Effect of Avatar Facial Expressiveness on Team Collaboration in Virtual Reality

Hugo Le Tarnec, Elisabetta Bevacqua, Olivier Augereau, Pierre de Loor

Effect of Avatar Facial Expressiveness on Team Collaboration in Virtual Reality

Hugo Le Tarnec
b<>com – Institut de recherche technologique &
ENIB, Lab-STICC UMR 6285 CNRS
Brest, France
Hugo.LETARNEC@b-com.com

Elisabetta Bevacqua
ENIB, Lab-STICC UMR 6285 CNRS
Brest, France
elisabetta.bevacqua@enib.fr

Olivier Augereau
ENIB, Lab-STICC UMR 6285 CNRS
Brest, France
olivier.augereau@enib.fr

Pierre De Loor
ENIB, Lab-STICC UMR 6285 CNRS
Brest, France
deloor@enib.fr

ABSTRACT

This paper investigates the impact of displaying the facial expressions of a user in real time on the performance of the task, the social presence and the behavioral changes of the users interacting in a Virtual Reality environment. To evaluate this approach, we conducted a study where the users collaborated to build a TV stand in dyad including a novice and an expert assistant. The collaborative task was divided into two independent phases: a face-to-face discussion phase without object manipulation, and a furniture assembly phase with object manipulation. Our results indicate that the proposed approach can increase social presence and lead to gaze behavior changes in multi-user environments during the face-to-face phase.

CCS CONCEPTS

• **Human-centered computing → Interaction design**; **Human computer interaction (HCI)**; **HCI design and evaluation methods**;

KEYWORDS

Virtual Reality, Collaboration, Non verbal communication, Facial expressions

ACM Reference Format:
Hugo Le Tarnec, Elisabetta Bevacqua, Olivier Augereau, and Pierre De Loor. 2024. Effect of Avatar Facial Expressiveness on Team Collaboration in Virtual Reality. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The realization of collaborative work in Virtual Reality (VR) has been a major challenge in recent years. The use of VR for collaborative tasks is motivated by the fact that users can interact with a virtual environment that closely resembles the real world, offering a high level of interactivity with its objects. This provides the advantage of shared visual information regarding the progress of a given task.

By collaboration, we mean the fact that two or more people agree on a common goal and commit to achieving it [29]. In order to enhance collaborative work, it is crucial to have a shared visual information about the task's progress, i.e. to have a shared visual workspace. Additionally, having information into the internal state of the interaction partner is also essential [12–14].

Moreover, VR offers an engaging solution that enables multiple individuals to interact within the same virtual space through the use of graphical representations known as avatars. This immersive environment allows users to engage in social interactions by leveraging various non-verbal cues, such as gestures, facial expressions, eye movements, and proxemic behavior [24]. The utilization of this non-verbal language is vital for effective interactions, as it conveys internal states, including emotional state [11], fatigue [20], stress [15], and even motivation for example. Consequently, this facilitates access to information pertaining to the internal state of the interaction partner.

The challenge of accurately capturing and conveying users' internal states through their non-verbal behaviors via avatars persists in the realm of VR. Additionally, the constraints imposed by limited technological capabilities, specifically the real-time transmission of large amounts of data, make it crucial to restrict data transmission. This limitation also serves to mitigate the ecological impact of digital technology. Therefore, it becomes imperative to identify the non-verbal behaviors that significantly contribute to improved interaction and collaborative work while discerning those that do not.

This research paper aims to investigate whether the real-time display of users' facial expressions on their avatar's face, which serves as a behavioral representation of internal states like stress [10], workload [30], or emotional state [11], can enhance collaborative work within a VR environment.

In Section 2, we review related work. In Section 3, we describe design and implementation decisions of the virtual environment. Section 4 details the experiment and Section 5 presents the results. Finally, we critically discuss these results in Section 6 and conclude in section 7.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 RELATED WORK

2.1 Social presence

During collaborative tasks, it is crucial for participants to perceive the presence of their fellow collaborators. This "feeling of being with another" [4] is commonly referred to as social presence and holds significant importance in the field of VR. Social presence is considered a sub-concept of presence, which is defined as the "feeling of being there" [22]. It should be noted that the concept of presence, in contrast to immersion, pertains to the psychological experience provided by the mediated system, while immersion focuses on the objective aspects derived from the inherent qualities of the technology [25].

