Cogn Process (2015) 16 (Suppl 1):S337–S342
DOI 10.1007/s10339-015-0672-2

SHORT REPORT

Virtual collaboration: effect of spatial configuration on spatial statements production

Lauriane Pouliquen-Lardy[1,2] · Franck Mars[2] · François Guillaume[3] · Isabelle Milleville-Pennel[2]

Published online: 25 July 2015

Abstract When guiding a remote collaborator in a virtual environment, people often take an addressee-perspective, which may have a high cognitive cost. In order to improve collaborative virtual environments, a better understanding of how operators share spatial information is needed. This work aimed to study the cognitive workload linked to spatial statements production in situations in which the relative positions of speaker, addressee and target were varied. Twenty-two participants were asked to give—in one go—instructions to a virtual collaborator on how to find a target in a 3D environment. The scene showed an avatar in the center of eight tables. Sixty-four configurations of avatar orientation (eight possibilities) and target location (on the eight tables) were tested. We measured the delay in starting the instruction once the target appeared, the instruction duration and the subjective evaluation of mental demand. Each instruction was classified according to the spatial reference frame used. The delay was influenced by the processing of spatial information in ego-centered and addressee-centered reference frames. All subsequent measures were determined by mental transformations in addressee-centered coordinates. One condition in particular, when the target was situated diagonally behind the addressee, gave rise to a higher mental demand for the speaker, which points to the investment made by the speaker in achieving the least collaborative effort. Further work should seek to develop efficient tools to facilitate spatial communication in situations that induce the most mental workload.

Keywords Virtual reality · Remote collaboration · Mental workload · Spatial communication

✉ Lauriane Pouliquen-Lardy
lauriane.pouliquen-lardy@irccyn.ec-nantes.fr

[1] IRT Jules Verne, Bouguenais, France

[2] IRCCyN, Nantes, France

[3] Airbus Group, Suresnes, France

Introduction

Immersive collaborative virtual environments (ICVEs) are used more and more to support remote collaboration in industry (Churchill et al. 2001), bringing about new constraints. For instance, when a collaborator guides someone in an ICVE, he can take the perspective of the guided person (Pouliquen-Lardy et al. 2014). While this strategy reduces the risk of misunderstanding, it requires perspective taking and mental rotations, which increase the cognitive workload. The aim of this study is to better understand spatial statements production in remote collaboration using ICVEs.

To locate a target, speakers use a frame of reference (FoR), i.e., a perspective and a coordinate system. The perspective can be that of the speaker (egocentric FoR), or it can be anchored in someone else's point of view or in a reference (e.g., an object). The coordinate system can be defined by characteristics that are either extrinsic to it, e.g., angles, hours and cardinal directions, or intrinsic to the reference, i.e., defined by the structural characteristics of the reference itself (McNamara 2003). For humans, intrinsic axes are the frontal (front/back), sagittal (right/left) and transversal (top/down) axes. A target's location relative to these axes influences cognitive processes: If gravity is observed, the transversal axis is preferred to the frontal and the sagittal axes (Bryant and Tversky 1999). Moreover, if the target is not aligned with one of those orthogonal axes, i.e., if it is oblique, the task is more difficult (McNamara 2003). More recently, Galati and Avraamides (2015) showed that the intrinsic characteristics of the layout itself, such as an axis of symmetry, influenced spatial dialogue.

Existing studies have used displays around which both parts of the dyad were placed. In our study, in order to approach the conditions of remote collaboration in large virtual environments, we placed the addressee in the layout itself. However, the speaker was able to see the entire scene in one go, without navigating within it. In particular, this new experimental display allowed us to test configurations in which the target was behind the addressee. Therefore, the goal was to better understand the types of situations in which configurations are most difficult for spatial instruction production and to identify strategies that remote coworkers use to be understood. Moreover, most previous studies have only measured reaction times or verbal strategies to assess spatial communication. The present study associates such measurements with the evaluation of mental demand.

