MAGIC: Manipulating Avatars and Gestures to Improve Remote Collaboration

Catarina Gonçalves Fidalgo
catarina.fidalgo@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal

October 2020

Abstract

Remote work plays an essential role in projects in which geographically separated people need to perform joint tasks. Apart from saving time and resources, collaborating remotely is helpful in situations when confinement is required. Life-sized face-to-face telepresence promotes the sense of ”being there” and improves collaboration by allowing an immediate understanding of nonverbal cues. However, when discussing shared 3D content in a face-to-face setting, having different points-of-view of the model paired with occlusions raises ambiguities in analysis, decreasing workspace awareness. In this dissertation, we introduce MAGIC, a novel telepresence approach that improves remote collaboration in shared 3D workspaces by allowing participants to communicate through nonverbal cues while sharing the same workspace perspective, integrating task-, person-, and reference-space seamlessly. To enable a face-to-face setting with shared perspective, we manipulate the remote participant’s representations and gestures so that they correctly apply to the local person’s reference space. To evaluate our approach, we developed a Virtual Reality prototype that combines the virtual spaces of two remote collaborators so that they can work together in the same worktable. We capture the collaborators’ hands and head positions in real-time and use them to animate fully rigged avatars through inverse kinematics. Results from a user evaluation suggest that MAGIC is effective in improving task performance and increase workspace awareness. Furthermore, manipulations improved the sense of presence between remote collaborators despite being unnoticed.

Keywords: Remote Collaboration; Workspace Awareness; Nonverbal Communication; Perception Manipulation; Virtual Reality.

I. INTRODUCTION

Many leading organizations have been operating globally for decades, and, nowadays, global working has become a universal concept. Statistics from *FlexJobs & Global Workplace Analytics* show that from 2005 to 2017, the number of people working remotely in the U.S. increased 159% as it allows for considerable savings in time and resources. Furthermore, the current global circumstances caused by the Covid-19 Outbreak demonstrated that collaborating remotely is of significant importance in situations where social confinement is required.

For remote collaboration to work, traditional approaches using voice and video have been used for decades. However, these technologies neglect essential aspects of interpersonal nonverbal communication such as gaze, body posture, and gestures to indicate objects referred to in speech, which are essential to a face-to-face meeting [1]. These limitations negatively affect how remote people interact with a local user, especially for operations that require 3D spatial referencing and action demonstration.

3D models are used in many fields, having a crucial role in areas that develop extremely complicated structures, such as the Aerospace Engineering Sector. The design process of aircraft components often requires a panel of experts to analyse all the 3D elements to guarantee no errors are made during their development, which requires a rigorous spacial comprehension of the 3D models.

Key new technology for enabling effective remote collaboration is Mixed reality (MR), which is growing increasingly capable and accessible every year. In fact, Embraer already takes advantage of a Virtual Reality Center, the CRV, which enables Embraer’s engineers to visualize the aircraft’s structure and systems during the development phase.The CRV allows for Design Review Technical Meetings where engineers examine the different aircraft 3D models to check for interferences in their integration in the aircraft, reducing the time spent on the development of any aircraft when compared to traditional 3D softwares.

Remote collaboration through MR allows teams to meet in life sized representations and spontaneously discuss a topic, as if sharing the same office, enabling a high level of co-presence. Indeed, previous research found that for a remote meeting to be closer to a co-located experience, it should rely on real size portrayal of remote people to maintain the sense of “being there” [2], [3] and, furthermore, people have higher task performance when able to communicate using full body gestures [4].

However, engineers still face some challenges in these type of virtual encounters. When virtually collaborating face-to-face, we have to deal with opposing points-of-view of the workspace, do not share the same forward-backwards orientation, there is no common orientation of right or left 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 [5] and causing tasks to be more laborious. For the purpose of analysing 3D content, a shared perspective is more effective and preferred when compared to an opposing point of view [6], [7]. Unfortunately, in the real world, it is impossible to collaborate face to face while sharing perspective.

In this article, we present MAGIC, a new approach on collaborative remote work that enables two people to engage in face-to-face collaborative work, while sharing each other's perspective of the workspace. In order to enable the share of perspective in a face to face setting, the real size virtual representations of collaborators are subtlety manipulated so that the remote person's gestures make sense to the local observer.

Our research attempts to prove the following hypotheses:

H1: *A face-to-face setting coupled with manipulations that enable perspective sharing improves workspace awareness.*

H2: *The manipulations applied to the remote person's representation are not noticeable by the local collaborator.*

H3: *The manipulations applied to the remote person's representation improve the feeling of co-presence.*

Results of our user study showed an improvement on people's capabilities of understanding deictic gestures by 13% using our approach when compared to the traditional face-to-face setting. Our manipulations were not detected by the users, nor they had any negative impact on their performance.

We contribute 1) *MAGIC* as an interactive space that integrates person-, task-, and reference spaces for face-to-face remote collaboration in mixed reality environments; 2) implementation details on manipulation techniques to reshape the gestures of people's virtual representations; 3) implementation details of the virtual office prototype we developed to study our approach; 4) a user study, evaluating the impact of our approach on workspace awareness.

