



Improving Collaborative Learning in Virtual Reality With Facial Expressions

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

► To cite this version:

Hugo Le Tarnec, Olivier Augereau, Elisabetta Bevacqua, Pierre de Loor. Improving Collaborative Learning in Virtual Reality With Facial Expressions. ACM international joint conference on pervasive and ubiquitous computing (UbiComp), Sep 2022, Atlanta, United States. 10.1145/3544793.3560387 . hal-03880377

HAL Id: hal-03880377

<https://hal.science/hal-03880377>

Submitted on 1 Dec 2022

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.

Improving Collaborative Learning in Virtual Reality With Facial Expressions

HUGO LE TARNEC, b<>com – Institut de recherche technologique & ENIB, Lab-STICC UMR 6285 CNRS, France

OLIVIER AUGEREAU, ENIB, Lab-STICC UMR 6285 CNRS, France

ELISABETTA BEVACQUA, ENIB, Lab-STICC UMR 6285 CNRS, France

PIERRE DE LOOR, ENIB, Lab-STICC UMR 6285 CNRS, France

This article presents an approach to improve collaborative learning in terms of performance & satisfaction through the generation of non-verbal behavior of users displayed on their avatar in virtual reality. Various works have focused on the behavioral realism of avatars, which can considerably improve interactions. The purpose of this paper is to investigate the impact of displaying the facial expressions of a user in real time on the performance of the task, the satisfaction and the behavioral changes of the users interacting in a virtual environment. To evaluate this approach, we carried out a study where the users collaborated to build a TV stand in dyad including a novice (the participant) and an expert assistant (the experimenter).

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

Additional Key Words and Phrases: Virtual Reality, Collaboration, Non verbal communication, Facial expressions

ACM Reference Format:

Hugo Le Tarnec, Olivier Augereau, Elisabetta Bevacqua, and Pierre De Loor. 2022. Improving Collaborative Learning in Virtual Reality With Facial Expressions. In *Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC '22 Adjunct)*, September 11–15, 2022, Cambridge, United Kingdom. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3544793.3560387>

1 INTRODUCTION

The realization of collaborative learning in Virtual Reality (VR) has been a major challenge in recent years. Indeed, one of the main reasons VR is used for learning and training is that it allows for individuals to be exposed to a virtual environment that resembles the real world with great interactivity with the objects and people that compose it [14]. In a pandemic context, more and more people are collaborating remotely and VR proposes an interesting solution that allows several people to interact and learn through the same virtual space by being embodied by a graphic representation called an avatar. Thanks to this, users can experience social interactions by exploiting the richness of social signals such as gestures, facial expressions, eye movements, etc. Being able to transcribe all one's internal states or to transcribe non-verbal behaviors correlating with those states, via the avatar, remains a current problem. Indeed, due to the limited technological capacities, ie to transmit a large quantity of data in real time, and the significant environmental and ecological issues, limiting the transmission of data remains essential. It is therefore necessary to identify the non-verbal behaviors that are important for better interaction and collaborative learning. This article contributes by exploring

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

Manuscript submitted to ACM

whether the display of user’s facial expressions in real time, considered as a behavioral representation of certain mental states such as stress [5], workload [19] or emotional state [6], allows an improvement of collaborative learning.

2 RELATED WORK

2.1 Collaboration and VR concepts

By collaboration, we mean the fact that two or more people agree on a common goal and commit to achieving it [18]. Collaborative learning is an educational approach to teaching and learning that involves groups of learners working together to solve a problem, complete a task, or create a product [13]. To improve collaborative learning, it is necessary to have shared visual information relating to the state of a task (shared visual workspace) and to have information relating to the internal state of the interaction partner [7–9]. It is, therefore, important that the user perceives the presence of the other collaborators. This “feeling of being with another” [3] is a concept known as social presence and is important in the field of VR. It is often highlighted and measured during interactions in a virtual environment. The concept of social presence is considered a sub-concept of presence, defined as the “feeling of being there”. This concept is seen as a psychological experience offered by the mediated system and differs from immersion—which designates what technology brings from a purely objective point of view through the intrinsic qualities of the mediated system [16]. To improve the interactions in VR and consequently, the social presence, various works have focused on factors such as the audio quality, the resolution of the screen, or even the age and the gender of the person [16] allowing variation to the social presence (see Fig. 1).

Because social presence often predicts positive communication outcomes, academic researchers and practitioners have shown great interest in studying factors that increase social presence. One of the most studied factors is the visual representation of the communication partner. Indeed, varying the appearance of the communication partner(s) can lead to better interaction and a change in the behavior of the interlocutors [16]. Because collaboration can be viewed as the intersection between communication, coordination, and cooperation [17], and because social presence has a direct effect on communication, we are interested in this notion of social presence and its use in our research. Indeed, we hypothesize that improving the feeling of social presence has a positive effect on collaboration.