Method

Twenty-two volunteers participated in the experiment (5 women and 17 men, mean age 25.5 years, age range 20–42 years). Participants received written instructions explaining they would be working with a remote collaborator. Their common goal was to destroy as many trapped bottles as possible. To do so, they would work within a virtual environment, in which they are both represented by avatars. The task was divided into two roles: the participant (*speaker*) gave the instruction and his collaborator (*addressee*) had to select the bottle. The speakers were told to be as accurate and fast as possible. They also knew that they would not have any feedback from their addressee, either verbally or visually. They were then placed in front of the immersive screen with a microphone and a keyboard. Speakers were only informed at the end of the experiment that, in reality, the addressee was an inanimate character.

The VE was composed of a single room containing eight tables organized in a circle. A bottle was placed on each (Fig. 1) and all bottles were identical. The avatar of the addressee was placed in the middle of the circle. Participants first navigated in the virtual scene to get a global representation of the place. They were then repositioned and would remain in these positions for the rest of the experiment. Seven training trials took place. After this, the experiment proper consisted of 64 trials (8 avatar orientations × 8 target positions) presented randomly and repeated three times. Each trial started with the display of the scene (avatar and tables). When ready, the participant pressed 0 on the keyboard to make the target appear (one of the tables changed color, see Fig. 1). To give the instruction, participants pressed the spacebar, holding it down while speaking. Once the spacebar was released, the scene disappeared. To start a new trial, participants pressed 0.

The starting time of the instruction (time 1) and the duration of the instruction (time 2) were automatically recorded. On the third repetition, the participants had to evaluate the mental demand of each configuration on a continuous scale. The question was "How mentally demanding was the task?", which is one of the subscales of the widely used NASA Task Load Index (Hart and Staveland 1988). The frame of reference (FoR) used in the instruction was coded according to the perspective (egocentered or addressee-centered) and the coordinate system (intrinsic or extrinsic) used, or if mixed perspectives or coordinate systems were used.

Different statistical analyses were performed depending on the dataset: Time 1, time 2 and mental demand were analyzed with repeated measures ANOVA. Tukey HSD post hoc tests were calculated when the ANOVA revealed an effect. The frames of reference were analyzed using a Chi-square test. The level of significance used was $p < .05$.

Results

Participants used different frames of reference to give their instructions: 79.7 % were addressee-centered, 6.8 % egocentered, 11 % used both ego- and addressee-centered FoR (doubling the information), and 2.5 % used other strategies (such as reference to the previous trial). Results showed that the FoR varied significantly according to the location of the target relative to the addressee [χ^2 (20; $N = 330$) = 79; $p < .001$]. Participants produced fewer mixed statements when the target was in the frontal axis of the addressee (0° and 180°) than for the other configurations (45°, 90° and 135°, see Fig. 2).

As most of the instructions were addressee-centered, and in order to provide some insight into the effects of configuration on spatial utterances, the data were analyzed in two ways: first, by taking into account the right/left reversal between the speaker and the addressee and second by considering the target's location relative to the addressee.

To test the effect of right/left reversal, we analyzed differences between conditions according to the laterality congruence between the speaker and the addressee (Fig. 3). Situations were considered as *Neutral* when the target was aligned with the frontal axis of either *the speaker*, *the addressee*, or *both of them*. In the remaining cases, situations were coded as *Congruent* when the target was on the

Fig. 1 Spatial configuration of the apparatus

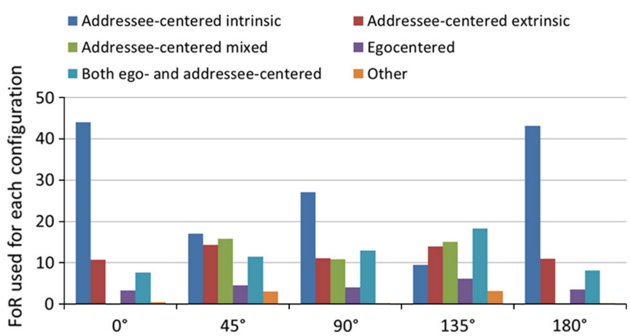


Fig. 2 Frames of reference used according to the location of the target relative to the addressee

same side as the addressee and speaker. Situations were considered as *Incongruent* when the target was located to the right of the addressee and to the left of the speaker, or vice versa.

