The Effects of Avatars, Stereo Vision and Display Size on Reaching and Motion Reproduction

Carlo Camporesi, *Student Member, IEEE,* and Marcelo Kallmann, *Member, IEEE*

Abstract—Thanks to recent advances on motion capture devices and stereoscopic consumer displays, animated virtual characters can now realistically interact with users in a variety of applications. We investigate in this paper the effect of avatars, stereo vision and display size on task execution in immersive virtual environments. We report results obtained with three experiments in varied configurations that are commonly used in rehabilitation applications. The first experiment analyzes the accuracy of reaching tasks under different system configurations: with and without an avatar, with and without stereo vision, and employing a 2D desktop monitor versus a large multi-tile visualization display. The second experiment analyzes the use of avatars and user-perspective stereo vision on the ability to perceive and subsequently reproduce motions demonstrated by an autonomous virtual character. The third experiment evaluates the overall user experience with a complete immersive user interface for motion modeling by direct demonstration. Our experiments expose and quantify the benefits of using stereo vision and avatars, and show that the use of avatars improve the quality of produced motions and the resemblance of replicated motions; however, direct interaction in user-perspective leads to tasks executed in less time and to targets more accurately reached. These and additional tradeoffs are important for the effective design of avatar-based training systems.

Index Terms—Virtual Reality, 3D Interaction, Avatars, Motion Capture, Perception, Training Systems.

✦

1 INTRODUCTION

Humans are highly social and possess exceptional skills for communication with other humans. A natural approach for immersive virtual reality systems is thus to rely on interactions that are as close as possible to how humans interact with each other. Animated characters and avatars often emerge as key elements for replicating human forms of communication, improving usability and accessibility to all types of users. The approach is promising for several applications in education, training and rehabilitation [15], [39].

Virtual characters are in particular useful when human–related tacit knowledge needs to be conveyed [31], like in motion-oriented training and rehabilitation [2], [35]. In these situations, a virtual coach or tutor can naturally demonstrate and monitor user performances, collecting important data for post-analysis. Human instructors are also important when designing training plans. Natural user interactions can be achieved with the concept of motion modeling by demonstration [10]: first, the expert human instructor demonstrates to the autonomous character how tasks should be performed, such that later the autonomous character can deliver the training material autonomously to users. Collaborative environments with remote participants can also be achieved [18],

• *C. Camporesi and M. Kallmann are with the School of Engineering, University of California, Merced, CA, 95343.*
E-mail: {ccamporesi,mkallmann}@ucmerced.edu

enabling a remote instructor to control a local avatar delivering training material to users.

In practice, achieving effective implementations of such training or rehabilitation systems requires key design choices to be made. For instance, while in some scenarios it may be useful for the user to see his or her own motions replicated in an avatar, in some other scenarios avatars may in fact distract the user from paying attention to the task at hand. The most appropriate configuration may also depend on hardware choices. For example, large full–scale screens, small desktop screens, and displays with stereo vision influence user performances in avatar–based scenarios in different ways. Achieving effective implementations of such systems therefore requires a deeper understanding of the tradeoffs involved among the many possible configurations.

In this paper we address some of these questions by investigating the effect of different system configurations on user performances. Because there are numerous variations that are possible, this paper focuses on specific configurations in three experiments that are particularly relevant to rehabilitation applications. Minimal user instrumentation is important in rehabilitation, thus we have not included experiments with head-mounted displays.

The first experiment focuses on reaching tasks (Figure 1-left). Reaching represents an important class of motions used in exercises for rehabilitation [33]. The experiment was designed to analyze the effect of different configurations on the reaching tasks. The

chosen configurations reflect typical choices that are made in practice and the obtained results provide new observations and quantified information on the tradeoffs between the varied conditions.
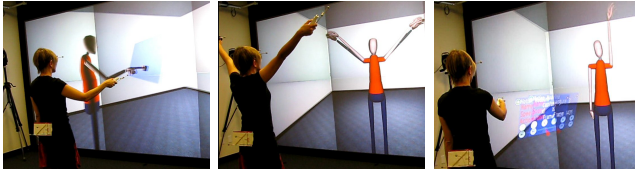

Fig. 1: Experiments: target reaching (left), motion reproduction (center), and motion modeling (right).

The second experiment was designed to analyze the ability to perceive and subsequently reproduce motions demonstrated by an autonomous virtual character (Figure 1-center). This scenario is important because motion reproduction is a key activity in several exercising and therapy applications. The motions of several participants were captured and compared under different conditions, providing new information on the effects of using avatars and stereo vision.

The third experiment is a usability study that analyzes user experiences in a complete immersive interface to model motions by demonstration (Figure 1-right). This study was selected in order to give insight in the usability of the approach in real applications, for the design of rehabilitation exercises or task-oriented motions for generic training scenarios.

The selected tasks and configurations expose new important tradeoffs between different forms of direct task execution and avatar-based visual feedback.

2 RELATED WORK

This paper evaluates specific scenarios with attention on the different ways of employing animated characters and avatars. The addressed factors have only been studied before in isolation, and not specifically addressing the impact on task execution.

