Impact of Multimodal Instructions for Tool Manipulation Skills on Performance and User Experience in an Immersive Environment

Cassandre Simon, Manel Boukli Hacene, Flavien Lebrun, Samir Otmane, Amine Chellali

Impact of Multimodal Instructions for Tool Manipulation Skills on Performance and User Experience in an Immersive Environment

Cassandre Simon*
IBISC Lab
Univ Evry Paris Saclay

Manel Boukli-Hacene†
IBISC Lab
Univ Evry Paris Saclay

Flavien Lebrun‡
IBISC Lab
Univ Evry Paris Saclay

Samir Otmane§
IBISC Lab
Univ Evry Paris Saclay

Amine Chellali¶
IBISC Lab
Univ Evry Paris Saclay

Figure 1: Illustration of the user study. A remote instructor transmits instructions to a local learner regarding the amplitude of a movement to be performed using four different combinations of modalities: verbal-visual, verbal-haptic, visual-haptic, and verbal-visual-haptic. Then the local learner attempts to replicate the requested movement as faithfully as possible.

Abstract

With the mentoring model, a mentee can learn technical skills under the supervision of more experienced peers who demonstrate their knowledge through several communication modalities. Supporting the mentoring model within shared immersive training simulators holds promise in enhancing mentor-mentee interactions and learning outcomes in a safe environment. However, efficient communication within these spaces remains an open issue. This work presents a user study that explores the combination of communication modalities (verbal-visual, verbal-haptic, visual-haptic, and verbal-visual-haptic) to convey instructions to learners on the amplitude of movements to perform during a tool-handling task in an immersive environment. The study aims to examine the impact of the four modality combinations on performance (speed and accuracy of movement replication), mental workload, and participants' user experience. The results show that participants achieved higher accuracy with the visual-haptic and verbal-visual-haptic conditions. Moreover, they performed the movements faster, and their movement trajectories were closer to the reference trajectories in the visual-haptic condition. Finally, the most preferred verbal-visual-haptic combination enhanced the users' sense of presence, co-presence, social presence, and learning experience. No impact on the mental workload was observed. These results suggest that combining haptic and visual modalities is the best suited for enhancing learners' performance. Adding the verbal modality can also improve the user experience in the immersive learning environment. These findings contribute to improving the design of immersive collaborative systems and pave the way for exploring novel avenues of research into the efficacy of multimodal communication for enhancing the mentoring-based acquisition of technical skills in VR. These tools hold promise for diverse applications, including medical simulation.

Keywords: Multimodal interactions, Mentorship, Remote collaboration, Immersive learning

Index Terms: Human-centered computing Virtual reality—Human-centered computing User studies—Human-centered computing Collaborative interaction—;

1 Introduction

New simulation methods, such as virtual reality (VR), have recently shown their efficacy in teaching technical skills, particularly in medicine [7,58]. Nevertheless, most existing simulators only allow autonomous learning with the trainees practicing the skills independently. This eliminates teachers' guidance and feedback, which has been shown to be necessary during the early stages of technical skills acquisition [1,31,46]. Several studies suggest that instructors play a crucial role in aiding learners in developing and sustainably improving their technical skills while reducing cognitive load [1,5,37]. To convey instructions or provide feedback to the learners, the instructors employ a variety of communication modalities, including verbal and non-verbal cues [15,53]. However, allowing the instructor to communicate efficiently with a learner in an immersive virtual

*e-mail: cassandre.simon@univ-evry.fr
†e-mail: manel.boukli-hacene@univ-evry.fr
‡e-mail: flavien.lebrun@univ-evry.fr
§e-mail: samir.otmane@univ-evry.fr
¶e-mail: amine.chellali@univ-evry.fr

environment remains an open issue.

This article explores the design of collaborative learning interactions and interfaces. These tools are intended to enable instructors to showcase their skills and guide learners in an immersive setting to enhance the technical skills transfer in a safe and controlled environment. To design such systems and suitable interaction techniques, it is essential to understand how experts transmit technical skills to learners and the impact of interaction modalities on this process.

This work is centered on training instrument-handling tasks in a collaborative virtual environment (CVE). Specifically, we investigate how a teacher can effectively assist learners in achieving the correct amplitude of tool movements using multimodal instructions (Fig. 1). Previous studies have explored the impact of different instructor-learner interactions using single communication modalities (verbal, visual, and haptic) on the learner's performance and learning experience [49], showing that each single modality has its strengths and limitations. This work builds on this previous research and aims to investigate the impact of combining communication modalities on performance and user experience. Our central hypothesis is that multimodal communication would improve the learner's performance and learning experience when receiving instructions in the CVE. The main research question is determining which combination of modalities best suits this context.

The principal contributions of this study are to provide insights into the roles played by each combination of modalities in the instructor-learner communication process and to extract guidelines for designing immersive learning environments that support multimodal interactions.

2 Related Work

2.1 The mentoring model and communication modalities

In the medical field, mentoring is a recognized model that combines theoretical lessons, observations, and practice under peer supervision, often involving skill transmission through demonstration [13, 22, 52]. This model fosters collaboration and interactions between instructors and learners. To learn motor skills, novices often begin with the observation step, in which the instructor demonstrates each part of the procedure. Goffman defines demonstration as the "performance of a task-like activity out of its usual functional context to allow someone who is not the performer to obtain a close picture of the doing activity" [20]. The demonstration is considered an effective learning method for a new complex motor skill [14]. To demonstrate the skill, the instructor will use instructions or feedback as a means of interaction. It is important here to distinguish between instructions, which are defined as guidelines provided by an expert whose aim is to help the learner integrate the necessary skills before performing the skill, and the concept of (augmented) feedback, which is defined as the information provided to learners to guide and correct their actions while or after they perform the skill [5, 24]. The literature often confuses these two concepts because both involve learner-expert interactions. However, as they occur at different stages of task completion, their impact on learning may differ. Our study will focus only on instructions.

To provide instructions, the teacher will use a variety of communication modalities: verbal, visual, and haptic [18, 38]. The verbal modality offers explicit information on task requirements, movement goals, and execution strategies and boosts motivation, self-efficacy, and engagement. However, verbal instructions may fall short due to motor skills involving elusive elements like movement dynamics and haptic sensations, making them challenging to articulate verbally [42, 47]. Conversely, the visual modality, involving imitation following physical demonstrations, is widely used in sports and medicine, empowering better communication of expertise during action execution [16, 47]. Yet, visual demonstrations lack internal process details, such as muscle activation patterns, joints, angles, or forces. Haptic communication involves physically guiding learners to perform the movement. However, its effectiveness varies based on individual perception, sensitivity, and adaptation, potentially impacting interpretation and performance [8, 9].

To summarize, mentoring is an effective model for teaching technical skills and enhancing the learning experience. The model is based on a "one-to-one" interaction and employs various communication modalities, enabling the teacher to offer personalized instructions and feedback to the learner. Consequently, when designing systems for teacher-learner interactions, it is crucial to account for each modality's specific characteristics. This study examines the interaction between instructors and learners in VR when teachers use the demonstration process to provide instructions. Our aim is to provide design recommendations by analyzing how the combination of communication modalities contributes to the mentoring process.