In what follows, we first examine related work and the design of our approach., and report on the user study and provide a discussion of the findings. Finally, we discuss design considerations for future face-to-face collaborative scenarios and directions for future work.

II. RELATED WORK

Previous research suggests that, to enhance remote collaborative experiences, staying aware of others and aware of the workspace are two key fundamental factors. MAGIC builds on top research in workspace awareness and sense of presence in virtual shared workspaces, virtual representations of people and perception manipulation.

II-A. Workspace Awareness

Workspace awareness can be defined as the *"up-to-the-moment understanding of another person's interaction with a shared workspace"* [8], and does not only relate to the contents and the immediate changes that occur in the workspace, but also to what actions people are doing. While maintaining awareness of what surrounds us in a real world workspace is something we take for granted, the same does not happen so easily in virtual collaboration systems. Hence, it has become increasingly evident that developing workspace awareness in remote collaboration systems is the key to making these types of encounters less awkward and more appealing to the public [8]. A real-time knowledge of another person's interactions and their effects on the workspace is essential to an effective collaboration. When working alone people are solely focused on completing the domain tasks required to achieve their goals. However, in a collaborative setting, meeting participants have to constantly carry out the collaboration tasks of communication and decision making, apart from their individual domain tasks. Hence, an inadequate awareness of the workspace makes people perform more challenging and awkward collaboration tasks which, in turn, causes the domain tasks to be more laborious.

To achieve workspace awareness, Gutwin and Greenberg [8] suggest that people achieve awareness naturally in the everyday world using mechanisms such as *feedthrough*, *consequential communication* and *intentional communication*.

Feedthrough is the ability to perceive how the artifacts within the workspace change as they are being manipulated, as well as if actors and artifacts are visible. Consequential communication relates to the perception of where people are looking (*gaze awareness*) and if the their actions can be understood by seeing the performance of that action (*visual evidence*). Lastly, intentional communication happens when people are able to include gestures to qualify verbal references to artifacts on the workspace (*deictic gestures*) and use *demonstrations* to convey actions. These mechanisms can all be achieved in co-located encounters, yet they are extremely difficult to acquire in remote collaboration because of current technological limitations. Low-fidelity representation of remote people and separation from task and person space are some of the factors that contribute to a decrease in workspace awareness and presence.

Our work follows the workspace awareness fundamentals to improve remote collaboration. We hypothesize that perception manipulation techniques can facilitate the access to the knowledge of what actions remote people are performing and their impact on the artifacts present on the shared workspace, improving consequential communication as well as feedthrough mechanisms in remote face-to-face encounters. Intentional communication can also be improved by these techniques, as they can make deictic gestures and other visual actions accurate for both the participants.

II-B. Awareness of Remote People

In virtual meetings the sense of presence of remote people has an important role in their capacity to communicate and collaborate. Previous research suggests that utilizing full or upper-body representations improves people's awareness [9], [10], since a richer vocabulary combining body language with speech can be used. Furthermore, understanding the other person's gaze [11], communicative gestures [12], [13] and deictics [14] are known to improve remote collaboration.

While traditional systems were successful in this goal using video and audio streaming [15], [16], recent developments in commodity depth cameras enabled 3D representations that permit a more reliable life-size scale portrayal of remote people [17], [2]. Three-dimensional avatars have allowed immersive experiences with a high sense of co-presence to be created [18], [19], mimicking more closely real-world face-to-face interactions.

II-C. Awareness in Shared Workspaces

Telepresence systems typically rely on a separation between *person space* and *task space*. However, Buxton [20] suggested that it is also important to meet in a shared space to collaborate. For this, Buxton identified the reference space as the shared locus where people can refer to portions of the workspace using gaze or deictic gestures. An early example of this concept is the Clearboard by Ishii et al. [15]. In Clearboard, two participants can 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 as if they were side-by-side. Junuzovic et al. [21] presented the Illumishare approach that combines physical and virtual objects on arbitrary surfaces enabling participants to collaborate in a common reference space, sharing the same point-of-view. Yet, participants cannot share gaze since the person space is limited. Alike Illumishare, BeThere [22] resorted to depth sensors and augmented reality to render the remote participants hand enabling deictic gestures. Leithinger et al. [23] proposed a physical telepresence system where two remote people can manipulate physical shapes in a face-to-face or corner-to-corner formation on the sides of a shared workspace. Li et al. [24] reiterated Ishii et al. [15] findings and suggested that to maintain awareness in face-to-face interactions, telepresence systems should resort to selective image reversal of text and graphics.Zillner et al. [3] proposed the 3D-board, which enabled interaction with a whiteboard while mirroring remote participants' representations on top of the content.Furthermore, Zillner et al. [3] demonstrated that a face-to-face telepresence approach allows for improved effectiveness when compared to side-by-side settings.Regarding 3D object manipulation, Benko et al. [25] introduced MirageTable, a system that brings people together face-to-face as if they were working on the same table and interact with physical objects. Participants share the same task space, although this approach did not seemingly avoid occlusions.

We conclude that face-to-face encounters provide benefits to remote collaboration. Nevertheless, there is still the need for further improvements to deal with ambiguities when people look at 3D objects from different perspectives.

