



# **Sharing spatial information in a virtual environment: How do visual cues and configuration influence spatial coding and mental workload?**

Isabelle Milleville-Pennel, Franck Mars, Lauriane Pouliquen-Lardy

## **► To cite this version:**

Isabelle Milleville-Pennel, Franck Mars, Lauriane Pouliquen-Lardy. Sharing spatial information in a virtual environment: How do visual cues and configuration influence spatial coding and mental workload?. Virtual Reality, 2020, 24 (4), pp.695-712. <10.1007/s10055-020-00430-0>. <hal-02502727>

**HAL Id: hal-02502727**

**<https://hal.science/hal-02502727v1>**

Submitted on 7 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# **Sharing spatial information in a virtual environment: how do visual cues and configuration influence spatial coding and mental workload?**

**Isabelle Milleville-Pennel** <sup>1, CA</sup>

**Franck Mars** <sup>1</sup>

**Lauriane Pouliquen-Lardy** <sup>2</sup>

<sup>1</sup> LS2N (Laboratoire des Sciences du Numérique de Nantes), UMR CNRS 6004, B.P. 92101  
F. 44321 Nantes Cedex 03, France.

E-mail address: [Isabelle.milleville@ls2n.fr](mailto:Isabelle.milleville@ls2n.fr)

ORCID: [0000-0002-2236-8346](https://orcid.org/0000-0002-2236-8346)

## **Abstract**

When sharing virtual collaborative environments, operators exchange spatial statements that refer to the objects' positions in the virtual space. If operators are to understand each other, they need to develop a common spatial frame of reference and then choose a space coding to describe the objects' positions. In this paper, we consider how the content of a virtual environment can influence communication between users. We designed two studies in which one participant (the speaker) had to indicate the position of one object to another participant (the addressee). The virtual environment was sometimes enriched by additional (proximal and distal) visual cues. In study 1, we considered statements production. We observed that the speakers most often used the avatar of their partner as a spatial reference to indicate a localization in the virtual space (i.e., addressee-centered coding) despite it increases their mental workload. Nevertheless, in complex situations, they also used distal cues to speak to the addressees, (i.e., exocentric coding of the space). In study 2, we considered statements comprehension. Addressee-centered and exocentric coding were used by the speakers in various spatial configurations to indicate the object position. We observed that exocentric coding is the most difficult to manage for the addressee. These results indicate that speakers implemented the principle of less collaborative effort by adopting a way of exchanging information based on an asymmetrical cognitive cost, taking into consideration each other's difficulties. This allows a balanced mental workload to be maintained between the two operators throughout the task.

**Keywords:** Common spatial frame of reference; Spatial coding; Mental workload; Least collaborative effort, Virtual environments

# 1. Introduction

The development of new methods of interaction proposed by virtual reality (VR) allows the creation of more and more diversified collaborative virtual environments (CVE). To be effective, CVEs need to allow the establishment of an accurate Common Frame of Reference (COFOR), i.e., a set of compatible (but not necessarily identical) representations of the virtual world that are continuously updated by collaborators and facilitate information sharing (Clark and Brennan 1991; Hoc 2001). The spatial COFOR (SpaCOFOR) is the part of the COFOR that relates to spatial cognition and determines the perception and comprehension of the action space (e.g. position, distance, and depth). It contains useful information for the localization of objects and other partners in the common space. It must also allow an intention of action or a displacement in the CVE to be quickly identified. The way in which collaborators create and develop an unambiguous SpaCOFOR is not a trivial question. As for the COFOR, the elaboration of the SpaCOFOR relies on different types of verbal and non-verbal communication. However, communication, particularly non-verbal communication, is highly constrained by the properties of the CVE.

## 1.1. CVE and the elaboration of the SpaCOFOR

Depending on its configuration, a CVE will impose some constraints on the means of communication usually used by operators to elaborate and update a common representation of the task and how to collaborate together (Heldal, Spante, Steed and Bengstoon 2005). This is particularly true concerning the elaboration of the SpaCOFOR. Indeed, non-verbal communication is mostly non-existent. One example of these constraints is that the partner's gaze direction is not available, even though this information could often be useful: as soon as someone looks at an object, person or place in space, it is easy to determine what she or he is referring to when speaking. A similar issue relates to deictic references (i.e., when one says "it is in this direction" whilst pointing a finger in that direction with that finger): this allows a person to clearly designate an object or direction, with no need to precisely describe its location. This type of communication become nearly unusable in CVEs because of their poor ability to represent users.

Another major limitation comes from the great difficulty in visualizing or understanding the point of view of another participant in the VE. The difficulty is often increased by the loss of sound localization. Another related difficulty is the spatial disorientation that often occurs

when navigating in a VE. This disorientation is mostly due to the lack of correspondence between the displacement in the real world and the displacement perceived in the VE.

## **1.2. Implications for the representation of space and communication**

In a context where it is difficult to share spatial representations, operators must draw hypotheses about the ability of their partner to identify the position of objects in the VE and to understand spatial statements. These suppositions will directly impact the way the operators describe the spatial environment when they speak together. Typically, two sorts of spatial descriptions can be used. The first one consists of coding the position of objects in space with respect to our own position (egocentric coding). The second one consists of coding the position of objects with respect to one another, independently of our position in space. It can be other-centric coding if another person is considered as the point of origin of the coding or exocentric coding if it is based on the other objects present in the CVE. Clark and Wilkes (1986) proposed the principle of least collaborative effort to account for the effort accepted by operators when choosing between all the available possibilities to describe a spatial configuration. This principle states that, during a conversation, both participants try to minimize the collaborative effort, i.e., the work that they accomplish between the beginning of a verbal exchange and its mutual acceptance. The objective is not to reduce the individual efforts but to adapt individual contributions to optimize the common effort. One could assume that this corresponds to a natural tendency of the individual to adopt the other's point of view when it seems useful for understanding. For example, Tversky and Hard (2009) have shown that simply introducing an observer into a spatial description task changes the proportions in which the speaker chooses an exocentric rather than an egocentric frame of reference. According to the authors “Perspective-taking is undoubtedly an effective strategy in social situations and might occur spontaneously in anticipation of social interaction. Consequently, seeing another person in a scene might prompt thinking about the world from the other’s perspective.” This can be seen as one of the fundamentals of the least collaborative effort principle.

The least collaborative effort principle was observed by Pouliquen-Lardy, Milleville-Pennel, Guillaume and Mars (2016) in a remote virtual collaborative task. Two operators, each with a specific role (either guide or manipulator), competency and knowledge of the environment, had to collaborate to complete a maintenance task in a virtual plane factory. The operators applied the principle of least collaborative effort by adopting a strategy for exchanging spatial information. When the guides took a particular perspective, they preferred to take the manipulator’s point of view. The required mental transformations resulted in an

asymmetrical cognitive workload that considered the difficulties experienced by the operators during the task. This resulted in a global reduction of mental workload for the dyad throughout the task.

This willingness to adopt an exocentric point of view depends on the participant's belief about the partner ability to contribute to the decrease of the collaborative effort (Duran, Dale and Kreuz 2011). Duran et al. (2011), showed that when a participant had to indicate the position of an object to a virtual social partner, the use of exocentric or egocentric point of view depended on the supposed ability to the partner to be aware of the participant position in space (in this case, the use of other-centric coding increased) or to share responsibility in working toward mutual understanding (in this case, the use of egocentric coding increased).

The choice of spatial coding may also be influenced by the respective positions of the operators in space since the difficulty to adopt an other-centric point of view increases with the discrepancy between the two points of views. For example, in Duran et al. (2011), when using an other-centric coding, the response-time to select the object designated by the partner (the speaker), increased from  $0^\circ$  (same orientation) to  $90^\circ$  (the speaker was positioned at  $90^\circ$  with respect to the addressee) and from  $90^\circ$  to  $180^\circ$  (the speaker was positioned at  $180^\circ$  with respect to the addressee). On the contrary, when using an egocentric coding, response time was all the time shorter (than with the other-centric coding) and increased only from  $0^\circ$  to  $90^\circ$ . The same sort of effect had been observed in a study by Michelon and Zacks (2006). The task was to determine if an object could be seen by another person. They used photography to show this person and a large number of objects placed on a table, with an exocentric point of view. The authors showed that the greater the differences between the points of view of the person and the participant, the more difficult it was for the participant to localize an object with respect to the other person. According to the authors, this can be explained by the fact that the participant had to align his/her egocentric point of view with that of the other person. Once this alignment was achieved, the participant had to localize the object with respect to this new imagined egocentric localization.

Finally, spatial coding can also be determined by the spatial configuration of the CVE itself. Galati and Avraamides (2015) showed that the intrinsic structure of the spatial scene (symmetry axis) could be used to choose a perspective for describing this scene to another person. The greater the alignment of some visual cues with an environmental axis, the more likely this axis would be chosen by the speaker. Moreover, the spatial configuration of the CVE combined with the position of the addressee in this CVE may also influence the speaker's choice of spatial coding. This was shown in a study by Pouliquen-Lardy, Mars, Guillaume and

Milleville-Pennel (2015), in which each participant received visual information on the position of an object in a CVE before being told to verbally indicate this position to his or her partner: the partner was represented by an avatar positioned at the center of eight tables arranged in circle. The object was placed on one of the eight tables. During the trials, the avatar's orientation with respect to the participant and the position of the object was manipulated. The spatial coding used by each participant to deliver the instruction could be egocentric (centered on the participant), or exocentric (centered on the avatar). The authors observed that the participants seldom used egocentric coding (7% of all statements) and favored other-centric coding (80% of all statements). In this study, participants did not know if their partners could see them. This is consistent with Duran et al. (2011) study who observed 74% of other-centric coding in the condition where participants were informed that their partner could not see their position in the VE (study 2 of their article). Another observation was that the position of the target with respect to the intrinsic direction of the partner (right/left; in front/behind; up/bottom) determined the difficulty of spatial coding. Several studies have already shown that spatial configurations are more or less difficult to manage depending on the target placement relative to the body axis: if gravity is respected, the sagittal plan (front-rear) is preferred to the frontal plan (left-right) and the frontal plan is preferred to the transversal plan (up-down; Bryant and Wright 1999; Bryant and Tversky 1999). If the target is positioned in a place that is not orthogonal to one of these plans (oblique direction), the level of difficulty increases (McNamara 2003). This was confirmed by Pouliquen-Lardy et al. (2015) who observed lower levels of mental workload (measured using the NASA-TLX; Cegarra and Mordago 2009) and less time needed to prepare the statements when the target was behind or in front of the avatar with respect to the other spatial configurations. Thus, the position of the target in the sagittal plan was a facilitating factor. On the contrary, the most difficult spatial configuration for the speaker was when the target was situated diagonally behind the partner.