The ANOVAs showed a significant effect on time 1 ($F(4,84) = 30{,}847$; $p < .001$), time 2 ($F(4,84) = 23{,}269$; $p < .001$) and mental demand ($F(4,84) = 20{,}247$; $p < .001$) (Fig. 3). Although the *Incongruent* trials gave rise to higher values than the *Congruent* ones, post hoc analysis did not reveal any differences between the two situations. For the three indicators, values for situations involving a lateralized target relative to the addressee (i.e., *Congruent*, *Incongruent* and *Neutral for the speaker*) were significantly higher than values for situations in which the target was in the frontal axis of the addressee (i.e., *Neutral for the addressee* and *Neutral for both of them*). Additional significant differences were observed on time 1 only: the speakers started their instruction earlier when the target was aligned with their own frontal axis than in incongruent situations. Moreover, when the target was aligned with both the addressee and the speaker, the instruction started earlier than when it was aligned only with the addressee.

The second set of analyses was performed on the target location relative to the addressee (Fig. 4). The ANOVAs revealed significant effects on time 1 ($F(4,84) = 23{,}766$; $p < .001$), time 2 ($F(4,84) = 26{,}196$; $p < .001$) and mental demand ($F(4,84) = 23{,}276$; $p < .001$).

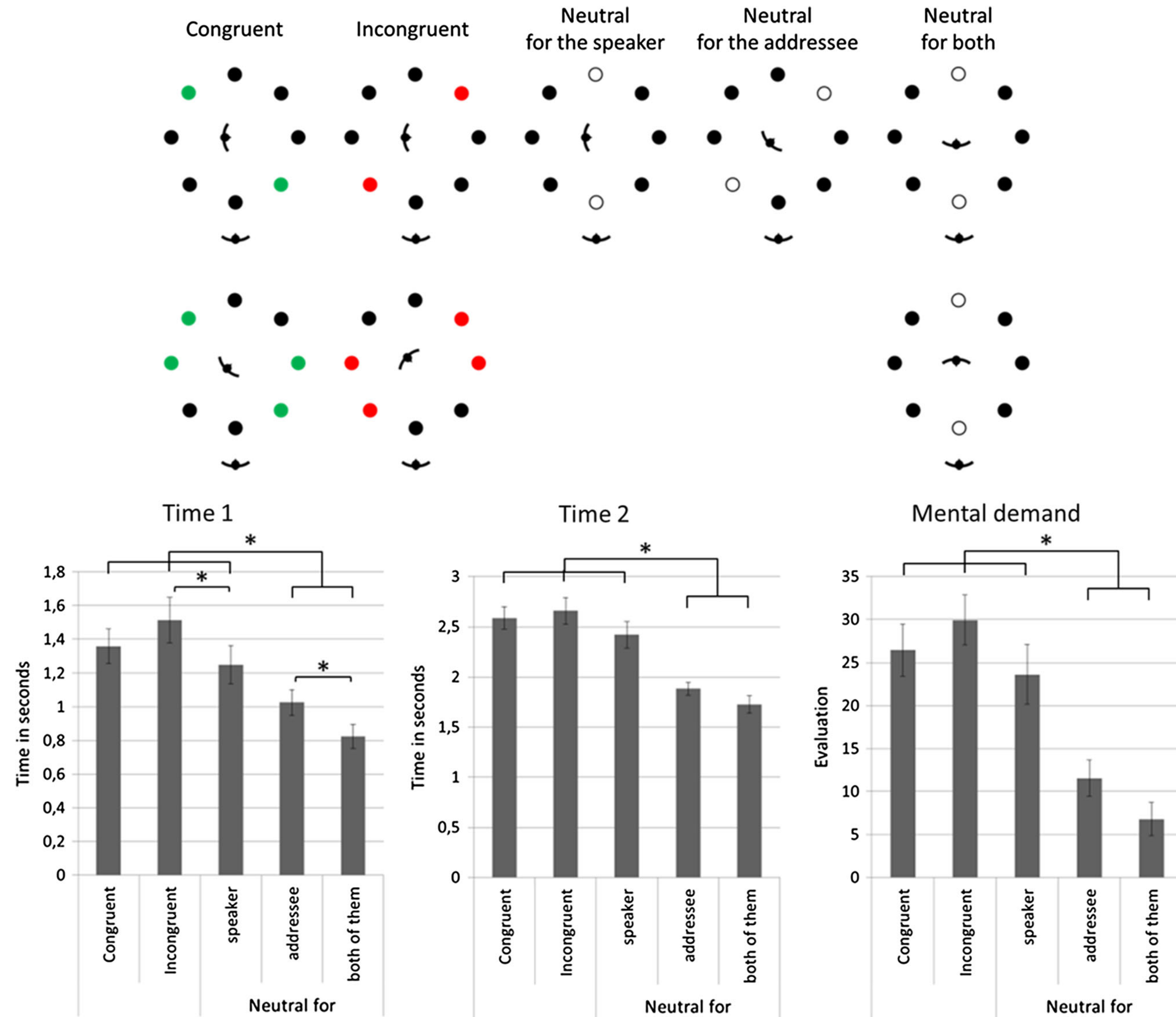
The post hoc analysis revealed similar patterns of differences for the three indicators: the 0° and 135° configurations gave rise to, respectively, significantly lower and higher values compared with the other configurations. For instruction duration and mental demand, the 45° and 90° configurations showed higher values than in the 180° configuration.