To measure social presence, different questionnaires have emerged since the first in 1976 published by Short and his team [28]. With renewed interest in the concept in the early 2000s for the evaluation of communication systems, various questionnaires appeared. Biocca, with various collaborators, has published various models and questionnaires widely used today [4, 5]. Makransky and his team [23] have developed a simplified social presence questionnaire composed of only 5 questions, while being conceptually valid and maintaining the reliability of the measurement. Social presence being a broad and subjective concept, real-time measurement through behavioral or physiological data is complicated. Nevertheless, some works have focused on behavioral indicators that would indicate social presence [8, 16]. This stems from the assumption that if a person is engaged in a social interaction, then he must feel that the other is socially present. Concerning physiological data, certain physiological indicators such as heart rate, skin conductance or brain activity make it possible to measure social psychological responses such as the processing of affect and motivation [6], and does not exclusively deal with the social presence. This may be due to the difficulty of identifying one coherent physiological response simply to gauge the presence of another, even though interaction with others may elicit various psychophysiological responses depending on the context and the interaction.

In order to enhance interactions in VR and, consequently, promote social presence, numerous studies have investigated factors such as audio quality, screen resolution, and even individual characteristics such as age and gender [25]. These factors have been shown to influence the level of social presence experienced by users (see Figure 1).

Due to the positive correlation between social presence and favorable communication outcomes, researchers and professionals in the field have displayed significant interest in examining factors that enhance social presence. Among these factors, the visual representation of the communication partner has been extensively investigated. This particular factor has demonstrated the potential to improve interaction and elicit behavioral changes in participants [25].

Since collaboration can be viewed as the intersection between communication, coordination and cooperation [26], and given the significant impact of social presence on communication, our research focuses on the exploration of this concept of social presence and its relevance in our study.

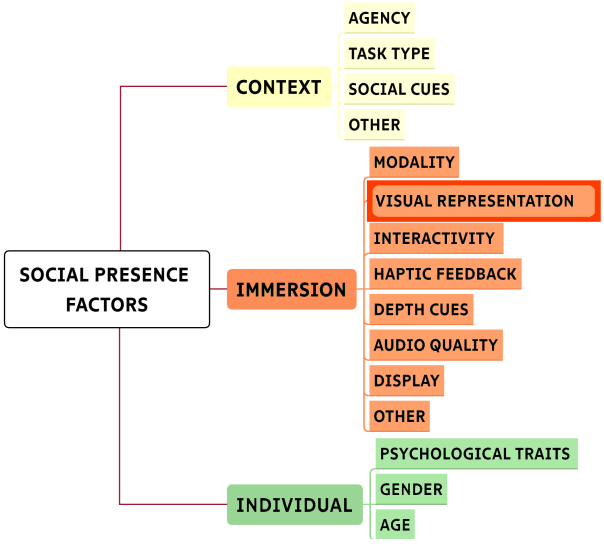

Figure 1: Social presence factors divided into three groups: Context, Immersion and Individual [25]. For our work, we are interested in the visual representation of the communication partner, classified in the "immersion" group.

2.2 Non-verbal behavior

Different aspects of the visual representation have been evaluated so far, such as the notion of behavioral appearance [2, 9, 32].This concept refers to the extent to which the virtual representation behaves as a real person would behave, i.e. replicate the non-verbal behavior of the user.

We focused on facial expressiveness because it allows us to give a lot of information about the state of the user [10, 11, 30]. For example, Bailenson and his team showed that adding an eye blink and a pupil dilation led to an increase in social presence [2]. Bente and his team have shown that varying the duration (2 vs 4 seconds) of gaze directed at agents has an impact on social presence [3]. The longer gaze duration leads to a higher level of social presence. In addition, various works, such as those of Casanueva and Blake [7] as well as Kang and his team [18] have shown that avatars with facial expressions allow for a higher feeling of social presence compared to static avatars. Krumhuber and his team [21] showed that changing facial dynamics significantly influences game participants' choices and decisions to cooperate with other players. In our study, the difference is that the environment is in virtual reality. Indeed, at a time when facial capture devices in VR are beginning to emerge, it is important to assess the impact of facial expressiveness on collaboration in VR.

2.3 Transformed Social Interaction

Even though much work has focused on reproducing non-verbal communication in virtual environments, it is possible, thanks to the almost infinite capabilities of VR, to artificially transform social interactions, namely to use non-imitative approaches, and go beyond what is is possible in an unmediated interaction. Hollan and Stornetta [17] hypothesized that non-imitative approaches might lead

Figure 2: The collaborative virtual environment during the discussion phase where the participant and the expert assistant prepare a strategy for assembling a piece of furniture.

to even better solutions than imitation of face-to-face mechanisms. The advantage of virtual environments is that it is possible to filter out non-verbal behaviors that are not essential for the interaction and which can interfere with communication, such as, for example, filtering out the fact that a person twirls their pen during a meeting. Moreover, due to the fact that virtual environments simulate the world separately for each user, it is possible to make the interaction specific for each user, and meet the requested needs. Bailenson and his team [1] introduced the concept of "Transformed Social Interaction" (TSI), i.e. generating non-verbal behaviors of the avatar in the virtual environment different from the behaviors that the user actually performs. For example, Roth and his team [27] presented a new approach for increasing social behaviors with the development of three visual transformations in the virtual environment: increased eye contact, joint attention and clustering. For example, joint attention was modeled with bubbles floating over an object when two people looked at it simultaneously. These transformations have increased the social presence perceived by users.