### **1.3. How to facilitate the elaboration of the SpaCOFOR**

In the previous sections, we have seen that, in order to minimize the collaborative effort, it is essential to have a good representation of the partner's position in the CVE in order to adapt the spatial coding to the spatial configuration of the VE. For this reason, it is essential for each participant to have a view of his or her partner. This is often achieved by means of an object that represents the partner in the virtual space. Ideally, this object should be lateralized so as to indicate its orientation in the virtual space (i.e., an avatar in the form of a 3D character, or compass rose). This solution often gives an approximation of the partner's point of view.

With regard to the actions of the partner, most of the time they are identifiable only through a change of the state of the VE, or the result of the actions on the objects (e.g., change of color, position, or size). Other solutions can be adopted, such as point-of-view sharing, the use of juxtaposed windows or the use of a stable, lateralized visual landmark in the VE. These solutions will be detailed in the following sections.

#### ***1.3.1. Sharing the same point of view***

The solution called WYSIWIS (What you see is what I see) can be adapted through a large number of more or less restrictive variants depending on nature of the sharing (i.e., partial or total sharing). This solution may reduce misunderstanding between collaborators, although it becomes difficult to use as the task often requires operators to work in different places at the same time, or when it is necessary to have different point of view of the same object at the same time.

#### ***1.3.2. Giving feedback on the other operator's point of view***

The solution uses juxtaposed windows to show the point of view and the actions of the partner in the CVE. Each operator is then able to describe the position of the objects in the CVE or to indicate an intention of action using his or her own (egocentric) point of view (Hindmarsh, Fraser, Heath, Benford, and Greenhalagh 1998; Spante, Schroeder, Axelsson and Christie 2004). Although it is very useful in some situations, this solution has the disadvantage of fragmenting the visualization, which can prevent the building of an appropriate SpaCOFOR and can lead to misunderstandings (Gaver, Sellen, Heath, and Luff 1993). Moreover, increasing the number of juxtaposed windows (more than two partners) may not only saturate the operator's visual space but also the available cognitive resources, to the detriment of the shared activity (Gaver et al. 1993). Thus, although such solutions can sometimes be useful, they do have some limits and can only be applied to specific collaborative activities.

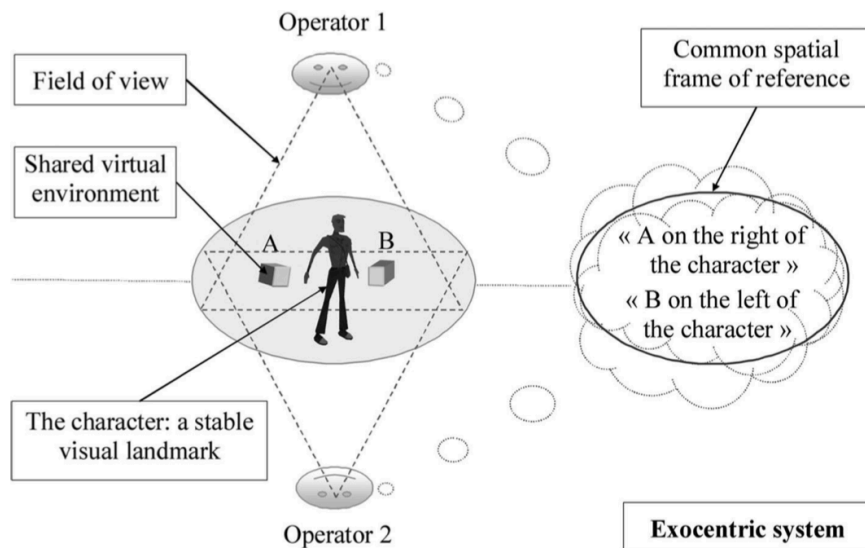
#### ***1.3.3. Stable lateralized visual landmark***

Alternatively, the content and appearance of the CVE could be modified in such a way that it implicitly leads to the perception of virtual space in a common way, with the same coding of space for each user and with minimal effort. One problem with some of the solutions proposed to help build the SpaCOFOR is that they often reinforce the use of egocentric spatial coding by the operator. On the contrary, it may be relevant to help the operator extract himself from the egocentric frame of reference and to spontaneously use exocentric spatial coding, which would be point-of-view independent and thus common to all users.

To this end, one solution would be to include contextual visual cues that spontaneously promote the use of exocentric spatial coding. Exocentric spatial coding can be particularly useful in the case of complex situations that require the adoption of different points of view and/or the manipulation of several objects at the same time. Objects and actions would thus be spatially located relative to common visual cues rather than from the point of view of each operator. Such stable visual landmarks would help the operator to describe the virtual space in an exocentric way, which would facilitate spatial communication by reducing the ambiguity and number of required mental rotations. Indeed, in this case, whatever the point of view of the different operators, only one mental rotation may be needed (i.e., one that allows the point of view of the landmark to be adopted) to localize an object with respect to the landmark.

The great advantage of using exocentric spatial coding is that the number of participants has no influence on the cognitive workload because one does not need to consider different point of views at the same time, nor is it necessary to continuously update the SpaCOFOR. Thus, participants can more easily build an efficient SpaCOFOR and then coordinate their mutual actions in the VE. This allows participants to allocate more cognitive resources to the task. In a study where several collaborative task were considered (WhoDo clue-finding game, word puzzle task, ...), Heldal and Tromp (2006) observed that participants often tend to use the avatar of their collaborator as a reference to indicate a spatial localization. In similar context, Chellali, Milleville-Pennel and Dumas (2013) proposed a paradigm of interaction based on the introduction of a Stable Lateralized Visual Landmark (SLVL) in the VE (figure 1). In this study, two participants collaborated to reproduce a number of geometric models using tetraminos. The model and tetraminos were then placed on a virtual table. For half the participants, a virtual character (avatar) was placed at the center of the scene. This avatar was lateralized (i.e., one right, one left, one front side, and one rear), thus creating a SLVL that was available at all times to facilitate exocentric spatial coding. For the other half of the participants, the avatar was not present. Results showed that the SVLM encouraged the participants to use exocentric spatial coding, whereas the participants favored an egocentric coding when the SVLM was absent. In addition, the presence of the SVLM had a positive effect on the collaborative task, with the direct consequence of a better understanding between partners.





**Fig.1** The SVLM by Chellali et al. (2013)

This study demonstrates that spatial coding in collaborative tasks can be directly related to the content of the CVE. This is in accordance with the results reported by Mou, McNamara, Rump and Xiao (2006), who showed that individuals rely more often on egocentric spatial coding when information allowing exocentric coding is inaccurate or ambiguous. For example, in a task that requires pointing movements toward different memorized objects in different conditions (leading to disorientation in space, or not), participants considered at the same time the spatial relations between objects (exocentric coding) and the objects' position with respect to themselves (egocentric coding). Nevertheless, they favored exocentric coding when the objects were ordered in space and their respective positions were easy to code. Conversely, in case of a non-ordered layout, the participants favored egocentric spatial coding. Thus, beyond the fact that they yield fewer mental rotations, the SVLMs may also disambiguate the relative positions between objects and, as a consequence, encourage the use of exocentric spatial coding to localize and memorize spatial configurations.

#### 1.4. Objectives

We showed that a SVLM could favor the elaboration of an accurate COFOR by making the spatial configuration of the VE less ambiguous. The interest of the SVLM is to promote the spontaneous use of exocentric coding of the visual scene, independent from the point of view of the CVE and common to operators, while reducing the cognitive costs associated with the multiple point of view changes usually encountered in CVE. Thus, such SVLM could reinforce

the decrease of the cognitive workload often associated with the least collaborative effort process. Nevertheless, introducing this sort of clue at the center of the visual scene is not easy, particularly when the content of the VEC is dense or when the users' avatars must be positioned in this place. An alternative would be to provide this sort of clue, but directly on the user's avatar. In this condition, the user's avatar would become an explicit SVLM that the speaker could use as a base for exocentric coding (as the users tended implicitly doing in Pouliquen-Lardy et al. 2015). Another alternative could be to use more distal cues in the surrounding area. These distal cues could help to make the VE more ordered and encourage the partners to use exocentric spatial coding, as observed in the study by Mou et al. (2006). Moreover, this could be used to facilitate information sharing between users and help to decrease mental workload. Yet to be determined is how the spatial configuration between users and the objects in the CVE may influence the use of these spatial cues for the communication and elaboration of the SpaCOFOR.

In order to address these questions, we carried out two complementary studies based on the experimental protocol used by Pouliquen-Lardy et al. (2015) but with new (distal and proximal) spatial cues placed in the CVE. We first considered the production of spatial utterances that aimed to indicate an object's position in a VE (with and without visual cues, Study 1). Then, we considered the understanding of these spatial utterances by an addressee (Study 2). The objectives were:

- To consider different types of contextual visual cues, independent of the observer's point of view, and to identify their impact on the coding of the spatial configurations (i.e., the relative orientation of the addressee, the speaker and the objects to localize in the CVE). For this, we considered the implications of these visual cues on egocentric, other-centric or exocentric coding during the production and understanding of spatial utterances.
- To assess the influence of the visual cues on the realization of the principle of least collaborative effort. To this end, we considered the mental workload for both the addressee and speaker, as well as the spatial coding spontaneously used by the speakers in all spatial configurations.