2.2 Interaction modalities in Collaborative Virtual Environments

CVEs are 3D spaces that enable multiple users to interact and work together even if they are at a distance [11]. These environments allow users to collaborate synchronously or asynchronously [11], to interact with virtual objects or artifacts [26], or to share knowledge and skills [19]. They can offer virtual platforms that foster participant collaboration for meetings, learning, or other group activities. They also provide the possibility of integrating several communication modalities. These environments are thus an appropriate means of facilitating instructor-learner interactions in an immersive setup. However, despite the promising advantages of CVE for supporting instructor-learner interactions, it is crucial to consider their current characteristics when designing VR learning systems. Indeed, they still have limitations, particularly regarding communication between users [12]. For example, it is sometimes difficult to faithfully reproduce partners' facial expressions, body movements, or gaze direction, which can hinder communication and collaborative interaction. In addition, partners do not necessarily share the same point of view and, therefore, are not necessarily aware of the ongoing activities of other participants [10]. These limitations can hinder the transmission of skills in the context of teaching interactions and must, therefore, be considered when designing related systems.

Lately, there has been a growing emphasis on the design of interactions and communication modes and their integration within CVE. Indeed, studies have shown that combining modalities can improve communication and collaboration in CVE [27, 34]. For example, verbal and visual modalities combined increase communication [56] and collaboration [21]. Other studies show that combined visual and haptic modalities can enhance performance [6, 53] and increase the feeling of presence [45]. Moreover, completing a collaborative task was faster with the combination of verbal and visual modalities and verbal and haptic modalities than with the combination of visual and haptic modalities [33]. In addition, combining the three modalities (haptic, verbal, and visual) leads to a better performance in collaborative tasks than using one modality or a combination of two modalities [28, 33].

The previous review indicates that multimodal interactions can enhance communication and collaboration in CVE. Nevertheless, the number of studies in this area of research is limited, particularly regarding the impact of combining communication modalities on the learning of technical skills [47].

2.3 Technologies to support mentor-mentee interactions

The use of CVEs and XR technologies for mentoring and facilitating the learning of new skills is gaining increasing attention as studies emphasize the need for interaction between mentors and mentees in CVEs to achieve favorable learning outcomes [3, 37, 48]. From this perspective, a study examined the influence of instructor feedback inside a laparoscopic VR simulator training [51]. In this study, participants received verbal and visual feedback from the instructor.

The findings indicate that participants who received feedback from an instructor during their training achieved proficiency in the task more quickly than those who did not receive feedback. Another study aimed to investigate whether verbal feedback provided by an expert is more efficient than self-generated feedback in assessing the effectiveness of movements when learning new surgical skills. The results show that verbal feedback from an expert led to lasting improvements in technical skill performance [40]. Another work investigated the effects of different feedback sources on acquiring and retaining a complex medical skill. The results show that experts' presence helped reduce cognitive load during practice [5]. In another study, the haptic modality was combined with visual and verbal modalities to instruct mentees in performing a biopsy procedure [9]. The findings indicate that incorporating haptic feedback led to enhanced performance among trainees compared to conditions involving only visual and verbal instructions. Similarly, Lu et al. [29] study shows that combining haptic and visual feedback improves task performance compared to visual feedback alone. These results indicate that when the benefits of each modality are effectively utilized, multimodal teacher-learner interactions promote the acquisition of complex tasks. Another study investigated the effects of different instruction modalities (visual, haptic, and verbal) on teaching tool manipulation skills [49]. The results showed that task completion times were significantly faster with haptic instructions, but the learners were more accurate with the visual instructions. In addition, the verbal modality increased the sense of copresence with the teacher [49].

The previous studies suggest that each modality has strengths and limitations and that multimodal communication improves collaboration and user experience in CVEs. However, incorporating VR and multimodal approaches to teaching and learning [39], particularly for developing technical skills, remains an area to be explored [30, 39, 47].

In summary, studies and theories on multimodal interactions indicate that combining multiple modes of communication can promote collaboration and the acquisition of motor skills. However, using these modalities depends on the skills to be taught [9] and the availability of these modalities [44]. Therefore, further investigation is required to comprehensively understand how each modality influences the acquisition of technical skills. While CVEs hold promise in enhancing teacher-learner interactions and mentoring approaches, it is crucial to meticulously assess their characteristics to prevent communication breakdowns and learning challenges. The presented work aims to contribute to this discussion by providing insights into utilizing a combination of modalities (visual-verbal, verbal-haptic, haptic-visual, and visual-verbal-haptic) within CVEs to instruct learners in effectively manipulating tools with the appropriate movement amplitude.

3 User Study

3.1 Study objectives and hypotheses

Our study investigates the impact of multimodal communication between an instructor and a learner in a CVE. We assess the impact of different modality combinations for instruction delivery during a manipulation task on the learners' performance and user experience. The used modalities include verbal, visual, and haptic modalities, leading to four different bimodal and three-modal combinations: verbal-visual, verbal-haptic, visual-haptic, and verbal-visual-haptic. The task involves replicating a tool movement conveyed by an instructor using one of these combinations. The differences between modalities are assessed by their impact on the participants' performance (speed and accuracy), workload, and subjective experience.

Our general hypothesis is that the four communication conditions will affect participants' performance, cognitive load, and subjective learning experience differently. More particularly, we expect that:

- **H1.** As suggested by previous studies [9, 29, 49], Conditions combining the haptic and visual modalities (i.e., visual-haptic and verbal-visual-haptic conditions) would improve the learners' performance compared to conditions where only one of them is combined with the verbal modality (i.e., verbal-visual and verbal-haptic conditions).
- **H2.** As suggested by previous research [47], the mental workload would decrease when instructions are provided in a multimodal way. Therefore, it is expected that the three-modal condition would decrease the perceived workload compared to the bimodal conditions.
- **H3.** Previous studies indicate that verbal communication is important to increase the sense of copresence and social presence [49]. Therefore, it is expected that conditions including the verbal modality would increase the user experience as measured using presence, social presence, copresence, and learning experience questionnaires.

3.2 Participants

A total of 32 participants took part in this study, including 6 females and 26 males, recruited among students, university staff, and external participants. No specific expertise criteria or prior experience with VR simulators were required. The participants' average age was 25.5 years (min = 20, max = 47), and all were right-handed. All participants had normal or corrected-to-normal vision, with 13 wearing corrective glasses during the experiment. Twenty-six of them had previous experience with VR headsets, including six regular users (using them once a week). Half of the participants used haptic devices in previous demonstrations or studies. The Research Ethics Committee (CER) of Université Paris-Saclay validated the experimental protocol. The study complied with the requisite ethical standards; all participants provided informed written consent before participating.

3.3 Experimental task

We defined a specific task to examine how the four communication modes impact the learners' ability to reproduce desired movements. This task consists of reproducing a tool manipulation by a learner following handling movement demonstrations by an instructor. It replicates pick and place tasks regularly used for training motor skills in VR [4, 43, 49]. To simplify the task, our focus was solely on the direction and amplitude of 2D movements. This simplification was necessary to limit bias related to a complex task whose instructions would have been challenging to provide using the three modalities. Consequently, the task involved moving a 3D sphere from its initial position to a final position along a single axis (either the X or Y axis). The experiment was structured into two distinct phases:

1. **Instruction Phase:** During this phase, the instructor's role was to provide guidance on the direction and amplitude of the movement to the participant. This was done under one of the four communication conditions. Participants were informed that the instructions were being delivered in real-time by a remote instructor located in another room, although, in reality, the instructions were pre-recorded. This approach helped control potential biases arising from differences in instructions received by each participant [49]. The choice to conceal the absence of the real instructor is justified by previous research indicating that the extent to which a virtual entity is perceived as being controlled by an actual person rather than by a computer (perceived agency) influences user experience and social presence in a VE [17, 36].
2. **Manipulation Phase:** In this phase, participants were tasked with independently replicating the demonstrated movements as quickly and as accurately as possible.