Evaluation of Immersive Systems The effects of immersive virtual reality on scientific visualization, data analysis and human interaction have been studied in different ways. Depth perception through stereoscopy has been demonstrated to reduce time and error, and to improve user performance in spatial tasks [28], [36]. A frequent problem in any type of virtual environment is distance misestimation [34], which has been detected in both real workspace measurements and egocentric distances. The reason for this behavior is not clear, and it has also been detected in head mounted displays (HMDs) and in stereoscopic wide screen displays (WSDs) [16], [37], [38]. Interestingly, Naceri et al. [24] have found distance underestimation to be higher in HMDs than in WSDs.

Display size has also been investigated, and large displays have been reported to be beneficial in spatial tasks [32]. In particular, Ball et al. [5] studied the effectiveness of large high-resolution displays for interactive data visualization, concluding that a large display is preferable because it minimizes the use of virtual tools for navigation control. Considering a display physical field of view (PFOV), it has been shown that a wider PFOV can yield significantly better performance than a smaller PFOV in hand-eye coordination tasks [1], and in search and comparison tasks [3], [5], [26]. PFOV has also a direct impact on spatial awareness, memory and presence [19]. Ni et al. [26] have conducted experiments showing that large displays and high resolutions improve user performance in search and comparison tasks.

Some previous studies have also considered multivariate evaluations of combined factors. Stereoscopy and head tracking have been found to have significant impact on spatial understanding [23], [28] but not necessarily on object manipulation tasks [25]. Display and interaction modes have been observed to significantly influence a user strategy and performance in a virtual reality game [22]. The results were in favor of real world settings (high-resolution display with user-perspective interaction) or simple game-like interaction (low resolution display with common mouse/keyboard interaction).

Evaluation studies are important to guide the development of effective rehabilitation applications, which have become particularly popular in a number of cases involving arm motions, such as in post-stroke rehabilitation [9], [13], reachable space measurement [14], etc. We provide in this paper new results on the effects of different configurations involving animated characters and avatars on reaching and motion reproduction tasks, which are important tasks in rehabilitation systems.

Animated Characters and Avatars The use of animated characters as mediators in virtual environments is a natural approach to replicate human forms of interactions. Human-human communication can however be highly complex, involving several multimodal processes [6]. In particular, psychologists have shown that people are remarkably skilled in recognizing the features of a person through his or her motions; for example, when identifying gender [17], emotions [4], or the identity of a known person from just a synthesized motion silhouette [12].

The subtleties of human perception with respect to virtual characters have been explored in different ways. Visual artifacts and unnatural animations have been observed to lead to negative user reactions [8], and user sensitivity to errors in synthesized human motions have been studied in the context of ballistic motions [29]. Other types of studies have targeted rendering styles [20] and how character appearance influences the perception of actions [11] and bodily emotions [21]. The use of avatars has also been investigated with respect to user embodiment, ownership and behavior [7], [27], [30].

TABLE 1: Summary of user groups for each experiment. Letters F and M in the second column specify the number of female and male participants. Experiments are labeled with the explained 3-letter acronyms.

Group	Participants	Experiment	Label	Description of the Corresponding Configuration
1	10 (F: 7 M: 3)	1	SLU	*Stereo vision. Large display. User-perspective direct interaction without the use of an avatar.*
		2	SLT	*Stereo vision. Large display. Tutor (as an autonomous virtual character) used to assist with the interaction.*
2	10 (F: 6 M: 4)	2	MLA	*Mono vision. Large display. Avatar of the user is displayed during the interaction.*
		1	SLA	*Stereo vision. Large display. Avatar of the user is displayed during the interaction.*
3	10 (F: 5 M: 5)	1	MLA	*Mono vision. Large display. Avatar of the user is displayed during the interaction.*
		3	SLA	*Stereo vision. Large display. Avatar of the user is displayed during the motion recording phase.*
4	10 (F: 5 M: 5)	2	SLA	*Stereo vision. Large display. Avatar of the user is displayed during the interaction.*
		1	MDA	*Mono vision. Desktop-based small display. Avatar of the user is displayed during the interaction.*
5	10 (F: 3 M: 7)	3	SLN	*Stereo vision. Large display. No avatar was used during the motion recording phase.*
		2	MLT	*Mono vision. Large display. Tutor (as an autonomous virtual character) used to assist with the interaction.*

Our work focuses on investigating how different ways of using avatars and animated characters influence the execution of motion-oriented tasks. The presented results expose tradeoffs not investigated before, related to task execution in user space versus avatar space and under different conditions.

3 EXPERIMENTAL DESIGN

In this session we describe the overall experimental design of the three reported experiments. The experiments are illustrated in Figure 1, and they are later described in detail in sections 4, 5 and 6.

3.1 Apparatus

The experiments were performed in our virtual reality lab and they were designed to run in a large immersive stereo vision display wall (UC Merced's Powerwall) or in a regular desktop machine.

The Powerwall visualization system is a retro-projected surface of $4.56m$ by $2.25m$ illuminated by twelve projectors (each 1024x768@60Hz) with circular passive polarization filters. The projectors are connected to a rendering cluster of six commodity Linux-based rendering nodes (Pentium Q9550 2.83GHz GeForce GTX 280 4Gb RAM) driven by a similar main machine controlling the virtual scene being displayed. The cluster is connected through a gigabit ethernet.

The virtual reality lab also contains an optical 10-camera Vicon motion capture system that provides sub-millimeter tracking precision. The system was used to track the user's head position (for user-perspective stereo rendering), the interaction device held by the user, and two other set of markers for tracking the free hand and the torso. The upper-body motion of the user was then reconstructed from the tracked information. The interaction device, used primarily for button input, was a Nintendo Wii-mote controller.