## **2. Study 1: Production of spatial statements as a function of the visual cues**

### **2.1 Method**

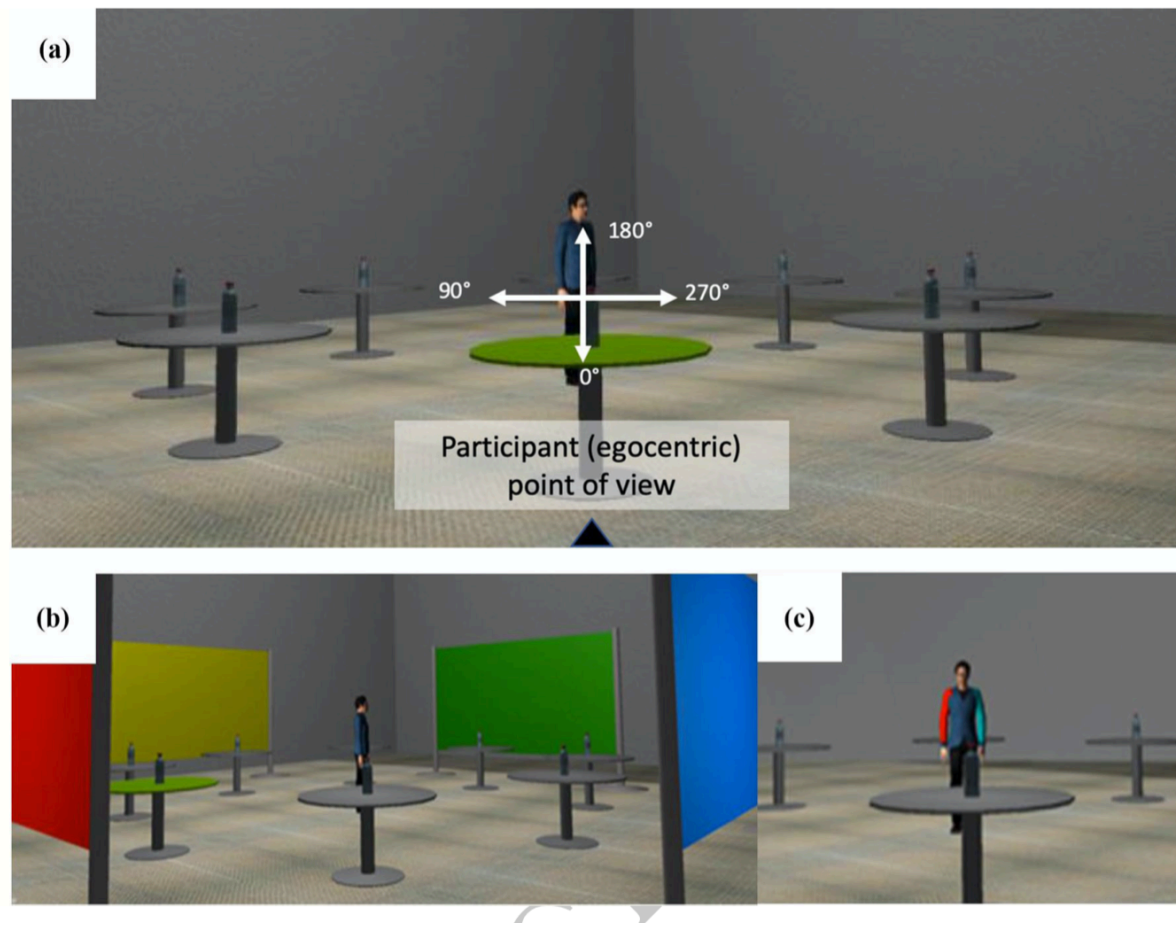
#### **2.1.1 Participants**

5 women and 17 men (all native French speakers) participated in this study. They were aged between 19 and 67 years (29.6 years of age on average).

#### **2.1.2 Apparatus**

This study was conducted on the virtual reality platform at Centrale Nantes (VIVRE.lab). It consisted of an immersive wall (1.8m\*2.4m) equipped with a double overhead projector. Participants wore stereoscopic glasses equipped with markers detected by eight infrared cameras in order to render motion parallax and reproduce head position in the CVE. As in Pouliquen-Lardy et al. (2015), the CVE was composed of a single room containing eight tables organized in circle. A bottle was placed on each table (figure 2). The tables and the bottles were identical (i.e., same size, color and shape). An avatar was placed at the center of the tables. For each trial, the avatar was oriented toward one of the tables. This avatar was supposed to represent a fictive partner in the CVE, but in reality, only the participant was present during the study. The participant was not informed that the avatar was not a real partner and he/she believed that the avatar represented a real participant but with which he/she could not discuss or share information. The participant had an egocentric point of view of the environment that allowed him/her to have a full point of view of the VE and its content (including, the avatar of the fictive partner). He/she was told that it was the same for the (fictive) partner. The participant's point of view relative to that of the avatar was manipulated (figure 2). The avatar could be oriented at 0°, 90°, 180° or 270° with respect to the participant (egocentric) point of view.

In some conditions (see procedure), four panels of different colors surrounded the tables and/or colored cues were placed on the arms of the avatar (figure 2). Each participant was equipped with a set of headphones and could use a keyboard placed in front of them to indicate their answer.



**Fig.2** The virtual environment: a) the target (green table), b) the distal cues (colored panels), and c) the proximal cues (lateralized avatar)

### 2.1.3 Method

The different steps of the study were explained to the participants when they arrived. Before the experiment began, the participants were administered the Santa Barbara Sense of Direction Scale (Hegarty, Richardson, Montello, Lovelace and Subbiah 2002) and the mental rotation test MRT-A © (Vandenberg and Kuse 1978) in order to assess their spatial abilities. Then participants were invited to navigate in the CVE to familiarize themselves. After this, the experiment proper started.

The task of the participants was to indicate to the collaborator (represented by the avatar at the centre of the table) the bottle he/she had to handle from among the eight bottles present in the VE. The participant was told that only he/she could receive information about the bottle and that his collaborator was the only one able to handle the bottle. The participant had to give oral instructions to his/her collaborator without interacting with him/her: thus, the instructions should be as clear as possible. Because this fictive collaborator was the addressee of the instructions, he/she will be referred to as such. Moreover, a time limitation was imposed (in

order to limit experimentation duration and to avoid too much discrepancies between participants): the participant had to give the instructions as soon as possible after the target was identified. The participant had no feedback about the success or the failure of the trial (and he/she could not see the avatar handling the bottle). Effectively, the objective was not to train the participants to give "the good answer" (there was no correct answer strictly speaking) but to evaluate the spontaneous coding they made of space according to the spatial configuration. Because the participant gave the instruction to the addressee, he/she will be referred as the "speaker".

The procedure during the trial is described below.

First, the speaker had to click on « 0 » on the keyboard placed in front of him. The virtual room then appeared. He/she could spend as much time as needed to assess the scene. After this, he/she had to once again click on the "0" and the color of one table changed to indicate the target (bottle). Once the target was identified, the speaker had to give his instruction as quickly and precisely as possible.

Four CVEs were defined depending on the kind of spatial cues they contained:

- The VEs had no specific spatial cues. This condition was named *No Cue*;
- Coloured cues were placed on the arms of the addressee (red on the right arm and green on the left arm; figure 2c); they were considered as proximal cues. This condition was termed *Lateralized addressee*.
- Four panels of different colours were placed in front of each wall of the virtual room. These cues were considered as distal cues. This condition was termed *Panels*.
- In the last condition, two sorts of cues (distal and proximal) were proposed at the same time. This condition was termed *Panels+ Lateralized addressee*.

Thus, the four VEs corresponded to the first independent variable, which was named *spatial cues* and consisted of four modalities: {No Cues, Lateralized addressee, Panels, Panels+ Lateralized addressee}.

The second independent variable was the *orientation of the addressee* with respect to the participant (with four modalities: {0°, 90°, 180°, 270°}).

The third independent variable was the *position of the target* with respect to the addressee with eight modalities, corresponding to the eight tables around him.

The combination of the three independent variables gave rise to 128 configurations (4\*4\*8), which were presented twice to the participant (256 trials). At the second presentation of each trial, the speaker had to indicate the mental workload associated with the task using the French version of the NASA-TLX (Cegarra and Mordago 2009). The NASA-TLX is a subjective,

multidimensional assessment tool that rates perceived workload in terms of six dimensions: mental demand, physical demand, temporal demand, effort, frustration and performance. This is a paper and pencil task. At the end of the main task, the participant scores each dimension from 0 to 100. A global score is then calculated by averaging the scores for the six dimensions). The order of presentation of the variable *spatial cues* was counterbalanced among the participants. For each modality of the variable *spatial cues*, the combinations of the other two variables (*orientation of the addressee* and *position of the target*) were successively presented in a random order.

For the purpose of this study, the following data were recorded and analyzed:

- At the beginning of the study: The results of the two questionnaires of mental rotation.
- For each trial: The speaker verbal description of the spatial localization of the bottle (spatial statements).
- After the second presentation of each trial: the speaker rating of mental workload.

## **2.2 Results**

The analysis of the questionnaire data revealed a negative correlation between the age of the participants and their mental rotation test results (MRT-A ®). It can be seen that as age increases, performance in this test is poorer ( $r=-0.54$ ;  $p<0.05$ ). No other correlation was observed with the variable considered in the study.

### **2.2.1 Coding of space as a function of the type of spatial cues**

We classified spatial statements into five categories:

- Statements that refer to the addressee's position: Addressee-centered Coding (which was a specific sort of exocentric coding and correspond to the other-centric coding mentioned by Duran et al. (2011));
- Statements that refer to the participant's position: Egocentric Coding;
- Statements that refer to the panels' position: Exocentric Coding;
- Statements that use several cues: Mixed Coding (the addressee and/or the participant and/or the panels),
- Statements that refer to other cues (for example cardinal coordinate): Other.