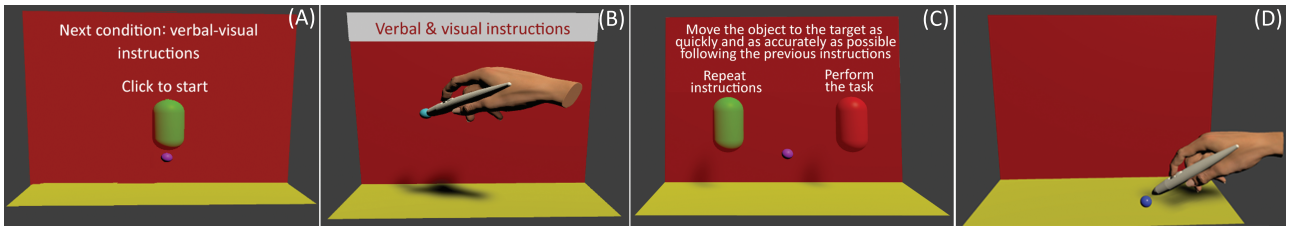

Figure 2: The virtual scenes: (A) The initial scene indicates the modality combination to be used in the next condition. (B) The instruction scene, where participants saw the sphere and the instructor's virtual hand (only for visual instructions). (C) The transition scene allows participants to repeat the previous instructions or to move to the manipulation phase. (D) The manipulation scene allows participants to perform the task.

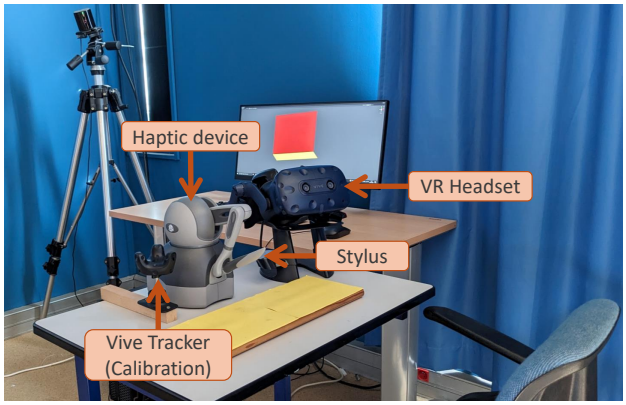

Figure 3: The Experimental apparatus included a VR headset, a haptic device (equipped with a stylus), and a Vive Tracker for calibration.

3.4 Apparatus and virtual environment

3.4.1 Physical Setup

The equipment (Fig. 3) includes a Vive Pro HMD to visualize the immersive VE. It has a resolution of 2880 × 1600 pixels in total, with a refresh rate of 90 Hz and a field of view of 110 degrees. Additionally, a Geomagic Touch haptic device, with a workspace of 160 mm x120 mm x70 mm, was utilized to manipulate the virtual objects and to receive the haptic instructions. The terminal component of this interface is a stylus featuring a button that participants press to grasp and release the virtual sphere. One Vive 3.0 tracker (attached to the table) was also used to calibrate the system and ensure the correct matching between the physical and virtual worlds. Indeed, the user's viewpoint and the positions of the haptic device and the table were tracked using the Vive's Lighthouse system.

3.4.2 Virtual Setup

The VE was created using Unity 3D with C# (version 2020.3.15f2) and the SteamVR plugin (version 1.17.3). The VE (Fig. 2) displays two perpendicular planes, measuring 44 cm x 15 cm, colored in yellow (horizontal) and red (vertical), respectively. During the instruction phase, these two planes were displayed for all communication conditions except the verbal-haptic condition (where a neutral grey screen was displayed). The environment also included a virtual tool, a 3D blue sphere, and a representation of the instructor's hand (Fig. 2-(B)). During the manipulation phase, the VE also included a blue sphere to be moved, a tool controlled by the participant via the haptic arm, and a virtual representation of their hand (Fig. 2-(D)). To enhance the perception of distance in the immersive environment, the virtual hand was attached to the tool and positioned to align with the participant's real hand; however, it was not animated. In addition, a wooden board colored yellow, with the same size as the virtual yellow horizontal plane, was placed on the table in front of the participants (Fig. 3). The virtual tool matched the color, size, and shape of the haptic arm's stylus and was positioned to align with it. Furthermore, the haptic device was placed so that when the participant touched the virtual plane with the tip of the virtual tool, they felt the collision between the wooden board and the haptic arm's stylus. Therefore, a calibration step was always required before the start of each experiment.

3.4.3 Virtual scenes and interactions

A series of virtual scenes were presented to the participants according to the experimental phase. The initial scene (Fig. 2-(A)) displayed a message explaining how instructions would be given to participants using the current communication condition. The scene also included a single 3D capsule. Participants had to put the 3D pointer inside this capsule using the haptic arm and press the device's button to start the instruction phase. In the instruction scene (Fig. 2-(B)), Only the two-colored planes were displayed (except for the verbal-haptic condition), along with a timing message to alert participants that instructions would be provided shortly. After each instruction, a transition scene (Fig. 2-(C)) with two 3D capsules was displayed, prompting participants to repeat the instruction or proceed to the manipulation phase. In the manipulation scene (Fig. 2-(D)), participants had to use the haptic arm to grasp the sphere (by pressing the stylus button) and move it according to the received instruction to the desired position. They had to press the stylus button again to release the sphere, indicating the end of the trial. A new scene then allowed them to move on to the subsequent trial. At the end of all trials in a condition, a transition scene appeared, instructing them to remove the headset to respond to the questionnaires.

3.5 Experimental design and conditions

The experiment followed a within-subjects design involving a single factor (the combination of communication modalities) with four conditions: verbal-visual, verbal-haptic, visual-haptic, and verbal-visual-haptic. The presentation order of conditions to participants was counterbalanced using a Latin square to mitigate potential learning effects. Each participant performed 14 trials for each condition, with movements on both the X and Y axes (left/right; up/down). To ensure that movements remained within the haptic arm's workspace limits, movement amplitudes varied between 4 and 8 cm and were randomly selected for each trial while being balanced across conditions. This resulted in a total of 1792 recorded trials (14 trials x 4 conditions x 32 participants). Three forms of instruction were used:

- **Verbal Instructions**: Instructions describing the movement's amplitude and direction were told verbally to participants. For example, the instruction for a movement of amplitude 8 cm to the left was: "Please move the sphere eight centimeters to the left". All the instructions were prerecorded on audio clips using the experimenter's voice and displayed on the headphones of the HMD.

- **Visual Instructions**: Instructions were conveyed through a 3D animation in the immersive environment representing the instructor's hand manipulating the tool. Following a Bezier curve, the virtual hand moved to grasp the sphere and place it in a varying final position for each trial.
- **Haptic Instructions**: Instructions were conveyed through the haptic arm grasped by the participants. The haptic arm moved from the starting to the final position, following the same curves as the visual condition, mimicking the instructor's task execution using the haptic device.

These three communication forms were then combined to build the four experimental conditions where the instructions were played simultaneously, conveying the same direction and amplitude:

- Condition 1: Verbal-Visual,
- Condition 2: Verbal-Haptic,
- Condition 3: Visual-Haptic,
- Condition 4: Verbal-Visual-Haptic.

It is to be noted that during the haptic and visual instructions, the 3D sphere and the haptic stylus's position were placed in different starting positions during the instruction and manipulation phases to prevent memorization of final positions. This way, participants were encouraged to memorize the movement's amplitude and direction rather than the sphere's or stylus's final position.

