International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2022)
J.-M. Normand and H. Uchiyama (Editors)

Comparing modalities to communicate movement amplitude during tool manipulation in a shared learning virtual environment

C. Simon[1] and S. Otmane[1] and A. Chellali [1]

[1]IBISC Lab, Univ Evry, Université Paris Saclay, France

Abstract

Shared immersive environments are used to teach technical skills and communicate relevant information. However, designing the appropriate interfaces and interactions to support this communication process remains an open issue. We explore using three modalities to communicate movement amplitude during tool manipulation tasks in a shared immersive environment. The haptic, visual, and verbal modalities were used separately to instruct a learner about the amplitude of the movements to perform in the 3D space. The user study comparing these modalities shows that instructions given through the visual modality permitted to decrease the distance estimation error. In contrast, the haptic modality helped the learners perform the task significantly faster. The verbal modality significantly increased the perceived sense of copresence but was the least preferred modality. This research contributes to understanding the importance of each modality when communicating spatial skills in a shared immersive environment. The results suggest that combining modalities could be the most appropriate way to transfer movement amplitude information to a learner by improving performance and user experience. These findings can enhance the design of immersive collaborative systems and open new perspectives for further research on the effectiveness of multimodal interaction to support learning technical skills in VR. Designed tools can be used in different fields, such as medical teaching applications.

CCS Concepts

• ***Human-centered computing*** → ***Collaborative interaction; Virtual reality;*** • ***Applied computing*** → *Collaborative learning;*

1. Introduction

Technical skills are used to perform specific tasks such as instrument handling in surgery. Mastering these skills requires performing the task unconsciously with great speed and accuracy. To reach the expertise level, novices must be taught how to perform the task correctly. During the early learning stages, teacher's instructions and guidance are necessary to improve skill acquisition [Sch87]. This is referred to as "augmented feedback"; information augmented by a teacher using verbal statements or visual aids rather than being received directly from the learner's senses [LSS94]. In some cases, the teacher takes the learner's hand to show the correct movement to perform [CDMP12]. Studies in the medical field have stressed the importance of this augmented feedback to improve satisfaction [MCM19], reduce medical errors [CMH*13], and improve technical skills acquisition [Lov18]. Recently, there has been increasing interest in using XR technologies to provide augmented feedback for motor learning [GK21]. However, designing tools supporting this learning model remains an open issue.

This work is part of a research project that aims to design collaborative interactions and interfaces allowing teachers to demonstrate their skills and guide learners in an immersive VR environment to better transfer technical skills. Allowing the teacher to interact with learners within a shared virtual environment permits, on the one hand, to support the guidance paradigm required during the early stages of technical skills learning and, on the other hand, to recreate real-world situations allowing learners to improve their skills in a safe and controlled environment. To develop such systems, it is essential to characterize better the communication modalities and design appropriate collaborative interaction techniques.

The current work focuses on instrument handling tasks. More particularly, we study how an instructor can guide a learner to move a tool with the correct amplitude. Three communication modalities are compared in a user study regarding their impact on performance and user experience. These modalities are based on those used to provide augmented feedback, namely the verbal, visual, and haptic modalities. The main contribution of this work is to inform on the role of each modality during the instructor-learner communication process for guidance in performing tool manipulations. Our primary hypothesis is that every modality contributes differently and plays a specific yet essential role in this communication process.

2. RELATED WORK

2.1. Teacher-learner interaction modalities

Interactions between the teacher and the learner to provide augmented feedback require using appropriate communication modal-

DOI: 10.2312/egve.20221270

ities. The choice of the communication modalities depends not only on the skills to be taught [CDMP12], but also on the availability of these modalities during the communication process [Sal04].

As discussed earlier, the teacher can use verbal, visual, and haptic communication modalities when teaching motor skills [GOT*98]. Research has shown that each modality has its strengths and weaknesses. For instance, the verbal modality is considered an immediate and easy means of communication. It permits sharing the correct rules for the skill execution. However, it could be insufficient to transfer the skill properties efficiently. Indeed, motor skills involve invisible elements such as tactile and haptic sensations, which can be difficult to describe verbally [Ras83].

On the other hand, the visual modality is used for learning by imitation. Teachers demonstrate the skills by showing a visual model of the movement to perform. This allows them to share their skills more efficiently since they communicate while acting. Learners observe the visual model and then try to reproduce the movement. This is a well-established model for motor skills learning [SRRW13], particularly in sports and surgery. However, visual demonstrations give an incomplete skill representation where only postures and movement dynamics are shown. The cause-and-effect of the movement principles cannot fully be demonstrated.

Finally, the haptic modality is used when a teacher takes the learner's hand to show the correct movement to perform. The direct contact between their hands permits the communication of the skill characteristics, such as the forces and movement dynamics. However, the contact between their hands can modify their perception of the movement and the intrinsic environment feedback. This may alter the primary task performance [CDMP11].

Hence, the characteristics of each modality need to be considered carefully when designing teacher-learner interactions. While the visual modality remains the primary communication approach, literature has shown that combining information from different sensory modalities can improve user performance and reduce the workload on human vision [BPG*06, WKZ*19]. In this case, the distribution of information over different modalities is superior to providing the same amount of information in one modality [Wic02]. Thus, if the strengths and weaknesses of each modality are considered during the design of user interactions, this can enhance the user experience by conveying additional information not easily obtained through another channel. This work aims to provide design recommendations for VR teacher-learner interaction systems by exploring the role of each modality in the communication process.

2.2. Technologies to support teacher-learner interactions

Shared virtual environments (SVE) allow multiple users to work together in the same virtual space [CSM12]. They are used in different applications such as industrial design, training, or entertainment [CSM12, CJD13, SS15]. These systems could be an appropriate medium for supporting teacher-learner interactions for learning, particularly when they support multiple communication modalities.

The design of interaction modalities and their combination in SVEs have received increased attention recently. For instance, audio and visual modalities were successfully combined to improve collaboration [GSXB11]. Haptic and visual modalities were also combined to improve task performance and social interactions [Sal04]. Recent studies also suggest a positive impact of haptics on visual attention [MF17]. Other studies have shown that combining the visual, haptic, and audio is even better for improving the users' performance in collaborative tasks compared to visual only or a combination of two modalities [MPEH14, GMG*09].