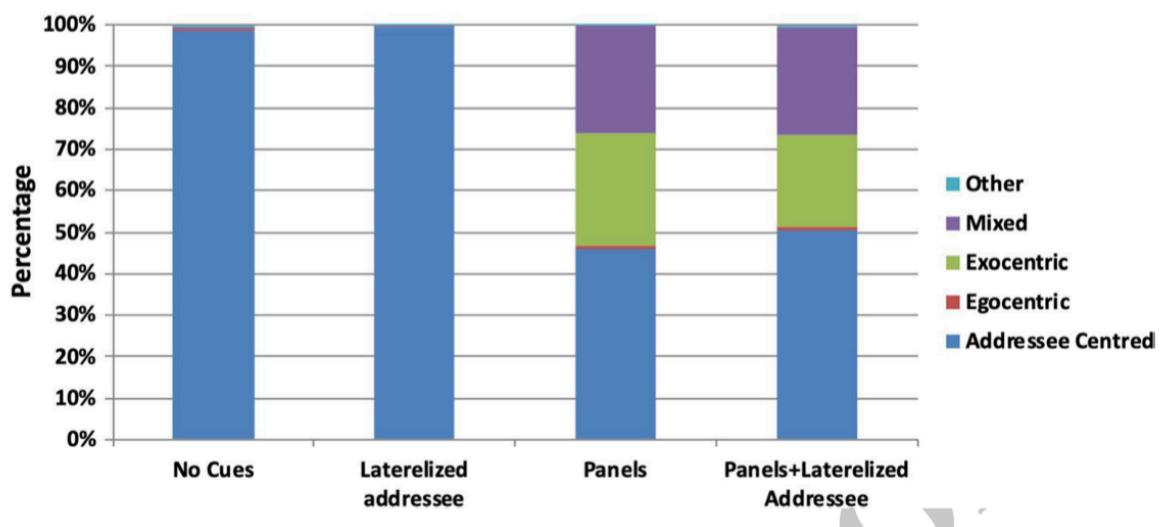
Preliminary inspection of the data revealed that two groups of participants could be distinguished: group 1 - those who never used the panels (7 participants); and group 2 - those who used the panels when they were available (15 participants). If we consider the order of

apparition of the *spatial cues* modalities, it appears that six of the seven participants who never used the panels began with a VE with no panels (*No Cues* or *Lateralized addressee*). Thus, we can suppose that these persons persevered using the strategy they initiated at the very beginning of the experiment.

With regard to group 2, in the *No Cues* and *Lateralized addressee* conditions, 99% of the speakers' statements could be categorized as *Addressee-centered* coding (figure 3). The few remaining statements were categorized as *Egocentric* coding ("in front of me") or referred to the previous trial ("same table").

When the panels were present (*Panels* and *Panels+ Lateralized addressee*), 45.8 % and 50.6 % of the statements were classed as *Addressee-centered* coding; 27% and 22.3% were classed as *Exocentric* coding and 26% and 25 % as *Mixed* coding, respectively.

Statistical analysis showed two significant effects: A change in the proportion of each spatial coding as a function of the position of the target with respect to the speaker ( $\chi^2(16) = 251, p < .001$ ), and a change in the proportion of each spatial coding as a function of the position of the target with respect to the addressee ( $\chi^2(12) = 263, p < .001$ ). The effect of the orientation of the addressee with respect to the speaker was not significant ( $\chi^2(12) = 3.14, p = .99$ ).

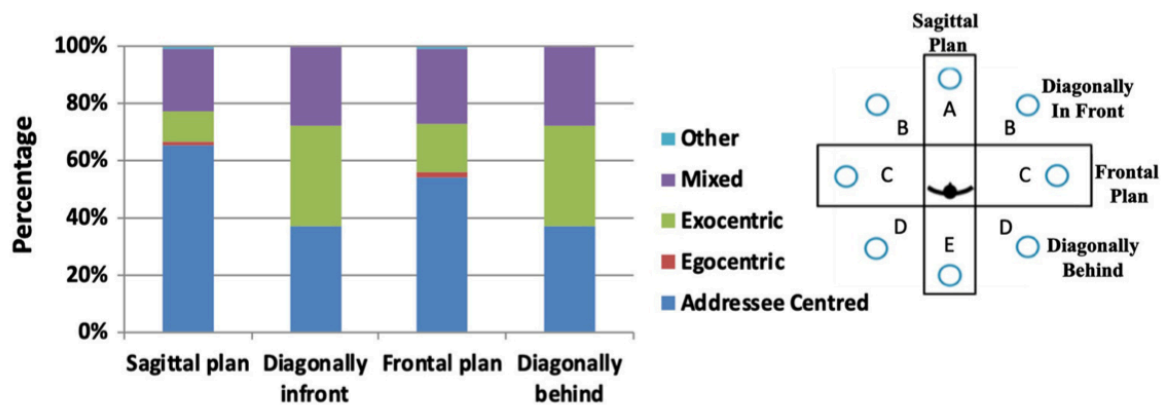


**Fig.3** Percentage of statements using each spatial coding as a function of available spatial cues for the participants who used the panels (group 2)

A detailed analysis of the spatial coding distribution in the two conditions with panels showed that the position of the target relative to the addressee was critical; however, this was not the case for the position of the target relative to the speaker. Indeed, speakers favored *Exocentric* coding when the target was within the diagonal of the addressee (targets B or D;  $\chi^2(4) = 0.03, p > .05$ ; figure 4). Conversely, they mostly used an *Addressee-centered* coding

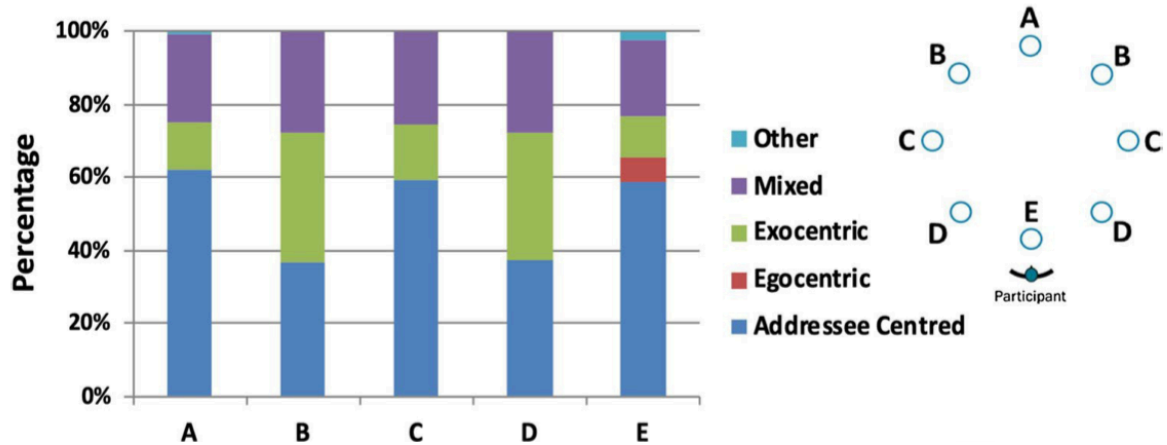
when the target was aligned with one of the intrinsic plans (sagittal or frontal) of the addressee (targets A or C;  $\chi^2(4) = 4.31$ ,  $p > .05$ ; figure 4). The only exception observed was for target E when compared to target A and C ( $\chi^2(8) = 59.22$ ,  $p < .0001$ ). In this case alone the speaker sometimes used Egocentric coding (figure 5). This happened in the spatial configuration in which the target was placed immediately in front of the speaker and in the frontal or sagittal plan of the addressee, depending on the trial. This suggests that the participant supposed that the addressee knew his position in the VE and that, consequently, it was easy for him to find the target with this indication.

Moreover, the diagonal configurations in front and behind the addressee gave rise to the same proportions of spatial coding ( $\chi^2(4) = 0.04$ ,  $p = .99$ ). However, the alignment with the two intrinsic planes of the addressee was not equivalent ( $\chi^2(4) = 12.50$ ;  $p < .05$ ). Indeed, although the statements were predominantly centered on the addressee in both cases, when the target was aligned with the sagittal plane, the participants produced more *Addressee-centered* statements and fewer *Exocentric* statements than when the target was aligned with the frontal plane. Looking in more detail at a comparison of the sagittal and frontal planes, *Addressee-centered* coding was statistically different for the two conditions as was *Exocentric* coding ( $\chi^2(1) = 363$ ,  $p < .001$ ).



**Fig.4** Percentage of statements using each spatial coding as a function of the position of the target with respect to the addressee, for the participants that used the panels (group 2)





**Fig.5** Percentage of statements of each spatial coding as a function of the position of the target with respect to the speaker, for the participants that used the panels (group 2)

### 2.2.2 Mental workload

The previous analysis of spatial coding showed that the position of the target was important mostly because of its orientation with respect to the addressee. For this reason, in our analysis of mental workload scores, conditions were regrouped so that the target position relative to the addressee was coded in the sagittal plane, the frontal plane, and in the front or behind diagonals. The three-factors repeated measures ANOVA (4 spatial cues\*4 Orientation of the addressee\*4 Position of the target with respect to the addressee) showed no effect of the spatial cues on mental workload ( $F(3,63) = 2.3$ ,  $p = .08$ ) and a statistically significant effect of the target position with respect to the addressee ( $F(3,63) = 30.86$ ,  $p < .001$ ). Post-hoc analysis (Table 1a) indicates that this effect was the result of a higher mental workload when the target was diagonally behind and a lower workload when the target was aligned with the sagittal plane. A significant effect of the orientation of the addressee was also observed ( $F(3,63) = 9.95$ ,  $p < .001$ ). Post hoc analysis of this effect (Table 1b) indicates that a lower mental workload was observed at  $0^\circ$  (i.e., when both addressee and speaker were oriented in the same direction). This increased when the addressee is turned right or left with respect to the speaker ( $90^\circ$  and  $270^\circ$ ); the workload is even higher when the addressee is facing the speaker ( $180^\circ$ ).