The desktop configuration consisted of the main node computer previously described, which was connected to a standard 32 inches display (1920x1080@60Hz), without stereo vision. In each activity the application of the experiment was running in full-screen.

3.2 Participants

Fifty participants took part on the experiments. The participants were divided in groups of 10 people randomly generated according to each experiment day and availability. In order to well cover all considered variations, each participant was assigned to perform two different experiments sequentially, with the order of execution rotated every five users. It is possible that the choice of reusing participants may have influenced familiarity and thus the results; however, we believe that this effect has been highly minimized due to the unrelated experiments and the varied execution order. The group assignments and system variations are summarized in Table 1.

The participants were undergraduate students selected randomly from a pool of students enrolled in the university's experiment management system (students of Engineering, Natural Sciences, Social Sciences or Humanities disciplines). The demographics varied from 18 to 25 years old and 26 participants were female. Because of hardware and tracking volume limitations few restrictions were imposed during the participant selection: color blind, stereo blind (monocular/flat-vision), motor impaired or taller than $1.85m$. Although the system required the use of the right hand during the reaching tasks, we did not enforce the requirement of having right-handed participants. Four participants were left-handed.

Three questions were used to estimate the familiarity of the participants with the involved technologies. A total of 46 out of the 50 participants considered themselves very familiar with electronic devices (smartphones, computers, tablets, etc.); 36 participants declared to be very good with first person shooting and role-play video games (games where avatars/characters are involved) and 3 knew or had already used an immersive user-perspective stereo vision system before.

3.3 Materials

Participants were required to wear or hold four objects with attached markers that were tracked by our optical tracking system: the stereo glasses, the Wii-mote controller (held with the right hand), a bracelet on the left hand, and a belt. These four tracked objects were needed to achieve user-perspective stereo

vision with calibrated real-virtual dimensions, and to reconstruct the user's upper-body motions in his or her avatar in real-time.

Before the start of each activity an instruction sheet was handed to the participant. The instructions consisted of text and pictures explaining the application scenario, the controls, and the task to be performed, in bulleted lists (but well detailed) explanations. At the end of each task participants were asked to fill a paper questionnaire with questions related to preferences, usability and user experience. Questions were both open and based on the Likert scale.

3.4 Procedure

Each participant session was organized in four phases: informal demographics questionnaire, introduction to the system and training, first activity, and second activity. Activities were performed in four steps: avatar-user calibration, activity learning, execution, and debriefing. The total time taken per user was around one hour, with short breaks allowed.

When ready, the participant was equipped with the trackers and positioned to execute a training scenario with the Powerwall display. The training scenario consisted of a simple user-perspective object manipulation environment that included floating panels with a virtual interaction pointer. The scenario allowed the user to manipulate virtual objects and get used to the system interface. In general participants took from 10 to 15 minutes training. The scenario included a virtual room extending the real laboratory room (same wall color, carpeting, etc.), and it was designed to minimize distractions from the task. This same background scenario was used in all experiments.

Following the training step, the instructions sheet for the next activity was handed to the participant. A summarized bulleted list of the task was also provided to help the participant memorize the task. Each activity involved minimal memorization of procedures, for example: place arms along the body, click button on the controller when ready, raise arm to drive the controller toward a target, click the button when satisfied, repeat when ready, etc. The participant was allowed to take the needed time reading the instructions and preparing for the task. Activities were each completed in about 5 minutes.

Each activity required a simple calibration procedure where the participant would perform a simple T-pose required for mapping his or her dimensions to the avatar, as described in previous work [10]. During the activity, participants were not allowed to step away from their initial placement and to communicate with the researcher. After each activity the participant then completed the follow-up questionnaire.

4 EXPERIMENT 1: REACHING TARGETS

The first experiment investigated the accuracy of reaching virtual targets under different configura-

tions. The variations included the avatar use, the screen size, stereo vision and the use of user-perspective direct interaction. Forty participants took part in the experiment and they were divided in group of 10 participants. Table 2 summarizes the four variations (with participant's gender balance), and Figure 2 illustrates the experiment.

Among the several combinations possible, only combinations that made sense in practice, and that could be reasonably implemented, were considered. For example, the small monitor configuration was not suitable for stereo vision because users had to perform the tasks standing and at a certain distance, and the stereo effect could be easily lost due the limited field of view during the interactions.

TABLE 2: Configurations of experiment 1 (G = Group, M = Male, F = Female).

Label	G	M	F	Screen	Stereo	Avatar	View
SLU	1	3	7	large	yes	no	first-person
SLA	2	4	6	large	yes	yes	third-person
MLA	3	5	5	large	no	yes	third-person
MDA	4	5	5	small	no	yes	third-person

Three variations of the experiment (SLA, MLA and MDA) included the user's avatar standing in front of a floating surface. The task consisted of reaching virtual targets spawning on the surface in front of the avatar. The flat semi-transparent surface had a vertical inclination of 20 degrees and the target objects (white cubes with red concentric circles) appeared on top of it one at a time. The avatar appearance was designed to be simplistic and with little facial detail, in order to drive the user's attention to the avatar's motions and to minimize perceptual distractions due to visualization artifacts or inexpressive gaze or face.