II-D. Manipulating People's Perception

Contrarily to what happens in the real world, it is possible in virtual environments to manipulate remote people's representations to achieve different goals.

Piumsomboon et al. [26] 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. [27] 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. [28] 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.

Recent work also suggests that the interpretation of deictic gestures can be significantly improved by using retargeting techniques that warp the pointing arm [29], [30].

In order to enable remote people to collaborate face-to-face in a shared 3D workspace while sharing each other's perspective, the representation of remote people needs to be distorted in such a way that remote and local participants spatial references match, providing effective collaboration.

In MAGIC, we introduce a new design space approach where we warp the remote collaborator's representation to guarantee that his gestures match the reference space when enabling a share of perspective. Similarly to [26], we apply inverse kinematics when redirecting the remote avatar's gestures.

In what follows, we describe our approach's concept and implementation, and then report on a user study evaluating our approach's benefits on the perception of remote people's gestures and nonverbal cues.

III. APPROACH

Following the assumptions of Gutwin and Greenberg [8] that peripheral tasks bring additional efforts in maintaining collaboration, we propose to diminish the shift of attention between the person space and the task space by creating a virtual environment where the two are integrated, improving workspace awareness. MAGIC avoids the problem of occlusions in the workspace by enabling its users to share perspective at all times, while maintaining possible the observation of the remote collaborator's gestures by rendering its virtual replica in front of the local user.

Next, we present a usage scenario for MAGIC, were our approach is used to facilitate the analysis of 3D model of a jet engine during a collaborative meeting between two aerospace engineers. We further provide detailed analysis of the impact of our work in workspace awareness, followed by an overview of our manipulation technique.

*III-A. **MAGIC Usage Scenario: A Jet Engine Design Review Meeting***

Kate and George, two aerospace engineers, are engaged in a remote meeting to discuss the development of a jet engine. George is developing the main engine while Kate is in charge of designing a Secondary Air System (SAS), and wants to show George her final design. Since they cannot have a face-to-face meeting, as they work for a multinational company and are currently in different countries, they scheduled a virtual meeting powered by the MAGIC approach. They meet facing each other across a virtual environment that includes life size depictions of their avatars in front of a worktable. George loads the CAD model of the project he has been working on, and the main system of the engine model appears in the shared workspace between them, above the worktable.

Halfway though the meeting, George points to direct Kate's attention to a change in the turbine configuration he would like to try. Kate notices however that the changes in the turbine will conflict with the simulations she ran indicating the best position for the (SAS) in the main engine.

Despite being face-to-face, both share the same perspective of the engine 3D model, and, using pointing gestures, Kate can indicate the optimal position she obtained for the (SAS) she is designing so that George takes that in consideration in the finalization of the main engine design. Figure 1 illustrates the remote meeting powered by MAGIC.

*III-B. **Improving Workspace Awareness***

In accordance with Ishii et al.'s [15] metaphors for how to communicate in a shared environment without separation between the task space and the person space, MAGIC assumes an "above-the- table" metaphor with life-size virtual representations of remote people.

In a face-to-face setting, participants can see each other face to face, being able to perceive physical gestures besides maintaining verbal communication, which contributes for a good sense of presence and increases awareness of other people's actions. Since the work happens above the same table, accessible by both collaborators, it is possible to use deictic references and pointing gestures to interact with the virtual objects in the workspace. This scenario appears to be ideal if it were not for the occlusions that can happen in the workspace, reducing its awareness.

Fig. 1. Kate uses deictic gestures to indicate the optimal position for the (SAS) she is designing. George understands the location she is pointing to and can keep designing the main engine with a clearer view of the system.

MAGIC tackles this problem by enabling the share of perspective between participants, preventing misunderstandings. However, to allow for a face-to-face setting with share of perspective, we cannot simply render the remote collaborator in front of the local user, as this turns the remote collaborator's gestures obsolete. To maintain the same reference space, MAGIC manipulates the remote person's representation so that there is a matching between the location the local observer perceives and the location to which the remote person is actually referring.

We believe MAGIC greatly contributes for the increase of workspace awareness in object centered three dimensional collaboration, as it integrates person-, task- and reference spaces, enabling collaborators to thoroughly express and perceive consequential communication, feedthrough, and intentional communication:

- *Consequential Communication*
  Since MAGIC adopts a life-size face-to-face arrangement, the remote collaborator is always visible, in a similar way to what happens in real life. Therefore, it is possible for the local person to observe his posture, body language and eye movement at all times, which brings a significant amount of information about what is happening in the workspace. Contrary to what would happen in a situation where two people are collaborating using video stream the need to oscillate the focus of attention between the person space and the task space is reduced (requiring fewer eye-movements and fewer body rotations away from the workspace).
- *Feedthrough*
  When the artifacts in the workspace are manipulated, they give feedback to both the person performing the action and the viewer. Thus, sharing the same perspective of the workspace allows object manipulations to remain identical in terms

of position and orientation in both the local and remote workspaces, ensuring feedthrough mechanisms are achieved.