**Table 1** Post hoc comparison with Newman-Keuls test

(a)	Sagittal	Diagonally in front	Frontal	Diagonally behind
Sagittal		0.00	0.00	0.00
Diagonally in front	0.00		0.46	0.01
Frontal	0.00	0.46		0.00
Diagonally behind	0.00	0.01	0.00	
(b)	0°	90°	180°	270°
0°		0.01	0.00	0.00
90°	0.01		0.03	0.28
180°	0.00	0.03		0.13
270°	0.00	0.28	0.13	

The tables indicate the probabilities of all effects for: (a) the position of the target with respect to the addressee and (b) the orientation of the addressee

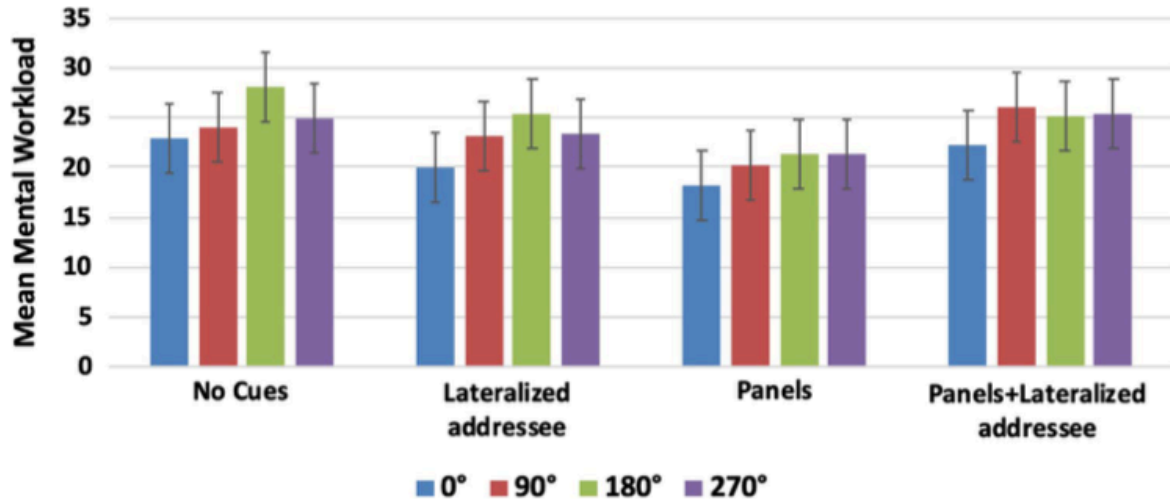
Moreover, the ANOVA showed significant interactions between the spatial cues and the orientation of the addressee on the one hand ( $F(9,189) = 2.8$ ,  $p < .01$ ), and between the spatial cues and the position of the target relative to the addressee on the other hand ( $F(9,189) = 4.34$ ,  $p < .001$ ). The following sections analyze these two interactions in greater detail.

### ***Interaction between the spatial cues and the orientation of the addressee***

Post-hoc analysis (Tukey-test, Within MSE = 85,110,  $df = 189.00$ ) of the interaction between the spatial cues and the orientation of the addressee showed differences in the pattern of results presented in figure 6 as a function of the addressee's orientation.

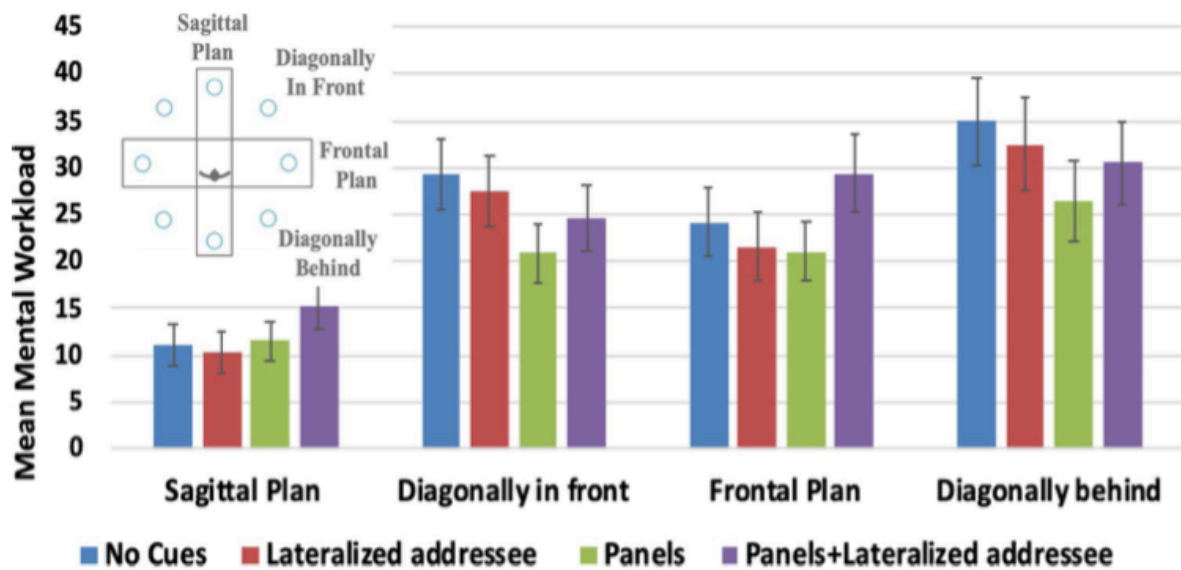
These results showed that the *Panels* condition was the only one to be unaffected by the changes in the orientation of the addressee. Moreover, this condition always required the smallest mental workload when compared with the *No Cues* and *Panels+Lateralized addressee* conditions, whatever the orientation of the addressee (all  $p < .01$ ). This was also the case for the *Lateralized addressee* condition. The only exception observed was for the orientation 0°, which did not differ from the other orientations in the *Panels* condition.

In brief, the mental workload was a function of addressee orientation only when no distal cues were present. The distal cues were found to be the most efficient way of reducing mental workload, particularly in the most complex condition (180° to the addressee), but only when presented alone in the *Panels* condition.



**Fig.6** Mean mental workload for the participants that used the panels. Error bars correspond to the standard error

*Interaction between the spatial cues and the target position with respect to the addressee*



**Fig.7** Mean mental workload in all the cues conditions as a function of the target position with respect to the addressee. Error bars correspond to the standard error

The post-hoc analysis of this interaction (Tukey-test, Within MSE = 40.547, df = 189.00) showed that the alignment of the target with the sagittal plan of the addressee was the spatial configuration that induced the smallest mental workload whatever the spatial cue condition (All  $p < 0.01$ ; Figure 7). In this configuration, distal cues (*Panels*) did not help more. On the contrary, the three other configurations were influenced by spatial cues, but not always

in the same way. The smallest mental workload was always observed in the *Panels* condition, although:

- When the target was in the diagonal, either in front of or behind the addressee, the mental workload was reduced in the *Panels* condition when compared with the *No Cues* condition ( $p < .001$  in both cases). This difference was also significant when the *Panels* condition was compared with the *Lateralized addressee* condition in the front diagonal condition ( $p < .05$ ). In these conditions, the *Panels+Lateralized addressee* gave rise to intermediate results, with no significant differences between any of the spatial cue condition.
- On the contrary, when the target was aligned with the frontal plane of the addressee, the findings for the *Panels* and *Lateralized addressee* conditions did not differ; Nevertheless, in these two conditions, we found a smaller mental workload than for the *Panels+Lateralized addressee* condition ( $p < .01$  in both cases).

## 2.3 Conclusion

The first study showed that the participants preferred to use the avatar of their partner as a spatial reference to indicate a localization in the virtual space. This was particularly true when the target was located in the sagittal plane of the partner. In the more complex situations (when the target was located diagonally in front of or behind the addressee, in particular), the speaker also used distal cues provided by the colored panels to speak to the addressee and thus, sometimes preferred exocentric spatial coding. We can hypothesize that this was determined by the mental workload associated with the mental rotations necessary to bring about addressee-centered coding in these conditions. Indeed, in agreement with Pouliquen-Lardy et al. (2015), we showed that this configuration is the most difficult for the speakers. Thus, the type of spatial cues available in the CVE influences not only the mental workload exerted by spatial coding but also the spatial statement that is produced.

The almost systematic decrease of mental workload when distal cues are present indicates that this condition offers the speaker the highest degree of flexibility when the type of spatial coding needs to be chosen. In this condition, the speaker may favour the least costly type of coding for him whatever the spatial configuration of the VE, including the addressee position and objects present in the space.

It remains to be determined if these choices are only the result of a strategy aimed at minimizing the cost of production for the speaker or if they are part of a more global strategy

that includes the speaker's partner in order to reduce the dyad's mental workload. A second study was set up to test this hypothesis in which the options available in the *Panels* condition of study 1 were replicated. This condition, identified as the most efficient to reduce the coding cost, offers options for both exocentric and addressee-centered coding. In the second study, the mental workload from the addressee's point of view was considered. The goal was to determine whether the speaker's choice of spatial coding strategy was consistent with the mental workload experienced by the addressee.

### **3. Study 2: Understanding of an utterance as a function of spatial coding**

#### **3.1 Method**

##### **3.1.1 Participants**

Sixteen participants (5 women and 11 men) participated in this study. None of the participants took part in the first study. All of them were native French speakers. They were aged between 18 and 59 years (24.2 years old on average).

##### **3.1.2 Apparatus**

The virtual environment used in this study was the same as the one used in the *Panels* condition in study 1. In this second study, the participant became the addressee and stood at the centre of the VE, at the same place as the addressee in the previous study. The eight targets were placed on the table around the participant. As before, the addressee could not see the (fictive) speaker who produced the spatial statements. As in the first study, the participant was not informed that the avatar was not a real partner and he/she believed that the avatar was a real participant but with which he/she couldn't discuss or share information.

##### **3.1.3 Procedure**

The same scenario was used, except that each participant was the addressee rather than the speaker. The different steps of the study were described to participants on arrival. Like in the previous study, in order to assess the spatial abilities of the participants, they were administered the Santa Barbara Sense of Direction Scale (Hegarty et al., 2002) and the mental rotation test MRT-A © (Vandenberg and Kuse 1978). Before beginning the trials, participants were invited to navigate in the VE to familiarize themselves with it. They were then asked to read the instructions. Their task was to identify the bottle indicated by the speaker. Participants were not informed that the speaker was not real. All spatial statements were pre-recorded and

delivered through loudspeakers. Participants had to indicate on the keyboard the number that corresponded to the chosen bottle. The trial took place as follows. First, participants had to click on the “0” of the keyboard placed in front of them (figure 8), which made the virtual room appear. They could spend as much time as was necessary to look at the VE. Then, they had to click once again on the “0” to hear the utterance. Once participants had identified the target, they had to give their answer as quickly as possible.