3 DESIGN AND IMPLEMENTATION

3.1 Collaborative environment

In all the experiments mentioned above [7, 18, 21], the tasks performed were oriented to encourage face-to-face interaction without object manipulation. Indeed, this type of task is a factor of social presence.

Nevertheless, while performing collaborative work in VR, the users will have to interact with the environment and the objects that compose it. Therefore, in our experiment, we wanted to break down the collaborative task into 2 independent phases: a face-to-face discussion phase without object manipulation and a phase with a shared visual workspace with object manipulation. We chose, for that, to have a TV stand assembly task between a novice (the participant) and an expert assistant. The role of the expert assistant is played by the experimenter. This task is divided into two:

1) a discussion phase (See Figure 2) where the participant and the expert assistant prepare a strategy for assembling a piece of furniture after the expert assistant has explained how the snapping of the parts works. During this phase, the expert assistant and the participant are located face to face at a distance of approximately two meters from each other in the virtual space.

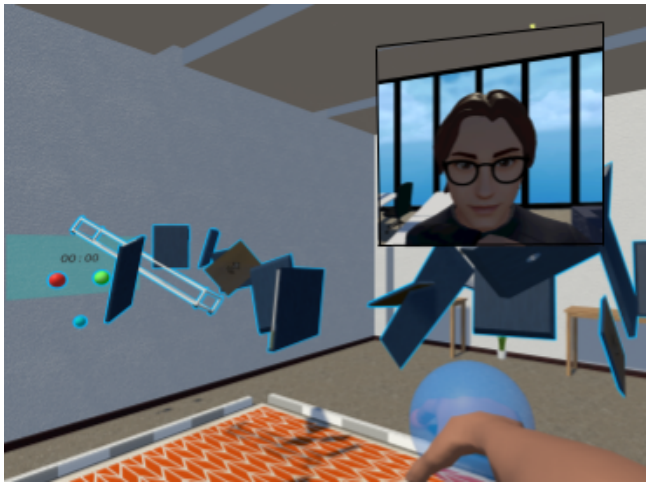

Figure 3: The collaborative virtual environment during the assembly phase of the furniture. The expert assistant gives instructions while the participant handles a piece of furniture. The HUD display screen allows the participant to view the face of the expert assistant's avatar in real time.

2) a furniture assembly phase (See Figure 3) with handling where only the participant can handle the pieces and assemble them. Handling is done without tools or/and screws. Indeed, the pieces stick together automatically when they are well placed. This feature is called "snapping". During this phase, a HUD display screen showing the face of the expert assistant's avatar is displayed. This screen allows participants to see the face of the expert assistant's avatar whenever they want. It is considered a Transformed Social Interaction (TSI) described in Section 2.3.

Our hypotheses are:

H1: *Participants with facial expressions will experience more social presence compared to those without facial expressions.*

H2: *Participants with facial expressions will spend more time looking at the face of the expert assistant's avatar than those without facial expressions.*

H3: *Participants with facial expressions will build the piece of furniture faster (speed of execution) than those without facial expressions during the assembly phase.*

The virtual environment is based on the Ngagement 1.0 framework developed in Unity 3D by b<>com[1]. This platform allows several users to simultaneously experience a shared environment in VR. For our experiment, we had to develop some additional features such as 3D modeling of the furniture with "snapping". This feature allows to stick one piece with another one when their relative positions are correct. Two HTC Vive Pro Eye tracker headsets with HTC Facial Tracker sensor, allowing to recover the facial movements of the lower face, with two HTC Vive Pro controller joysticks allowing the

[1]b<>com website https://b-com.com/en

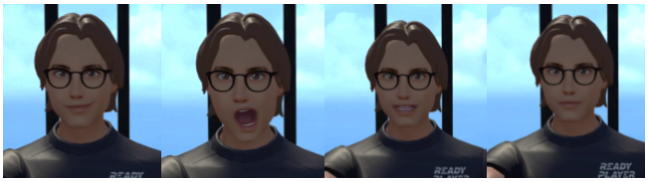

Figure 4: Facial expressions on a Ready Player Me avatar animated with blendshapes based on facial movements from HTC facial tracker and eye tracking sensors.

control of the movement of the arms and hands. Two HTC Station VIVE Pro 2.0 bases were used for the experiment.