2.2 Non-verbal behavior

Different aspects of the visual representation have been evaluated so far, such as the notion of behavioral appearance, defined as the extent to which the virtual representation behaves as a real person would behave. Much work has focused on reproducing behavioral appearance in virtual environments. For example, Bailenson and his team showed that adding an eye blink and a pupil dilation led to an increase in social presence [1]. Bente and his team have shown that varying the duration of gaze directed at agents has an impact on social presence [2]. In addition, various works, such as those of Casanueva and Blake [4] as well as Kang and his team [10] have shown that avatars with facial expressions allow for a higher feeling of social presence compared to static avatars. Krumhuber and his team [12] showed that changing facial dynamics significantly influences game participants’ choices and decisions to cooperate with other players.

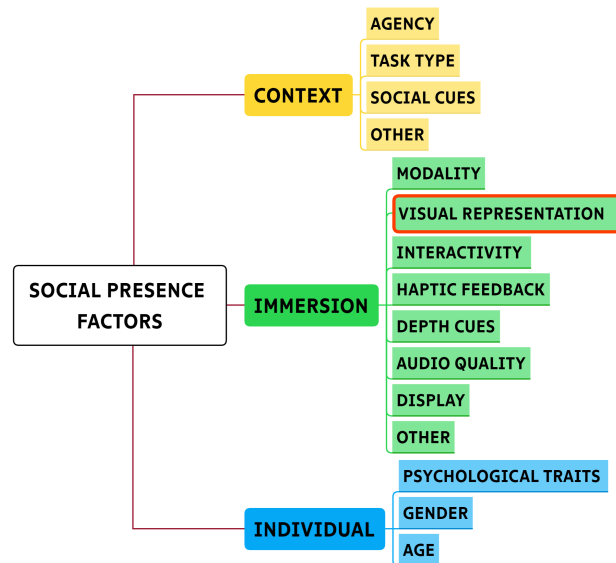


Fig. 1. Social presence factors divided into three groups: Context, Immersion and Individual [16].

3 DESIGN AND IMPLEMENTATION

3.1 Collaborative environment

In all the experiments mentioned above, 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 learning in VR the students will have to interact with the environment and the objects that compose it. Therefore, in our experiment we wanted an object manipulation phase. We chose 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. Only the participant can handle the pieces and assemble them.



Fig. 2. The collaborative virtual environment during the assembly phase of the furniture. The expert assistant gives instructions while the participant handles a piece of furniture

Our main hypotheses are:

- H1: Participants with facial expressions will build the piece of furniture faster (speed of execution) than those without facial expressions.
- H2 (a-b): Participants with facial expressions will experience more (a) satisfaction, (b) social presence compared to those without facial expressions.
- H3: Participants with facial expressions will spend more time looking at each other compared to those without facial expressions.

The virtual environment is based on a framework developed by B<>com¹, which is an Institute of Research and Technology that provides innovations to companies that want to increase their competitiveness through digital technology. This platform allows several users to simultaneously experience the immersion in VR in a shared environment. For our experiment, we had to develop some additional features such as 3D modeling of the furniture with "magnetism". This feature allows to stick one piece with another one when their relative positions are correct.

3.2 Avatar appearance

For the experiment, the objective was to display the facial expressions of each user on their avatars in real time. To collect behavioral data on the participants' faces, an HTC facial tracker sensor² and an eye tracker integrated into the HTC vive pro eye made it possible to recover 41 key points on the face (32 on the lower part and 9 on the upper part). Ready Player Me avatars³ were used as they integrate blendshapes, allowing the display of facial expressions. Each value of the blendshapes was defined in relation to the value of the different key points of the face (See Fig. 3).