The utterances could be based on two different types of spatial coding:

- One type of coding was centered on the addressee and used the intrinsic properties of the addressee (*Addressee-Centered* coding).
- One type of coding was exocentric (*Exocentric* coding) and used colored panels (exactly the same as in the first study).

For each type of coding, two trials were proposed in a different order to describe each of the 8 targets positions in the five orientations of the addressee (cf. figure 9).

Three independent variables were thus manipulated:

- The type of spatial coding (two modalities: *Addressee Centered* and *Exocentric*).
- The position of the target with respect to the addressee (eight modalities, corresponding to the eight tables around the addressee).
- The orientation of the addressee with respect to the speaker was also controlled (five modalities: 0°, 45°, 135°, 180°, 270°), because it was one of the variables considered in the first study. However, we did not expect this variable to have any influence because it was not supposed to have an impact on comprehension of the statements. Indeed, in this study, none of the statements used the egocentric frame of reference for the speaker. Thus, only the position of the target with respect to the addressee's axes of symmetry was supposed to influence comprehension of the statements.

Thus, 80 conditions ( $2 \times 5 \times 8$ ) were presented twice to each participant (a total of 160 trials). On the second presentation of each condition, participants had to indicate the mental workload associated with the task (using the NASA-TLX French version put forward by Cegarra and Mordago 2009). The order of the trials was defined as follows. All target positions were presented successively and randomly for a given spatial coding and orientation of the addressee. Thus, for orientations 2 and 4 (figure 9), the diagonally placed targets were located

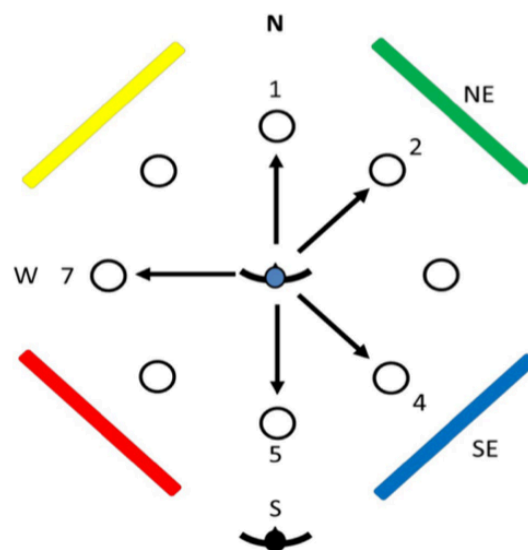
between two panels, whereas for orientations 1, 5 and 7, they were in front of a single panel. The 5 Addressee Orientations and the 4 Space Coding were counterbalanced among the participants. Each change in type of spatial coding was preceded by five training trials. If the participant made a mistake when clicking on the keyboard, the trial was canceled and repeated at the end of the block.

For the purpose of this study we analyzed the following data:

- At the beginning of the study: The results of the two questionnaires of mental rotation.
- For each trial: The addressee verbal description of the spatial localization of the bottle.
- After the second presentation of each trial: the addressee rating of mental workload.



**Fig.8** Representation of the target position on the keyboard used by participants to give their answers

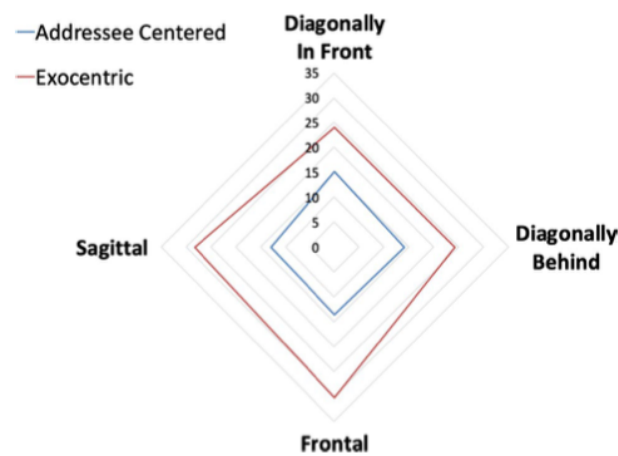


**Fig.9** Addressee positions with respect to the speaker (represented by the five arrows). Participants (represented by a blue head) stood at the center of the eight tables

### 3.2 Results

As in the previous study, there was also a negative correlation between the age of participants and SCOT test results ( $r = .89$ ,  $p < .001$ ) and the mental rotation test results (MRT-A ®). It can be seen that as age increased, performance in this test was poorer ( $r = -.54$ ,  $p < .05$ ). No other correlation was observed with the variable considered in the study.

During the first study, we observed that the target position with respect to the addressee was a determining factor for the choice of the spatial coding and mental workload of the speaker. For consistency with the previous analysis, the mental workload was analyzed as a function of the target position relative to the addressee.



**Fig.10** Mental workload as a function of spatial coding and the target position with respect to the addressee

A three-factor repeated measures ANOVA was computed (2 space coding \* 5 orientation of the addressee \* 4 target positions with respect to the addressee: *sagittal plane*, *frontal plane*, *diagonally in front* and *diagonally behind*). As expected, no effect of the orientation of the addressee was observed ( $F(4,60) = 0.39$ ,  $p > .05$ ). We observed a statistically significant effect of spatial coding on mental workload ( $F(1,15) = 14.42$ ,  $p < .005$ ), which resulted in a significantly lower mental workload for *Addressee Centered* statements and a higher mental workload for *Exocentric* statements (figure 10).

We also observed a significant effect of the target position with respect to the addressee ( $F(3,45) = 3.70$ ,  $p < .05$ ). The target in the frontal plane resulted in a higher mental workload than the *diagonally in front* and *diagonally behind* positions (table 2). Nevertheless, this result must be considered at the light of the interaction between the type of spatial coding and the target position with respect to the addressee ( $F(4,45) = 8.64$ ,  $p < .001$ ). This interaction mostly indicates an influence of the target position on the mental workload associated with *Exocentric*



coding (figure 10 and table 3): the mental workload was lower when the target was *diagonally* oriented (behind and in front) than when it was in the *sagittal plane* (in front and behind) and in the *frontal plane* (figure 10). With regard to *Addressee-centered* coding, no differences were observed between the different target positions with respect to the addressee; furthermore, the mental workload was always lower than in the *Exocentric* coding condition.

**Table 2** Pairwise comparisons for the target position with respect to the addressee

	Frontal	Diagonally behind	Sagittal	Diagonally in front
Frontal		0.02	0.23	0.05
Diagonally behind	0.02		0.17	0.53
Sagittal	0.23	0.17		0.23
Diagonally in front	0.05	0.53	0.23	

	Addressee-Centered				Exocentric			
	Frontal	Diagonally behind	Sagittal	Diagonally in front	Frontal	Diagonally behind	Sagittal	Diagonally in front
<b>Addressee-Centered</b>								
Frontal		0.54	0.72	0.30	0.30	0.00	0.00	0.00
Diagonally behind	0.54		0.59	0.38	0.00	0.00	0.00	0.00
Sagittal	0.72	0.59		0.26	0.00	0.00	0.00	0.00
Diagonally in front	0.30	0.38	0.26		0.00	0.00	0.00	0.00
<b>Exocentric</b>								
Frontal	0.00	0.00	0.00	0.00		0.00	0.29	0.00
Diagonally behind	0.00	0.00	0.00	0.00	0.00		0.00	0.88
Sagittal	0.00	0.00	0.00	0.00		0.00		0.01
Diagonally in front	0.00	0.00	0.00	0.00	0.00	0.88	0.01	

### 3.3 Conclusion

This second study showed that, for the comprehension of statements, spatial coding is more difficult to manage when based on *Exocentric* coding. In terms of the influence of the spatial configuration of the objects in the VE, the target position with respect to the addressee had no influence on mental workload when *Addressee-centered* coding was used whereas it influenced the mental workload in the case of an *Exocentric* coding.

To make the link between the mental workload of the speaker and addressee, and to determine the extent to which the theory of least collaborative effort may be validated by our results, one further step is needed. In the first study, we showed that the VE could influence the choice of spatial coding and the mental workload associated with the production of a spatial statement. The speaker's preferred type of coding was either centered on the addressee or was exocentric: an environment that allowed the use of these two types of spatial coding was the most efficient to facilitate the task. In the second study, the mental workload associated with the addressee's understanding of the two types of spatial coding was assessed. It was demonstrated that exocentric coding was the most difficult to understand.

Within the theoretical framework of the least collaborative effort, the choice of a particular type of spatial coding results from a compromise between the mental workload needed for the production and comprehension of spatial utterances. The following analyses aim to determine whether the hypothesis made by the speaker about the difficulty of the task for his/her partner (as a function of his/her orientation and the position of the target) are relevant and whether his/her choice of type of coding actually allows his/her mental workload to be reduced, along with that of the addressee. To this end, the mental workload needed to produce a specific type of spatial coding in a given addressee/target configuration will be linked to its counterpart in terms of comprehension.

## **4. Validation of the principle of least collaborative effort**

In this section, we will compare the data about mental workload for the production and comprehension of the same sort of spatial utterance. When comparing two strictly identical situations in terms of VE content and spatial coding, it is possible to determine whether the mental workload resulting from coding carried out by the speaker is proportionally inverse to the workload resulting from the addressee's comprehension. For this reason, we decided to compare the *Panels* conditions of the two studies (production and comprehension) for the same orientations of the addressee.