3.2 Avatar appearance

In this experiment, the goal was to render the real-time facial expressions of each user on their avatar. To capture behavioral data related to participants' facial movements, an HTC Facial Tracker sensor [2] and an integrated eye tracker within the HTC Vive Pro Eye were employed. These technologies allowed the retrieval of 41 key points on the face (32 on the lower part and 9 on the upper part). Ready Player Me avatars [3] were utilized, as they incorporate blendshapes that enable the display of facial expressions. Each blendshape value was determined based on the corresponding key point values on the face (See Figure 4). Additionally, the eye tracker facilitated the replication of users' gaze direction. To achieve accurate representation of facial expressions and gaze direction, a calibration phase was conducted.

3.3 Augmentation

During the assembly phase, a HUD display screen showing the face of the expert assistant's avatar is displayed. This screen allows the participant to see the face of the expert assistant's avatar in all circumstances (See Figure 3). We added this screen because during the pre-test, we noticed that the participants were so focused on the pieces of furniture and never looked at the expert assistant's avatar. This may also be due to the expert assistant's avatar being located a bit far from the participant avatars relative to the resolution of the headset. This screen allows the participant to look at the pieces and the face of the expert assistant's avatar in a quasi-simultaneous way. This screen makes it possible to assess whether people are not looking at the expert assistant's avatar due to the fact that they have to make a physical effort to look at the faces of others or simply due to the fact that facial expressions are unnecessary during the assembly phase.

4 EXPERIMENT

4.1 Design

This study employed a between-subjects design, comparing two conditions: "Without facial expressions" and "With facial expressions". Participants engaged in a task where they assembled a TV

[2]VIVE Facial Tracker https://www.vive.com/fr/accessory/facial-tracker
[3]Metaverse Full-Body Online 3D Avatar Creator — Ready Player Me. https://readyplayer.me/.

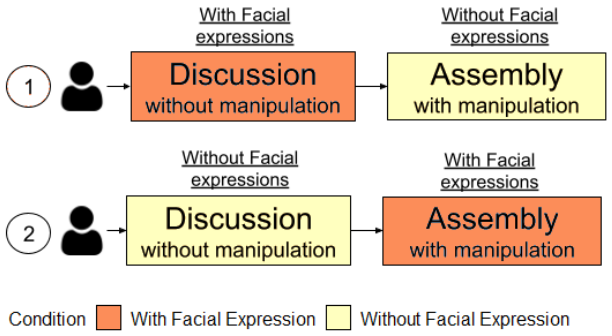

Figure 5: The experiment in between/inter design broken down into two phases discussion/assembly with alternation of the starting condition.

stand with an expert assistant. The participant and the expert assistant were physically situated in separate rooms. Participants were unaware of the true purpose of the experiment, while the expert assistant, played by the experimenter, had full knowledge of it.

4.2 Task

The task was organized into dyads, consisting of an expert assistant and a participant. Their task was to assemble a TV stand within a maximum time limit of 20 minutes. The participant handled the individual parts and performed the assembly, while the expert assistant provided instructions. Notably, only the expert assistant had access to a visual representation of the fully assembled furniture. The participant was instructed to remain within a designated area, while the expert assistant remained in their original position.

4.3 Procedure

The participant gave informed consent. He was given a random number from 1 to 2 to know which condition was going to be assigned first to ensure correct relationship measurements (Without facial expressions, With facial expressions). We asked the participant to complete the pre-study questionnaires (demographic questionnaire, VR experience and cybersickness questionnaire). The participant was informed about the upcoming task using a script-based spoken instruction.

For the first part of the experiment (discussion), we equipped the subject with a VR headset (HTC Vive Pro with Eye tracker with HTC Facial Tracker), two controllers (HTC Controller Vive Pro). During this phase the participant and the expert assistant prepare a strategy for assembling a piece of furniture after the expert assistant has explained how the snapping of the pieces works. After this phase, after removing the VR headset, a social presence questionnaire is completed by the participant.

Then, for the second part of the experiment (manipulation), we re-equipped the subject with the VR headset. We gave him about 1 minute of acclimatization time in a practice room, to become familiar with the use of the controllers for object manipulation. During this manipulation, an easy-to-use furniture (three pieces) can be manipulated. After that, the participant was taken to the virtual

test environment. We started data recording and experimentation with an oral "go". The dyad (participant + expert assistant) had 20 minutes to complete the TV stand. If it was finished before, the experience ended there. The participant could raise their hands during the experiment if he encountered technical problems.

After the experiment, we asked the participant to complete the post-experimental questionnaires (social presence + Cybersickness). The participant could add general comments on the experiment. The overall experience lasted about 45 minutes. The experiment was conducted at the European Center for Virtual Reality[4].