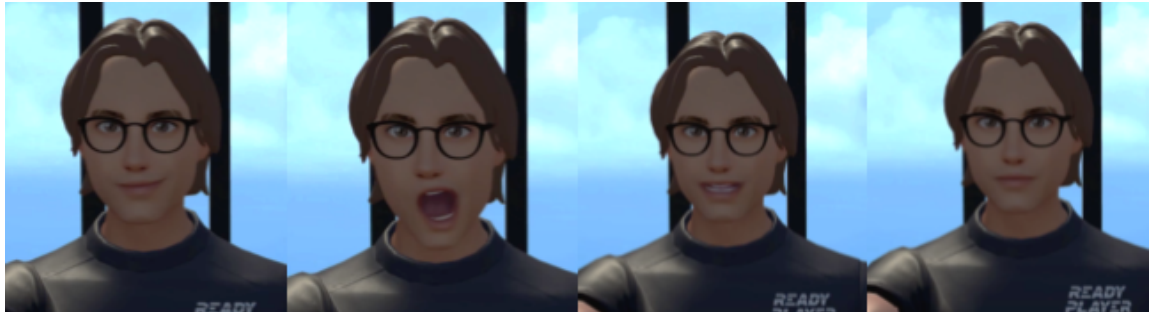


Fig. 3. Facial expressions on a Ready Player Me avatar

4 METHODS

4.1 Design

This prestudy was conducted according to a between-subjects design comparing the conditions “without facial expressions” and “with facial expressions”. The participants experienced the assembly of a TV stand with an expert assistant. The participant and the expert were located into two different rooms. The participants were ignoring the real purpose of the experiment while the expert was aware of it.

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

²VIVE Facial Tracker <https://www.vive.com/fr/accessory/facial-tracker>

³Metaverse Full-Body Online 3D Avatar Creator – Ready Player Me. <https://readyplayer.me/>.

4.2 Task

The participants were grouped as dyad with an expert assistant. They were asked to assemble a TV stand as quickly as possible. The participant manipulates the parts to put them together, while the expert assistant gives the instructions. Only the expert assistant has a visual on the assembled piece of furniture. The participant was instructed to stay in a demarcated area while the expert assistant remained in their original position.

4.3 Procedure

The participant gives informed consent and receives a random number from 1 to 2 to know which condition would be applied to ensure correct relationship measures (Without facial expressions, With facial expressions). Before starting the experiment we asked the participant to complete three questionnaires related to demography, virtual reality experience and cybersickness. The participant was informed about the upcoming task using a script-based spoken instruction. Then, we equipped the subject with a VR headset (HTC Vive Pro with Eye tracker with HTC Facial Tracker), two controllers (HTC Controller Vive Pro). Then, the participant has about 1 minute of acclimatization in a practice room, to become familiar with the use of the controllers. During this manipulation, an easy-to-use furniture (three pieces) can be manipulated. After that, the participant is taken to the virtual testing environment. We started data recording and experimentation with an oral "go". The dyad had 15 minutes to complete the TV stand. If it was over before, the experience ended there. The participants could raise their hand during the exhibition if they encountered technical problems. After the experience we asked the participant to complete the post-experimental questionnaires (social presence, satisfaction, and general comments on the experience). The overall experience lasted about 30 minutes. The experiment was conducted at the European Center for Virtual Reality⁴.

4.4 Apparatus

The virtual environment was implemented in Unity 3D using the Ngagement 1.0 framework, developed by b<>com, based on a server-client network architecture. Two HTC Vive Pro Eye tracker headsets with HTC Facial Tracker, allowing to recover the facial movements of the lower face, with two HTC Vive Pro controller joysticks allowing the control of the movement of the arms and hands. Two HTC Station VIVE Pro 2.0 bases were used for the experiment.

4.5 Measures

4.5.1 Subjective Measures. We measured Social Presence [15] (likert scale with 1=strongly disagree, 5=strongly agree) and teamwork satisfaction [20] (likert scale with 1=strongly disagree, 5=strongly agree) to test the impact of facial expressions (H2). In addition, participants had the opportunity to add qualitative comments regarding the experience.

4.5.2 Objective Measures. We collected two types of objective data. The first concerns task performance. There was no possibility of making a construction error because of the magnetism of the furniture parts, we only recovered the total assembly time of the furniture ("Building time" in seconds) (H1). The second 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 he is looking at. From these data, we calculated the "Time gaze expert" in percentage. It corresponds to the time to have looked at the expert divided by the total time of the experiment (H3)

⁴European Center for Virtual Reality <https://cerv.enib.fr>

4.5.3 Control Measures. To assess whether other factors influence subjective and objective measures, we administered a demographic questionnaire and the Simulator Sickness Questionnaire [11] before and immediately after the task.

5 RESULT

Eight participants took the experiment. We removed two of them from the analysis due to technical problems during the experiment. The sample was composed of six participants (two women and four men). The conditions were assigned randomly. The average age was 24.16 years. Three participants were employees and three were students.

The first results showed that the participants are not looking at the expert assistant while manipulating the piece of furniture. This could be verified by the behavioral data of the participants' gaze. Indeed, the gaze time on the face of the expert's avatar was 3.06% of the total time. On average, the furniture assembly time was 7min34sec in the condition without facial expressions and 7min04sec with facial expressions (the difference is not significant). Results of the objective measures are presented in Fig. 4.