3.6 Experimental Procedure

The average experiment duration was 75 minutes per participant. Fig. 4 details the experimental procedure. The first step was introducing the study's objective to the participants and the equipment they would use. Next, they were asked to read and sign a consent form to participate in the study. Before entering the experimental room, they met the "false" instructor sitting in an adjacent room (Fig. 1) and were told he would give them the instructions remotely. Then, when arriving at the experimental room, they were given an instruction sheet detailing how the prototype functions, the actions to be performed, and what was expected of them. The subsequent step was to complete a demographic questionnaire, after which participants were asked to put on the immersive headset, to begin with the first condition. Once the 14 trials of the condition were completed, participants were instructed to remove the headset to respond to the NASA - TLX questionnaire and the questionnaire evaluating their sense of presence, social presence, co-presence, and learning experience during the previous condition. Once this step was completed, they were required to put on the headset again to start the trials of the following condition and repeat the same cycle described earlier. After completing the trials and answering all questionnaires for all conditions, they were asked to respond to a questionnaire comparing the four conditions based on various criteria.

3.7 Measurements and data analyses

Both objective and subjective measurements were used in this experiment. The objective measurements assessed the participants' ability to replicate the requested movement and included three metrics: the average manipulation time for all trials, the mean distance estimation error, and the quality of tool trajectory. The manipulation time calculation for each trial started when the participants picked up the sphere and ended when they placed it in the final position. Lower values indicate better performance. The distance estimation error is calculated as the average Euclidean distance, in centimeters, between the final position of the sphere's center and the desired position (based on the instructed amplitude) for all trials. Lower values indicate better performance. Finally, we assessed the participants' tool trajectories. In our study, we employed Dynamic Time Warping (DTW) [57] to assess the similarity between the trajectories demonstrated by the instructor and those performed by the participants. DTW is an algorithm designed to identify the optimal alignment between two trajectories. It proves to be more successful in finding trajectories' similarity than conventional methods such as measuring point-to-point Euclidean distance. The lower the DTW distance, the more similar the two trajectories are. We thus computed the DTW distance (using Python and the DTAIDistance [1] module) between participant trajectory and their respective reference trajectory for every participant and every trial for each condition. We then compared the mean distances of each condition.

The subjective measurements consisted of participants' responses to various questionnaires. These included a seven-point Likert scale questionnaire assessing the sense of presence, social presence, co-presence, and the participants' learning experience (Table 1). The questions were extracted from questionnaires used in peer-reviewed international publications and adapted for our study. The sense of presence was measured using five questions (Q1-Q5) from the questionnaire of Nowak and Biocca [35]. The sense of social presence was measured using two questions (Q6-Q7) also from the questionnaire of Nowak and Biocca [35]. The sense of copresence was measured using eight questions. Three of them (Q8-Q10) were extracted from the questionnaire of Nowak and Biocca [35], and five (Q11-Q15) from the questionnaire of Basogan et al. [2]. Finally, the learning experience was measured using one question from the questionnaire of Simon et al. [49]. The NASA-TLX [23] assessed the participants' mental workload while performing the task. Finally, the participants responded to a comparison questionnaire requiring them to rank the combination of modalities according to the same eleven classification criteria used in the study of Simon et al. [49], from the most preferred to the least preferred.

The SPSS software (IBM Corp., Armonk, NY, USA) was used for data analyses, employing the relevant statistical tests. Our analyses were performed with a confidence level of 95%, and in cases where corrections were applied, we have reported the adjusted p-values.

4 Results

4.1 Objective measurements

For each participant, we averaged each measure per condition, leading to 128 (4 conditions, 32 participants) values per measure.

4.1.1 Normality tests

The Shapiro-Wilk tests were used to check the normal distribution of data. The results indicate that all the distance estimation errors and the completion times data were normally distributed. In contrast, only the DTW distance data for Visual-Haptic and Verbal-Visual-Haptic conditions were normally distributed.

4.1.2 Distance estimation error

Following the normality test results, a one-way repeated measure ANOVA was used to test the effect of modality combination on the mean distance estimation errors. The results (sphericity assumed, $p = .21$) show a significant main effect of modality combination on distance estimation errors ($F_{(3,93)} = 10.56$, $p < .001$, partial $\eta^2 = .25$; Fig. 5). The post-hoc pairwise comparisons with Bonferroni correction show that the mean distance estimation errors were significantly lower in the Verbal-Visual-Haptic condition compared to the Verbal-Visual (p = .001) and Verbal-Haptic (p = .006) conditions. The mean distance estimation errors were also significantly lower in the Visual-Haptic condition compared to the Verbal-Visual (p = .008) and Verbal-Haptic (p = .002) conditions. No other significant differences are observed.

[1] https://pypi.org/project/dtaidistance/

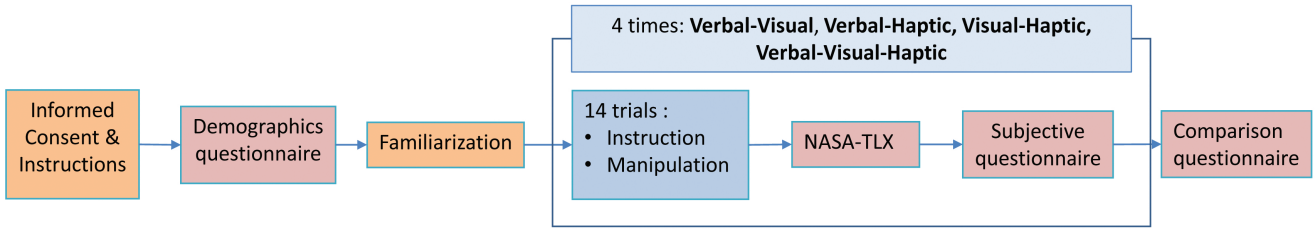

Figure 4: Detailed experimental procedure.

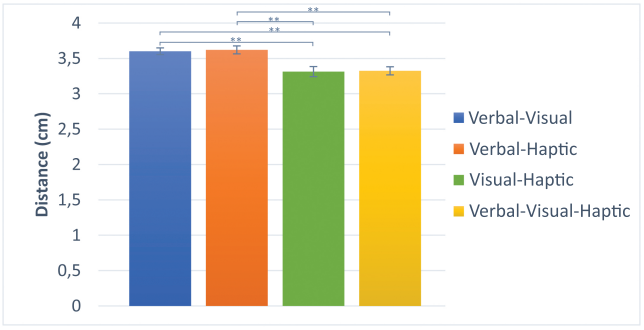

Figure 5: Average amplitude estimation errors for each modality combination (error bars represent the standard error; ∗∗ = p < .01)

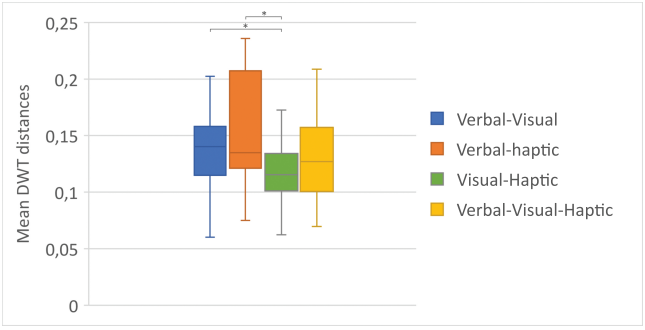

Figure 7: Average DTW distance for each modality combination (error bars represent the standard error; ∗ = p < .05)

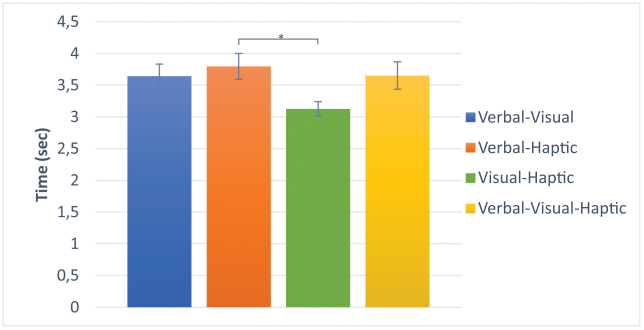

Figure 6: Average manipulation times for each modality combination (error bars represent the standard error; ∗ = p < .05)

