximal occluded pointing* in *face-to-face scenarios* for targets anywhere in the 3D space reachable by the user.

2.4 Discussion

We found approaches [30, 36, 48, 76] successful in ensuring all three types of workspace awareness mechanisms for 2D artifacts, and approaches [7, 10, 47] that support 3D artifacts, but have a limited

capacity to ensure intentional communicational cues, hindering the use of pointing gestures to refer the workspace. To the best of our knowledge there is no approach for remote face-to-face object-centered collaboration with 3D content that allows all three types of workspace awareness mechanisms.

Analysing Ishii et al. [36]’s view on the problem of how to communicate in a shared environment without separation between the task and the personal space, we concluded that the *over the table metaphor* would be the most suitable for object-centered collaboration. We believe that by enabling a share of perspective between collaborators, we can tackle the challenge of mismatched views of the task space and occlusions present in approaches with a similar metaphor [7, 10, 47]. We resort to manipulations of the remote person’s representation, as supported by [34, 64], to enable a face-to-face meeting paired with a shared perspective of the workspace. Regarding the interpretation of pointing gestures in CVEs, we build on top of previous work [49, 63, 73, 74] exploring retargeting techniques to correct distal pointing interpretation in 1D and 2D, to address the problem of proximal occluded pointing in 3D face-to-face scenarios. Similarly to [22, 46, 49, 56, 63], we apply inverse kinematics when redirecting the remote avatar’s gestures.

3 MAGIC

Following the assumptions of Gutwin and Greenberg [26] that peripheral tasks bring additional efforts in maintaining collaboration, we propose to diminish the shift of attention between the personal space and the task space by creating a virtual environment to integrate the two, improving workspace awareness. To achieve that, MAGIC employs body representation manipulations in ways unnoticed by users to improve the understanding of pointing gestures used in the shared space, letting users focus on the main collaborative task. The present work addresses collaboration with two users only. While this is arguably the simplest case, future work could explore the extension of our work’s principles to accommodate multiple people. In the following, we detail the design and implementation of MAGIC.

3.1 Improving Workspace Awareness

MAGIC assumes an “above-the-table” metaphor [36] with life-sized virtual representations of remote people in front of one another with the workspace between them, as depicted in Figure 1. In a face-to-face f-formation, participants can perceive natural gestures besides maintaining verbal communication, which contributes to presence and increases awareness of other people’s actions. This reduces the separation between task space and personal space. Hence, it enables the use of pointing gestures when interacting with the workspace’s virtual artefacts. Yet, the 3D virtual model’s occlusions can diminish workspace awareness. To tackle this problem, MAGIC guarantees that the two participants share the same point of view of the workspace at all times. Sharing the same perspective in object-centred remote collaboration allows users to have a mutual understanding of the task space, allowing for the use of natural communication, easing the pressure of engaging in lengthy verbal cues about task-specific points of interest [19].

However, to allow for a face-to-face setting with the same perspective in shared 3D workspaces, we can’t only render the remote person in front of the local collaborator. This naive approach makes the remote collaborator’s gestures spatially incorrect as their gestures will not match the reference space. To maintain the same reference space, MAGIC manipulates the remote person’s representation so that there is a matching between the local observer’s pointing location and the point of interest to which the remote person is referring. Therefore, MAGIC aims to increase workspace awareness in object-centered three-dimensional collaboration by improving the pointing agreement between two collaborators. It integrates person-, task- and reference spaces, enabling collaborators

to thoroughly express and perceive consequential communication, feed-through, and intentional communication.

Consequential Communication: Since MAGIC adopts a life-sized face-to-face arrangement, the remote collaborator is always visible, similarly to traditional face-to-face interactions. Therefore, the local person can observe posture and body language, inferring significant information about what is happening in the workspace. This way, remote collaborators do not need to switch back-and-forth between the person space and the task space (requiring fewer eye-movements and no body rotations away from the workspace).

Feedthrough: When manipulating artifacts in the workspace, they provide feedback information to both the person performing the action and the observer. Thus, sharing the same perspective of the workspace allows direct object manipulations to retain position and orientation in both the local and remote workspaces, assuring that feedthrough.

Intentional Communication: MAGIC assures that gestures in the remote workspace are correctly converted to the local reference space. And since both participants have the same understanding of the workspace artifacts’ whereabouts, pointing gestures performed remotely remain meaningful when converted, allowing collaborators to communicate using natural language and gesture freely and accurately.

3.2 Manipulating users’ virtual representation

To combine the advantages of both being face-to-face and sharing perspective, both participants stand on the same location, sharing the same point of view of the 3D model, but see a manipulated version of their partner. These manipulations include mirroring along left-right (z-axis), similarly to Zillner et al. [76] followed by an adjustment of the remote person’s overall pose to correct for the depth of the interaction.

In Figure 2, we illustrate the different steps necessary to accomplish the manipulations that enable our approach, where a remote collaborator intentionally communicates the position of a workspace artifact using a proximal pointing gesture.

In their local space, the remote collaborator points to an area of interest within the shared workspace to communicate a specific area of the 3D artifact (Figure 2 [A]). In a veridical face-to-face scenario, the remote pointing gesture is occluded from the local participant (Figure 2 [B]). With MAGIC, we follow the following steps to make the remote pointing gesture visible by the local user.

1. Workspace Local Copy: We first proceed to the rendering of the workspace artifact from the same PoV as the remote user their local space. In the local space, without any manipulation, the remote user’s gesture highlights a different area of the task space (Figure 2 [C]).

2. Mirror: Similarly to Clearboard [36] and 3D-Board [76], we then mirror the remote person’s representation (Figure 2 [D]). A local observer can correctly understand the remote person’s interactions in the workspace’s left-right axis after mirroring. There is a common understanding of right, left. Body language matches horizontally in the local reference space.