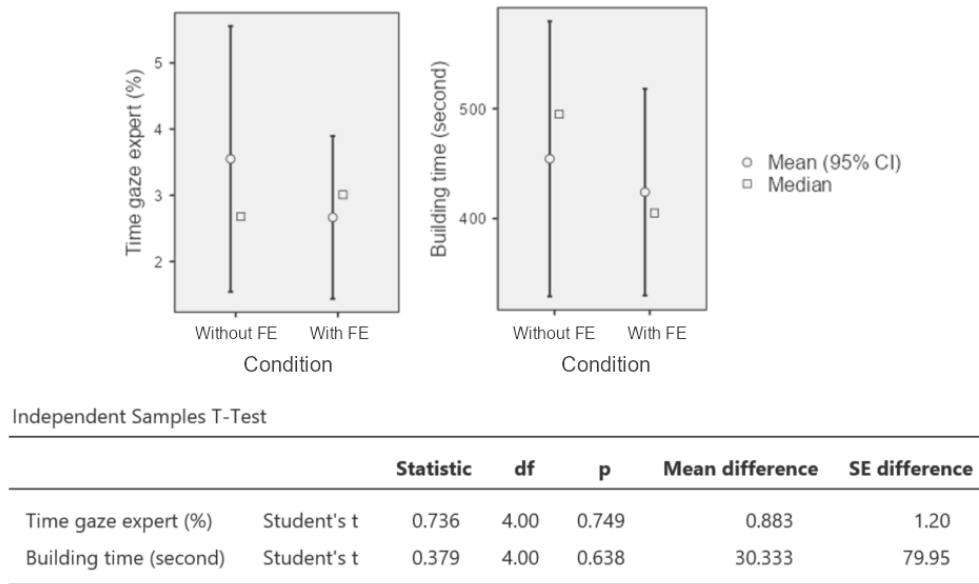


Fig. 4. Results from the objective measures. Descriptive plots show the mean and the median of the Time gaze expert and the Building time for each condition (Without FE : Without Facial Expressions , With FE : With Facial Expressions)

6 DISCUSSION AND LIMITATION

The first results showed that the participants were so focused on the pieces of furniture, that they did not look at the expert assistant. Additionally, this may also be due to the expert assistant avatar being located a bit far from the participant avatars relative to the resolution of the headset. For this reason, we decided to stop the prestudy. To address the problem of the gaze not directed towards the head of the avatar of the expert assistant, we are planning to add a

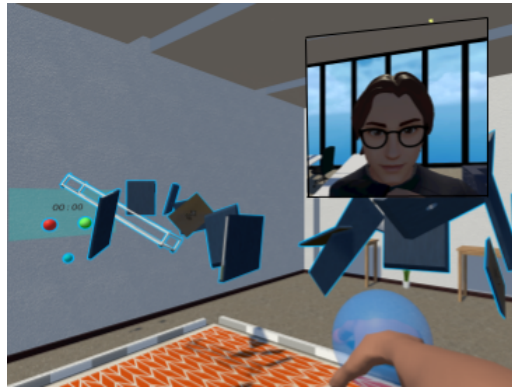


Fig. 5. Collaborative virtual environment during the furniture assembly phase with a screen displaying the face of the expert assistant

screen in the top right corner of the participant’s field of vision where the face of the expert assistant is displayed (see Figure 5).

7 CONCLUSION AND FUTURE WORK

In this paper, we presented the design, implementation and a beginning of evaluation of a concept for the improvement of collaborative learning in Virtual Reality. 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 improves collaborative learning in terms of performance and in terms of satisfaction. The results showed that the participants did not look at the face of the assistant expert but spent the majority of the time looking at the pieces of furniture. 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”. This is surely due to the fact that the participants were not looking at the expert. To address this issue, we are planning to do a new experiment with a screen in the top right corner of the participant’s field of vision where the face of the expert assistant is displayed so that it is visible in all circumstances.

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.

REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.

- [5] 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.
- [6] Paul Ekman. 1993. Facial expression and emotion. *American psychologist* 48, 4 (1993), 384.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] Marjan Laal and Seyed Mohammad Ghodsi. 2012. Benefits of collaborative learning. *Procedia-social and behavioral sciences* 31 (2012), 486–490.
- [14] Elinda Ai-Lim Lee and Kok Wai Wong. 2008. A review of using virtual reality for learning. *Transactions on edutainment I* (2008), 231–241.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] Candace L Sidner. 2016. Engagement, emotions, and relationships: on building intelligent agents. In *Emotions, Technology, Design, and Learning*. Elsevier, 273–294.
- [19] 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.
- [20] Hungwei Tseng, C Wang, H Ku, and Ling Sun. 2009. Key factors in online collaboration and their relationship to teamwork satisfaction. *Quarterly Review of Distance Education* 10, 2 (2009), 195–206.