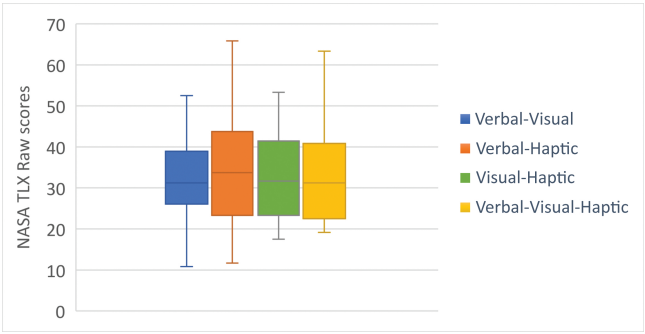

Figure 8: NASA-TLX (raw scores)

4.1.3 Manipulation times

Following the data normality test results, a one-way repeated measure ANOVA was used to test the effect of the modality combination on the mean completion times. The results (sphericity assumed, $p = .21$) show a significant main effect of the modality-combination on the mean completion times ($F_{(3,93)} = 3.50$, $p < .019$, partial $\eta^2 = .10$; Fig. 6). The post-hoc pairwise comparisons with Bonferroni correction show that the mean completion times were significantly lower in the Visual-Haptic condition than the Verbal-Haptic condition (p = .017). No other significant differences are observed.

4.1.4 Trajectories

Following the normality test results, a Friedman non-parametric test was used to test for the effect of the modality combination. The results show a main effect of the modality combination on the mean DWT distances between the participants' trajectories and the instructor's trajectories ($\chi^2 = 7.91$, p = .04; Fig. 7). The pairwise Wilcoxon signed-rank tests with Bonferroni correction show that the participants' trajectories were significantly closer to the instructor's trajectories in the Visual-Haptic condition than in the Verbal-Visual (p = .02) and in the Verbal-Haptic (p = .01) conditions. No other significant differences were observed.

4.2 Subjective measurements

4.2.1 Perceived workload

The Friedman test indicates no significant main effect of the modality combination on the raw NASA-TLX scores ($\chi^2 = 3.94$, p = .26; Fig. 8). The test also shows no significant main effect of the modality combination on any of the sub-scales of the NASA-TLX (mental demands, physical demands, temporal demand, own performance, effort, and frustration).

4.2.2 Subjective questionnaire

For the subjective questionnaire, each dimension was investigated separately by calculating the mean score of the related questions. Then, Friedman tests were used to study the main effect of modality combinations on these mean scores. The results of the Friedman tests are reported in Table 2, and data are shown in Fig. 9.

Table 1: Items of the subjective questionnaire on presence, social presence, copresence and learning experience. Rating scales range from 1 to 7.

Q#	Question Text
	Presence
Q1	How involving was the experience?
Q2	How intense was the experience?
Q3	To what extent did you feel like you were inside the environment you saw/heard?
Q4	To what extent did you feel immersed in the environment you saw/heard?
Q5	To what extent did you feel surrounded by the environment you saw/heard?
	Social presence
Q6	To what extent was this like a face-to-face meeting?
Q7	To what extent was this like you were in the same room with the instructor?
	Copresence
Q8	The instructor was intensely involved in our interaction.
Q9	The instructor communicated coldness rather than warmth
Q10	To what extent did you feel isolated from the instructor in the VE?
Q11	To what extent did you have a sense of being with the other person?
Q12	To what extent were there times during which the computer interface seemed to vanish, and you were directly working with the instructor?
Q13	To what extent did you forget about the instructor, and concentrate only on doing the task as if you were the only one involved?
Q14	To what extent were you and the instructor in harmony during the course of the performance of the task?
Q15	Overall rate the degree to which you had a sense that there was an instructor interacting with you, rather than just a machine?
	Learning Experience
Q16	To what extent do you think you can learn new skills in this application?

The pairwise Wilcoxon signed-rank tests with Bonferroni correction for each dimension are reported hereafter. The participants' sense of presence was significantly higher in the Verbal-Visual-Haptic condition than in the Verbal-Haptic (p = .01). The mean score was also higher than that of the Verbal-Visual (p = .07) and the Visual-Haptic (p = .07) conditions, but the effect was marginal. No other significant differences were found.

The participants' sense of social presence was significantly higher in the Verbal-Visual-Haptic condition than in the Verbal-Visual condition (p = .03). No other significant differences were found.

The participants' sense of copresence was significantly higher in the Verbal-Visual-Haptic condition than in the Verbal-Visual (p = .03) and the Visual-Haptic (p = .006) conditions. The mean score was also higher in the Verbal-Haptic than in the Visual-Haptic (p = .06) condition, but the effect was marginal. No other significant differences were found.

Finally, the participants felt that they had a better learning experience in the Verbal-Visual-Haptic condition than in the Verbal-Visual condition (p = .04). No other significant differences were found.

Table 2: Friedman tests for the subjective measurements

Dimension	χ^2	P-values
Presence	15.33	.002
Sociale presence	10.96	.012
Copresence	13.97	.003
Learning experience	11.63	.009

4.2.3 Subjective comparison

For the subjective comparison, each dimension was investigated separately by comparing the ranking of each condition to the others. The Friedman tests were used to assess the effect of modality combinations on these rankings for each dimension. The results of these tests are reported in Table 3 and data shown in Fig. 10.

The significant pairwise Wilcoxon signed-rank comparisons with Bonferroni correction for each dimension are reported hereafter.

The results show that participants ranked the Verbal-Visual-Haptic condition as the easiest to understand instructions (Q1), the most appropriate (Q2), the most accurate (Q3), and the most pleasant (Q4) to receive instructions as compared to the other conditions ($p<.01$). In addition, it was ranked as the easiest to memorize movements (Q6), the easiest to replicate the movement (Q7), the most educational (Q8), and the most engaging (Q9) communication form as compared to the other conditions ($p<.01$). Finally, it was ranked as the most efficient to receive instructions (Q10) and was ranked overall as the most preferred communication method (Q11) as compared to the other conditions ($p<.01$). On the other hand, it was ranked significantly less disturbing than the Visual-Haptic condition (p = .03) and marginally less disturbing than the Verbal-Haptic condition (p = .09). No other significant differences were observed.

5 Discussion

This research investigated how various combinations of modalities impact the communication of movement amplitude when instructing tool manipulation in CVE. This study gives numerous significant results.

5.1 Performance

The distance estimation error was the primary factor in evaluating the learner's performance after receiving the instructions. The results show that the visual-haptic and the verbal-visual-haptic combinations are the most accurate in conveying movement amplitude, resulting in a substantial decrease in errors when estimating amplitude. In light of the findings from previous work [49], the visual modality is the most accurate for communicating movement amplitude. Following this, the haptic modality emerges as the next most accurate, with the verbal modality being the least accurate of the three. Thus, it is unsurprising that the visual-haptic combination was found to be the most efficient in the current study. Moreover, instructions transmitted with the haptic and visual modalities are similar. Indeed, they gave information about the form of the trajectory instead of a simple absolute distance (in the case of the verbal instructions). The "spatiotemporal rule" holds that stimuli presented in spatiotemporal proximity have a higher probability of being combined to form a perception of a physical event [32]. In our brain, the responses of multi-sensory neurons (neurons integrating data coming from different modalities) are increased in case of spatiotemporal congruence and decreased otherwise [54]. The visual and haptic instructions are more spatiotemporally related than they are to the verbal instructions. The spatiotemporal similarity between visual and haptic instructions provides another explanation for the improved accuracy achieved by their combination. The similar accuracy obtained with the combination of all three modalities suggests that the information received through the verbal modality was less utilized by the participants to replicate the trajectory than that received through the two other modalities. This also aligns with previous research indicating that incorporating haptic instructions led to enhanced performance among trainees compared to conditions involving only visual and verbal instructions [9].