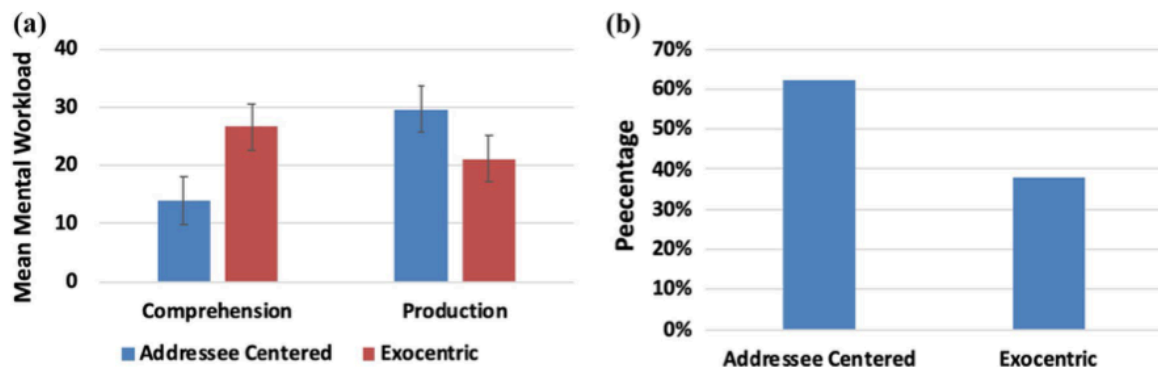
### **4.1 Method**

We considered the orientation of the addressee that were common to the two studies for all target positions: 0°, 180° and 270°. In order to compare the same types of statements, the results obtained for the *Exocentric* and *Addressee-centered* statements in the production study were compared with those obtained from the comprehension study.

This analysis can include all participants from study 2 (16), but only some of those who took part in study 1. Indeed, since the participants of study 1 were free to choose their preferred spatial coding, they did not systematically produce *Exocentric* and *Addressee Centered* statements. As a consequence, only those participants who had fewer than 25% of missing data in *Exocentric* and *Addressee-centered* statements for all the conditions were kept in the analysis (six participants for *Exocentric* coding and eight for *Addressee-centered* coding). For this reason, only non-parametric analyses were performed: we computed a Friedman ANOVA for repeated measures and post hoc comparisons were analyzed with a Wilcoxon test.

## 4.2 Results

As observed in study 1, the participants we selected for this analysis produced a majority of addressee-centered utterances (figure 11b). It appears that addressee-centered utterances are less difficult to understand than exocentric ones ( $t(15) = 2.44$ ,  $p > .05$ ), whereas the amount of mental workload was not statistically different for the two types of spatial coding during production ( $t(7) = 1.24$ ,  $p > .05$ ; figure 11a). This result is congruent with the higher percentage of addressee-centered utterances produced by the speaker and is in alignment with the least collaborative effort principle.



**Fig.11** a) Mental workload for comprehension and production statements with regard to *Addressee Centered* and *Exocentric* types of coding. Error bars correspond to the standard error. b) Number of utterances for *Addressee Centered* and *Exocentric* types of coding

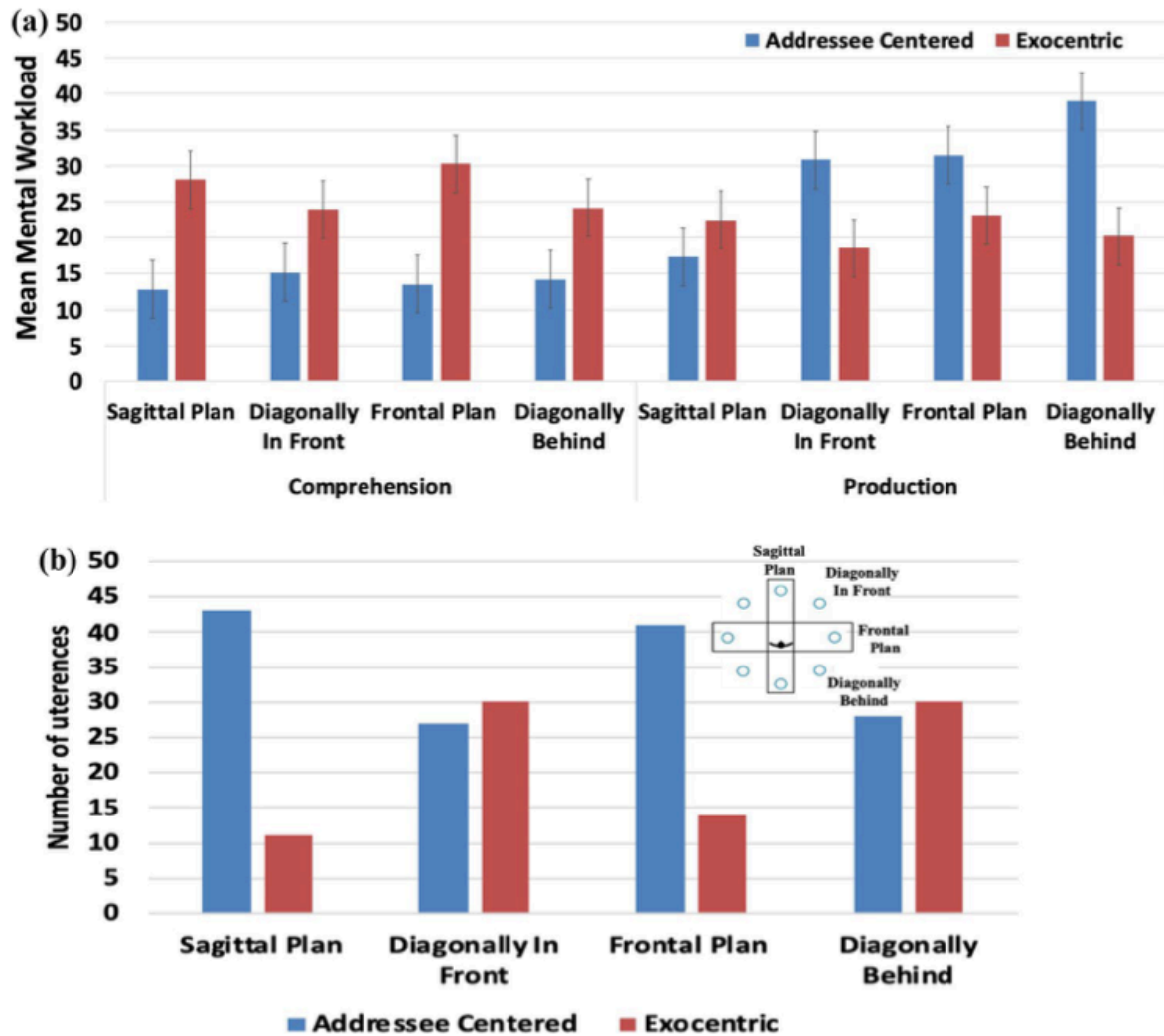
A more detailed analysis that takes into account the spatial configuration of both the target and the addressee showed noteworthy differences between *Exocentric* and *Addressee-Centered* coding instances (for production and comprehension, respectively: Friedman Chi Sqr( $N = 16$ ,  $df = 7$ ) = 61.11,  $p < .001$  and Friedman Chi Sqr( $N = 6$ ,  $df = 7$ ) = 17.13,  $p < .05$ ).

During the production task (figure 12a), the mental workload for the *Exocentric* statements did not differ as a function of the target/addressee configuration (Friedman Chi

Sqr( $N = 6$ ,  $df = 3$ ) = 7.29, ns), in contrast to the *Addressee-centered* statements (Friedman Chi Sqr( $N = 8$ ,  $df = 3$ ) = 16.75,  $p < .001$ ). Indeed, the cost associated with *Addressee-centered* coding was smaller for a target placed in the sagittal plane of the addressee than for the other conditions ( $z = 2.38$ ,  $z = 2.38$  et  $z = 2.52$  all  $p < .05$  for diagonally in front, frontal and diagonal behind planes, respectively). On the other hand, mental workload was higher when the target was placed diagonally behind ( $z = 2.36$  and  $z = 2.10$ ,  $p < .05$  for diagonally in front and frontal planes, respectively).

In the comprehension task, we observed the reverse pattern of results. The cognitive cost of comprehension did not differ as a function of the target/addressee configuration for *Addressee-Centered* statements (Friedman Chi Sqr( $N = 16$ ,  $df = 3$ ) = 3.71,  $p > .05$ ), whereas significant differences were observed for *Exocentric* statements (Friedman Chi Sqr( $N = 16$ ,  $df = 3$ ) = 28.09,  $p < .001$ ). The *diagonal* (in front and behind) configurations were the least costly (diagonally in front/sagittal plane:  $z = 2.99$ ,  $p < .01$ ; diagonally in front/frontal plane:  $z = 2.99$ ,  $p < .01$ ; diagonally behind/sagittal plane:  $z = 2.89$ ,  $p < .01$ ; diagonally behind/frontal plane:  $z = 3.15$ ,  $p < .01$ ) and did not differ from each other ( $z = 0.36$ , ns). The *frontal plane* condition was the most difficult condition; indeed, mental workload was even higher in this configuration than in the sagittal plane ( $z = 2.21$ ,  $p < .05$ ).

Nevertheless, whatever the target/addressee configuration in the comprehension task, the *Addressee-centered* statements always yielded a lower mental workload than for *Exocentric* statements (sagittal plane:  $z = 3.18$ ,  $p < .01$ ; diagonally in front:  $z = 2.37$ ,  $p < .05$ ; frontal plane:  $z = 3.30$ ,  $p < .001$ ; diagonally behind:  $z = 2.89$ ,  $p < .01$ ).



**Fig.12** a) Mental workload for the comprehension and production tasks for addressee-centered coding and exocentric coding as a function of the target position. Error bars correspond to the standard error. b) Number of utterances for addressee-centered and exocentric coding as a function of target position

### 4.3 Conclusion

From this study, it appears that the participants often used spatial coding that was most costly for them in situations where the reverse was true for their partner. Figure 12 clearly shows that the selection of spatial coding was determined more by the spatial configuration of the target/addressee than by the cognitive workload needed for the production of spatial statements. The distribution of spatial coding (figure 12b) was obviously in favor of addressee-centered coding when the target was aligned with the intrinsic plans of the addressee, whereas it was equally distributed between the two types of coding when the target was diagonally positioned with respect to the addressee. The choice of addressee-centered coding in the sagittal plane can be easily explained by the lower costs of comprehension and production in this configuration. Indeed, when an object located in the sagittal plane is coded in an addressee-

centered space, the mental rotations are not necessary because, in this condition, it is only necessary to identify the front and behind of the object and, indeed, is easy to do so.