SVEs were also used for teaching new skills to a remote trainee using visual guidance for industrial applications [CDG*19]. A recent work also presented a new technique for effective and stable teacher-learner viewpoint sharing [LRMC*20]. Another study explored Mixed Reality viewpoint sharing between a remote expert and multiple local trainees [LKLH20]. Other studies [CDMP11, LAP*18, FMS*19, WCC16] have proposed augmented feedback paradigms in VR based on a single communication modality (either haptic or visual). However, the amount of research in this area remains limited, particularly in studying technical skills learning [SRRW13]. The few existing studies suggest that if the workload is high in one modality, skills should be communicated through another modality (according to the modality appropriateness hypothesis [WW80]) or in a multimodal way [HR09]. This was shown to reduce the learner's workload and enhance skills transfer [SRRW13]. However, while SVEs are promising technologies to support teacher-learner interactions, some of their characteristics must be considered carefully since they can lead to communication failure and, thus, learning issues.

In general, communication in these spaces is different from face-to-face communication [SS15] because some basic communication features are hard to use directly through these mediated technologies [CMS*11]. For instance, body movements, touch, facial expressions, and gaze require using cumbersome devices to reproduce them faithfully in SVEs. Moreover, partners do not necessarily share the same viewpoint, are not necessarily aware of their partner's activities, and interact with distant objects [CMPD13]. These characteristics affect communication between partners [SS15] and must be considered carefully during the design process.

To summarize, research and theories in multimodal interactions suggest that multimodal communication or modality substitution can enhance collaboration and motor skills learning. Workload reduction mainly explains this positive impact due to the intuitiveness of multimodal interactions and the distribution of information processing across modalities. This, in turn, leads to better learning. However, research in this area remains insufficient to understand the impact of each modality on technical skills learning. Moreover, the current characteristics of SVEs may impact communication and motor skills learning. The work presented here aims to contribute to this discussion by providing information on how the verbal, visual, and haptic modalities can be used in SVEs to instruct learners on movement amplitudes during tool manipulation.

3. USER STUDY

This exploratory study aims to investigate each modality's impact on communication between a teacher and a learner during tool manipulation. The focus will be made on the learners receiving the instructions. Therefore, the instructions will be given by the same teacher (an experimenter) while varying the communication modalities. The teacher will instruct the learners on performing a tool movement with the correct amplitude using one modality at a time. Then, the learners must replicate the movement as accurately and quickly as possible. It is expected that the three modalities will differ in their impact on performance, workload, and user experience.

The user performance will be measured using task completion time and movement accuracy. The cognitive workload will be measured through a standard questionnaire. Finally, the learners will subjectively assess the quality of their collaboration with the instructor and compare the modalities according to different aspects.

3.1. Participants

A total of twenty-one (21) participants (18 males and three females) took part in this study. They were recruited from university students and staff. The mean age was 29.23 (σ = 8.94, min = 20, max = 53). Nineteen were right-handed, and two were left-handed. They all had normal or corrected to normal vision. Sixteen have previously used HMDs, with six being regular users (once a week). Six participants used haptic devices before this experiment (mainly in demonstrations or previous user studies). The institutional ethics committee of Université Paris Saclay (CER-PS) approved the experimental protocol beforehand, and all the participants were presented with informed written consent before participating.

3.2. Experimental task

A new task was defined to investigate the impact of teacher-learner communication modalities on learning tool manipulation skills. It mimics a tool manipulation task with an instructor guiding a learner before this manipulation. To simplify the task, we have chosen to focus only on the amplitude of the movement by giving instructions on the distance to be traveled by the tool. The movement had to be performed only on one axis (X, Y, or Z-axis). Thus, the task consisted of picking up a small 3D sphere from a starting position and dropping it at a final position using a virtual tool. The amplitude of the movement to perform was unknown for the participant who received indications about it from the instructor using three different modalities: verbal, visual, and haptic. Therefore, the task was divided into two phases: (1) the instruction phase: during which the experimenter instructed the participant on the amplitude and the direction of the movement to perform through one modality (according to the experimental condition), and (2) the manipulation phase: during which the participant had to replicate on his/her own the described movement with the correct amplitude.

3.3. Experimental setup and virtual environment

The apparatus consisted of a Vive Pro HMD with a tracker for visualization and matching between the physical and real worlds (Figure 1). A Geomagic Touch haptic device was used to interact with the virtual objects and receive the haptic instructions.

A simple VE was developed using Unity3D with C# and the SteamVR plugin. During the instruction phase in the haptic and the verbal conditions, the VE was composed of two plane surfaces (44cm x 15cm), one horizontal colored in yellow and one vertical colored in red (Figure 1). In the visual condition, the VE also

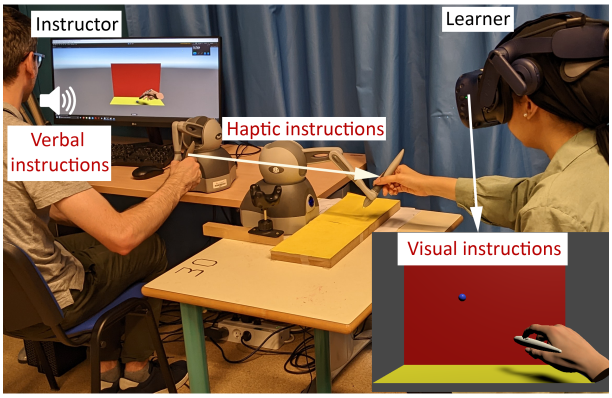

Figure 1: *Apparatus and experimental setup. The apparatus includes a haptic device and an HMD. The participant was seated next to the experimenter, and both were holding their own haptic device. The virtual environment consists of a yellow and a red plane. The participants also saw a blue sphere and the instructor's virtual hand (only in the visual condition) or their own virtual hand (only during the manipulation phase) holding a tool. The tool matched the haptic device handle in position, shape, and size. The virtual hand was attached to the tool and was collocated with the participant's hand. The yellow wooden board matched the virtual yellow plane in position and size.*

included a small blue sphere, a virtual tool, and a virtual hand representing the instructor's hand avatar (Figure 1). During the manipulation phase, the VE also included the sphere to be moved, the tool controlled by the participant through the haptic device, and the virtual hand representing this time his/her hand avatar (Figure 1). The hand avatar was attached to the tool and collocated with the participant's real hand. As suggested by the literature, this avatar was added to improve distance estimation in the immersive environment [LLB*11, PRKI10, RIKA08]. However, it was not animated, as this was shown to have no impact on motor skills learning [RCO21]. To further improve distance estimation in the immersive setup, the VE was set up to replicate as faithfully as possible the physical one [KCS17, KWS*09]. Thus, a wooden yellow board corresponding to the virtual yellow plane was put on the table (Figure 1). The virtual and physical planes had the same size and were collocated from the participant's perspective. In addition, the virtual tool was collocated with the device handle and had the same size and shape. Besides, the haptic device was positioned such that when the participants touched the virtual plane with the virtual tooltip, they felt the resistance of the physical plane with the device stylus. This matching between the real and virtual worlds required calibration before each experimental session.