3. Adjust Arm Pose: Since we are working in a 3D workspace, the mirroring by itself is insufficient, and depth needs to be corrected. We re-position the pointing arm along the forward-backward axis to correct for the depth of the interaction (Figure 2 [E]). The transformation vector, \vec{z} , applied to the remote avatar’s pointing finger, is calculated through the difference in position from the remote collaborator’s fingertips in his local space, and the remote avatar’s fingertips after the mirroring process (Figure 3). We note that all the arm segments need to move when changing the pointing finger’s end position. We need to displace the upper arm, the forearm, and the wrist according to the finger’s new position. We can calculate these movements using Inverse Kinematics constrained by shoulder, elbow and wrist joints. As we focus on proximal pointing, the finger

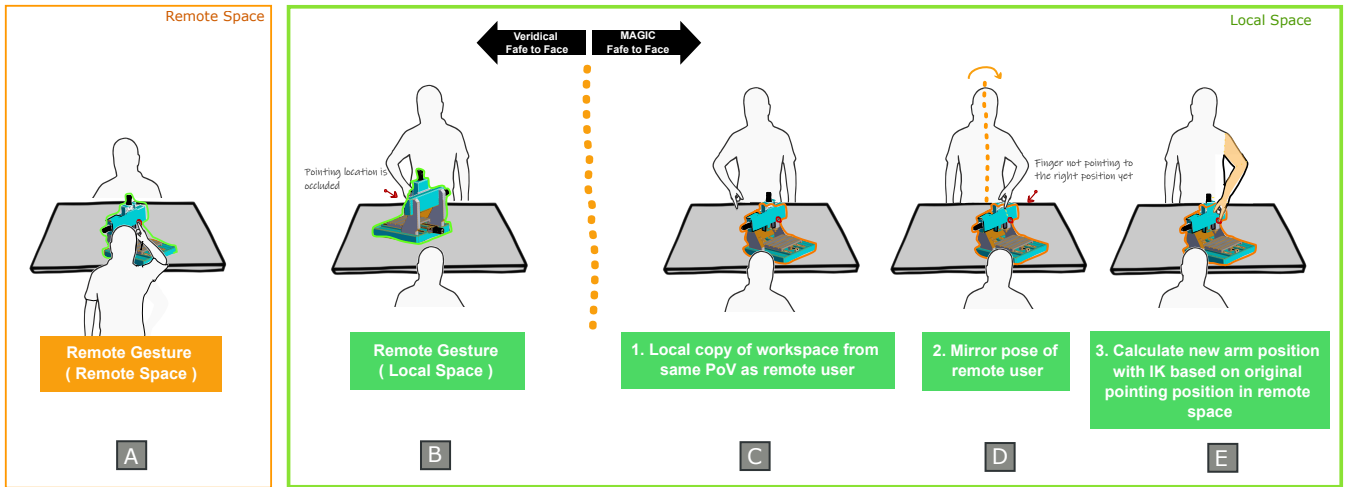


Figure 2: Illustration of the different steps involved in implementing our approach.

direction itself does not influence understanding of position as long as the fingertip is in the right place [60].

Additionally, if the remote person is in a position relative to the worktable where, in order to match the workspace’s pointing position, requires stretching of the arm beyond the arm length, we add an additional step: moving the remote person’s avatar further along the forward-backward axis. An arm that is too long could make the avatar representation weird for the local observer, breaking the user’s sense of presence [52]. As we want to adjust the avatar’s position, p_a , according to the remote collaborator’s index fingertip position, p_f , we compute the remote avatar’s head position as a linear function of the updated fingertip position:

$$p_a = m \cdot p_f + b, \quad (1)$$

where m and b are calculated based on the limit points of the shared space considered (the table dimensions in our case).

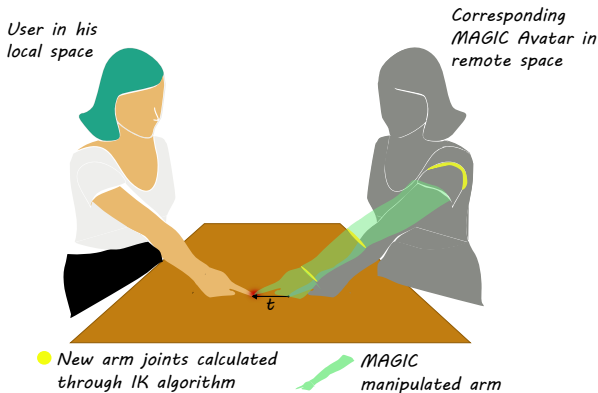


Figure 3: Manipulated avatar (right) corresponding to a pointing user in his local space (left). Transformation \vec{t} is calculated through the difference in position from the remote collaborator’s fingertips in his local space (left), and the remote avatar’s fingertips after the mirroring process (right - grey). \vec{t} is applied to the remote avatar’s fingertip in combination with IK algorithm to calculate new arm joint’s positions.

4 EVALUATION

We conducted a user study to determine whether our approach improves face-to-face collaboration in 3D object-centered remote collaboration settings when both participants share the same perspective, when compared to a *veridical* face-to-face. Therefore, our hypothesis throughout the experiment is:

H: A face-to-face setting coupled with manipulations that enable perspective sharing improves the understanding of proximal pointing gestures, increasing workspace awareness.

Additionally, as we designed our approach to be used implicitly, we also wanted to confirm that it did not distract users from the main task. Furthermore, we wanted to assess whether the shared point of view with corrected gestures impacts the feeling of presence of the remote collaborator.

To test our hypothesis, we conducted a within-subjects user study where we asked pairs of participants to complete a collaborative pointing task under two different conditions:

1. MAGIC Face-to-Face: Participants can see their partner located in front of them, both share the “illusion” of standing on opposite sides of the workspace but are in fact sharing the same point of view. Remote participant’s representation is manipulated in order to enable the illusion.

2. Veridical Face-to-Face: Participants can see their partner located in front of them, each one standing on opposite sides of the model, with opposing points of view of the workspace, as it occurs in “traditional” real life face-to-face interactions.

These two conditions were designed to allow us to compare MAGIC interactions with “traditional” face-to-face collaboration, with the goal of improving face-to-face collaboration specifically. We did not compare or address mirrored side-by-side settings, since these prevent visualizing both the workspace and the opposite participant, making it difficult to take advantage of gaze direction, posture and body language cues. Additionally, in order to use proximal pointing, side-by-side collaborators need to invade each other’s personal space which is avoided if standing face to face as the table delimits each other’s personal spaces. Otherwise, collaborators need to stick to distal pointing which has been shown [63] to induce errors in MR if uncorrected. Hence, we discarded side-by-side configurations and did not consider it in our experiments.