The task made the user to control his or her avatar's right hand index finger tip towards the center of the current target cube, being as accurate as possible. The upper-body motions of the user were directly mapped in real-time to the avatar. The avatar was thus mimicking the user's motions and, since the user's point of view was from behind the avatar, no motion mirroring was implemented.

The user's point of view was from a lateral/anterior position, such that the whole working surface and the avatar's right arm motion was clearly visible at all times. Proximity and full visibility are important because otherwise the user would experience an additional cognitive that could impact task performance.

The three avatar-based variations differed from each other only by the type of visualization. The first group (SLA) worked with a large screen with user perspective stereo vision enabled, the second group (MLA) worked with the large visualization surface without user perspective vision (only simple mono vision), and the third group (MDA) performed the task in front of a desktop display without stereo vision. In
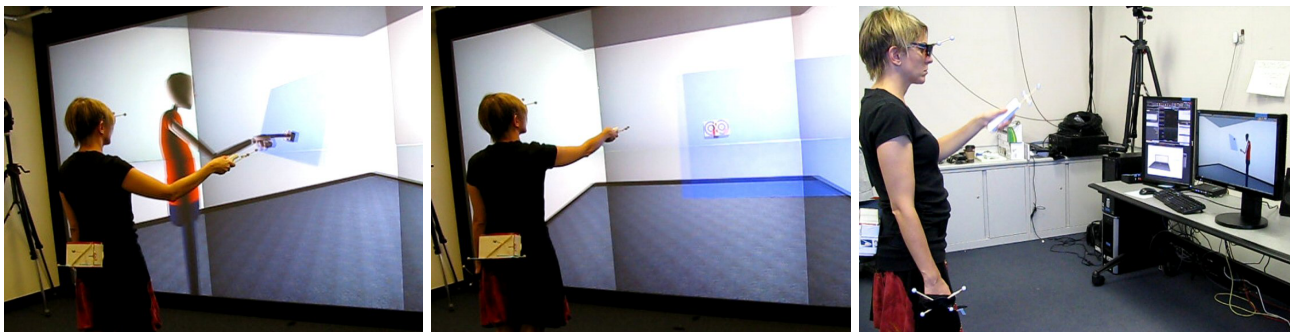
Fig. 2: Experiment 1 investigated the influence of avatars, stereo vision and display size on user performances during reaching tasks. Variations (see Table 2): SLA and MLA (left); SLU (center); and MDA (right).

this last variation the users were placed at $1.5m$ from the main screen, which was placed at a comfortable height. This setting was designed to emulate a user interacting with an inexpensive tracking device such as Microsoft Kinect or similar in a possible home setup. The distance selected is the optimal distance that would grant to the sensor enough field of view to optimally track the user's body.

Users performing the variation with the avatar and stereo vision (SLA) were able to perceive the scene in a spatially calibrated and metrically correct fashion. In this variation they were placed in front of the screen at a distance that allowed them to perceive the character at the approximate distance of $1m$ away.

The fourth variation, user-perspective direct interaction (SLU), was similar to SLA but with the virtual avatar not being used and instead the user directly interacted with the virtual scene. The working plane and targets were thus perceived by the participants as directly floating in front of them, enabling the participants to directly perform pointing actions toward the targets. The virtual pointer was rendered in the scene always floating at $10cm$ in front of the interaction controller and the user was asked to drive the virtual pointer towards the center of each target.

The task execution started with the participant standing in a comfortable rest position with arms down along the body. A new target would appear by pressing a button in the Wii-mote interaction device. The user was then required to move his or her right arm until the avatar's index finger would touch the center of the target. The user was asked to pay particular attention on precisely reaching the target's center. When the user was satisfied with the positioning, pressing the interaction button again would complete the task. These steps were then repeated ten times per user. The targets appeared in five different locations, regularly distributed in the surface, but not following any pattern so that the targets were perceived to be randomly placed.

4.1 Data Collected

In all variations except SLU, the motions of the avatar were collected and saved for analysis. Motion files were represented as time series of joint angles expressed locally in the hierarchical skeletal representation of the avatar. Given the initial scaling calibration parameters, it was possible to reconstruct both global positions and relative distances to virtual objects such as for measuring the error with respect to reaching the center of the targets. For each motion collected, the global trajectory generated by the fingertip of the right arm of the avatar was extracted for analysis.

For the user-perspective direct interaction variation (SLU) the body motion data were not used and instead the time-stamped global positions and orientations of the virtual pointer were extracted and collected per action performed. The motions of the participants and the virtual pointer trajectory were recorded at 30 frame per second for all performed activities.

Finger and pointer trajectories were segmented and clustered according to the five targets reached. We have noticed that trajectories exhibited two distinctive phases: an approach phase and an adjustment phase. Given a motion M, let t_{ph} denote its phase transition time point. The approach phase of the motion M_{ap} is the initial part of the motion where the user quickly moved his/her arm towards the target, and the adjustment phase M_{ad} is when the user spent time to adjust the end-effector on the target center as accurately as possible. Transition point t_{ph} was manually annotated per motion, considering the first occurrence of a sudden deceleration or change of direction in the trajectory. We have used a rule that analyzes if the group of frames inside 1 cm diameter sphere (usually 10 to 15 frames) show a sudden change in trajectory direction, of more than 45 degrees. In very few cases in SLU this point was not observed and the t_{ph} value received the full duration of the motion.