For each trial, several virtual scenes were successively displayed to the participants depending on the experimental phase: in the starting scene, the participants had to read an instruction message explaining how the instructions would be given to them using the current modality. After that, they were put inside a transition scene (including a single 3D capsule) and asked if they were ready to start the subsequent trial. Using the haptic device, they had to touch the capsule and push the button to start the instruction phase. During the instruction phase, only the two colorful planes were displayed with a timeout text message warning them that the instructor would give them the instruction shortly. After each instruction, a transition scene including two 3D capsules was displayed, asking the participants to repeat the instruction if necessary. In this case, they had to touch the right capsule and push the button to reset the virtual scene. Then, the timeout message was displayed again before giving them the instruction. Otherwise, they could move to the manipulation phase by touching the left capsule and pushing the button. The virtual scene was then reset for the manipulation phase. In this case, the sphere was displayed in a new random position. The participants then used the haptic device to grab the sphere (by pushing the upper button of the device) and move it according to the received instruction. Once the sphere reached the desired position, they had to push the button again to release the sphere. Finally, the end scene asked them to start the subsequent trial. After the last trial for one condition, they were asked to remove the HMD and answer the following questionnaire.

3.4. Experimental design and conditions

The experiment followed a within-subjects design including one factor (instruction modality) with three conditions: verbal, visual, and haptic modalities. The presentation order of the conditions was counterbalanced to avoid any learning effect. Each participant performed twelve trials for each condition, with four movements on each of the XYZ-axes in both directions (left/right; top/down; forward/backward). To stay within the haptic device workspace, the movement amplitudes ranged from 3 to 13 centimeters (4cm and 13cm left;5cm and 13cm right; 3cm and 10 cm up; 4cm and 8cm down; 3cm and 6cm forward; 3cm and 5cm backward). These amplitudes were randomly picked up for each trial and balanced between conditions. A total of 756 trials (12 trials x 3 conditions x 21 participants) were recorded for this experiment.

The three experimental conditions are based on the paradigms commonly used for teaching motor skills in different application domains (such as sports and surgery). They provide the participants with the same information (amplitude and direction of movements). However, the information was encoded differently depending on the characteristics of each modality as described hereafter.

3.4.1. Visual condition

The instructions in the visual condition were communicated through a 3D animation of a realistic virtual hand manipulating the tool (Figure 1). The virtual hand and tool appeared on the right-hand side of the participant's view for each trial. They were then moved to grab the sphere and place it at the final position (that was changed for each trial), simulating the experimenter's hand movement when performing the task. This is based on the visual demonstration paradigm observed in different motor skills teaching situations. The participant watched the demonstration from the instructor's first-person perspective using the HMD. No additional instructions were given in this condition. To avoid a memorization effect of the sphere position in this condition, the sphere was positioned in different starting positions during the manipulation and the instruction phases. Thus, the participants had to memorize the amplitude and the direction of the movement rather than the final position of the sphere.

3.4.2. Verbal condition

In the verbal condition, the experimenter communicated the amplitude through voice instructions. For each trial, he asked the participant to move the ball according to one direction (left/right, up/down, forward, backward) and specified the movement amplitude in centimeters (e.g., "five centimeters to the left"). This is the most straightforward teaching paradigm observed frequently in motor skills learning [MCS14]. When listening to verbal instructions, the participant was wearing the HMD and watching the basic static scene. No other instructions were given in this condition.

3.4.3. Haptic condition

In the haptic condition, the participants had to grab the haptic arm using their right hand. Then, the haptic arm was moved from the starting position to the end position simulating the experimenter performing the same task using the arm. The movement of the haptic device provides the learner with the amplitude and the direction of the movement to perform. This paradigm is observed when the teacher guides the learner's hand when performing a movement. Similar to the verbal condition, the participants were wearing the HMD and watching the basic static scene while the haptic device guided their hands. No other instructions were given in this condition. Again, to avoid a memorization effect of the tool position in this condition, the haptic device was positioned in different starting positions during the manipulation and the instruction phases. Thus, the participants had to memorize the amplitude and the direction of the movement rather than the final spatial position of the arm.

All the instructions were pre-recorded to avoid any bias related to the experimenter. Thus, the experimenter had to read a script in the verbal condition. In the visual condition, the virtual hand movements were pre-recorded and displayed to the participant. Finally, the haptic arm movements were also pre-recorded and displayed to the participant in the haptic condition. However, to evaluate the collaborative learning experience, the participants were told that the (visual and haptic) instructions were being given in real-time by the instructor (similar to the verbal instructions). For that purpose, the experimenter was seated next to the participant and simulated the interactions using a second haptic device (Figure 1).

3.5. Experimental procedure

The experimental protocol is summarized in Figure 2. First, the experimenter informed the participants about the purpose of the study and the apparatus they would use. After that, participants had to read and sign the informed consent. Then, they were asked to read an instruction sheet describing how the prototype works, the actions to perform, and what is expected from them. They were then asked to complete a pre-test questionnaire to control interpersonal differences in distance estimation in the real world. This paper-pencil questionnaire consisted of eight different distance estimation trials displayed on the same paper sheet (mean distances to estimate = 5.75 centimeters). Each trial consisted of a line with a starting point and a written instruction indicating the distance and direction to position the endpoint using the pencil (e.g., "put a mark 8 cm down").

Four lines were positioned horizontally and four vertically. The following step was to fill in the demographics questionnaire. Finally, the participants were seated comfortably and wore the HMD to perform a familiarization task, which aimed to help them understand the VE and how to use the haptic device to interact with it.

After that, the actual experiment started. After performing the twelve trials for one condition and before beginning the trials of the following condition, the participants had to remove the HMD and answer the NASA TLX (Task Load Index) and a quality of collaboration questionnaire related to the previous condition. After that, they wore the HMD again and performed the twelve trials for the following condition. These steps were repeated for each modality. After finishing the last condition's trials and questionnaires, they had to fill in a comparison questionnaire.