4.4 Measures

4.4.1 Subjective Measures. We measured a 5-question social presence questionnaire (likert scale with 1 = strongly disagree, 5 = strongly agree) to test the impact of facial expressions to investigate H1. We have selected the social presence dimension from the multimodal presence questionnaire of Makransky and his team [23]:

- Q1: *I felt like I was in the presence of another person in the virtual environment.*
- Q2: *I felt that the people in the virtual environment were aware of my presence.*
- Q3: *The people in the virtual environment appeared to be sentient (conscious and alive) to me.*
- Q4: *During the simulation there were times where the computer interface seemed to disappear, and I felt like I was working directly with another person.*
- Q5: *I had a sense that I was interacting with other people in the virtual environment, rather than a computer simulation.*

In addition, participants had the opportunity to add qualitative comments regarding the experience.

4.4.2 Objective Measures. We collected two types of objective data. The first type of data was behavioral data. We collected eye tracking data, which allows us to retrieve the behavior of the user's gaze and know which object or which avatar they are looking at. From these data, we calculated the "Time gaze expert assistant" in percentage. It corresponds to the time to have looked at the expert assistant's avatar divided by the total time of the experiment (H2). For the assembly phase, the gaze time corresponds to the gaze time of the face of the expert assistant's avatar added to the gaze time of the HUD display screen showing the face of the expert assistant's avatar.

The second type of data concerns task performance. There was no possibility of making a construction error because of the snapping of the furniture parts, we only recovered the total assembly time of the furniture ("Building time" in seconds) (H3).

4.4.3 Control Measures. To assess whether other factors influence subjective and objective measures, we administered a demographic questionnaire and the Simulator Sickness questionnaire [19] before and immediately after the task.

4.5 Participants

The participants were recruited through a registration link distributed on university and company mailing lists. A total of 29 participants took part in the experiment. All analyzes were done with the data of the 29 participants, except for the analysis of the eye tracking data, where 2 participants were withdrawn due to a technical problem. Thus, for hypothesis H2, data from 27 participants were used. The sample was composed of 15 women and 14 men. The average age was 27.76 years. The conditions were assigned randomly. The number of previous VR experiences did not differ significantly between the conditions.

[4]European Center for Virtual Reality https://cerv.enib.fr

5 RESULT

The presentation of the results is divided into two phases corresponding to the two phases of the task.

5.1 Discussion phase

5.1.1 Subjective Results. To test hypothesis *H1* and see the impact of facial expressions on social presence during the discussion phase, we conducted a Wilcoxon-Mann-Whitney test on the averages of the 5 questions of the social presence questionnaire presented in the section 4.4.1. The results showed a significant difference in perceived social presence between the conditions "Without facial expressions" (M = 3.84, SD = 0.52, "M" represents the mean and "SD" represents the standard deviation.) and "With facial expressions" (M = 4.26, SD = 0.43, p = .02) (See Figure 6).

5.1.2 Objective Results. To test hypothesis *H2* and see the impact of facial expressions on gaze behavior during the discussion phase, we conducted independent samples t-tests, in compliance with the test application conditions. The results showed a significant difference in the conditions "Without facial expressions" (M = 0.44, SD = 0.13)

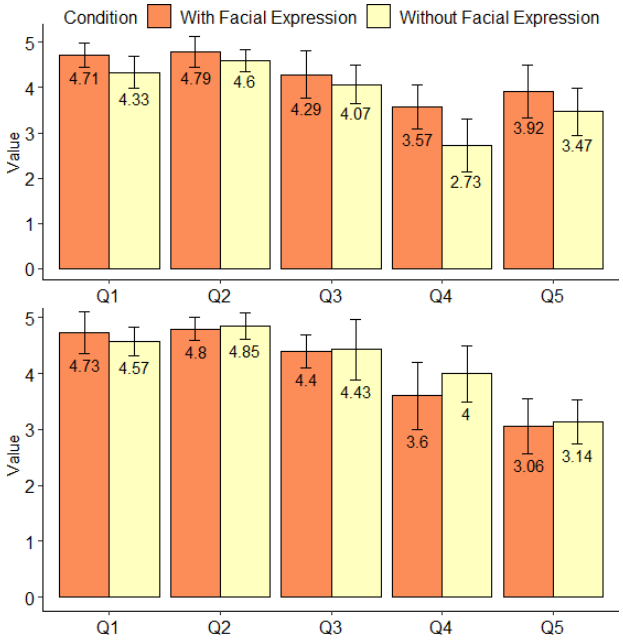

Figure 6: Top: Result of the social presence questionnaire during the discussion phase. Bottom: Result of the social presence questionnaire during the assembly phase.