Conversely, in the other three conditions, addressee-centered coding was costlier for the speaker, and clearly easier for the addressee. One can wonder why the speaker would favor addressee-centered coding when the target is in the frontal plan and yet be indifferent about choosing between the two types of spatial coding when the target is placed diagonally to the addressee. One possible explanation is that the speaker's choice is based on a representation of the cognitive workload (for the addressee) resulting from the use of exocentric coding when the target was in the frontal plane rather than when placed diagonally: in other words, the speaker favors addressee-centered coding when he/she knows for sure that it is less costly for both him and the addressee (sagittal plane) or when he/she supposes that it is, at least, less costly for the addressee. In the case of the frontal plane, it is a common understanding that it is easy to identify his/her right and left. Conversely, the speaker feels that he/she can use a less costly coding (exocentric coding) when it is less obvious that using addressee-centered coding will benefit the addressee. When the target is placed diagonally it can be difficult to identify a position on the basis of addressee-centered coding ("in front of you and on your right in the middle" or "between the one in front of you and the one on your right"). In these conditions, using the panels can sound like a good solution, particularly when one panel is located just behind the target. It can reduce ambiguity for the addressee and make the task easier for the speaker ("the one just in front of the yellow panel"). This can also explain the more marked alternation between the two types of coding in this condition; the speaker chooses the frame of reference according to the positioning of the visual context and its relevance for spatial coding. A clear explanation is thus offered as to why the speaker would favor addressee-centered coding, with the associated risk of an additional cost for himself/herself. This is just because the speaker is aware that he/she makes the task easier for the addressee. When the target is placed diagonally, the effort is less justified: the speaker feels authorized to choose exocentric coding, which allows the balance between the cognitive cost of the two partners to be restored. This is a perfect example of least collaborative cost.

## **5. General discussion**

### **5.1. Influence of the spatial configuration of VE content on mental workload**

The two studies presented in this paper confirm the results obtained by Pouliquen-Lardy et al. (2015) Duran et al. (2011) and McNamara (2003) about the influence of objects configuration on spatial coding. They showed that, for speakers, the spatial configurations diagonally in front and behind are the most complex to manage, either with or without visual cues (even if these cues were useful for reducing mental workload in these conditions). In the best cases, visual cues reduce the level of difficulty to a level just equivalent to one of the other conditions. Thus, this spatial configuration, which is particularly complex to handle, should be avoided whatever the amount and quality of visual cues in the VE.

### **5.2. Use of the visual cues available in the VE**

We hypothesized that the presence of stable visual cues would allow speakers to describe the space in an exocentric frame of reference that would be completely independent from their point of view in the VE. Indeed, we supposed that it would be very useful to reinforce, thanks to colored visual cues, the perception of the laterality of the avatar (standing for the addressee). This could have favored the speakers' tendency to spontaneously use the addressee or another character in the VE as a basis for an exocentric spatial coding (Chellali et al. 2013). Yet, the results of the first study showed that an ability to reinforce the "laterality" of the addressee with proximal cues was of little use to the speaker. Indeed, it was deemed to be even harmful when distal exocentric cues were available to organize space. Presumably, in this case there was competition between the two types of cues (each based on a distinct color code), which may have yielded an additional cost for the speaker. Conversely, distal-colored cues were efficient in reducing mental workload. Studies 1 and 2 showed that these cues make it possible to homogenize mental workload during the comprehension phase in the most complex spatial configurations and to significantly reduce production costs in these same configurations. They have the particular advantage of offering a less costly coding solution for the speaker by allowing him to avoid overly complex mental rotations. The analyzes carried out in section 4 make it possible to better understand the interest that the presence of these cues may have for the speaker but also, indirectly, the addressee. This is the subject of the discussion in the next section, along with the principle of least collaborative effort.

### **5.3. Mental workload and the principle of least collaborative effort.**

The principle of least collaborative effort proposed by Clark and Wilkes (1986) postulates that, in a conversation, the two participants try to minimize the collaborative effort, not by reducing individual effort, but more by reducing joint effort and maximizing individual effort. In Study 1, we saw no effect of the addressee's orientation on the choice of spatial coding, although there was an effect on mental workload. This implies that the speakers did not take this into account when choosing the type of spatial coding of space, even though the orientation of the addressee had repercussions on the mental workload that this represented for them. On the other hand, they did take into account the position of the target relative to the addressee. Thus, the coding choices were guided by cost-specific considerations for the addressee. This was also observed in the study by Pouliquen-Lardy et al. (2016). It is interesting to note that the speaker chose this strategy without even having to interact with the interlocutor. The speaker's choice was based solely on his personal representation of the difficulties related to the task. The operators implemented the principle of least collaborative effort by adopting a way of exchanging information based on asymmetrical cognitive cost, taking into consideration each other's (supposed) difficulties. Thus, the mental workload needed for the two operators can be balanced throughout the task. This is consistent with Duran et al (2011) study, which showed that participants knowledge about "what the partner is supposed to know or see" has direct influence on the strategy used by the speaker to code spatial location. As demonstrated by Duran, Dale and Galati (2016), all these examples of coding's plasticity contribute to confirm that spatial communication is a dynamical and multifactorial process implying the interaction of many factors like social cues, spatial configuration, previous experience and partners' common ground... In our study, as well as that by Pouliquen-Lardy et al. (2016), the result of this "implicit betting" was positive because the speaker had a good representation of the difficulties of the task as well as the (supposed) skills of the addressee. One can suppose that any collaboration would be much less favorable if the speaker was mistaken in this supposition. Therefore, it is essential to promote among the operators a reciprocal knowledge of the space that can be directly perceived by each of them, as well as a good representation of its content and evolution according to movements and manipulations carried out by each of them in the VE.



## 6. Conclusion

Reinforcing exocentric cues in the VE is a good way to maximize the performance of the collaborators by maximizing both the overall and individual cognitive costs. However, this reinforcement must not be carried out in an arbitrary manner but must take into account not only the information already presented in the VE, but also the partners' respective positions and the quality of their shared knowledge about the spatial configuration and content of the VE.

## References

- Bryant, D. J., & Tversky, B. (1999). Mental representations of spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25:137-156.
- Bryant, D.J., & Wright, W.G. (1999). How body asymmetries determine accessibility in spatial frameworks. *Q J Exp Psychol A*, 52(2), 487-508.
- Cegarra, J., & Morgado, N., (2009) Étude des propriétés de la version francophone du NASA-TLX. In: Cahour, B., Anceaux, F., Giboin, A. (Eds.), *EPIQUE 2009: 5ème Colloque de Psychologie Ergonomique*. Nice, France, pp. 233-239.
- Chellali, A., Milleville-Pennel, I., & Dumas, C. (2013). Influence of Contextual Objects on Spatial Interactions and viewpoints sharing in Virtual Environments. *Virtual Reality Journal*, 17(1), 1-15.
- Clark, H.H., & Brennan, S.E. (1991). Grounding in Communication. In L.B. Resnick, J. Levine, S.D. Teasley (Eds.), *Perspectives on Socially Shared Cognition*. Washington D.C.: APA, pp. 127-149.
- Clark, H.H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1-39.
- Duran, N., Dale, R., & Galati, A. (2016). Toward integrative dynamic models for adaptive perspective taking. *Topics in Cognitive Science*, 8, 761-779.
- Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition*, 121, 22-40.
- Galati, A., & Avraamides, M. N. (2015). Social and Representational Cues Jointly Influence Spatial Perspective-Taking. *Cognitive science*, 39, 739-765. DOI: 10.1111/cogs.12173.
- Gaver, W.W., Sellen, A., Heath, C., & Luff, P. (1993) *One is not enough: multiple views in a media space*. In: Proceedings of INTERCHI, pp 335-341.
- Hegarty, M., Richardson, A.E., Montello, D.R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, 30, 425-447.
- Heldal, I., Spante, M., Steed, A., & Bengtsson, S. (2005). Successes and failures in co-present situations. *Presence: Journal of teleoperators and virtual environments*, 14(5), 563-579.
- Heldal, I., & Tromp, J. (2006). The usability of collaborative virtual environments and methods for the analysis of interaction. *Presence: Journal of teleoperators and virtual environments*, 15(6), 655-667.
- Hindmarsh, J., Fraser, M. Heath, C. Benford, S., & Greenhalg, C. (1998). *Fragmented Interaction: Establishing mutual orientation in virtual environments*. In CSCW'98, pp.217-226.
- Hoc, J.M. (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations. *Int J Hum Comput Stud*, 54, 509-540. doi:10.1006/ijhc.2000.0454.

- McNamara T.P. (2003) How Are the Locations of Objects in the Environment Represented in Memory?. In: Freksa C., Brauer W., Habel C., Wender K.F. (eds) *Spatial Cognition III. Spatial Cognition 2002*. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 2685. Springer, Berlin, Heidelberg.
- Michelon, P., & Zacks, J.M. (2006) Two kinds of visual perspective taking. *Percept Psychophys*, 68, 327-337.
- Mou, W., McNamara, T. P., Rump, B., & Xiao, C. (2006). Roles of egocentric and allocentric spatial representations in locomotion and reorientation. *Journal of Experimental Psychology: Learning Memory and Cognition*, 32, 1274-1290.
- Pouliquen-Lardy, L., Mars, F., Guillaume, F., & Milleville-Pennel, I. (2015). Virtual collaboration: effect of spatial configuration on spatial statements production. *Cognitive Processing*, 16, S337-S342.
- Pouliquen-Lardy, L., Milleville-Pennel, I., Guillaume, F., & Mars, F. (2016). Effect of role distribution on the common spatial frame of reference in a remote collaborative task. *Virtual Reality*, 20(4), 213-220. DOI 10.1007/s10055-016-0294-8.
- Spante, M., Schroeder, R., Axelsson, A.-S., & Christie, M. (2004). How Putting Yourself into the Other Persons Virtual Shoes Enhances Collaboration. In *Proceeding of the 7th International Workshop on Presence*, Valencia, Spain, pp. 190-196.
- Tversky, B., & Hard, B.M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition* (110), 124-129. doi:10.1016/j.cognition.2008.10.008
- Vandenberg, S.G., & Kuse, A.R. (1978). Mental Rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599-601.