3.6. Measurements and data analyses

Objective measurements consisted of the mean manipulation time for all trials (from picking the sphere to placing it at the final position) and the mean distance estimation error calculated as the mean Euclidean distance (in centimeters) between the final position of the sphere (center) and the desired position (based on the target amplitude) for all trials.

Subjective measurements consisted of the responses to the five-scale Likert quality of collaboration questionnaire (evaluating the sense of presence, copresence, and the quality of communication with the instructor; Table 1), the NASA TLX [HS88], and the modality comparison questionnaire (Table 2). For the quality of the collaboration questionnaire (Table 1), the questions (Q1-Q8) are inspired by questionnaires used in peer-reviewed international publications [NB03, LRG*17, GDLM15]. We have also proposed other questions to serve the purpose of our study (Q9-Q11). The comparison questionnaire asked the participants to rank the three modalities from the most preferred to the least preferred according to eleven classification criteria (Table 2).

Table 1: *Items of the questionnaire on the quality of collaboration.*

Q#	**Question text** : To what extent...
Q1	did you feel immersed in the environment you saw?
Q2	did the instructor try to create a sense of closeness with you?
Q3	was this like being in the same room with the instructor?
Q4	was this similar to a face-to-face meeting?
Q5	did you feel isolated from the instructor in the VE?
Q6	did you feel you were connected with the instructor?
Q7	did the instructor communicate warmth rather than coldness?
Q8	did you feel the instructor tried to help you?
Q9	do you think you can learn new skills in this application?
Q10	did you feel you were correctly understanding the instructions given to you?
Q11	did you feel you were correctly performing your tasks?

All data analyses were performed using the SPSS software (IBM Corp., Armonk, NY, USA) with the appropriate statistical tests. We have used a confidence level of 95% for all our statistical analyses. When corrections were applied, the adjusted p-values are reported.

Table 2: *items of the comparison questionnaire.*

Q#	Classification criteria (from most to least preferred)
Q1	Easiest modality to understand the received instructions
Q2	Most appropriate modality to receive instructions
Q3	Most accurate modality to receive instructions
Q4	Most pleasant modality to receive instructions
Q5	Most disturbing modality when receiving instructions
Q6	Easiest modality to memorize the received instructions
Q7	Easiest modality to reproduce the received instructions
Q8	Most educational modality for learning spatial skills
Q9	Most engaging modality for learning spatial skills
Q10	Most efficient modality for learning spatial skills
Q11	Overall most preferred modality

4. RESULTS

4.1. Distance estimation error

The first analysis compared the distance estimation performance between the real and virtual worlds. The Shapiro-Wilk normality test indicates that the mean distance estimation errors for the haptic condition were not normally distributed. Therefore, we have used Pearson's correlation test for the mean distance error estimation between the pre-test and the verbal and visual conditions and Spearman's correlation test between the pre-test and the haptic condition. The results show no significant correlation in distance estimation errors between the visual ($p = 0.66$) and the haptic ($p = 0.055$) conditions on one side and the pre-test on the other. In contrast, the mean errors between the pre-test and the verbal condition were moderately correlated ($r = 0.585$, $p = 0.005$).

After that, we compared the mean distance estimation error between conditions. Since the data were not normally distributed, the Friedman non-parametric test was used for means comparison instead of the one-way repeated measure ANOVA. The results show a significant main effect of modality on distance estimation error ($\chi^2 = 12.61$, $p = 0.002$; Figure 3). The Wilcoxon signed ranks pairwise tests with Bonferroni correction show that the mean distance estimation error was significantly lower in the visual condition than in the verbal ($p = 0.003$) and haptic ($p = 0.006$) conditions. The difference between the two other conditions was not significant.

4.2. Completion time

All data were normally distributed. The one-way repeated measure ANOVA (sphericity assumed) shows a significant main effect of modality on manipulation times ($F_{(2,40)} = 11.26$, $p < 0.001$; Figure 3). The post-hoc tests with Bonferroni correction show that the mean completion time was significantly lower in the haptic condition than in the verbal ($p = 0.0$) and visual ($p = 0.018$) conditions. The difference between the verbal and visual conditions was marginal ($p = 0.07$).

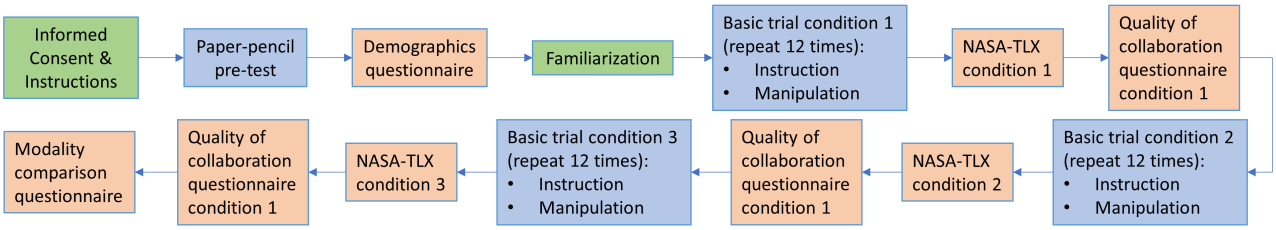

Figure 2: *Experimental protocol.*

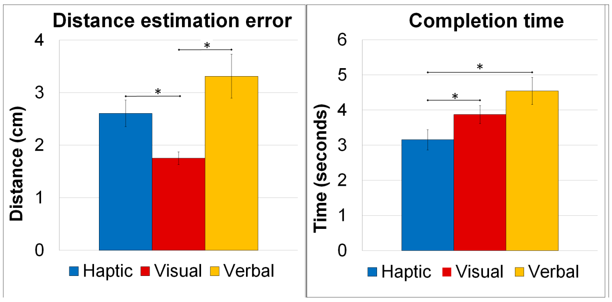

Figure 3: *Distance estimation error and completion time results.*

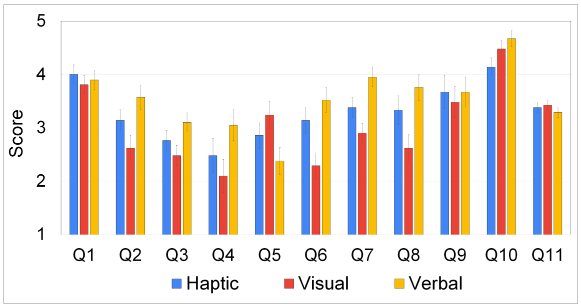

Figure 4: *Quality of collaboration questionnaire mean scores.*

4.3. Subjective measures

A non-parametric Friedman test was used for data analyses. Each question was investigated separately for the quality of collaboration and the comparison questionnaires. When a significant main effect was found, the Wilcoxon signed-rank test with a Bonferroni correction was used for the post-hoc pairwise comparisons.