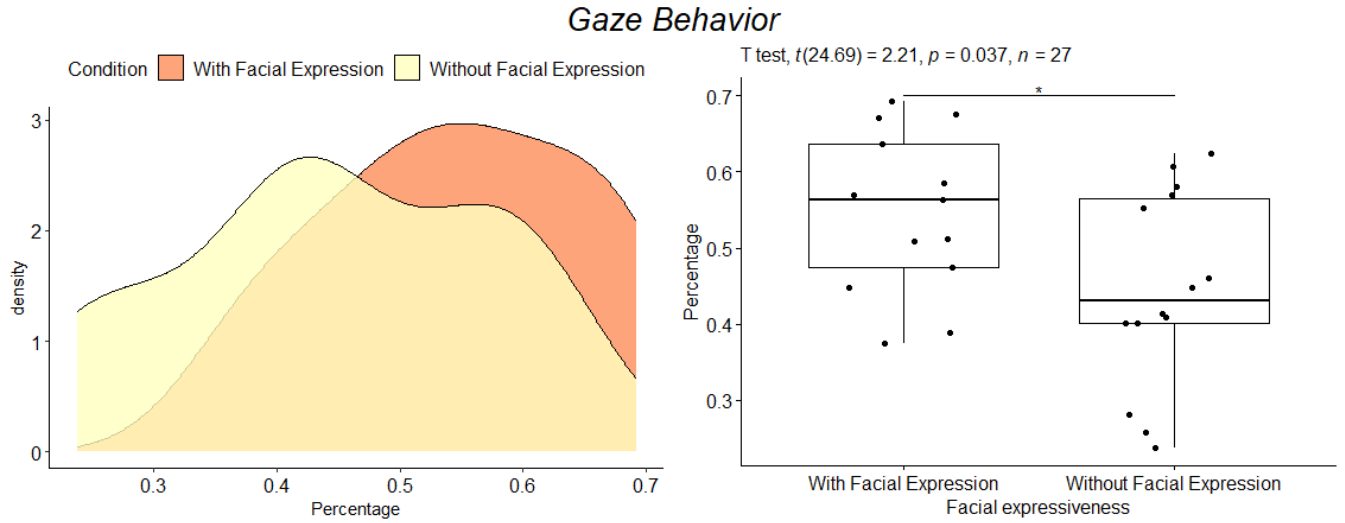

Figure 7: Results from the objective measures. On the left, the distribution representation curve shows the distribution of percentage of participants' gaze time on the face of the expert assistant's avatar during the discussion phase. On the right, the boxplot shows the distribution of each percentage of participants' gaze time on the face of the expert assistant's avatar during the discussion phase, * indicates p < .05 for two Sample t-test

and "With facial expressions" (M = 0.55, SD = 0.11, $t(24.69) = 2.21$, p = .037) (See Figure 7).

5.2 Assembly phase

5.2.1 Subjective Results. To test hypothesis *H1* and see the impact of facial expressions on social presence during the assembly phase, we conducted a Wilcoxon-Mann-Whitney test. The results showed a non significant difference in perceived social presence between the conditions "Without facial expressions" (M =4.44, SD = 0.48) and "With facial expressions" (M = 4.28, SD = 0.48, p = .38) (See Figure 6).

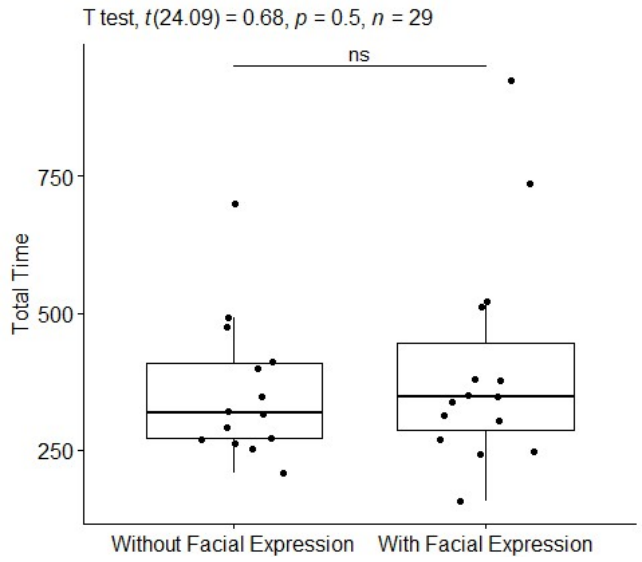

Figure 8: Results from the time performance in seconds during the assembly phase. "ns" indicates a non-significant difference and p > .05 for two Sample t-test

5.2.2 Objective Results. To test hypothesis *H2* and see the impact of facial expressions on gaze behavior during the assembly phase, we conducted independent samples t-tests. The results showed a non significant difference in the conditions "Without facial expressions" (M = 0.073, SD = 0.08) and "With facial expressions" (M = 0.071, SD = 0.03, $t(15.24) = -0.09$, p = .93). To test hypothesis *H3* and see the impact of facial expressions on task performance, we conducted independant samples t-tests. The results showed a non significant difference in task performance between the conditions "Without facial expressions" (M = 358.61, SD = 129.38) and "With facial expressions" (M = 401.05, SD = 200.87; $t(24.09) = 0.68$, p = .5). Results of the task performance are presented in Figure 8.