In order to investigate the performances we have considered several descriptive parameters such as trajectory durations, phase transition times, average velocities and accelerations, and distances to the target. Distances to the target were measured from the end-effector position (character's fingertip or virtual cursor) to the target center. An extract of the descriptive statistic analysis is shown in Table 3.

We have also considered trajectory profiles on dis-

TABLE 3: Experiment 1 descriptive statistics extract. Notation: t_e: overall trajectory duration, d_e: distance to the target at t_e, v_{avg}: average velocity, a_{avg}: average acceleration, t_{ph}: time of phase change, t_{phr}: time of phase change relative to t_e, d_{ph}: distance to the target at t_{ph}, d_σ: standard deviation of the target distance during M_{ad}. Except for the last row, the shown values are mean values with the standard deviation in parenthesis.

p.	unit	SLA	MDA	MLA	SLU
t_e	s	5.803 (2.06)	6.943 (1.34)	5.963 (1.24)	2.526 (1.28)
d_e	m	0.029 (0.04)	0.080 (0.31)	0.100 (0.04)	0.005 (.004)
v_{avg}	m/s	0.282 (0.08)	0.249 (0.02)	0.271 (0.04)	0.657 (0.12)
a_{avg}	m/s^2	3.420 (2.75)	2.287 (1.59)	3.043 (2.74)	19.03 (7.24)
t_{ph}	s	1.923 (0.69)	2.333 (0.46)	2.229 (0.37)	1.414 (0.36)
t_{phr}	%	38.22 (10.1)	38.88 (10.9)	41.73 (8.00)	64.24 (14.2)
d_{ph}	m	0.100 (0.06)	0.124 (0.05)	0.162 (0.03)	0.024 (0.01)
d_σ	m	0.064	0.085	0.074	0.016

Fig. 3: Left: each bar represents the average trajectory duration t_e for each variation in experiment 1. The horizontal segment mark depicts t_{ph} in relation to the average t_e. Right: each bar represents the average target distance d_e. In both graphs, the vertical line ranges show the standard error.

tance to the target, velocity, and acceleration at every frame. In order to generate comparable data we have uniformly time-warped and normalized the profile samples. The readings were grouped by participant repetition and by target with the purpose of analyzing if the target location would affect performances. Since the targets were placed in comfortable reaching positions we did not observe any significant difference related to target placement.

4.2 Results

From our initial analysis we have clearly found, as expected, that users were faster and more accurate when using the direct interaction configuration using user-perspective stereo vision (SLU). Figure 3 shows the differences in the average time, phase transition time and final distance to the target. In SLU users were almost six times more accurate than the best solution using avatars SLA (on the range of $5 \pm 4mm$ against $3 \pm 0.4cm$ in SLA) and they were twice as fast (around $2.5 \pm 1.3s$ against $5.5 \pm 2.0s$ average time). Considering the difference in phase transition time (t_{ph}) SLU participants were 5 times closer to the target's center already at point t_{ph}, in comparison to SLA, which is the best avatar-based configuration. At the end of the task they were 6 times closer to the targets than in SLA. This fact explains the main difference in velocity and the interval velocity variance during the reaching phase with a subsequently shorter adjustment period. Figure 4 shows the normalized velocity profiles resulting from each trajectory clustered and averaged by variation where, clearly, SLU is faster than the other three methods and t_{ph} occurs closer to the t_e.

Considering the three variations with avatar interaction, Table 3 reveals that the configuration with stereo vision (SLA) offers the best results. Although the overall reaching time and velocity is similar and the adjustment phase can be comparable (only slightly
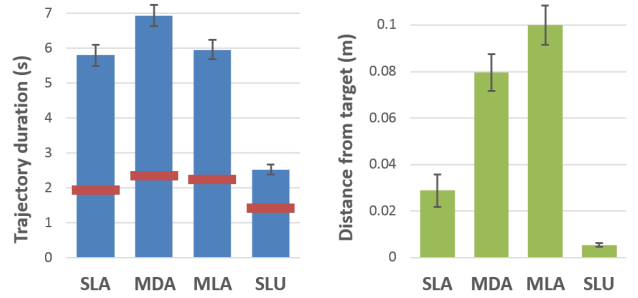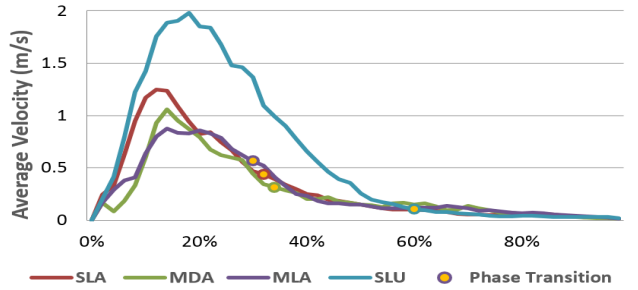

Fig. 4: Normalized velocity profiles grouped by variations (circles depict t_{ph}).

better in SLA), SLA participants generated motions that are three times more accurate than the other two variations. Moreover, the reaching phase during SLA led to a shorter adjustment phase where the overall standard deviation distance was lower. Surprisingly, it is possible to observe that the two variations with 2D vision are very similar in terms of accuracy, with the accuracy achieved in MLA being slightly lower than in MDA. But the group using the small desktop display (MDA: $0.249 \pm 0.02m/s$) performed each task slower than the group interacting with the large display (MLA: $0.271 \pm 0.04m/s$).