4.3.1. Perceived workload

The Friedman test shows no main effect of modality on raw TLX scores (45.11±17.23; 43.41±15.99, 45.67±14.15; respectively for the verbal, visual and haptic conditions; $\chi^2 = 0.5$, $p = 0.77$).

4.3.2. Quality of collaboration questionnaire

The results are shown in Figure 4 and the Friedman tests are reported in Table 3. The Wilcoxon signed-rank tests for questions with significant main effects are reported hereafter. The participants found the experience closer to a face-to-face meeting (Q4) in the verbal than the haptic conditions ($p = 0.03$). The participants felt they were less connected with the instructor (Q6) in the visual than in the haptic ($p = 0.045$) and the verbal ($p = 0.006$) conditions. The participants felt the instructor communicated warmth rather than coldness (Q7) more in verbal than visual conditions ($p = 0.012$). The participants felt the instructor tried to help them (Q8) more in verbal than visual conditions ($p = 0.021$). The participants felt they better understood the instructions (Q10) in the verbal than in the haptic conditions ($p = 0.024$). No other significant differences were found.

4.3.3. Comparison questionnaire

The results are shown in Figure 5 and the Friedman tests are reported in Table 3. The Wilcoxon signed ranks pairwise tests with Bonferroni correction for questions with significant main effects are reported hereafter. The visual modality was rated by 42.8% of participants as the easiest to receive instructions (Q3). Instead, only 19% of participants ranked the verbal modality as the easiest. The difference between these modalities is marginal ($p = 0.069$). No significant difference is found for the haptic modality. The haptic modality was ranked as the most disturbing modality to receive instructions (Q5) by 71.4% of participants, against 9.5% for the visual modality and 19% for the verbal modality. These differences are statistically significant between the haptic and visual modalities ($p = 0.003$) and between the haptic and visual modalities ($p = 0.018$). The haptic modality was ranked as the easiest for memorizing instructions (Q6) by only 9.5% of participants. It was ranked second and third, respectively, by 28.6% and 61.9% of participants. These values are significantly lower than those of the verbal modality ($p = 0.021$). The haptic and visual modalities difference was marginal ($p = 0.066$). No participant ranked the verbal modality as the most educational modality for learning spatial skills (Q8). Besides, 71.4% of participants ranked it third. The ranking was significantly lower than the visual ($p < 0.001$) and haptic ($p = 0.003$) modalities. No significant difference is found between these two modalities. The visual modality was ranked as the most engaging for learning spatial skills (Q9) by 42.9% of participants. Instead, the verbal modality was ranked as the most engaging by only 14.3% of participants. The difference between the verbal and the visual modalities was marginal ($p = 0.069$). No other significant differences were found. Finally, only 4.7% of participants ranked the verbal modality as the most efficient modality for learning spatial skills (Q10). This ranking was significantly lower than the visual ($p = 0.009$) and haptic ($p = 0.027$) modalities. No other significant differences were found.

Table 3: *Friedman test for quality of collaboration questionnaire (QCQ) and the comparison questionnaire (CQ)*

Q#	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
QCQ: χ^2 (p)	2.0 (0.36)	2.98 (0.22)	3.22 (0.19)	**7.89 (0.019*)**	3.36 (0.18)	**12.23 (0.002*)**	**7.96 (0.019*)**	**10.14 (0.006*)**	0.17 (0.91)	**9.48 (0.009*)**	2.0 (0.36)
CQ: χ^2 **(p)**	1.81 (0.40)	2.0 (0.36)	**6.0 (0.05*)**	0.095 (0.95)	**14.38 (0.001*)**	**19.23 (0.01*)**	5.42 (0.066)	**16.1 (0.001*)**	**6.0 (0.05*)**	**10.66 (0.005*)**	3.71 (0.15)

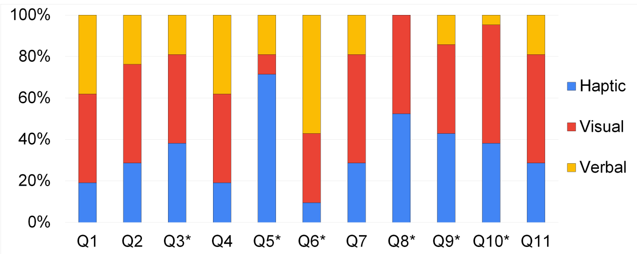

Figure 5: *Preferred modality (percentage of participants choosing the modality as the most preferred) for each question.*

5. DISCUSSION

The study aimed to investigate the impact of three modalities on communicating movement amplitude during tool manipulation to a learner in a SVE. There are several key findings from this study.

5.1. Performance

First, the correlation tests show a moderate correlation between the distance estimation errors in the verbal condition and those obtained during the paper-pencil pre-test. In the pre-test, the participants read the instructions. This suggests that distance estimation instructions are processed similarly when reading and listening. This is not the case for the haptic and visual instructions, whose values are not correlated with the pre-test. This suggests differences between modalities regarding the processing of information received through each of them. This can also be related to the differences in information that can be communicated through each modality. Indeed, the verbal and written instructions permitted to communicate the same information about the movement amplitude (ex. “5 cm to the left”). On the other hand, the haptic and visual modalities permitted communicating additional information such the movement dynamics and forces. Further investigations are needed to understand better what and how information is processed through each modality.