On the other hand, when the verbal instructions were combined with only one other modality, the participants were less accurate.

Table 3: Friedman tests for the comparison questionnaire

Q#	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
χ^2	28.27	33.45	25.42	12.67	30.93	32.39	32.39	33.82	27.37	29.06	24.56
P-values	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

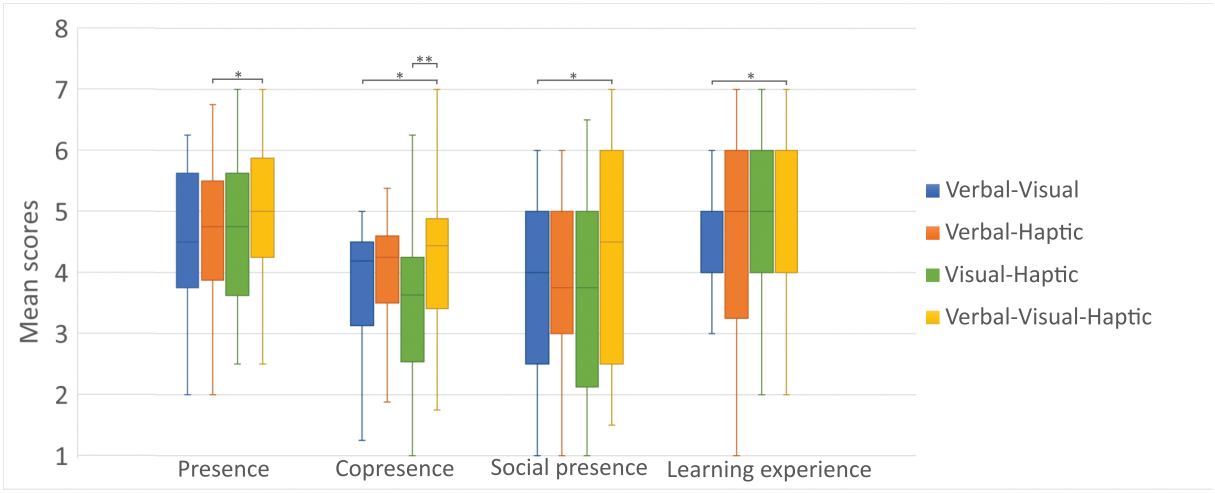

Figure 9: Mean scores for the questionnaire on presence, social presence, copresence, and learning experience ($*$ = p $<$ .05; $**$ = p $<$ .01)

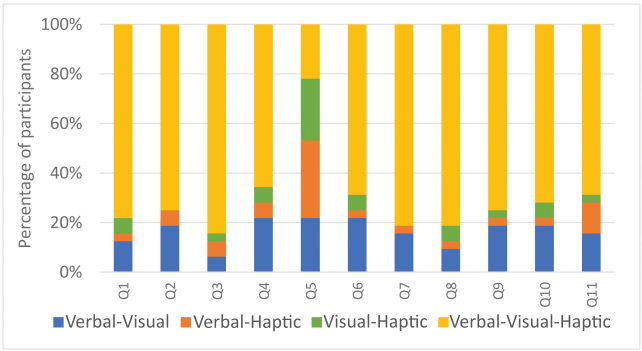

Figure 10: Preferred modality combination for each criterion (percentage of participants choosing the combination as the most preferred)

This suggests that the verbal instructions may interfere with information received from the other combined modality (visual or haptic), leading to a decreased accuracy performance.

The movement trajectories results further confirm this hypothesis. Indeed, the results showed that the participants' tool trajectories were closer to the reference trajectories with the combination of visual-haptic modalities than with the other combinations. As mentioned previously, verbal instructions did not inform on tool trajectories. Therefore, when these instructions were provided (during three of the experimental conditions), participants seemed to have focused more on reaching the target and less on replicating the trajectories provided by the other modalities (visual or haptic). On the other hand, when verbal instructions were not provided (visual-haptic condition), the participants tried to replicate the movement trajectory and, at the same time, reach the target. This led to more optimized trajectories and more accurate movements, which were also executed faster. Indeed, the results revealed that participants replicated the movement faster with the visual-haptic combination. In previous research [49], instructions conveyed through the haptic modality resulted in a lower movement completion time. The visual instructions were also found to decrease the completion times compared to verbal instructions. In our experiment, all the conditions that included verbal instructions increased the completion times. Verbal instructions may be easier to memorize than visual and haptic instructions. Indeed, a given amplitude (numerical value) is less complex than a visual or haptic hand trajectory. Participants may have had to execute their actions quickly when the verbal instructions were absent (in the visual-haptic condition) while the information was still "fresh" in their memory.

These findings validate our hypothesis H1, suggesting that combining visual and haptic instructions leads to the best performance.

5.2 Subjective Measurements

The evaluation of the users' experience involved three measurements. The first measurement focused on the participants' perceived workload. The second measurement consisted of a subjective assessment of interactions with the instructor, including the sense of presence, copresence, social presence, and learning experience. Finally, the third measurement was used to compare the four conditions.

Regarding the NASA-TLX, previous research [47] has suggested that multimodal learning reduces cognitive load due to a distribution of information processing. The results of the NASA-TLX did not reveal any significant difference in scores across the modality combinations, suggesting that participants did not perceive any variations in mental workload. A recent meta-review including 556 studies found that the average score for the raw NASA-TLX was 42, while the average score in 72 VR-based studies (mainly education and healthcare applications) was 41 [25]. Based on unimodal instructions, the closest work to our current study [49] also reported similar values (a mean raw score of 42). The mean raw scores of the NASA-TLX in our study (32.94, 34.66, 33.88, 34.34, respectively for the Verbal-Visual, Verbal-Haptic, Visual-Haptic, and Verbal-Visual-Haptic conditions) are lower than these reported values. While conducting a systematic comparison between these values is impossible, our work may suggest that multimodal instructions lead to decreased workload. However, further studies are needed to compare the impact of unimodal and multimodal communication on cognitive workload.

The non-significant difference between the modality combina-

tions may be attributed to the simplicity of the given amplitude instructions (one direction and one amplitude), which did not require significant mental effort to be comprehended. Further investigations using more complex instructions will be necessary to better understand the potential impact of modality combinations on workload. Therefore, hypothesis H2 is rejected.

The results of the subjective questionnaire indicate that the participant experienced a higher sense of presence, copresence, and social presence and a better learning experience with the combination of the three modalities. This result is in line with previous research indicating that adding new modalities increases presence [45] and thus, the more modalities, the higher the presence of participants. Moreover, Slater et al. [50] explain that presence and copresence often tend to co-vary, so when users experience a higher sense of presence, they also tend to experience a more heightened sense of copresence. In the work of Simon et al. [49], the verbal modality also increased the sense of copresence compared to the visual-only and the haptic-only modalities. However, our study differs from this previous work in that the instructor was located remotely. While the results show that the sense of copresence was the lowest in the visual-haptic combination (the only combination without the verbal modality), the difference was only significantly different with the three-modality combination (and marginal with the verbal-haptic combination). This does not permit us to confirm our hypothesis H3 that verbal instructions are essential in improving the user experience in CVE. However, it suggests that combining the three modalities is even better to improve the user experience and interactions with a remote instructor.