Discussion

The aim of this study was to better understand the processes at work in the production of spatial statements during a remote collaborative task. The results showed that the position of the target relative to the addressee is a major component in these processes.

Firstly, we observed that participants mostly used instructions centered on their addressee. This confirms previous findings whereby, in some situations, speakers spontaneously change their perspective (Pouliquen-Lardy et al. 2014). This phenomenon may be linked to the collaborative dimension of the task: In order to minimize the risk of misunderstanding and to ensure the success of the task, the speaker may accept a higher cognitive workload (Schober 1995). This effect may have been amplified in our experiment because the speaker had no feedback from the addressee.

However, all situations did not yield the same difficulty; some situations were performed faster and perceived as easier than others. The main factor was the target's location relative to the intrinsic axes of the addressee (McNamara 2003; Bryant and Tversky 1999). Indeed, participants started their statements faster, gave briefer instructions and perceived the situations as easier when the target was aligned with the frontal axis of the addressee (*Neutral for*



	Cong.	Incong.	N. speak.	N. addre.	N. both
Cong.		0,18	0,52	< 0,01	< 0,01
Incong.	0,18		< 0,01	< 0,01	< 0,01
N. speak.	0,52	0,00		0,02	< 0,01
N. addre.	0,00	< 0,01	0,02		0,04
N. both	< 0,01	< 0,01	< 0,01	0,04	