The results also indicate that the visual modality is the most accurate for communicating movement amplitudes. It permitted to reduce the distance estimation errors compared to the two other conditions. This is also supported by the comparison questionnaire, where the participants found the visual modality more accurate, although the difference with the other conditions was marginal. These findings are in line with previous research suggesting that visual augmented feedback can be a more effective learning strategy [SRRW13]. Vision is described as a spatial sense and adept at interpreting spatial information [Fre74]. This may have played an essential role in improving the participants’ distance estimation performance. On the other hand, haptics is adept at sensing movements [Fre74]. The haptic guidance strategy used in our study is based on position control, where the learner passively follows the movement. This strategy has been previously shown to be effective for learning temporal information of a movement [RSRW15]. However, it is less efficient for learning spatial movement aspects than visual guidance and haptic path control guidance, where the learner’s movement is corrected whenever he/she makes spatial errors following the movement path [SRGS00]. This last haptic strategy provides real-time and immediate augmented feedback to the learner and can be explored in the future as a teaching strategy for spatial information. For instance, using our system, the teacher can follow the learner’s movement with the second haptic device and correct the trajectory when errors occur.

The difference in completion time was a secondary criterion for comparing the modalities. The results show interesting findings. Indeed, the participants performed the task faster after receiving instructions through the haptic modality. Furthermore, they evaluated this modality as the most difficult for memorizing the instructions. This may explain why they performed the task faster in this condition. They could have followed the instructions quickly after receiving them while still “fresh” in their memory. However, this hypothesis will require more investigations to be confirmed.

5.2. Cognitive workload

Reducing workload is crucial for choosing the appropriate learning modality [SRRW13]. The NASA TLX shows no significant difference in scores among modalities. This suggests that participants did not experience differences in mental workload. Hence, modalities do not affect the task complexity when communicating simple distance information. This may be explained by the fact that the given distance instructions are easy to understand and do not require a subsequent mental effort to be processed. Further investigations with more complex instructions will be necessary to understand the potential impact of communication modalities on workload.

5.3. Perceived quality of collaboration

Regarding the quality of the collaboration questionnaire, only five questions have shown significant differences. Generally, the verbal modality was the most preferred one. Participants found the virtual experience closer to the real-world meeting when using this modality. They also felt a stronger connection with the instructor during the haptic and verbal conditions than during the visual one. They perceived that the instructor was trying to help them more in the haptic and verbal modalities. Besides, they found the verbal instructions more straightforward to understand. This shows a contrast between the subjective and the objective measurements. The participants' preference for the verbal modality might be related to the fact that the experimenter was physically close to them, and the verbal instructions came directly from the "real-world". In contrast, the visual and haptic instructions were provided in the immersive VE through the virtual hand or the haptic device. Hence, the sense of copresence (Q4, Q6, Q7, Q8) and the quality of the learning experience (Q10) could have been influenced by these different instruction sources: direct (for verbal) and mediated (for haptic and visual). It will be interesting to compare the three conditions with the same level of mediation (for instance, with an instructor located in another room or with pre-recorded voice messages displayed on an HMD).

On the other hand, the haptic modality generated a stronger connection (Q6) and helpfulness (Q8) than the visual modality. These dimensions refer to the feeling of establishing rapport by the partner, as discussed by Gratch and Lucas [GDLM15]. The lower scores in the visual modality could be explained by the fact that the non-animated hand could have decreased the fidelity of the partner's avatar, which has already been reported to have a negative impact on this dimension [LRG*17, GCC*20, BR21]. In addition, the system did not provide the instructor avatar's face and body, which is also critical to increase the sense of copresence [LRG*17, GCC*20, BR21]. On the other hand, the haptic modality has been reported to increase the feeling of closeness and intimacy with another person [CDMP11, SRGS00]. This may have contributed to improving the related scores for this modality.

5.4. Subjective comparison of modalities

The comparison questionnaire results contrast those of the quality of the collaboration questionnaire but are in line with the performance results. The participants felt they were learning more with the haptic and visual modalities, which were perceived to be more effective in receiving spatial information. Finally, the participants rated the verbal modality as the least engaging. The verbal modality was generally the least preferred by the participants. In addition, the differences between the visual and haptic modalities were not significant. This suggests that they were generally accepted in the same way. Finally, the haptic modality was felt to be more disturbing. This may be related to the fact that this is a new means of communication that the participants have never experienced before. Nevertheless, this did not impact either their performance or their user experience. Therefore, this modality can be further explored to improve communication in immersive teaching environments.

6. CONCLUSION

This work is part of a research project aiming to design collaborative interaction modalities in shared immersive VEs for learning technical skills in collaboration with an instructor. The work presented here focused particularly on transferring spatial information to the learner. Three modalities were compared: verbal, visual, and haptic modalities. The user study results indicate that the instructions received through the visual modality increased the movement replication accuracy in a tool manipulation task by reducing the distance estimation error compared with the verbal and haptic modalities. In addition, the haptic modality permitted replication of the instructions faster than the two other modalities. On the other hand, the verbal modality increased the sense of copresence and the perceived quality of the learning experience. The visual modality was perceived to be more adapted to learning and memorizing spatial information. The haptic modality was the most disturbing and hardest for memorizing spatial instructions. The verbal modality was generally the least preferred.

These results give various insights into the design of collaborative interactions for spatial skills learning in SVE. Indeed, they suggest that each modality can bring additional features to improve the learning experience and performance and that multimodal interactions could be the most appropriate approach. Hence, we plan to study the impact of combining modalities on the learning experience and performance in the future. Besides, the verbal modality may not be suited for complex spatial instructions (eg. curves). Therefore, other and more complex types of spatial instructions could be explored (for instance, instructions on how to correctly orient a tool). This may suggest using other modalities or a different combination of modalities. Finally, a more complex VE closer to real-world setups could add to the task's complexity and improve the generalization of the results.

Finally, while our long-term goal is to study teacher-learner communication in immersive learning environments, two limitations of the current work can be highlighted. First, the present work did not investigate the impact of modalities on learning outcomes. This would have required conducting a longitudinal study with pre-post and retention tests. Conducting such a study is costly and requires careful preparation. Before conducting such a study, we wanted in the current work to acquire a clearer picture of the benefits and drawbacks of each modality for communication. Based on the findings of the present work, we plan to conduct a longitudinal study to investigate the impact of each modality on the learning outcomes.

Second, the current work was focused on investigating the impact of communication modalities only from the learner's perspective. While this is intended to control the experiment, it will be important in the future to investigate the effect of using these modalities on teachers and how the technologies can help them better share their skills.

This will help us design more appropriate user interfaces supporting the transfer of technical skills between a teacher and a learner in SVEs.

7. ACKNOWLEDGMENTS

This work was funded by the French National Research Agency (ANR-20-CE33-0010 Show-me) and by the Paris Ile-de-France Region (grant # 20002717).