This is further confirmed by the results of the comparison questionnaire, indicating that participants preferred the three-modality combination. They perceived this combination as the most appropriate, pleasant, engaging, and educational. These findings align with the observation that utilizing all three modalities contributes to an enhanced learning experience. Furthermore, participants indicated that receiving instructions by combining all three modalities was associated with greater ease of comprehension and efficiency in instruction reception. Additionally, this combination was the most accurate. This is consistent with the objective measurements and also with the presence and social presence results. It is worth noting that while the visual-haptic condition led to better performance (better trajectories and faster completion times), it was less preferred than the three-modal combination and was found to be more disturbing. This suggests that adding verbal instructions is helpful to improve user experience as long as these instructions do not interfere with those coming from other modalities.

6 Conclusion

This study is part of a research project aimed at integrating the concept of mentoring to acquire technical skills within CVE. Mentoring is a model widely employed in different disciplines at the beginning of the learning process, providing learners with expert instructions and feedback. This study evaluated the effect of four communication modalities (verbal-visual, verbal-haptic, visual-haptic, and verbal-visual-haptic) on transferring spatial information to learners when handling tools. The findings indicate that the visual-haptic combination and verbal-visual-haptic combination were the most effective in reducing distance errors and increasing movement replication accuracy compared to the other combinations. The visual-haptic combination also allowed for faster instruction replication and a better movement trajectory. However, participants indicated that the visual-haptic combination was more complex for memorizing spatial instructions and more disturbing than the three-modal combination. In addition, the three-modal combination increased the participants' sense of presence, social presence, copresence, and perceived quality of the learning experience. This combination was also perceived as the most suitable for learning and memorizing spatial information. In general, the visual-haptic combination was the least preferred among participants.

These results give valuable insights for designing collaborative interactions to enhance the acquisition of tool manipulation skills inside a CVE. Indeed, they further confirm that each modality brings distinct advantages for improving the learning process and that employing a multimodal communication strategy is optimal. The visual-haptic combination is demonstrated to be the most appropriate to increase learners' performance when receiving movement amplitude instructions. Adding verbal instructions can be helpful to improve the user experience and interactions with a remote instructor. However, such instructions may degrade the learners' performance by interfering with information received through the other modalities. Therefore, we suggest adding verbal interactions to improve communication with the instructor while avoiding using them to provide specific instructions on tool movements. For instance, verbal interactions could be used to encourage the learners or give them feedback on their performance.

Future studies will investigate how combining modalities impacts learning outcomes and performance when dealing with more complex motor tasks in a CVE. In addition, employing more sophisticated shared immersive environments that closely replicate real-world scenarios can increase the complexity of tasks and improve the relevance of the study's findings.

To conclude, our study aimed to acquire insights into communication between instructors and learners within immersive learning environments. Nevertheless, some limitations must be noted. First, our study did not investigate the influence of communication modalities on learning outcomes, highlighting the need for a longitudinal study encompassing pre-post and retention assessments. Our primary objective in this research was to gain a deeper comprehension of the strengths and weaknesses of each combination of communication modalities before embarking on such a comprehensive study. Given the present findings, our future works include conducting a longitudinal study to examine the effects of each combination of modalities on learning. Secondly, the study exclusively focuses on the learner's perspective when evaluating the influence of communication modalities. While this approach was essential for controlling the experiment, it is crucial in future research to explore how these combinations of modalities affect instructors and how technology can facilitate their ability to convey their skills effectively [41, 55]. This investigation will enable us to design more efficient user interfaces to facilitate the transfer of technical skills from an instructor to a learner in immersive shared environments.

Acknowledgments

The authors wish to thank all the volunteers who participated in the study. This work was supported by a grant from the French National Research Agency (ANR-20-CE33-0010 Show-me).

References

[1] M. J. Barrington, D. M. Wong, B. Slater, J. J. Ivanusic, and M. Ovens. Ultrasound-guided regional anesthesia: how much practice do novices require before achieving competency in ultrasound needle visualization using a cadaver model. *Regional Anesthesia & Pain Medicine*, 37(3):334–339, 2012.

[2] C. Basdogan, C. Ho, M. A. Srinivasan, and M. Slater. An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7:443–460, 2000.

[3] C. W. Borst, N. G. Lipari, and J. W. Woodworth. Teacher-guided educational vr: Assessment of live and prerecorded teachers guiding virtual field trips. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*, pages 467–474. IEEE, 2018.

[4] D. Brickler, M. Volonte, J. W. Bertrand, A. T. Duchowski, and S. V. Babu. Effects of stereoscopic viewing and haptic feedback, sensorymotor congruence and calibration on near-field fine motor perception-

action coordination in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 28–37. IEEE, 2019.

[5] D. Cecilio-Fernandes, F. Cnossen, J. Coster, A. D. C. Jaarsma, and R. A. Tio. The effects of expert and augmented feedback on learning a complex medical skill. *Perceptual and motor skills*, 127(4):766–784, 2020.

[6] D. Chang, K. V. Nesbitt, and K. Wilkins. The gestalt principle of continuation applies to both the haptic and visual grouping of elements. In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, pages 15–20. IEEE, 2007.

[7] A. Chellali, W. Ahn, G. Sankaranarayanan, J. Flinn, S. D. Schwaitzberg, D. B. Jones, S. De, and C. G. Cao. Preliminary evaluation of the pattern cutting and the ligating loop virtual laparoscopic trainers. *Surgical endoscopy*, 29:815–821, 2015.

[8] A. Chellali, C. Dumas, and I. Milleville-Pennel. Influences of haptic communication on a shared manual task. *Interacting with Computers*, 23(4):317–328, 2011.

[9] A. Chellali, C. Dumas, and I. Milleville-Pennel. Haptic communication to support biopsy procedures learning in virtual environments. *Presence: Teleoperators and Virtual Environments*, 21(4):470–489, 2012.

[10] A. Chellali, I. Milleville-Pennel, and C. Dumas. Influence of contextual objects on spatial interactions and viewpoints sharing in virtual environments. *Virtual Reality*, 17(1):1–15, 2013.

[11] E. F. Churchill and D. Snowdon. Collaborative virtual environments: an introductory review of issues and systems. *virtual reality*, 3:3–15, 1998.

[12] G. Convertino, H. M. Mentis, A. Slavkovic, M. B. Rosson, and J. M. Carroll. Supporting common ground and awareness in emergency management planning: A design research project. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(4):1–34, 2011.

[13] C. Dabo and S. Connac. Les représentations des étudiants tuteurs en masso-kinésithérapie sur le dispositif de tutorat entre pairs. *Recherches en éducation*, 46, 2022.

[14] M. Desmurget. *Imitation et apprentissages moteurs: des neurones miroirs à la pédagogie du geste sportif*. Groupe de Boeck, 2006.

[15] O. Ehmer. Synchronization in demonstrations. multimodal practices for instructing body knowledge. *Linguistics Vanguard*, 7(s4):20200038, 2021.

[16] O. Ehmer and G. Brône. Instructing embodied knowledge: multimodal approaches to interactive practices for knowledge constitution. *Linguistics Vanguard*, 7(s4):20210012, 2021.

[17] J. Fox, S. J. Ahn, J. H. Janssen, L. Yeykelis, K. Y. Segovia, and J. N. Bailenson. Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Human–Computer Interaction*, 30(5):401–432, 2015.