- *Intentional Communication*
  Once the remote person's gestures are correctly converted to the local reference space, deictic references and gestural demonstrations can be used safely, with the guarantee that the two collaborators always share the same understanding of the workspace.

III-C. *Manipulation of virtual people's representations*

To combine the advantages of both being face-to-face and sharing perspective, both users stand on the same location, sharing the same point of view of the 3D model, but see a manipulated version of their partner.

In Figure 2, we illustrate the different steps involved in implementing the manipulations that enable our approach:

A) A collaborator points to an area of interest in Location A;
B) In Location B, without any manipulation, the gesture of the remote collaborator highlights a different area of the 3D model;
C) Similarly to Clearboard [15] and 3D-Board [3], we introduce a mirror-reversal of the remote person's representation. With the mirroring put in place, the local observer can correctly understand the remote person's interactions in the workspace's left-right axis. There is a common understanding of right and left and, body language and gaze direction match horizontally with the local reference space. However, since we are working in a three dimensional workspace, the mirroring by itself is insufficient and depth needs to be corrected.
D) Re-positioning of the pointing arm along the forward-backward axis to correct for the depth of the interaction. It is to be noticed that when changing the end position of the pointing finger, all the arm has to move accordingly, creating changes at the level of the upper arm, the forearm and the wrist that are calculated through an Inverse Kinematics algorithm.
E) Adjustment of the embodiment's position. If the remote person is in a position relative to the worktable that requires stretching of the arm to match the pointing position in the local workspace, the remote person's embodiment is moved further along the forward-backward axis, An arm that is too long could make it weird for the local observer, breaking the sensation of immersion.

In order to evaluate the validity of our approach, we developed a prototype where the previous manipulations could be applied. We describe the prototype's implementation process in the following section.

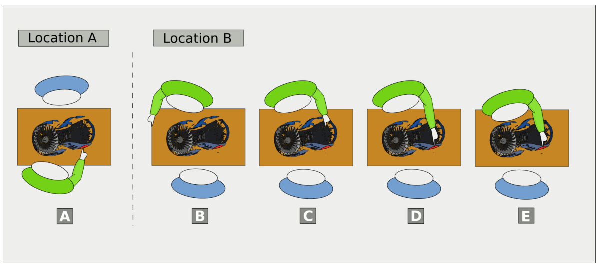

Fig. 2. Illustration of the different steps involved in implementing our approach.

IV. IMPLEMENTATION

In order to create a virtual environment that would allow us to study our approach, we developed a prototype using the Unity cross-platform game engine.

Our prototype combines the virtual spaces of both collaborators to allow them to work together sharing the same worktable in a face-to-face formation.

The prototype features an abstract avatar to embody each collaborator. We choose a far from detailed representation to fit different genders and body-types and focus on nonverbal communication cues. Our prototype involves three main modules: avatars' animation from user tracking motion capture; transmission of local data to the remote collaborator's space; and manipulation of the remote avatar embodiment in the local space

IV-A. *Tracking and Avatar Animation*

In order to allow the users to see the virtual space and track their movements, we opted to use the Oculus Rift VR hardware package, composed of an HMD, two touch controllers and two optical sensors.

We used the inbuilt Constellation Tracking System from the Oculus sensors to track the users' head and hands.This information was used to animate an abstract avatar model, which acted as the user's embodiment once he entered the virtual space.

Following Zibrek et al. [31] findings suggesting manipulations should rely on realistic behaviours, we animated the avatars in a way that replicated the users' movement as closely as possible.

For this purpose, we implemented the FABRIK solver [32], an inverse kinematics algorithm, to the avatar's embodiment. This algorithm allows us to determine an appropriate set of configurations for different joints so that the end-effector moves to the desired position as smoothly and accurately as possible.

Thus, we can restrict the position of the arm joints (shoulder, elbow and wrist) so that the avatar's hand is rendered in the position of the touch controller with an adequate arm movement.We opted to do an upper body representation since there is no information on how the legs and feet are moving.

IV-B. *Data Transmission*

In order for the local collaborator to be rendered in front of the remote collaborator, data about the local user's VR hardware information had to be transmitted to the remote space

To transmit the data, we opted for implementing a UDP communication protocol as our application was very time-sensitive, and a TCP communication protocol would induce latency in our prototype as it stops the streaming of data for error handling and packet retransmission in case of loss.

The message we sent through our UDP network included information on the different body parts of the local person's avatar that were needed to animate the person's avatar in the remote space: head, right hand, left hand, right fingertip and left fingertip.

Once the information was available in the remote space, we proceed to its manipulation.

IV-C. Movement Manipulation

The manipulations we employ to the remote collaborator's representation allow it to be rendered in front of the local user guaranteeing that his gestures make sense to the local observer.