6 DISCUSSION AND LIMITATION

Regarding hypothesis *H1* during the discussion phase, we found a significantly higher perception of social presence in the "With facial expressions" condition compared to the "Without facial expressions". This implies that the visual facial expressions displayed on the avatar can increase the feeling of being with another when people have a classic face-to-face interaction.

Regarding hypothesis *H2* during the discussion phase, we found a significant difference in the percentage of gaze time at the expert assistant's avatar between the "With facial expressions" condition and the "Without facial expressions" condition. This implies that visual facial expressions displayed on the avatar leads to a behavioral change in the participants. Indeed, participants in the "With facial expressions" condition will tend to look more at the face

of the expert assistant's avatar compared to the "Without facial expressions" condition.

However, for the assembly phase, regarding hypothesis *H1, H2, H3*, we did not find significant differences between the "With facial expressions" condition compared to the "Without facial expressions". The results showed that the participants do not look at the expert assistant's avatar when handling the furniture compared to the discussion phase, despite the fact that a HUD display screen was added, making it easy to see the face of the expert assistant's avatar. This could be verified by the behavioral data of the participants' gaze. Indeed, the gaze time during the assembly phase on the face of the expert assistant's avatar (real avatar head + HUD display screen) was only 7% of the total time compared to the 52% of the discussion phase.

Hypotheses H1 and H2 are partially validated (only in the face-to-face discussion phase). Participants with facial expressions felt more social presence and spent more time looking at the expert assistant's avatar compared to those without facial expressions. Hypothesis H3 is not validated. Participants with facial expressions did not build the furniture faster (speed of execution) than those without facial expressions during the assembly phase.

With regard to our overall research goal, we can therefore state that facial expressions can be beneficial for face-to-face interaction experiences, but which has no effect during the manipulation phase, even if the facial information is displayed on a screen to be clearly visible in all circumstances.

Thus, it can be concluded that depending on the task (face-to-face discussion/ manipulation) it is not necessary to transmit facial expressions, especially in a context where limiting the data transfer is fundamental. We can hypothesize that, in the case of shared attention between the pieces of furniture and the communication partner (the expert assistant), the participants, in order not to lower the performance of the task, tend towards different sensory input (sight/hearing) [31]. As the sensory input regarding awareness of the positions of the pieces can only be visual, participants will tend to use another available channel (hearing) to interact with the expert assistant.

7 CONCLUSION AND FUTURE WORK

In this paper, we presented the design, implementation and an evaluation of a concept for the improvement of collaborative work in VR. We have developed a furniture assembly environment where the expert assistant's facial expressions are captured using an HTC face tracker sensor and reproduced in real time on his avatar. We hypothesized that displaying facial expressions on the expert assistant's avatar improves collaborative work in terms of performance and in terms of social presence during the discussion and the assembly phase.

During the discussion phase, the results of the social presence questionnaire and the gaze time showed a significant difference between the two conditions.

During the assembly phase, the results showed that the participants did not look at the face of the expert assistant's avatar but spent the majority of the time looking at the pieces of furniture, even if the facial information is displayed on a screen to be clearly visible in all circumstances. Indeed, the results of the satisfaction and social presence questionnaires and the results of the furniture assembly time showed no difference between the conditions "Without facial expressions" and "With facial expressions".

The restitution of the facial expressions of the users on their avatar, with very high fidelity and restoring the micro-expressions, making it possible to deduce the internal state such as stress for example, is still difficult to conceive at the present time. In our future work we want to develop a model capable of detecting and analyzing user behaviors from a more general point of view to infer their internal state. Then, display a representation of this inference in the form of an augmentation/transformation that would maintain interaction and improve collaboration.

ACKNOWLEDGMENTS

This work is conducted within the eXtendR project founded by the French National Agency for Research (ANR) in b<>com research institute. We would like to thank b<>com for their support and the CONTINUUM (ANR-21-ESRE-0030) project for the use of material and equipment necessary for the experiments.

REFERENCES

[1] Jeremy N Bailenson, Andrew C Beall, Jack Loomis, Jim Blascovich, and Matthew Turk. 2004. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators & Virtual Environments* 13, 4 (2004), 428–441.

[2] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. 2001. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments* 10, 6 (2001), 583–598.

[3] Gary Bente, Felix Eschenburg, and Lisa Aelker. 2007. Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. In *Presence 2007: Proceedings of the 10th Annual International Workshop on Presence, October 25-27, 2007, Barcelona, Spain*. Citeseer, 207–14.

[4] Frank Biocca, Chad Harms, and Judee K Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments* 12, 5 (2003), 456–480.