4.1 Task

We paired participants and asked them to complete a series of outlining tasks using proximal pointing. Each participant was assigned

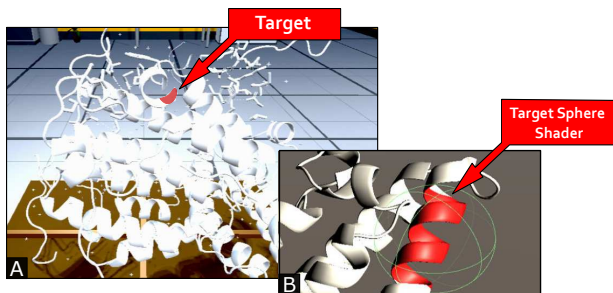


Figure 4: A) Abstract representation of a protein used for the task with red highlighted target to be indicated by the participants. B) Target sphere target.

a different role in the beginning of the experiment, either *Demonstrator* or *Interpreter*. The *Demonstrator*'s role was to communicate a highlighted area of the task model to the *Interpreter* by outlining it with his pointing finger. The *Interpreter*, who could not see the highlighted area, was required to follow the *Demonstrator*'s gestures and perform an outline of the interpreted area.

For the 3D model containing the targets participants had to indicate, we adopted an abstract representation of a protein, which was placed above the shared table (Figure 4), visible to both participants. We employed an abstract 3D model so that it would be hard for participants to describe targets using only words and would be compelled to using pointing gestures. Participants could talk with each other freely throughout the experiment.

4.2 Procedure

All evaluation sessions followed the same structure, and each lasted for about 50 minutes. Each session gathered two participants and the experiment moderator. In response to the global circumstances caused by the COVID-19 outbreak, we undertook extra safety measures. Afterwards, participants completed a demographic profile questionnaire and a consent form. We then introduced the evaluation, where we explained conditions, tasks, and roles. Before executing the main task, participants had a training period to familiarize themselves with the environment and the hardware, which consisted of three outlining tasks.

To start each matching task, the Demonstrator was asked to press a button on the controller, and once the workspace appeared, the target to communicate was shown as a small part of the workspace highlighted in red (Figure 4 A). The Demonstrator had to communicate that red target to their partner. To do so, the Demonstrator was asked to outline it with his finger while pressing a button from the controller, which would leave a green trail in the area outlined (Figure 8 A). This was only visible to the Demonstrator. The outline stopped once the button was released, ending the Demonstrator's turn. Then, the Interpreter proceeded the same way, pressing a button while outlining his interpretation of the volume indicated by the Demonstrator, and finishing his turn by releasing the button. Once an iteration finished, the target shown to the Demonstrator changed and the procedure was repeated. The Demonstrator was shown a set of 16 targets to communicate to their partner, one a time. Upon completion, we asked participants to answer a user preference questionnaire regarding the condition experienced. Then, participants switched roles and performed the same task using a different set of 16 targets.

Hence, each pair of participants performed under both conditions, and each participant experimented both roles (2 Conditions x 2 Roles x 16 Targets). As participants performed a similar task in more than one condition, we created four different sets of 16 targets and counterbalanced them to avoid carryover effects. Each target group comprised similar spots of the model that differed enough

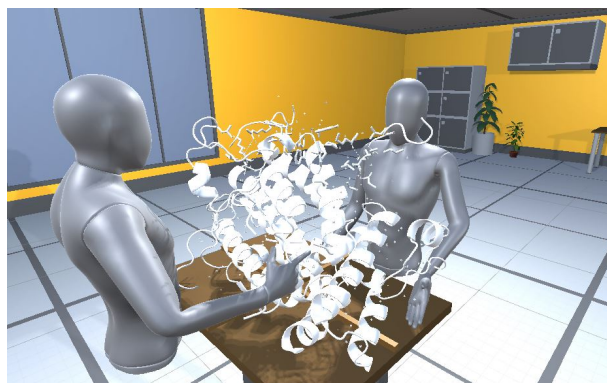


Figure 5: Our Prototype Virtual environment, featuring two collaborators and the abstract model used for the task.

to guarantee that participants would not remember them from the precedent task but were similar enough to ensure consistency of tasks between participants and conditions. The order of conditions was also balanced.

4.3 Evaluation Prototype

We implemented our prototype using Unity3D. Although we envision our approach as relevant for both AR and VR, we conducted our evaluation in VR to minimize external factors and avoid problems resulting from matching real and virtual content in current AR hardware, and have a fully controllable experience. We created a simplistic office space with a tabletop and a 3D model between both users (see Figure 5). The prototype features an abstract avatar to embody each collaborator. We chose a humanoid abstract representation to avoid gender- and body-bias while focusing on nonverbal communication cues. We used an Oculus Rift CV1 as the visualization tool, and we used the controllers' positions and rotations input to animate the avatars using inverse kinematics [4, 5]. These data are then transmitted to the remote participant's prototype to animate the local avatar and display it on the remote participant's HMD. Whenever participants placed their hands inside the workspace, the virtual hand automatically adopted a pointing pose.

4.4 Setup and Apparatus

Our experiment used a controlled laboratory environment with no contact with the exterior, occupied by the two participants and the moderator. The physical setup consisted of two stations separated

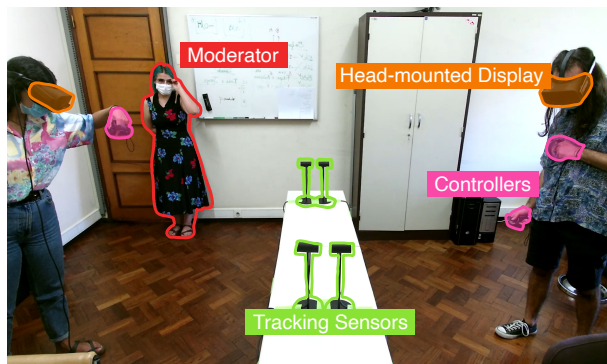


Figure 6: Setup for the experiment: two users in different tracked stations wearing an Oculus Rift HMD and using Touch Controllers. Moderator was present at all times.

by a physical divider. As participants were physically co-located, they could communicate verbally among themselves. Each station comprised a desktop running the participant’s application, and an Oculus Rift set (Oculus HMD, two Touch Controllers, and two-position trackers), as displayed in Figure 6.