- *Mirroring*
  We obtained the mirrored representation of the remote collaborator by inverting it along the x axis using the scale properties of Unity. The remote head is matched directly to the mirrored avatar without further manipulation, while we attribute the properties from the right hand (and index fingertip) to the left avatar hand, and vice versa.
- *Repositioning the pointing arm*
  To guarantee the remote avatar's hands point to the exact same space as the remote person's hands in his local space, we reposition the pointing arm along the transformation shown in Equations (2) and (1):

  $$\vec{t} = rf_p - af_p \tag{1}$$

  $$w_p = w_{p_0} + \vec{t} \tag{2}$$

  Where rf_p and af_p are the positions of the remote collaborator and remote avatar index fingertips, respectively, and w_{p_0} and w_p are the positions of the avatar's wrist before and after the transformation.
- *Adjust embodiment's position*
  There are scenarios where, in order for the remote avatar to reach its collaborator intended position, his arm would have to stretch beyond it's length. To avoid such a stretch, the avatar's position is adjusted in depth We apply the correction to the head position, as the body will move accordingly. The transformation of the avatar's position according to the location the remote collaborator's pointing finger follows linear equation (3):

  $$ah_{p_z} = m \cdot rf_{p_z} + b \tag{3}$$

  Where ah_{p_z} is the avatar's head position along the z axis, and rf_{p_z} is the remote collaborator's index fingertip position along the z axis, and $m = -0.823529412$ and $b = -0.26$.

V. EVALUATION

We conducted a user study in order to validate our research statement. Our main goal was to check whether people can maintain a shared understanding of the workspace while interacting in a face-to-face formation and sharing the same perspective of the workspace at the same time.

V-A. Design

The user evaluation aimed to show that when two remote collaborators work together with three dimensional models in a face to face disposition, if a collaborator indicates a specific area of interest in the shared workspace, the observing collaborator can identify it better if both of them share the same point of view. Additionally, we wanted to find out if the manipulations employed to the remote collaborator's representation to enable the shared perspective affected the feeling of co-presence between users.We carried out a user study where pairs of participants were asked to complete a collaborative "matching" task on an abstract 3D model, under two different conditions:

- *Face-to-face with our approach* (**MAGIC**): Participants were in opposite sides of the model but shared the same point of view of the workspace.
- *Veridical face-to-face* (**Veridical**): Participants were in opposite sides of the model with opposing points of view of the workspace, as in real life face-to-face interactions.

Participants were not made aware of the condition under which they were performing.

Each of the participants was assigned a different role: *Demonstrator* or *Interpreter*:

- The *Demonstrator* can see a red highlighted area of interest in the workspace. His 'job' is to communicate that area to the Interpreter by outlining it with his pointing finger.
- The *Interpreter* cannot see the highlight. Following the Demonstrator's gestures only, the Interpreter uses one of his pointing fingers to outline the interpreted area.

Once the workspace appeared, a small part of the model on top of the table would change color to red, and the Demonstrator had to communicate that red target area to his partner. In order to do so, the Demonstrator was asked to outline it with his finger while pressing the A or X Buttons from the Touch Controllers, depending on the hand he choose to use, which would leave a green trail in the area outlined

To highlight different parts of the abstract model, we developed a shaderthat features a "target sphere" that, when applied to an object, has the ability to change the parts of the object it intersects with to red

The Demonstrator was shown a set of 16 targets to communicate to his partner, one a time.

V-B. Procedure

All evaluation sessions followed the same structure and lasted for about 50 minutes. Participants started by fulfilling a profile questionnaire and a consent form. They were then introduced to the evaluation, where we explained conditions, tasks, and roles. Participants were able to pose any questions they might have on the experiment Before begging the experiment, participants took part in a training session that consisted of the execution of four practise matching tasks for familiarization with the environment and hardware. After, participants performed under both conditions. For each condition, upon completing the set of the fist 16 tasks, participants switched roles and executed the task again with a different set of 16 targets. After completion, they were asked to answer a questionnaire regarding that condition.

To avoid carryover effects, we created four different sets of sixteen targets - T1, T2, T3 and T4 - and counterbalanced them, along with the order of conditions.

*V-C. **Metrics***

The user study included both objective and subjective measurements. To evaluate task performance, we recorded the time participants took to complete the tasks, as well as their accuracy, measured as the percentage of intersection between the two areas of interest outlined by the participants.

As for user preference, we developed a User Preference Questionnaire, to be answered after completion of matching task under each condition. This questionnaire contained a set of statements regarding the feeling of co-presence between participants, attentional allocation and perceived message understanding.We also included an open question regarding if participants were able to identify any strange behaviors, to evaluate if participants noticed the manipulation we were employing to their remote partner's avatar and asked participants any suggestions they had to better the experiment plus any comment they would like to share with us.

V-C1 ***Accuracy of the matching task***

To measure the accuracy of the matching task, we opted to calculate the intersection between the two matching zones of interest: the one the Demonstrator outlined, and the one the Interpreter perceived. To find the percentage of intersection, we generated the 3D solid that forms from the convex set comprising all the points that constitute the outline of the zone of interest. For that purpose, we implemented a QuickHull Algorithm that returns a list of vertices, triangles and normals that can be directly converted into a Unity mesh.

Once we obtained the Demonstrator's Solid, S_{Dem}, and the Interpreter's Solid, S_{Int}, we computed the list of points that result from their intersection, using a CSG Library. From this set of points, we applied again the QuickHull algorithm to compute the Intersection mesh, S_I.