References

[BPG*06] BURKE J. L., PREWETT M. S., GRAY A. A., YANG L., STILSON F. R., COOVERT M. D., ELLIOT L. R., REDDEN E.: Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. In *Proceedings of the 8th international conference on Multimodal interfaces* (2006), pp. 108–117. doi:10.1145/1180995.1181017. 2

[BR21] BENDA B., RAGAN E. D.: The effects of virtual avatar visibility on pointing interpretation by observers in 3d environments. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality* (Oct. 2021), IEEE. doi:10.1109/ismar52148.2021.00019. 8

[CDG*19] CHÉNÉCHAL M. L., DUVAL T., GOURANTON V., ROYAN J., ARNALDI B.: Help! i need a remote guide in my mixed reality collaborative environment. *Frontiers in Robotics and AI 6* (2019), 106. URL: https://www.frontiersin.org/article/10.3389/frobt.2019.00106, doi:10.3389/frobt.2019.00106. 2

[CDMP11] CHELLALI A., DUMAS C., MILLEVILLE-PENNEL I.: Influences of haptic communication on a shared manual task in a collaborative virtual environment. *Interacting With Computers 23* (2011), 317–328. doi:10.1016/j.intcom.2011.05.002. 2, 8

[CDMP12] CHELLALI A., DUMAS C., MILLEVILLE-PENNEL I.: Haptic communication to support biopsy procedures learning in virtual environments. *Presence: Teleoperators and Virtual Environments 21* (2012), 470–489. doi:10.1162/pres_a_00128. 1, 2

[CJD13] CHELLALI A., JOURDAN F., DUMAS C.: Vr4d: an immersive and collaborative experience to improve the interior design process. In *Joint Virtual Reality Conference of EGVE and EuroVR (JVRC)* (2013), Joint Virtual Reality Conference of EGVE and EuroVR (JVRC), p. 61–65. 2

[CMH*13] COLE S. J., MACKENZIE H., HA J., HANNA G. B., MISKOVIC D.: Randomized controlled trial on the effect of coaching in simulated laparoscopic training. *Surgical Endoscopy 28*, 3 (Nov. 2013), 979–986. doi:10.1007/s00464-013-3265-0. 1

[CMPD13] CHELLALI A., MILLEVILLE-PENNEL I., DUMAS C.: Influence of contextual objects on spatial interactions and viewpoints sharing in virtual environments. *Virtual Reality 17*, 1 (2013), 1–15. doi:10.1007/s10055-012-0214-5. 2

[CMS*11] CONVERTINO G., MENTIS H. M., SLAVKOVIC A., ROSSON M. B., CARROLL J. M.: Supporting common ground and awareness in emergency management planning: A design research project. *ACM Transactions on Computer-Human Interaction (TOCHI) 18*, 4 (2011), 1–34. doi:10.1145/2063231.2063236. 2

[CSM12] CHURCHILL E. F., SNOWDON D. N., MUNRO A. J.: *Collaborative virtual environments: digital places and spaces for interaction*. Springer Science & Business Media, 2012. 2

[FMS*19] FENG Y., MCGOWAN H., SEMSAR A., ZAHIRI H. R., GEORGE I. M., PARK A., KLEINSMITH A., MENTIS H.: Virtual pointer for gaze guidance in laparoscopic surgery. *Surgical endoscopy* (2019), 1–7. doi:10.1007/s00464-019-07141-x. 2

[Fre74] FREIDES D.: Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin 81*, 5 (1974), 284–310. doi:10.1037/h0036331. 7

[GCC*20] GAMELIN G., CHELLALI A., CHEIKH S., RICCA A., DUMAS C., OTMANE S.: Point-cloud avatars to improve spatial communication in immersive collaborative virtual environments. *Personal and Ubiquitous Computing 25*, 3 (July 2020), 467–484. doi:10.1007/s00779-020-01431-1. 8

[GDLM15] GRATCH J., DEVAULT D., LUCAS G. M., MARSELLA S.: Negotiation as a challenge problem for virtual humans. In *International Conference on Intelligent Virtual Agents* (2015), Springer, pp. 201–215. doi:10.1007/978-3-319-21996-7_21. 5, 8

[GK21] GEISEN M., KLATT S.: Real-time feedback using extended reality: A current overview and further integration into sports. *International Journal of Sports Science* (Nov. 2021), 174795412110510. doi:10.1177/17479541211051006. 1

[GMG*09] GARCÍA A. S., MOLINA J. P., GONZÁLEZ P., MARTÍNEZ D., MARTÍNEZ J.: A study of multimodal feedback to support collaborative manipulation tasks in virtual worlds. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (2009), pp. 259–260. doi:10.1145/1643928.1643994. 2

[GOT*98] GILLESPIE R. B., O'MODHRAIN M., TANG P., ZARETZKY D., PHAM C.: The virtual teacher. In *Proceedings of the ASME Dynamic Systems and Control Division* (1998), vol. 64, American Society of Mechanical Engineers, pp. 171–178. doi:10.1115/imece1998-0250. 2

[GSXB11] GUTWIN C., SCHNEIDER O., XIAO R., BREWSTER S.: 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* (2011), pp. 85–94. doi:10.1145/1958824.1958838. 2

[HR09] HECHT D., REINER M.: Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental brain research 193*, 2 (2009), 307–314. doi:10.1007/s00221-008-1626-z. 2

[HS88] HART S. G., STAVELAND L. E.: Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52. Elsevier, 1988, pp. 139–183. doi:10.1016/s0166-4115(08)62386-9. 5

[KCS17] KELLY J. W., CHEREP L. A., SIEGEL Z. D.: Perceived space in the HTC vive. *ACM Transactions on Applied Perception 15*, 1 (Nov. 2017), 1–16. doi:10.1145/3106155. 3

[KWS*09] KUNZ B. R., WOUTERS L., SMITH D., THOMPSON W. B., CREEM-REGEHR S. H.: Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking. *Perception 71*, 6 (Aug. 2009), 1284–1293. doi:10.3758/app.71.6.1284. 3

[LAP*18] LIN C., ANDERSEN D., POPESCU V., ROJAS-MUNOZ E., CABRERA M. E., MULLIS B., ZARZAUR B., ANDERSON K., MARLEY S., WACHS J.: A first-person mentee second-person mentor ar interface for surgical telementoring. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR 2018)* (2018), 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR 2018), p. 3–8. doi:10.1109/ISMAR-Adjunct.2018.00021. 2