4.5 Participants

We recruited 12 participants (3 female, 9 male) through convenience sampling in our institution. Participants were not compensated monetarily for participating in the user study. Participants’ ages ranged from 22 to 26 years ($M = 23.6$; $SD = 1$). All participants knew their partners. One participant reported being left-handed. Three participants indicated previous experience with using video-conferencing platforms at least once a day, eight at least once a week, and one reported a monthly use. Three participants reported never having experienced VR environments, while the other nine participants reported rarely using such environments.

5 RESULTS

Results from our user study suggest that MAGIC significantly improves *pointing agreement* in face-to-face collaboration settings without negatively affecting the time to perform a task. Furthermore, it improves co-presence and awareness of interactions performed in the shared space.

5.1 Collected Data and Metrics

During the user evaluations we collected information on both Task Performance and User Preferences.

TASK PERFORMANCE To evaluate task performance, we measured the task error by determining the *pointing agreement* between both participants. Additionally, we measured the task completion time.

5.1.1 Pointing Agreement

We define *Pointing Agreement* as a measure of what two users perceive in common when using pointing gestures in a shared 3D space. We evaluate the pointing agreement between two users by calculating the *Jaccard Similarity Coefficient* [38]. The Jaccard Similarity Coefficient, J , is a metric usually used in understanding the similarities between sample sets, and is formally defined as the size of the intersection divided by the size of the union of the sample sets, as follows:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}. \quad (2)$$

We apply this to the 3D sample sets of points that constitute the zone of interest pointed by each person. We chose this metric since it is able to characterize all possible scenarios in our experiments, as shown in Figure 7. Hence, the higher the Pointing Agreement, the less error was present in identifying the region indicated by the task demonstrator and vice versa. We note that, as we chose to designate parts of an abstract object, a centroid-based approach would not be accurate enough to show agreement as the target shapes are not spherical. Merely comparing 3D points would only measure the error distance and not highlight what participants had commonly perceived.

To measure J for each outlining task, we generate, for each participant, the convex set comprising all the points that constitute the outline of the zone of interest (Figure 8 A). This enables us to calculate the volume outlined by the Demonstrator, S_{Dem} (Figure 8 C), the one the Interpreter perceived, S_{Int} (Figure 8 C), and the volume of intersection between both, S_I (Figure 8 D).

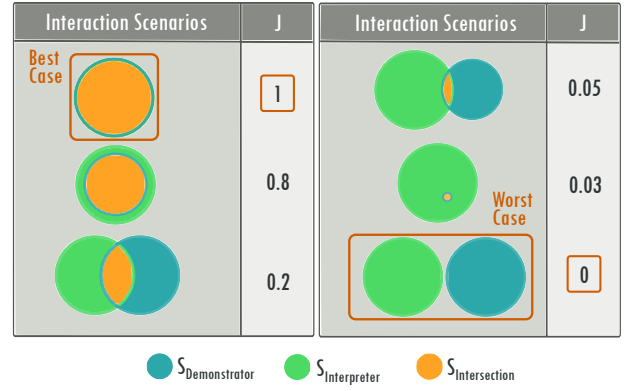


Figure 7: Different possible scenarios for outlines from best to worst case scenarios, and correspondent J . Blue, green and orange represent the Demonstrator’s Solid, S_{Dem} , Interpreter’s Solid, S_{Int} , and the Intersection of both, S_I

We use an adapted QuickHull Algorithm¹ to create S_{Dem} and S_{Int} , and a *Constructive Solid Geometry (CSG) Library*² to retrieve the intersection between the two, S_I . We calculate the volume of the meshes, $V_{S_{Int}}$, $V_{S_{Dem}}$, and V_{S_I} , by adding the volumes of all tetrahedra that compose the mesh. The tetrahedra joins each triangle in the mesh with the vertex at the origin [14]. From expression 2, J becomes:

$$J(S_{Int}, S_{Dem}) = \frac{|S_{Int} \cap S_{Dem}|}{|S_{Int}| + |S_{Dem}| - |S_{Int} \cap S_{Dem}|} \quad (3)$$

$$= \frac{V_{S_I}}{V_{S_{Int}} + V_{S_{Dem}} - V_{S_I}}.$$

USER PREFERENCES We also conducted a post-test assessment employing a user preference questionnaire that participants answered after completing the outlining task under each condition. We selected three sub-scales from the original social presence questionnaire proposed by Harms et al. [28]. This questionnaire included 13 statements regarding co-presence (i.e., the feeling of being together

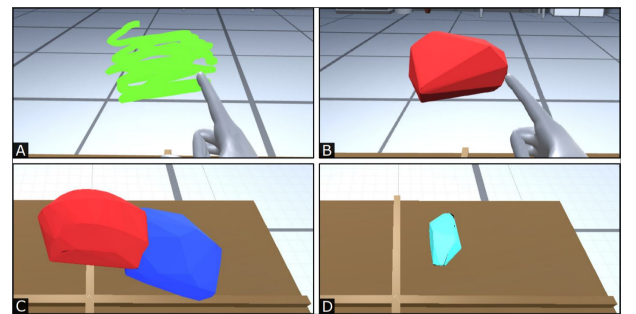


Figure 8: Calculating the intersection of the volumes outlined by both participants: (A) Outline of zone of interest; (B) Corresponding Convex Hull; (C) S_{Dem} (red) and S_{Int} (blue); (D) Resulting intersection, S_I .