By comparing its volume with S_{Dem}'s volume, we can compute their percentage of intersection, $I_{\%}$:

$$I_{\%} = \frac{V_{S_I}}{V_{S_{Dem}}} \cdot 100 \tag{4}$$

*V-D. **Evaluation Communication Protocol***

During the experimental session, we had to guarantee that information on the current stage of the experiment was known both by the Demonstrator and the Interpreter's Machines.

We implemented a new network relying on a TCP communication protocol as we needed a reliable, ordered, and error-checked delivery of data for the experiment to keep running smoothly.

The Demonstrator's Machine, sends information to the Interpreter's Machine on when it should display the workspace or not and on when the participants should switch roles.

The Interpreter's Machine, sends to the Demonstrator a string containing the set of trail points used by the Interpreter to outline the area of interest, so that a percentage of intersection can be computed.

*V-E. **Setup and Apparatus***

Our experiment took place a room with no contact with the exterior, occupied by the two participants and the experiment supervisor only. The physical setup consisted of two separated stations composed by a desktop running the participant's application, and an Oculus Rift set (Oculus HMD, two Touch Controllers and two position trackers).

*V-F. **Participants***

Our subject group included five pairs of participants (10 total, 3 female). Participants' ages ranged from 23 to 24 years (M = 23,3; SD = 0,67). All participants declared having University Education and knowing their partner. Only 1 participant reported using his left hand as a dominant hand. Two participants indicated the use of video-conferencing platforms, such as Skype, Google Hangouts or Zoom, at least once a day, seven at least once a week and one reported daily use. As for virtual reality environments three participants reported they never used virtual reality environments and the other seven participants reported rarely using this type of environment.

VI. RESULTS

Throughout the evaluation experiments we collected information on Task Performance and User Preferences.

To analyse the gathered data under both different conditions, we first used the Shapiro-Wilk testto check for data normality. In case the gathered data had a normal distribution, we employed a Paired T-Test in order to determine if statistical significance was observed, and for non normally distributed data, we employed a Wilcoxon signed-rank test.

*VI-A. **Task Performance***

Running a Shapiro-Wilk test for the medium percentage of intersection obtained by each participant, we concluded that our accuracy results were normally distributed. We proceeded similarly with the medium completion time obtained by each participant and also

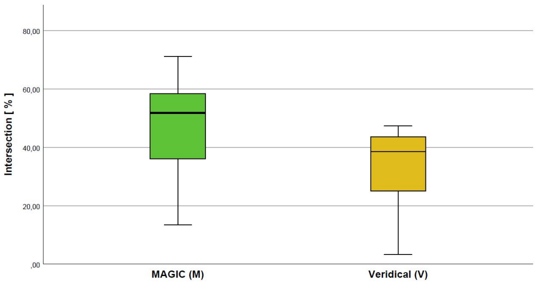

Fig. 3. Tasks' accuracy as intersection percentage for each condition.

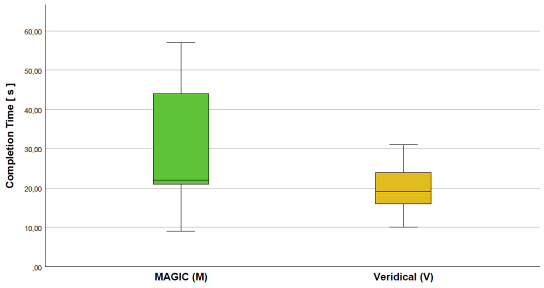

Fig. 4. Tasks' completion time for each condition.

concluded completion time data was normally distributed.

Results of the paired t-test indicated that the percentage of intersection was higher for our approach (*M* = 46.6, *SD* = 19.6) than for the Veridical condition (*M* = 33.8, *SD* = 13.8), t(9) = -2.31, p = .047 . As $p = .047 < .05$, we proved statistical significance of the results, and the hypothesis that observable differences were caused by random variation was refuted.

For completion time, the paired t-test showed no statistically significant differences in task completion time between our approach (*M* = 28.0, *SD* = 15.9) and the Veridical condition (*M* = 19.8, *SD* = 5.8), t(9) = -2.00, p = .076.

Results for task performance metrics are presented in Figures 3 and 4.

VI-B. User Preferences

When it comes to User Preferences, our questionnaires contained a set of statements to be answered on a 6-point Likert-scale, from Strongly Disagree (1) to Strongly Agree (6). For Likert Scales, it follows from definition that data can not be normally distributed as its values are discrete. Therefore, no normality test was employed. The questionnaires and associated results are summarized in Table I.

Questionnaires showed no statistically significant differences between the two approaches in terms of Attentional Allocation and Perceived Message Understanding, although a tendency for a better message understanding under MAGIC. However, five participants reported in their observations that targets were easier to understand when performing the experiment under MAGIC, and two participants pointed out they had to move less in order to understand what their partner was pointing to. In terms of feeling of co-presence, the Wilcoxon test reported a significant increase in the feeling of presence of the remote partner in the virtual environment under MAGIC (Statement 1.2.), where p = .025.