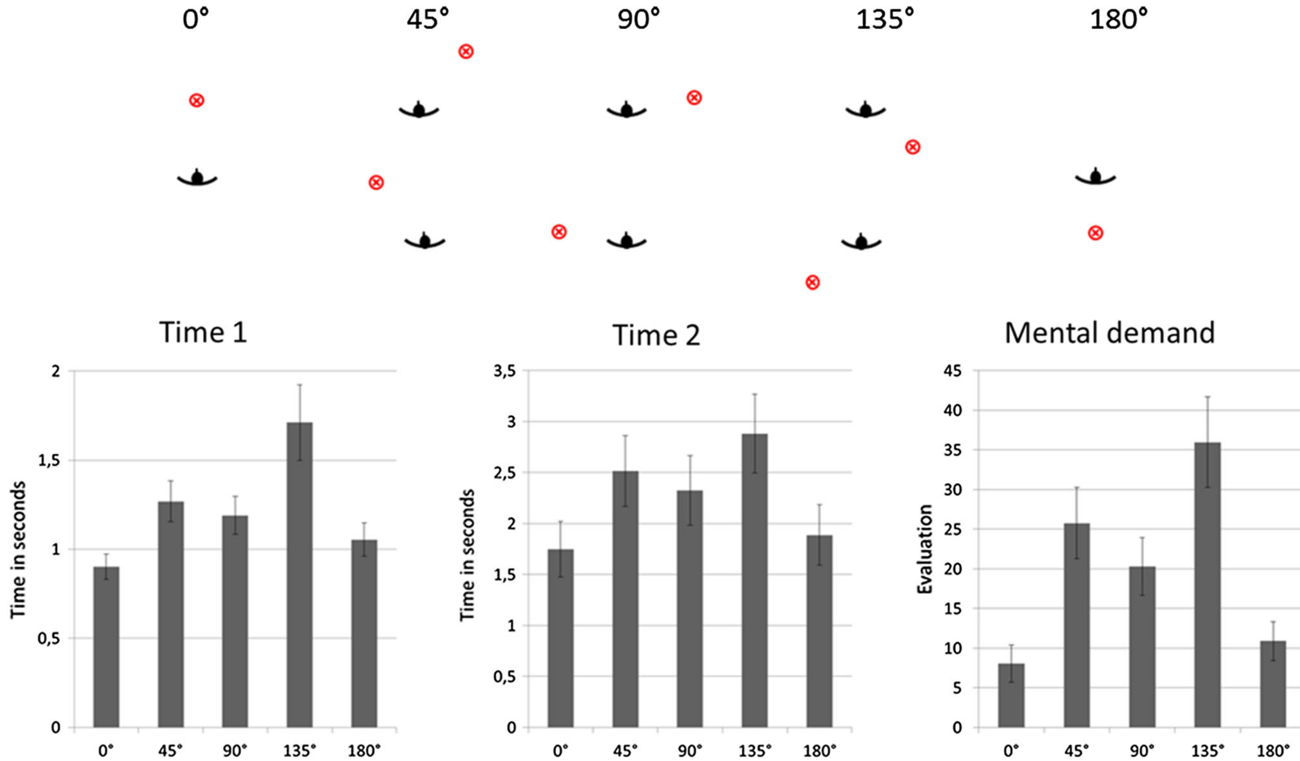
	Cong.	Incong.	N. speak.	N. addre.	N. both
Cong.		0,98	0,66	< 0,01	< 0,01
Incong.	0,98		0,31	< 0,01	< 0,01
N. speak.	0,66	0,31		< 0,01	< 0,01
N. addre.	< 0,01	< 0,01	< 0,01		0,71
N. both	< 0,01	< 0,01	< 0,01	0,71	

	Cong.	Incong.	N. speak.	N. addre.	N. both
Cong.		0,80	0,90	< 0,01	< 0,01
Incong.	0,80		0,27	< 0,01	< 0,01
N. speak.	0,90	0,27		< 0,01	< 0,01
N. addre.	< 0,01	< 0,01	< 0,01		0,55
N. both	< 0,01	< 0,01	< 0,01	0,55	

Fig. 3 *Top* The experimental conditions coded as a function of the target's lateral position relative to the frontal planes of the speaker and addressee. *Bottom* Effect of spatial configuration on the starting time of the instruction (time 1), the duration of the instruction (time 2) and mental demand (mean values and standard deviation, *tables* show *p* values given by Tukey HSD tests)

the addressee and *for both* vs. other conditions; 0° and 180° vs. other conditions). We observed that this phenomenon was enhanced for the delay in starting the instruction (time 1) if the frontal axis of the addressee was aligned with the frontal axis of the speaker (*Neutral for both* vs. *Neutral for the addressee*). This result echoes the study by Galati and Avraamides (2015), which showed the positive effect of the alignment of several cues on spatial descriptions.