¹Quickhull Algorithm for Generating 3D Convex Hulls - an Implementation in Unity, Oskar Sigvardsson (Accessed: 2020-05-15): <https://github.com/OskarSigvardsson/unity-quickhull>

²Constructive Solid Geometry (CSG) for Unity in C#, Andrew Perry (2020-07-22): <https://github.com/omgwtfgames/csg.cs/blob/master/Assets/CSG/Plugins/CSG/CSG.cs>

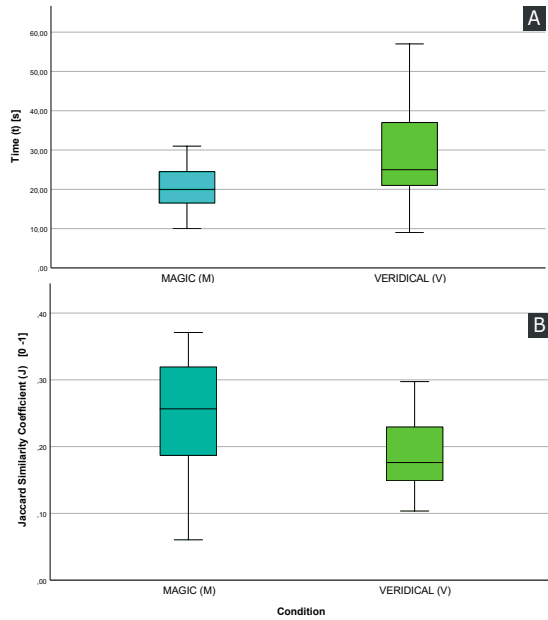


Figure 9: Task performance results for each condition: A) Completion Time, t . B) Jaccard Similarity Coefficient, J .

in the same shared space), attentional allocation (i.e., amount of attention given to each other), and perceived message understanding, to be answered on a six-point Likert-scale, from Strongly Disagree (1) to Strongly Agree (6). We also included two open questions at the end of the questionnaire: one that queried users if they were able to identify any strange behaviors (“Did you notice any strange behavior during the session?”), to evaluate if participants noticed the manipulation we were employing to their remote partner’s avatar, and another for further comments/improvements about the techniques and the general user study (“Observations and Suggestions”).

5.2 Analysis

TASK PERFORMANCE A Shapiro-Wilk test showed that results for both metrics were normally distributed. A paired t-test indicated that J was statistically significantly higher for MAGIC ($M = 0.24$, $SD = 0.03$) than for the Veridical condition ($M = 0.18$, $SD = 0.02$), $t(11) = -2.899$, $p = .014$. As for task completion time, we found no statistically significant differences between conditions, $t(11) = -2.00$, $p = .07$. Figure 9 shows results for task performance metrics.

USER PREFERENCES In terms of co-presence, the Wilcoxon test reported a significant difference for Statement 1.2 ($Z = -2.121$, $p = .034$), revealing an increase in feeling of co-presence of the remote partner in the virtual environment under MAGIC. Questionnaires showed no statistically significant differences between either approach in terms of Attentional Allocation and Perceived Message Understanding. However, we observed a tendency for better message understanding under the MAGIC condition. Furthermore, when performing under MAGIC as Condition 2, three participants reported in the “Observations and Suggestions” open question that targets were easier to understand, and one participant pointed out that they had to “move less than previously” to understand what their partner was pointing to. One participant also pointed that targets were harder to understand in the second part of the experiment when performing under Veridical as Condition 2. Details on the questionnaires and associated results can be found in the Appendix, Figure ??.

5.3 Discussion

Results obtained from the statistical analysis of the collected data confirm our research hypothesis, H . Results showed a higher pointing agreement for both participants when sharing perspective, yielding a better understanding of the collaboration tasks. Our approach improved pointing agreement by 6%, representing a relative improvement of 33.3% compared to the Veridical baseline condition. Participants also reported a better understanding of the target location under our approach, with some commenting that they had to move less to understand their partner’s pointing gestures. However, the questionnaire results did not show any statistically significant differences in Perceived Message Understanding statements between either condition. This could be due to a false sensation of agreement that was more prevalent under the Veridical condition, i.e., people thought they perceived their partner’s actions correctly when in fact, they failed the outlining task. Therefore, pointing agreement dropped under the Veridical condition even though participants felt a similar personal performance.

Additionally, results showed no statistically significant differences in time completion under either condition. We hypothesize this to be due to a false sensation of agreement during the veridical condition that led participants to just move on to the next task even though they “failed” the latter.

As both conditions employed a life-sized representation of the remote avatar in a face-to-face configuration, we can say consequential communication was ensured under both conditions. Yet, when sharing perspective, there was an increase of feedthrough mechanisms, as the manipulations of artifacts present in the workspace were visible to the local observer at all times, contrary to what happened in the Veridical condition. The increased agreement on the pointing task also highlights enhanced intentional communication under MAGIC, as people showed a better capacity to use and understand pointing gestures. Hence, we argue that MAGIC generally improves workspace awareness in terms of feedthrough, consequential, and intentional communication mechanisms.

As an additional finding, we should note that answers from the User Preferences Questionnaire reported a significant difference for Statement 1.2, revealing participants felt that their partner was more present in the virtual environment under MAGIC. We argue that MAGIC’s improvement in co-presence might be due to the fact that since the actions the remote collaborator performs in the workspace are both more visible and better understood, people feel their partner is more present in the workspace.

Regarding possible uncanny valley effects, our avatar was deliberately chosen as an abstract androgynous avatar, as these avatars are less likely to affect presence and embodiment significantly in unnatural behaviors [61]. We directly asked participants in the User Preferences Questionnaire whether they noticed any “strange” behaviors when performing the tasks. None of the participants reported noticing anything strange under MAGIC or the Veridical conditions, which leaves us to assume that the local collaborator did not notice distortions of the remote avatar, hence not revealing significant uncanny valley effects. However, we acknowledge that this could be due to the unrealistic avatar in itself, and that our distortions could potentially cause significant effects on presence and embodiment in more realistic avatars. To tackle this problem, the distortions we perform on the user’s movements could be optimized in the future by manipulating all points of a cloud representation instead of manipulating the three arm joints we used in our current implementation.

We also observed that participants performed the tasks using pointing gestures only and used verbal cues for keeping track of the partner’s understanding of their performed actions (with sentences like “This is the area I want to show you,” “Done!”), typically with one sentence before starting outlining and one sentence when finishing. As we developed our approach to tackling the problem of

the complexity of trying to explain relative positions through words, considering pointing comes as such a natural behavior to humans that avoids confusion arising from speech, we assumed that if one can easily point and say “here,” one would not feel the necessity to use “to your left/ right,” eliminating communication errors related to handedness. Indeed, no participant used relative position words such as left and right during our experiments.