VI-C. Discussion

As both conditions employed a life-sized representation of the remote avatar in a face to face configuration, 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, contrarily to what happened in the veridical condition. Results showed an increased percentage of intersection between the areas of interest outlined by participants when sharing perspective, showing a better understanding of the collaboration tasks. Indeed, MAGIC improved accuracy by 12.8% when compared to the veridical condition. The increase of accuracy of the pointing task is explained by an enhancement in intentional communication, as people showed a better capacity to use and understand deictic gestures. The number of occlusions was also reduced comparing to when participants saw opposite sides of the model, which explains why participants reported having to move less in order to understand what their partner was referring to under MAGIC. It is to be noticed however that the questionnaire results did not show any statistically significant differences in Perceived Message Understanding statements between both conditions. This could be due to a ”false” sensation of understanding that was more prominent under the veridical condition, i.e. people thought they perceived their partner's actions correctly when in fact they failed the matching task. Evaluating workspace awareness in terms of the previous three communication mechanisms, we might say that our approach improved workspace awareness, proving our first hypothesis, **H1**.

We also hypothesized that the manipulations applied to the remote person's avatar were not noticeable by the local collaborator. We directly asked participants if they noticed any strange behaviors when performing the tasks. As non of the participants reported noticing anything strange under any condition, we assume that participants remained unaware of the experiment's manipulations. Results on completion time showed no statistically significant differences in task completion time under both conditions. Hence, MAGIC does not have a negative impact in the time collaborators take to complete the shared task. Since participants did not noticed manipulations, collaboration through MAGIC remains straightforward, with no added complexity to the task. There is no need for participants to learn a new interaction method as it appears to them that they are collaborating in a traditional way. Thus, we proved **H2**.

When it comes to the sense of co-presence enabled

Question	MAGIC	Veridical
1. Co-Presence		
1.1. I felt present in the virtual environment.	6 (0.75)	6 (1)
1.2. I felt that my partner was present in the virtual environment.*	6 (0)	5 (0.75)
1.3. I felt that my presence was evident to my partner.	6 (1)	5 (1.75)
2. Attentional Allocation		
2.1. I remained focused on my partner throughout our interaction.	6 (1)	6 (1)
2.2. I remained focused on the workspace throughout our interaction.	6 (1)	6 (1)
2.3. I felt that my partner could remain focused on me throughout our interaction.	5.5 (1)	5 (2)
2.4. I felt that my partner could remain focused on the workspace throughout our interaction.	5 (0.75)	5 (0.75)
3. Perceived Message Understanding		
3.1. I could understand my partner's actions.	5 (0.75)	5 (0.75)
3.2. I felt that my partner could understand my actions.	5 (0)	5 (1)
3.3. It was easy to understand what was happening in the workspace.	5 (0.75)	5 (2)
3.4. I felt that I was pointing to where I wanted to point.	5.5 (1.75)	5 (1)
3.5. It was easy to outline an area.	5 (0.75)	4 (1)
3.6. It was easy to understand the area outlined by my partner.	5 (0.75)	4 (1.75)

TABLE I

RESULTS OF THE USER PREFERENCE QUESTIONNAIRES FOR BOTH MAGIC AND VERIDICAL CONDITIONS. * INDICATES STATISTICAL SIGNIFICANCE. THE VALUES INCLUDE MEDIAN AND INTER-QUARTILE RANGE (IQR) IN PARENTHESES.

by our approach, answers from the User Preference Questionnaire showed no statistically significant differences in the sense of presence felt by the local collaborator itself. Yet, participants felt that their partner was more present in the virtual environment under MAGIC. We believe MAGIC improves the sense of co-presence since it increases awareness of the actions the remote collaborator is performing in the workspace. As people have a better understanding of what actions their partner is performing, they feel their partner is more present in the workspace. Hence, we validate our last hypothesis, **H3**.

As results show improved accuracy without disadvantageous effects on completion time, we can argue that MAGIC increases task performance. Additionally, we confirmed that the manipulations we employ to the representation of the remote collaborator are not noticeable and do even improve the sense of presence of the remote collaborator in the workspace. We can then complete this discussion by saying that we validated our research statement.

VII. CONCLUSIONS AND FUTURE WORK

In today's society, remote work is increasingly common and plays an essential role in projects in which geographically separated people need to perform joint tasks. Video-conference and tele-presence technologies allow verbal and visual communication, however, they fail in other aspects of nonverbal communication such as gaze, deictic gestures and other nonverbal cues.Virtual meetings make it possible to create a remote meeting closer to the co-located experience, as people can meet face-to-face with life-sized embodiments, which promotes the sense of presence as well as workspace awareness. Face-to-face virtual encounters still face, however, problems in enabling certain types of remote tasks, such the analysis of three dimensional models. As collaborators do not share the same forward-backwards orientation and there is no common orientation of right or left, taking advantage of nonverbal communication is still difficult. Also, contrary points-of-view paired with occlusions that result from people standing in front of each other can lead to serious communication missteps.