[5] Frank Biocca, Chad Harms, and Jenn Gregg. 2001. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*. 1–9.

[6] Jim Blascovich. 2000. Using physiological indexes of psychological processes in social psychological research. *Handbook of research methods in social and personality psychology* (2000), 117–137.

[7] Juan Casanueva and Edwin Blake. 2001. The Effects of Avatars on Co-presence in a Collaborative Virtual Environment. In *Annual Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT), September 25-28, 2001, Pretoria, South Africa*. 19–28.

[8] Deborah A Coker and Judeek Burgoon. 1987. The nature of conversational involvement and nonverbal encoding patterns. *Human Communication Research* 13, 4 (1987), 463–494.

[9] Sara Dalzel-Job. 2015. Social interaction in virtual environments: the relationship between mutual gaze, task performance and social presence. (2015).

[10] David F Dinges, Robert L Rider, Jillian Dorrian, Eleanor L McGlinchey, Naomi L Rogers, Ziga Cizman, Siome K Goldenstein, Christian Vogler, Sundara Venkataraman, and Dimitris N Metaxas. 2005. Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, space, and environmental medicine* 76, 6 (2005), B172–B182.

[11] Paul Ekman. 1993. Facial expression and emotion. *American psychologist* 48, 4 (1993), 384.

[12] Darren Gergle, Robert E Kraut, and Susan R Fussell. 2004. Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of language and social psychology* 23, 4 (2004), 491–517.

[13] Darren Gergle, Robert E Kraut, and Susan R Fussell. 2013. Using visual information for grounding and awareness in collaborative tasks. *Human–Computer Interaction* 28, 1 (2013), 1–39.

[14] Darren Gergle, Carolyn P Rosé, and Robert E Kraut. 2007. Modeling the impact of shared visual information on collaborative reference. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1543–1552.

[15] Dimitris Giakoumis, Anastasios Drosou, Pietro Cipresso, Dimitrios Tzovaras, George Hassapis, Andrea Gaggioli, and Giuseppe Riva. 2012. Using activity-related behavioural features towards more effective automatic stress detection. (2012).

[16] C Heeter. 1992. Being There: the subjective experience of presence, presence: teleoperators and virtual environments. (1992).
[17] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 119–125.
[18] Sin-Hwa Kang, James H Watt, and Sasi Kanth Ala. 2008. Communicators' perceptions of social presence as a function of avatar realism in small display mobile communication devices. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*. IEEE, 147–147.
[19] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.
[20] Jarek Krajewski, David Sommer, Udo Trutschel, Dave Edwards, and Martin Golz. 2009. Steering wheel behavior based estimation of fatigue. In *Driving assessment conference*, Vol. 5. University of Iowa.
[21] Eva Krumhuber, Antony SR Manstead, Darren Cosker, Dave Marshall, Paul L Rosin, and Arvid Kappas. 2007. Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion* 7, 4 (2007), 730.
[22] Kwan Min Lee. 2004. Presence, explicated. *Communication theory* 14, 1 (2004), 27–50.
[23] Guido Makransky, Lau Lilleholt, and Anders Aaby. 2017. Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior* 72 (2017), 276–285.
[24] Divine Maloney, Guo Freeman, and Donghee Yvette Wohn. 2020. " Talking without a Voice" Understanding Non-verbal Communication in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–25.
[25] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. 2018. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI* (2018), 114.
[26] Sam Redfern and Niall Galway. 2002. Collaborative virtual environments to support communication and community in internet-based distance education. *Journal of Information Technology Education. Research* 1 (2002), 201.
[27] Daniel Roth, Constantin Klelnbeck, Tobias Feigl, Christopher Mutschler, and Marc Erich Latoschik. 2018. Beyond replication: Augmenting social behaviors in multi-user virtual realities. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 215–222.
[28] John Short, Ederyn Williams, and Bruce Christie. 1976. *The social psychology of telecommunications*. Toronto; London; New York: Wiley.
[29] Candace L Sidner. 2016. Engagement, emotions, and relationships: on building intelligent agents. In *Emotions, Technology, Design, and Learning*. Elsevier, 273–294.
[30] Richard T Stone and Chen-Shuang Wei. 2011. Exploring the linkage between facial expression and mental workload for arithmetic tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55. SAGE Publications Sage CA: Los Angeles, CA, 616–619.
[31] Christopher D Wickens. 1984. The multiple resources model of human performance: Implications for display design. (1984).
[32] Yuanjie Wu, Yu Wang, Sungchul Jung, Simon Hoermann, and Robert W Lindeman. 2021. Using a fully expressive avatar to collaborate in virtual reality: Evaluation of task performance, presence, and attraction. *Frontiers in Virtual Reality* 2 (2021), 641296.