[LKLH20] LEE G., KANG H., LEE J., HAN J.: A user study on view-sharing techniques for one-to-many mixed reality collaborations. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020). doi:10.1109/vr46266.2020.00054. 2

[LLB*11] LEYRER M., LINKENAUGER S. A., BÜLTHOFF H. H., KLOOS U., MOHLER B.: The influence of eye height and avatars on egocentric distance estimates in immersive virtual environments. In *ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (2011), ACM Press. doi:10.1145/2077451.2077464. 3

[Lov18] LOVELL B.: What do we know about coaching in medical education? a literature review. *Medical education 52* (2018), 376–390. doi:10.1111/medu.13482. 1

[LRG*17] LATOSCHIK M. E., ROTH D., GALL D., ACHENBACH J., WALTEMATE T., BOTSCH M.: The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (2017), Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, p. 1–10. doi:10.1145/3139131.3139156. 5, 8

[LRMC*20] LIN C., ROJAS-MUNOZ E., CABRERA M. E., SANCHEZ-TAMAYO N., ANDERSEN D., POPESCU V., NOGUERA J. A. B., ZARZAUR B., MURPHY P., ANDERSON K., DOUGLAS T., GRIFFIS C., WACHS J.: How about the mentor? effective workspace visualization in ar telementoring. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Mar. 2020), 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE. doi:10.1109/vr46266.2020.00040. 2

[LSS94] LEE T. D., SWINNEN S. P., SERRIEN D. J.: Cognitive effort and motor learning. *Quest 46*, 3 (1994), 328–344. doi:10.1080/00336297.1994.10484130. 1

[MCM19] MACKENZIE C., CHAN T. M., MONDOUX S.: Clinical improvement interventions for residents and practicing physicians: A scoping review of coaching and mentoring for practice improvement. *AEM Education and Training 3* (2019), 353–364. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/aet2.10345, doi:10.1002/aet2.10345. 1

[MCS14] MENTIS H. M., CHELLALI A., SCHWAITZBERG S.: Learning to see the body: supporting instructional practices in laparoscopic surgical procedures. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2014), pp. 2113–2122. doi:10.1145/2556288.2557387. 4

[MF17] MOLL J., FRID E.: Using eye-tracking to study the effect of haptic feedback on visual focus during collaborative object managing in a multimodal virtual interface. In *SweCog 2017, October 26–27, Uppsala, Sweden* (2017), Högskolan i Skövde, pp. 49–51. 2

[MPEH14] MOLL J., PYSANDER E.-L. S., EKLUNDH K. S., HELLSTROM S.-O.: The effects of audio and haptic feedback on collaborative scanning and placing. *Interacting with computers 26*, 3 (July 2014), 177–195. doi:10.1093/iwc/iwt031. 2

[NB03] NOWAK K. L., BIOCCA F.: The effect of the agency and anthropomorphism of users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoper. Virtual Environ. 12* (2003), 481–494. URL: http://dx.doi.org/10.1162/105474603322761289, doi:10.1162/105474603322761289. 5

[PRKI10] PHILLIPS L., RIES B., KAEDI M., INTERRANTE V.: Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments. In *IEEE Virtual Reality Conference* (Mar. 2010), IEEE. doi:10.1109/vr.2010.5444802. 3

[Ras83] RASMUSSEN J.: Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions in Systems, Man, and Cybernetics 13* (1983), 257–266. doi:10.1109/TSMC.1983.6313160. 2

[RCO21] RICCA A., CHELLALI A., OTRNANE S.: The influence of hand visualization in tool-based motor-skills training, a longitudinal study. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)* (Mar. 2021), IEEE, IEEE, pp. 103–112. doi:10.1109/vr50410.2021.00031. 3

[RIKA08] RIES B., INTERRANTE V., KAEDING M., ANDERSON L.: The effect of self-embodiment on distance perception in immersive virtual environments. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2008), VRST '08, Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology, ACM, p. 167–170. URL: http://doi.acm.org/10.1145/1450579.1450614, doi:10.1145/1450579.1450614. 3

[RSRW15] RAUTER G., SIGRIST R., RIENER R., WOLF P.: Learning of temporal and spatial movement aspects: A comparison of four types of haptic control and concurrent visual feedback. *IEEE Transactions on Haptics 8*, 4 (Oct. 2015), 421–433. doi:10.1109/toh.2015.2431686. 7

[Sal04] SALLNÄS E.-L.: *The Effect of Modality on Social Presence, Presence and Performance in Collaborative Virtual Environments*. Tech. rep., KTH, Stockholm, 2004. 2

[Sch87] SCHMIDT R. A.: *Motor control and learning: A behavioural emphasis*, 2 ed. Human Kinetics, Champaign, IL, 1987. 1

[SRGS00] SALLNÄS E.-L., RASSMUS-GRÖHN K., SJÖSTRÖM C.: Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction 7* (Dec. 2000), 461–476. doi:10.1145/365058.365086. 7, 8

[SRRW13] SIGRIST R., RAUTER G., RIENER R., WOLF P.: Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review 20*, 1 (2013), 21–53. doi:10.3758/s13423-012-0333-8. 2, 7

[SS15] STEED A., SCHROEDER R.: Collaboration in immersive and non-immersive virtual environments. In *Immersed in Media*. Springer, 2015, p. 263–282. doi:10.1007/978-3-319-10190-3_11. 2

[WCC16] WANG J., CHELLALI A., CAO C. G.: Haptic communication in collaborative virtual environments. *Human factors 58*, 3 (2016), 496–508. doi:10.1177/0018720815618808. 2

[Wic02] WICKENS C. D.: Multiple resources and performance prediction. *Theoretical issues in ergonomics science 3*, 2 (2002), 159–177. doi:10.1080/14639220210123806. 2

[WKZ*19] WOLF E., KLÜBER S., ZIMMERER C., LUGRIN J.-L., LATOSCHIK M. E.: " paint that object yellow": Multimodal interaction to enhance creativity during design tasks in vr. In *2019 International Conference on Multimodal Interaction* (2019), pp. 195–204. doi:10.1145/3340555.3353724. 2

[WW80] WELCH R. B., WARREN D. H.: Immediate perceptual response to intersensory discrepancy. *Psychological bulletin 88*, 3 (1980), 638. doi:10.1037/0033-2909.88.3.638. 2