[18] R. B. Gillespie, M. S. O'Modhrain, P. Tang, D. Zaretzky, and C. Pham. The virtual teacher. In *ASME International Mechanical Engineering Congress and Exposition*, volume 15861, pages 171–178. American Society of Mechanical Engineers, 1998.

[19] N. J. Glaser and M. Schmidt. Usage considerations of 3d collaborative virtual learning environments to promote development and transfer of knowledge and skills for individuals with autism. *Technology, Knowledge and Learning*, 25(2):315–322, 2020.

[20] E. Goffman. *Frame analysis: An essay on the organization of experience.* Harvard University Press, 1974.

[21] C. Gutwin, O. Schneider, R. Xiao, and S. Brewster. Chalk sounds: the effects of dynamic synthesized audio on workspace awareness in distributed groupware. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 85–94, 2011.

[22] W. S. Halsted. The training of the surgeon. *Bull Johns Hop Hosp*, pages 267–275, 1904.

[23] S. G. Hart and L. E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[24] J. Hattie and H. Timperley. The power of feedback. *Review of educational research*, 77(1):81–112, 2007.

[25] M. Hertzum. Reference values and subscale patterns for the task load index (tlx): a meta-analytic review. *Ergonomics*, 64(7):869–878, 2021.

[26] J. Hindmarsh, M. Fraser, C. Heath, S. Benford, and C. Greenhalgh. Fragmented interaction: establishing mutual orientation in virtual environments. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 217–226, 1998.

[27] A. Kapur, G. Tzanetakis, N. Virji-Babul, G. Wang, and P. R. Cook. A framework for sonification of vicon motion capture data. In *Proceedings of the 8th International Conference on Digital Audio Effects (DAFX-05)*, pages 20–22, 2005.

[28] J.-H. Lee and C. Spence. Assessing the benefits of multimodal feedback on dual-task performance under demanding conditions. *People and Computers XXII Culture, Creativity, Interaction 22*, pages 185–192, 2008.

[29] K. Lu, G. Liu, and L. Liu. A study on haptic collaborative game in shared virtual environment. In *Fifth International Conference on Machine Vision (ICMV 2012): Computer Vision, Image Analysis and Processing*, volume 8783, pages 425–431. SPIE, 2013.

[30] D. Martin, S. Malpica, D. Gutierrez, B. Masia, and A. Serrano. Multimodality in vr: A survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–36, 2022.

[31] H. M. Mentis, A. Chellali, and S. Schwaitzberg. Learning to see the body: supporting instructional practices in laparoscopic surgical procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2113–2122, 2014.

[32] M. A. Meredith. On the neuronal basis for multisensory convergence: a brief overview. *Cognitive brain research*, 14(1):31–40, 2002.

[33] J. Moll, E.-L. S. Pysander, K. S. Eklundh, and S.-O. Hellström. The effects of audio and haptic feedback on collaborative scanning and placing. *Interacting with computers*, 26(3):177–195, 2014.

[34] K. Nesbitt et al. *Designing multi-sensory displays for abstract data*. PhD thesis, School of Information Technologies, 2003.

[35] K. L. Nowak and F. Biocca. The effect of the agency and anthropomorphism of users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoper. Virtual Environ.*, 12:481–494, 2003.

[36] C. S. Oh, J. N. Bailenson, and G. F. Welch. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5:409295, 2018.

[37] S. Ojala, J. Sirola, T. Nykopp, H. Kröger, and H. Nuutinen. The impact of teacher's presence on learning basic surgical tasks with virtual reality headset among medical students. *Medical education online*, 27(1):2050345, 2022.

[38] V. Papageorgiou and P. Lameras. Multimodal teaching and learning with the use of technology: Meanings, practices and discourses. *International Association for Development of the Information Society*, 2017.

[39] S. Philippe, A. D. Souchet, P. Lameras, P. Petridis, J. Caporal, G. Coldeboeuf, and H. Duzan. Multimodal teaching, learning and training in virtual reality: a review and case study. *Virtual Reality & Intelligent Hardware*, 2(5):421–442, 2020.

[40] M. C. Porte, G. Xeroulis, R. K. Reznick, and A. Dubrowski. Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *The American journal of surgery*, 193(1):105–110, 2007.

[41] Y. Rahman, S. M. Asish, N. P. Fisher, E. C. Bruce, A. K. Kulshreshth, and C. W. Borst. Exploring eye gaze visualization techniques for identifying distracted students in educational vr. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 868–877. IEEE, 2020.

[42] J. Rasmussen. Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions in Systems, Man, and Cybernetics*, 13:257–266, 1983.

[43] A. Ricca, A. Chellali, and S. Otmane. Influence of hand visualization on tool-based motor skills training in an immersive vr simulator. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 260–268. IEEE, 2020.

[44] E.-L. Sallnäs. *The effect of modality on social presence, presence and performance in collaborative virtual environments*. PhD thesis, KTH, 2004.

[45] E.-L. Sallnäs, K. Rassmus-Gröhn, and C. Sjöström. Supporting presence in collaborative environments by haptic force feedback. *ACM*

Transactions on Computer-Human Interaction (TOCHI), 7(4):461–476, 2000.

[46] R. A. Schmidt, T. D. Lee, C. Winstein, G. Wulf, and H. N. Zelaznik. *Motor control and learning: A behavioral emphasis*. Human kinetics, 2018.

[47] R. Sigrist, G. Rauter, R. Riener, and P. Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review*, 20:21–53, 2013.

[48] A. L. Simeone, M. Speicher, A. Molnar, A. Wilde, and F. Daiber. Live: The human role in learning in immersive virtual environments. In *Symposium on spatial user interaction*, pages 1–11, 2019.

[49] C. Simon, M. Boukli-Hacene, S. Otmane, and A. Chellali. Study of communication modalities to support teaching tool manipulation skills in a shared immersive environment. *Computers & Graphics*, 117:31–41, 2023.

[50] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder. Small-group behavior in a virtual and real environment: A comparative study. *Presence*, 9(1):37–51, 2000.

[51] J. Strandbygaard, F. Bjerrum, M. Maagaard, P. Winkel, C. R. Larsen, C. Ringsted, C. Gluud, T. Grantcharov, B. Ottesen, and J. L. Sorensen. Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial, 2013.

[52] A. D. Udani, T. E. Kim, S. K. Howard, and E. R. Mariano. Simulation in teaching regional anesthesia: current perspectives. *Local and regional anesthesia*, pages 33–43, 2015.

[53] H. S. Vitense, J. A. Jacko, and V. K. Emery. Multimodal feedback: an assessment of performance and mental workload. *Ergonomics*, 46(1-3):68–87, 2003.

[54] M. T. Wallace, M. A. Meredith, and B. E. Stein. Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *Journal of neurophysiology*, 69(6):1797–1809, 1993.

[55] J. W. Woodworth, D. Broussard, and C. W. Borst. Redirecting desktop interface input to animate cross-reality avatars. In *2022 IEEE conference on virtual reality and 3D user interfaces (VR)*, pages 843–851. IEEE, 2022.

[56] F. Wu, J. Thomas, S. Chinnola, and E. S. Rosenberg. Exploring communication modalities to support collaborative guidance in virtual reality. In *2020 IEEE conference on virtual reality and 3d user interfaces abstracts and workshops (VRW)*, pages 79–86. IEEE, 2020.

[57] G. Wyvill, C. McPheeters, and B. Wyvill. Data structure for *soft* objects. *The Visual Computer*, 2(4):227–234, Aug. 1986.

[58] E. Yiannakopoulou, N. Nikiteas, D. Perrea, and C. Tsigris. Virtual reality simulators and training in laparoscopic surgery. *International Journal of Surgery*, 13:60–64, 2015.