In general, MAGIC improves pointing agreement, thus increasing task performance. Furthermore, MAGIC generally improves workspace awareness regarding feedthrough-consequential- and intentional- communication. Additionally, MAGIC distortions of remote avatars are not noticeable or impairing. Finally, MAGIC positively impacts the presence of remote collaborators in the workspace. Thus, we can affirm that “*a face-to-face setting coupled with manipulations that enable perspective sharing improves the understanding of pointing gestures, increasing workspace awareness.*”.

5.4 Limitations

Our experiments assumed static participant positions around the shared workspace. Further research is needed to see whether these results scale to moving (or more than two) participants.

Since our work falls into the category of a shared perspective *with the twist of maintaining the advantages of a face-to-face formation*, we were interested in comparing this with traditional face-to-face formations. However, we acknowledge that a comparison with a traditional share of perspective could have also been interesting.

Additionally, our approach was tested through an observational task, where users did not have the chance to manipulate the object. We wanted to introduce our new paradigm of view sharing and test its feasibility in a simple case still representative of design review scenarios: pointing to regions of interest of a fixed 3D object for the discussion of relevant points.

Our approach conceptually guarantees that there is a share of the same perspective of the workspace for both users. In our use case and implementation, this workspace was static, and only the remote user’s movements were manipulated. We believe that applying similar manipulations to a dynamic workspace will guarantee that both users keep the same perspective of the workspace, hence generalizing to interactive scenarios. We envision future work to include manipulation of 3D objects with MAGIC.

We should also note that our approach doesn’t avoid hand occlusions. Future work could address this problem by manipulating posture to ensure visibility at any time by, e.g., always bringing the hand to the top layer of rendering.

The experiments focused on objects commensurable with participants’ apparent sizes. An open question remains on how to explore the trade-offs between apparent size and level of detail.

Furthermore, MAGIC does not work with distal pointing since the manipulated pointing gestures only reference the correct place when the user touches the desired area. It could be interesting to extend our approach to enable distal pointing by dynamically changing the user representation depending on the task at hand [50], exploring different *Beyond Real Interactions* [1] such as extending the user’s arm in order to facilitate reaching an object as inspired by works such as in the Go-Go technique [57].

Our technique does not ensure verbal communication between participants located in separate physical spaces, despite it being crucial in the course of the collaboration. During our experimental sessions, participants could talk to each other since they shared the same room. Our findings suggest that co-localization only impacted verbal communication between participants, implying that our technique should be as effective for audio-enabled remote collaboration.

Additionally, during our studies, we asked pairs of participants to repeatedly point to the exact position of an object for internal validity, so that we could understand if MAGIC increases understanding of pointing gestures. A less controlled experiment will be valuable to

evaluate further MAGIC’s impact in a more general collaborative object-centered task.

It is also worth mentioning that the experimental sessions were conducted during the pandemic which presented challenges in gathering participants, 12 total. Nevertheless, our results are statistically significant.

Finally, our results show an increase of J from a medium value of 0.18 in the veridical condition to a medium value of 0.24 with MAGIC, representing an improvement of 33.3%. These illustrate situations where the pointing agreement is close to what is depicted in Figure 7 - 3rd left. Although this value has room for improvement, we believe this increase to be relevant in this context.

6 CONCLUSIONS AND FUTURE WORK

We introduced MAGIC, an approach to improve remote collaboration by enhancing pointing agreement in object-centered collaboration in virtual environments. Our approach allows people to be in a face-to-face f-formation, promoting the sense of co-presence, and improving object-related communication accuracy. MAGIC addresses the problem of different viewpoints and the natural occlusions that arise from it, by allowing people to view the same workspace PoV while subtly altering the remote participant gestures to match the same reference space. Our work is the first to support face-to-face collaboration with a shared viewpoint without apparent avatar manipulation or resorting to additional tools to perform the communication (e.g., pointers).

Results from a user study show that MAGIC improved participants’ pointing agreement during tasks using pointing gestures and increased the sense of co-presence, while being unnoticeable to users. These findings validate our initial hypothesis and clear the way for future research on remote face-to-face meetings in virtual environments. Although our study was conducted in VR, we envision our approach to be relevant for general Mixed Reality scenarios. However, this generalization to MR was not formally evaluated, remaining as future work.

Future work could extend our approach to multi-user settings and apply MAGIC to more complex real-life tasks, expanding to dynamic workspaces where users can manipulate the objects under analysis. Moreover, as we could see that subtle manipulations of pointing gestures had such positive impacts, future work should focus on more complex body representations. Interestingly, gaze and posture can communicate different intents in non-verbal ways, including cooperativeness, social status, and turn-taking, among others [3]. We would be interested in exploring whether we can manipulate such cues by subtly warping a remote speaker’s gaze and posture (e.g., correcting one’s gaze and posture to promote likeness and cooperation). Additionally, while other works have explored other ways of improving the use of pointing gestures in collaborative environments by complementing these with the use of pointers and sketches [16,40], we were interested in improving “raw pointing” to mimic real-world interactions better. Whether our approach could be further improved by complementing it with external tools remains future work. Finally, we designed our approach to manipulate remote user movements imperceptibly. However, we would like to explore whether perceptible exaggerated distortions have the potential to improve collaboration. Many inspiring examples come from the *Commedia dell’Arte* soap operas, cartoons, mime, and many other forms of communication through exaggeration, and more generally looking at user interfaces as theatre [43].

ACKNOWLEDGMENTS

This work is financed by *Fundação para a Ciência e a Tecnologia* (Portuguese Foundation for Science and Technology) through grants 2022.09212.PTDC (XAVIER), UIDB/50021/2020 and Carnegie Mellon Portugal grant SFRH/BD/151465/2021 under the auspices of the UNESCO Chair on AI&XR.