Considering the data gathered from the post-experiment questionnaire (expressed in a Likert scale from 1 to 7), we noticed that participants belonging to MDA expressed a lower level of confidence, in terms of accuracy and awareness of being precise, in comparison to the participants using the large display (the system accuracy perceived in MDA was 3.6 while in the other three variation was higher that 5.0). Most of the participants believed that the limitation was given by the setup and they needed extra care to perform the task precisely. It is interesting to notice that, on the contrary, group MLA (6.1) expressed a level of confidence similar to SLU (6.0) and higher that SLA (5.1), while their real averaged accuracy level was
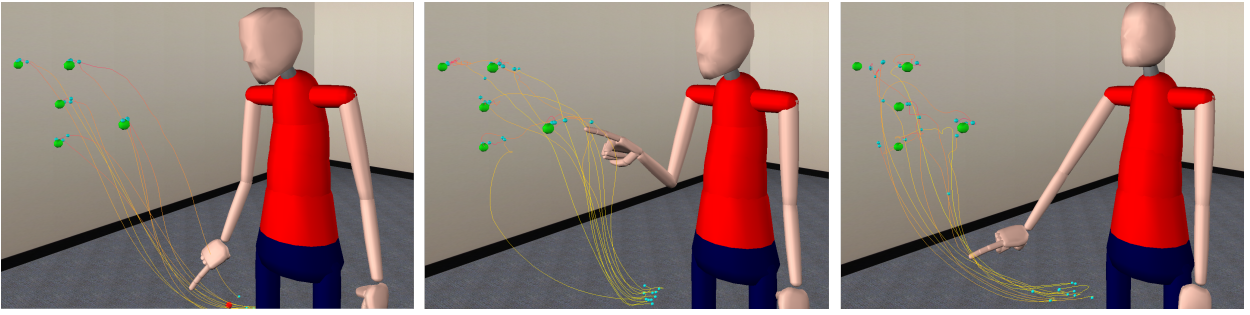
Fig. 5: Example trajectories collected from one participant in experiment 1. Large (green) spheres represent the targets, and small (blue) spheres represent the start, the end, and point t_{ph} in each trajectory. The show trajectories, in left-right order, were collected in configurations SLU, SLA and MLA.

similar to MDA and SLA (around 6.0cm of precision).

To support these findings we performed a between subjects One-Way analysis of variance (ANOVA) to compare the effect of stereo vision, avatar and display size on the participant's performances (expressed by the following dependent factors: t_e, t_{ph}, t_{phr}, d_e, d_{ph}, v_{avg}, a_{avg} during phases M_{ap} and M_{ad}) in SLA, SLU, MLA, and MDA conditions. The test for normality, examining standardized skewness and the Shapiro-Wilks test, indicated the data to be statistically normal. The Homogeneity test (Levene's test) also reported non–significant variance between the groups. An alpha level of $.05$ was used for the analysis and the post-hoc analysis was performed using a standard Tukey-HSD test for comparison of means and Bonferroni's correction.

The analysis showed that there was a statistically significant difference between groups for several factors. Considering the averaged time t_e to reach each target ($F(3, 36) = 15.909, p < .001$), averaged velocity v_{avg} ($F(3, 36) = 22.169, p < .001$), averaged acceleration a_{avg} ($F(3, 36) = 11.428, p < .001$), phase transition time t_{ph} ($F(3, 36) = 6.992, p < .001$) and relative t_{phr} ($F(3, 36) = 12.46, p < .001$), SLU resulted in faster and shorter motions with respect to the other three variations (Means and Standard Deviations are reported in Table 3). Considering the distance to the target d_e at the end of the averaged trials ($F(3, 36) = 15.253, p < .001$) the ANOVA showed that there is a significant separation between the performances of users under stereo vision (SLU: Mean (M) $= 0.005$, Standard Deviation (SD) $= 0.004$; SLA: $M = 0.029$, $SD = 0.04$) with respect to mono vision (MLA: $M = 0.100$, $SD = 0.04$; MDA: $M = 0.080$, $SD = 0.31$). Similarly, the same subdivision could be found already during the transition phase d_{ph} ($F(3, 36) = 16.974, p < .001$).

Even though the subdivision between the use of large visualization system versus the desktop setup seemed to be affecting performances, it could not be significantly stated.

4.3 Discussion

Based on the results, our observations, and on the comments of the participants we have drawn several inferences from the evaluation of reaching accuracy.

As expected, users were faster and more accurate when in user-perspective vision with direct interaction (SLU). Users were 6 times more accurate and 2 times faster than the second best case scenario (SLA). In addition, participants were 4 times closer to each target already at the end of the approaching phase and consequently they needed less time to reach each target. These numbers indicate that the use of the avatar increased the cognitive load of the participants, since they had to transfer their arm movement attention to avatar space. Considering these findings we can conclude that direct interaction with user-perspective stereo vision is a better choice for precision tasks that depend on environment constraints such as reaching for a target.

We can also observe a significant difference in accuracy among the configurations employing avatars. Users were 3 times more accurate with the use of stereo vision than in mono visualization. In addition, users also approached targets more accurately (20% closer), resulting in adjustment motions more focused around the target center area. The execution times and the overall execution velocities were similar across the three variations using the avatar. User perspective stereo vision seemed to improve the space understanding even when the task is transferred to avatar's space.