Finally, the most difficult configuration was when the target was diagonally behind the speaker. The peculiarity



	0°	45°	90°	135°	180°
0°		< 0,01	0,01	< 0,01	0,43
45°	< 0,01		0,91	< 0,01	0,12
90°	0,01	0,91		< 0,01	0,55
135°	< 0,01	< 0,01	< 0,01		< 0,01
180°	0,43	0,12	0,55	< 0,01	

	0°	45°	90°	135°	180°
0°		< 0,01	< 0,01	< 0,01	0,82
45°	< 0,01		0,56	0,04	< 0,01
90°	< 0,01	0,56		< 0,01	0,01
135°	< 0,01	0,04	< 0,01		< 0,01
180°	0,82	< 0,01	0,01	< 0,01	

	0°	45°	90°	135°	180°
0°		< 0,01	< 0,01	< 0,01	0,91
45°	< 0,01		0,47	0,02	< 0,01
90°	< 0,01	0,47		< 0,01	0,04
135°	< 0,01	0,02	< 0,01		< 0,01
180°	0,91	< 0,01	0,04	< 0,01	

Fig. 4 *Top* The experimental conditions coded as a function of the target's orientation relative to the addressee. *Bottom* Effect of the target location relative to the addressee on the starting of the instruction (time 1), the duration of the instruction (time 2) and mental demand (*tables* show *p* values given by Tukey HSD tests)

of these situations is that they require both a reversal of the addressee's orientation and an assessment of the target's lateral position. These situations led to an ambiguity, as found in the following statement: "the target is behind you on your left'. Will the addressee understand "the target is currently on your left" or "turn around, and then the target will be on your left"? To reduce this ambiguity, participants adopted a strategy consisting of doubling the information: They mixed the frames of reference to ensure mutual understanding. This strategy yielded a higher cognitive workload. However, when the protagonists are able to adjust their strategies and to correct the action thereafter, it is possible that the dyad supports this effort differently (Galati et al. 2013).

To conclude, this study suggests that to facilitate remote communication, some spatial configurations should be privileged and others avoided. To make the task easier, it seems important to align the frontal axis of the co-workers with the direction of the target. Nevertheless, if it is impossible to avoid situations when the target is laterally behind the addressee, it is easier to: first, ask the addressee to turn around and, second, describe the target's location relative to the addressee's new orientation.

This study is a first step in the comprehension of spatial statements for mutual understanding in remote collaboration using ICVEs. Our next objective will be to assess the impact of the speaker's instructions on the addressee's understanding as a function of the frame of reference used and the target's position.

Acknowledgments This study is part of the PIVIPP project managed by IRT Jules Verne (French Institute in Research and Technology in Advanced Manufacturing Technologies for Composite, Metallic and Hybrid Structures). The authors wish to associate the industrial and academic partners of this project, respectively Airbus Group, Airbus and IRCCyN.

References

Bryant DJ, Tversky B (1999) Mental representations of perspective and spatial relations from diagrams and models. J Exp Psychol Learn Mem Cogn 25:137

Churchill EF, Snowdon D, Munro AJ (2001) Collaborative virtual environments. Springer, Berlin

Galati A, Avraamides MN (2015) Social and representational cues jointly influence spatial perspective-taking. Cogn Sci 39(4):739–765

Galati A, Michael C, Mello C, Greenauer NM, Avraamides MN (2013) The conversational partner's perspective affects spatial memory and descriptions. J Mem Lang 68:140–159

Hart S, Staveland L (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock P, Meshkati N (eds) Human mental workload. Advances in psychology, vol 52. Elsevier, Amsterdam, pp 139–183

McNamara TP (2003) How are the locations of objects in the environment represented in memory? In: Freksa C, Brauer W, Habel C, Wender KF (eds) Spatial cognition III. Springer, Berlin, pp 174–191

Pouliquen-Lardy L, Milleville-Pennel I, Guillaume F, Mars F (2014) How role distribution influences choice of spatial reference frames in a virtual collaborative task. Paper presented at Spatial Cognition 2014, Sep 2014, Bremen, Germany. Universität Bremen / Universität Freiburg, SFB/TR 8 Report, pp 106–108

Schober MF (1995) Speakers, addressees, and frames of reference: whose effort is minimized in conversations about locations? Discourse Process 20:219–247