REFERENCES

- [1] P. Abtahi, S. Q. Hough, J. A. Landay, and S. Follmer. Beyond being real: A sensorimotor control perspective on interactions in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.
- [2] M. J. Adams, Y. J. Tenney, and R. W. Pew. Situation awareness and the cognitive management of complex systems. *Human factors*, 37(1):85–104, 1995.
- [3] R. K. d. Anjos, M. Sousa, D. Mendes, D. Medeiros, M. Billingham, C. Anslow, and J. Jorge. Adventures in hologram space: Exploring the design space of eye-to-eye volumetric telepresence. In *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, New York, NY, USA, 2019. ACM.
- [4] A. Aristidou, Y. Chrysanthou, and J. Lasenby. Extending FABRIK with model constraints. *Comput. Animat. Virtual Worlds*, 27(1):35–57, Jan. 2016.
- [5] A. Aristidou and J. Lasenby. FABRIK: A fast, iterative solver for the inverse kinematics problem. *Graph. Models*, 73(5):243–260, 9 2011.
- [6] M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson. 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*, pages 1968–1979. ACM, 2016.
- [7] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):616–625, 2013.
- [8] M. M. Bekker, J. S. Olson, and G. M. Olson. Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In *Proceedings of the 1st Conference on Designing Interactive Systems: Processes, Practices, Methods, & Techniques*, DIS '95, pages 157–166, New York, NY, USA, 1995. ACM.
- [9] S. Benford, J. Bowers, L. E. Fahlén, C. Greenhalgh, and D. Snowden. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 242–249, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.
- [10] H. Benko, R. Jota, and A. Wilson. Miragetable: Freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 199–208, New York, NY, USA, 2012. ACM.
- [11] P. Brunet and C. Andujar. Immersive data comprehension: visualizing uncertainty in measurable models. *Front. in Robotics & AI*, 2:22, 2015.
- [12] W. Buxton. Telepresence: Integrating shared task and person spaces. In *Proceedings of graphics interface*, pages 123–129, 1992.
- [13] M. Cai, S. Masuko, and J. Tanaka. Gesture-based mobile communication system providing side-by-side shopping feeling. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, IUI '18 Companion, New York, NY, USA, 2018. Association for Computing Machinery.
- [14] Cha Zhang and Tsuhan Chen. Efficient feature extraction for 2d/3d objects in mesh representation. In *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, volume 3, pages 935–938 vol.3, 2001.
- [15] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros. Everybody dance now. <https://arxiv.org/abs/1808.07371>, 2018.
- [16] M. Contero, F. Naya, J. A. Jorge, and J. F. C. Pastor. CIGRO: A minimal instruction set calligraphic interface for sketch-based modeling. In V. Kumar, M. L. Gavrilova, C. J. K. Tan, and P. L'Ecuyer, editors, *Computational Science and Its Applications - ICCSA 2003, International Conference, Montreal, Canada, May 18-21, 2003, Proceedings, Part III*, volume 2669 of *Lecture Notes in Computer Science*, pages 549–558. Springer, 2003.
- [17] T. Dodds, B. Mohler, and H. Bühlhoff. A communication task in hmd virtual environments: Speaker and listener movement improves communication. In *23rd Annual Conference on Computer Animation and Social Agents (CASA 2010)*, pages 1–4, 2010.
- [18] C. Elvezio, M. Sukan, O. Oda, S. Feiner, and B. Tversky. Remote collaboration in ar and vr using virtual replicas. In *ACM SIGGRAPH 2017 VR Village*, SIGGRAPH '17, New York, NY, USA, 2017. Association for Computing Machinery.
- [19] M. Feick, T. Mok, A. Tang, L. Oehlberg, and E. Sharlin. Perspective on and re-orientation of physical proxies in object-focused remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 281. ACM, 2018.
- [20] T. Feuchtner and J. Müller. Extending the body for interaction with reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5145–5157, 2017.
- [21] A. Genest and C. Gutwin. *Characterizing Deixis over Surfaces to Improve Remote Embodiments*, pages 253–272. Springer London, London, 2011.
- [22] M. Gleicher. Retargetting motion to new characters. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 33–42. ACM, 1998.
- [23] D. Gotsch, X. Zhang, T. Merritt, and R. Vertegaal. Telehuman2: A cylindrical light field teleconferencing system for life-size 3d human telepresence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 522:1–522:10, New York, NY, USA, 2018. ACM.
- [24] S. Greenberg, C. Gutwin, and A. Cockburn. Awareness through fisheye views in relaxed-wysiwiw groupware. In *Graphics interface*, volume 96, pages 28–38, 1996.
- [25] S. Greenberg, C. Gutwin, and M. Roseman. Semantic telepointers for groupware. In *Proceedings Sixth Australian Conference on Computer-Human Interaction*, pages 54–61. IEEE, 1996.
- [26] C. Gutwin and S. Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3-4):411–446, 2002.
- [27] E. T. Hall. *The hidden dimension*. Doubleday & Co, 1966.
- [28] C. Harms and F. Biocca. Internal consistency and reliability of the networked minds measure of social presence. 2004.
- [29] C. Hecker, B. Raabe, R. W. Enslow, J. DeWeese, J. Maynard, and K. van Prooijen. Real-time motion retargeting to highly varied user-created morphologies. In *ACM Transactions on Graphics (TOG)*, volume 27, page 27. ACM, 2008.
- [30] K. Higuchi, Y. Chen, P. A. Chou, Z. Zhang, and Z. Liu. Immerseboard: Immersive telepresence experience using a digital whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2383–2392, 2015.
- [31] J. Hindmarsh, M. Fraser, C. Heath, S. Benford, and C. Greenhalgh. Fragmented interaction: establishing mutual orientation in virtual environments. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 217–226, 1998.
- [32] J. Hindmarsh, M. Fraser, C. Heath, S. Benford, and C. Greenhalgh. Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction*, 7(4):477–509, 2000.
- [33] J. Hindmarsh and C. Heath. Embodied reference: A study of deixis in workplace interaction. *Journal of Pragmatics*, 32(12):1855–1878, 2000.
- [34] A. H. Hoppe, F. van de Camp, and R. Stiefelhagen. Personal perspective: Using modified world views to overcome real-life limitations in virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 577–578, 2018.
- [35] A. H. Hoppe, F. van de Camp, and R. Stiefelhagen. Shisha: Enabling shared perspective with face-to-face collaboration using redirected avatars in virtual reality. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW3), jan 2021.
- [36] H. Ishii and M. Kobayashi. Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 525–532, New York, NY, USA, 1992. ACM.
- [37] H. Ishii, M. Kobayashi, and K. Arita. Interactive design of seamless collaboration media. *Communications of the ACM*, 37(8):83–98, 1994.
- [38] P. Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.
- [39] S. Junuzovic, K. Inkpen, J. Tang, M. Sedlins, and K. Fisher. To see or not to see: A study comparing four-way avatar, video, and audio conferencing for work. In *Proceedings of the 17th ACM International Conference on Supporting Group Work*, GROUP '12, pages 31–34,