With respect to the difference between small and large displays we cannot state any significant conclusion. The data however show a trend towards the conclusion that the groups using large displays do not gain any benefits in terms of accuracy, while their perception of being accurate can be slightly compromised. Participants performing reaching tasks in avatar space with mono vision and small display (MDA) were, on average, 10% slower than users performing the same task using a large display (MLA). Similarly, the trajectories generated were 15% longer. However, participants using the small display showed an increase in precision of about 25% in comparison with the users using the large display. Participants reported that the task performed in this condition was uncomfortable and they needed extra care and attention. On the contrary, participants using the large display felt overconfident on judging their perfor-

mances. They believed to have precisely reached targets, and spent less time during the adjustment phase, resulting in a less accurate final position. This misestimation and overconfidence resulted in a shorter adjustment phase than the variation using the small display (around 7% less), consequently resulting in a less accurate final position. In light of these findings, it is interesting to notice a trend towards the conclusion that reaching interactions are affected by display size differently than other types of interactions, since Ni et al. [26] have reported that large displays improve user performances during navigation and element search in virtual environments. A deeper evaluation should be performed in order to isolate and possibly quantify how different types of interactions are affected by different screen sizes.

We have also investigated the visual aspects of the generated trajectories and body motions. Trajectories generated using the virtual pointer were smooth with a typical "S-shape" defined by the user raising and approaching the target from a frontal point (Figure 5-left). In most cases the user did not pay attention to the environment and intersections with the supporting virtual blue plane would often occur during the approach phase.

Considering the trajectories generated from the avatar motions, we noticed that SLA resulted in more natural (human-like) motions. Participants paid more attention to driving the fingertip of the character towards the target from a frontal position and carefully avoided the supporting surface (Figure 5-center). The pattern observed was: users first raised their hands with an elbow flexion to naturally avoid the virtual plane, and then approached the target from the front. In the variations adopting mono vision (Figure 5-right), on average, participants did not consider the virtual space occupied by the avatar. The observed pattern was: the avatars arm was first raised to the target's height, without bending the elbow, the avatars hand was then retracted until the fingertip was in front of the target, and then adjusted towards the target position. A mid-term result can be observed in configuration SLU (Figure 5-left).

These observations show that the coupling of avatar and stereo vision was optimal in having users pay attention to the upper-body motion displayed by the avatar. Users made the effort to produce a realistic motion instead of simply focusing on maneuvering a pointer to reach targets.

5 EXPERIMENT 2: MOTION REPRODUCTION

The second experiment investigated if and how avatars and user-perspective stereo vision affected the spatial understanding of motions to be reproduced. The experiment had two phases: the demonstration phase and the reproduction phase. In the demonstration phase a blue virtual character (the "tutor")

appeared in front of the user and demonstrated a predefined upper-body motion. Later in the reproduction phase the user was then asked to reproduce the observed motion (see Figure 6).
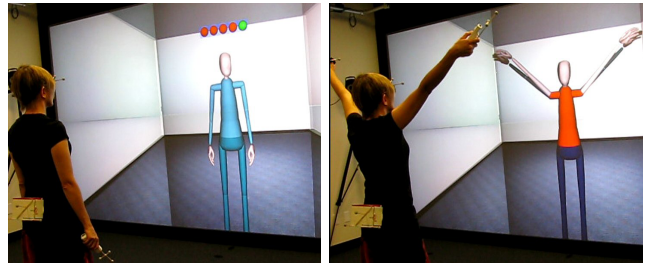

Fig. 6: Experiment 2 investigated the influence of avatars and stereo vision during motion observation (left image) and reproduction (right image).

Before each motion demonstration participants were instructed to memorize the motion they were going to observe, and to pay attention to details like: motion speed, arm key poses, final height of the hands, and torso orientation. The participants were allowed to watch the demonstrated motion up to three times but they were not allowed to move and simulate the task with their bodies.

The demonstrated motion, or reference motion (M_r), was designed to be simple and not ambiguous. It consisted of three arm raises. Each raise started by raising both arms simultaneously from the rest posture until the hands surpassed the head, with the elbows straight, and then the arms would return to the rest posture. First a lateral raise with arms parallel to the coronal plane was performed, then followed a frontal raise parallel to the tutor's sagittal plane, and then followed a raise exactly in-between the lateral and frontal raises.

During the reproduction phase the participants then reproduced the motions together with the virtual tutor. When the user's avatar was employed, a red avatar was displayed mimicking the user's motions in real-time as in a virtual mirror. In this case the tutor avatar was rendered transparently, overlapping (in fact slightly behind) the user's avatar. The tutor and avatar motions were still clearly distinguishable. Figure 6-right shows both avatars being displayed, with the tutor's arms visible slightly below the arms of the user's red avatar.

In order to prepare the participants to promptly start the reproduction fairly in sync with the tutor, a five seconds traffic light was displayed. The participants were informed that the virtual tutor would start to move immediately after the green light.

Similarly to the previous experiment, forty participants took part on the experiment and it was performed in four variations (10 participant per variation), as described in Table 4. The variations covered the joint combination of employing or not the avatar

and the user-perspective stereo vision.

TABLE 4: Configurations of experiment 2 (G = Group, M = Male, F = Female).