In this article, we introduced MAGIC, a design space for MR environments that enables people to engage in 3D object-centered face-to-face collaboration in a shared workspace while sharing the same perspective of the workspace. To enable the share of point of view in a face-to-face arrangement, the remote participant's virtual representation is subtly transformed so that gestures performed in the remote workspace are virtually correct in the local workspace. MAGIC also integrates person space, task space, and reference space, minimizing the need for meeting participants to constantly switch attention between the remote collaborator's person space and the workspace.

We conducted a user study with 10 participants in order to find if MAGIC could in fact improve workspace awareness and task performance by comparing the performance of the users when doing a pointing matching task with MAGIC versus a veridical baseline condition (face to face with opposing points of view). In terms of task performance, results showed that MAGIC improved the accuracy of tasks using pointing gestures without affecting the task completion time. Results also suggested that the remote person's representation manipulations were not noticeable by the participants, improving the sense of co-presence. Hence, we validated our research statement.

We believe that MAGIC opens a door for interesting future work, some related to the nature of our approach and other related to a more technical aspect. Using depth cameras for capturing the real embodiment of the user to be rendered in the virtual environment could be employed in order to study the real sense of presence induced by our approach. Also, our user study focused on the accuracy of pointing tasks since other deictic references could not be accessed using our hardware only. It would be interesting to focus on multi-modal deictic references by combining detailed gaze with gestures. Future work could also target the extension of our approach to a multi-user set-up. Finally, our approach was designed in order to manipulate the remote user movements in imperceptible ways, we believe however that perceptible exaggerated distortions should be studied to ascertain whether they have the potential to either improve or hinder collaboration.

REFERENCES

[1] Steve Benford, John Bowers, Lennart E Fahlén, Chris Greenhalgh, and Dave Snowdon. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 242–249, 1995.

[2] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew 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, 2016.

[3] Jakob Zillner, Christoph Rhemann, Shahram Izadi, and Michael Haller. 3d-board: a whole-body remote collaborative whiteboard. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 471–479, 2014.

[4] T Dodds, B Mohler, and H Bülthoff. 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.

[5] Maurício Sousa, Daniel Mendes, Rafael K dos Anjos, Daniel Simões Lopes, and Joaquim 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.

[6] Martin Feick, Terrance Mok, Anthony Tang, Lora Oehlberg, and Ehud 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*, pages 1–13, 2018.

[7] Morgan Le Chénéchal, Thierry Duval, Valérie Gouranton, Jérôme Royan, and Bruno 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.

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

[9] Steve Benford, John Bowers, Lennart E Fahlén, Chris Greenhalgh, and Dave Snowdon. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 242–249, 1995.

[10] William Buxton. Telepresence: Integrating shared task and person spaces. In *Proceedings of graphics interface*, volume 92, pages 123–129. Canadian Information Processing Society Toronto, Canada, 1992.

[11] David Nguyen and John 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.

[12] Mathilde M. Bekker, Judith S. Olson, and Gary 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.

[13] David Kirk, Andy Crabtree, and Tom Rodden. Ways of the hands. In *ECSCW 2005*, pages 1–21. Springer, 2005.

[14] Aaron Genest and Carl Gutwin. *Characterizing Deixis over Surfaces to Improve Remote Embodiments*, pages 253–272. Springer London, London, 2011.

[15] Hiroshi Ishii and Minoru 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*, pages 525–532, 1992.

[16] Osamu Morikawa and Takanori Maesako. Hypermirror: toward pleasant-to-use video mediated communication system. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 149–158, 1998.

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

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

[19] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 741–754, 2016.

[20] Bill Buxton. Mediaspace–meaningspace–meetingspace. In *Media space 20+ years of mediated life*, pages 217–231. Springer, 2009.

[21] Sasa Junuzovic, Kori Inkpen, Tom Blank, and Anoop Gupta. Illumishare: sharing any surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1919–1928, 2012.

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

[23] Daniel Leithinger, Sean Follmer, Alex Olwal, and Hiroshi 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*, pages 461–470, 2014.

[24] Jiannan Li, Saul Greenberg, Ehud Sharlin, and Joaquim Jorge. Interactive two-sided transparent displays: designing for collaboration. In *Proceedings of the 2014 conference on Designing interactive systems*, pages 395–404, 2014.

[25] Hrvoje Benko, Ricardo Jota, and Andrew Wilson. Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 199–208, 2012.

[26] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas, and Mark 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.

[27] Maurício Sousa, Daniel Mendes, Rafael K dos Anjos, Daniel Simões Lopes, and Joaquim 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.

[28] 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.

[29] Sven Mayer, Jens Reinhardt, Robin Schweigert, Brighten Jelke, Valentin Schwind, Katrin Wolf, and Niels 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.

[30] Maurício Sousa, Rafael Kufner dos Anjos, Daniel Mendes, Mark Billinghurst, and Joaquim Jorge. Warping deixis: Distorting gestures to enhance collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[31] Katja Zibrek, Elena Kokkinara, and Rachel McDonnell. The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role? *IEEE transactions on visualization and computer graphics*, 24(4):1681–1690, 2018.

[32] Andreas Aristidou, Yiorgos Chrysanthou, and Joan Lasenby. Extending FABRIK with model constraints. *Comput. Animat. Virtual Worlds*, 27(1):35–57, January 2016.