- New York, NY, USA, 2012. ACM.
- [40] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, and M. Billinghurst. Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. ACM.
- [41] D. Kirk, A. Crabtree, and T. Rodden. Ways of the hands. In *ECSCW 2005*, pages 1–21. Springer, 2005.
- [42] A.-K. Krolovtich and L. Nilsson. 3d visualization for model comprehension a case study conducted at ericsson ab. B.S. thesis, 2009.
- [43] B. Laurel. *Computers as Theatre*. Pearson Education, Inc., New Jersey, 2nd edition, September 2014.
- [44] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi. Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, pages 9–12. IEEE, 2016.
- [45] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics and Interactive Applications*, SA '17, New York, NY, USA, 2017. ACM.
- [46] J. Lee and S. Y. Shin. A hierarchical approach to interactive motion editing for human-like figures. In *Siggraph*, volume 99, pages 39–48, 1999.
- [47] D. Leithinger, S. Follmer, A. Olwal, and H. Ishii. Physical telepresence: Shape capture and display for embodied, computer-mediated remote collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 461–470, New York, NY, USA, 2014. ACM.
- [48] J. Li, S. Greenberg, E. Sharlin, and J. Jorge. Interactive two-sided transparent displays: Designing for collaboration. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, pages 395–404, New York, NY, USA, 2014. ACM.
- [49] S. Mayer, J. Reinhardt, R. Schweigert, B. Jelke, V. Schwind, K. Wolf, and N. Henze. Improving humans' ability to interpret deictic gestures in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.
- [50] J. McIntosh, H. D. Zajac, A. N. Stefan, J. Bergström, and K. Hornbæk. *Iteratively Adapting Avatars Using Task-Integrated Optimisation*, page 709–721. ACM, New York, NY, USA, 2020.
- [51] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago press, 1992.
- [52] M. Mori. The uncanny valley: The original essay by masahiro mori. *IEEE Spectrum*, 2012.
- [53] D. Nguyen and J. Canny. Multiview: Spatially faithful group video conferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 799–808, New York, NY, USA, 2005. ACM.
- [54] S. Orts-Escolano, C. Rhemann, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 741–754, New York, NY, USA, 2016. ACM.
- [55] T. Pejisa, J. Kantor, H. Benko, E. Ofek, and A. Wilson. 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*, pages 1716–1725. ACM, 2016.
- [56] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13, 2018.
- [57] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pages 79–80, 1996.
- [58] S. Razzaque, Z. Kohn, and M. C. Whitton. Redirected Walking. In Short, editor, *Eurographics 2001*. Eurographics Association, 2001.
- [59] T. Rhee, S. Thompson, D. Medeiros, R. Dos Anjos, and A. Chalmers. Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE transactions on visualization and computer graphics*, 26(5):1923–1933, 2020.
- [60] V. Schwind, S. Mayer, A. Comeau-Vermeersch, R. Schweigert, and N. Henze. Up to the finger tip: The effect of avatars on mid-air pointing accuracy in virtual reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '18, page 477–488, New York, NY, USA, 2018. Association for Computing Machinery.
- [61] M. Shin, S. J. Kim, and F. Biocca. The uncanny valley: No need for any further judgments when an avatar looks eerie. *Comput. Hum. Behav.*, 94:100–109, 2019.
- [62] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Macioccoi. Bethere: 3d mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 179–188, New York, NY, USA, 2013. ACM.
- [63] M. Sousa, R. K. dos Anjos, D. Mendes, M. Billinghurst, and J. Jorge. Warping deixis: Distorting gestures to enhance collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 608:1–608:12, New York, NY, USA, 2019. ACM.
- [64] M. Sousa, D. Mendes, R. K. d. Anjos, D. S. Lopes, and J. Jorge. Negative space: Workspace awareness in 3d face-to-face remote collaboration. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pages 1–2, 2019.
- [65] M. e. a. Speicher. 360anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality. *Proc. ACM Hum.-Comput. Interact.*, 2(EICS), jun 2018.
- [66] A. Tang, C. Neustaedter, and S. Greenberg. Videoarms: embodiments for mixed presence groupware. *People and Computers XX—Engage*, pages 85–102, 2007.
- [67] T. Tanimoto. An elementary mathematical theory of classification and prediction, ibm report (november, 1958), cited in: G. salton, automatic information organization and retrieval, 1968.
- [68] D. G. Tatar, G. Foster, and D. G. Bobrow. Design for conversation: Lessons from cognoter. *International Journal of Man-machine studies*, 34(2):185–209, 1991.
- [69] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, and M. Adcock. A technique for mixed reality remote collaboration using 360 panoramas in 3d reconstructed scenes. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, New York, NY, USA, 2019. Association for Computing Machinery.
- [70] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock. Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, OzCHI '18, page 406–410, New York, NY, USA, 2018. ACM.
- [71] L. Tong, A. Serna, S. Pageaud, S. George, and A. Tabard. It's not how you stand, it's how you move: F-formations and collaboration dynamics in a mobile learning game. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 318–329, 2016.
- [72] S. Whittaker. Things to talk about when talking about things. *Human-Computer Interaction*, 18(1-2):149–170, 2003.
- [73] N. Wong and C. Gutwin. Where are you pointing? the accuracy of deictic pointing in cves. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1029–1038, 2010.
- [74] N. Wong and C. Gutwin. 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*, pages 1377–1387, 2014.
- [75] E. Wood, J. Taylor, J. Fogarty, A. Fitzgibbon, and J. Shotton. Shad-owhands: High-fidelity remote hand gesture visualization using a hand tracker. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*, ISS '16, pages 77–84, New York, NY, USA, 2016. ACM.
- [76] J. Zillner, C. Rhemann, S. Izadi, and M. Haller. 3d-board: A whole-body remote collaborative whiteboard. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 471–479, New York, NY, USA, 2014. ACM.