Label	G	M	F	Stereo	Avatars
SLA	4	5	5	yes	user's avatar and virtual tutor employed
SLT	1	3	7	yes	only tutor employed, no avatar
MLA	2	4	6	no	user's avatar and virtual tutor employed
MLT	5	3	7	no	only tutor employed, no avatar

5.1 Data Collected

The full motions performed by the participants were recorded for each variation. For each motion, we extracted the trajectories generated by the wrist joints in global coordinates. We denote t_e as the duration of a trajectory (motion) in time, and t_{p1} and t_{p2} as the time values that divide the trajectories in the three distinctive motion phases: lateral raise M_l, frontal raise M_f and intermediate raise M_i. Values t_{p1} and t_{p2} were manually annotated for each motion and in the reference motion.

The performances were analyzed in terms of time, distance and velocity differences when comparing each recorded trajectory against M_r, and after aligning the trajectories according to each cyclic phase. The time difference per phase, here denoted as phase synchronization time, was calculated by subtracting t_{p1}, t_{p2} and t_e from their M_r counterparts. The obtained differences are denoted as t_{d1}, t_{d2} and t_{de}.

Distance and velocity errors were calculated using the following procedure: each performed trajectory was subdivided and time-aligned (by uniform time warping) with the corresponding segments M_l, M_f and M_i of the reference motion. Each trajectory segment was also uniformly re-sampled. We denote the re-sampled trajectories as S_l, S_f and S_i. Each trajectory was compared with the corresponding samples in M_r in order to achieve the error values. Distance errors between corresponding samples are denoted as S_{dl}, S_{df} and S_{di}; and velocity errors are denoted by S_{vl}, S_{vf} and S_{vi}.

Since the reference motion was designed to be simple and symmetric, as expected, the difference between the left and the right samples was not significant. For this reason we omit the disjoint analysis between the left and right samples and report only the averaged left-right investigation.

5.2 Results

As shown in Table 5, in our initial descriptive statistics we averaged the participant's readings by category (Variations) and performed a joint analysis between the independent categories (Factors).

From the grouped variations sub-table (Table 5-left) we can infer that participants in the variations with the avatar (MLA and SLA) showed a better phase

synchronization than users in the other variations. This effect is also reflected by the velocity error and partially by the distance error. In this last case, SLA still shows the best result while MLA users were less accurate. Similarly we can infer the same pattern from the joint factors sub-table (Table 5-right) where, except for the distance error, during variations using avatars and variations using stereo vision, users had the best performances (time phases synchronization and less velocity profile error).

This can also be observed in the trajectories depicted in Figure 7. When stereo vision was used the motions were closer to the reference motion, in particular, phase M_i (arms raising diagonally) was better perceived by the participants. Moreover, looking at the trajectories generated in variation SLA, compliance to the reference motion is maintained similarly to the case of using stereo but without the avatar (SLT), however the height and key poses were better respected in SLA.

Considering the velocity profiles (Figure 8) we notice that both variations showing the avatar were overall closer to the reference motion profile than the other variations. It is also interesting to notice that in all variations participants anticipated the start of the first phase. This fact might have been driven by the implementation choice of using the virtual traffic light, as they felt driven to start right after the green light appeared instead of waiting for the virtual tutor to start the motion.

Analyzing the average distance compliance with M_r we can infer that participants using the system in mono vision performed in the least accurate fashion, $10cm$ difference on average per section. On the contrary, if we consider the factors combination sub-table this subdivision disappears. An explanation for this behavior can be given considering the joint performance of mono vision and without the avatar, which represents a non-optimal configuration.

From the data collected in the post-experiment questionnaire participants equally rated, in all the variations, their confidence of having perceived the motion correctly. In terms of reproduction, participants under the conditions without the avatar felt to be slightly less accurate in terms of position and speed (on average 15% less). Conversely, the presence or absence of stereo vision did not affect the users' level of confidence.

After this preliminary study a two-factor (2x2) ANOVA (SS Type III) was performed to evaluate the effect of the visualization type and the presence of the avatar on the user performances represented by each accuracy measures: t_{d1}, t_{d2}, t_{de}, S_{dl}, S_{df}, S_{di}, S_d (average of the previous three values), S_{vl}, S_{vf}, S_{vi}, and S_v (average of the previous three values). The test for normality, examining standardized skewness and the Shapiro-Wilks test indicated the data to be statistically normal. In order to meet the Homogene-

TABLE 5: Experiment 2 descriptive statistics summary. Averaged groups performances (left) and factor combinations (right): t_{de}, t_{d1} and t_{d2} denote the end of phase difference in time between the user's motion phase and the reference motion phase M_r; S_{vl}, S_{vf} and S_{vi} denote the average velocity and S_{dl}, S_{df} and S_{di} denote the distance error per phase from the reference motion M_r. The shown values are mean values with the standard deviation in parenthesis.

|	SLA	SLT	MLA	MLT		stereo	mono	avatar	tutor
phase sync. (s)					**phase sync. (s)**				
t_{d1}	0.384 (0.297)	0.702 (0.486)	0.568 (0.490)	1.591 (0.963)	t_{d1}	0.543 (0.425)	1.079 (0.910)	0.476 (0.406)	1.146 (0.871)
t_{d2}	0.533 (0.191)	0.836 (0.689)	0.908 (0.441)	2.320 (1.232)	t_{d2}	0.684 (0.516)	1.614 (1.156)	0.720 (0.383)	1.578 (1.234)
t_{de}	0.396 (0.280)	1.243 (0.973)	1.167 (1.037)	2.923 (2.076)	t_{de}	0.819 (0.821)	2.045 (1.892)	0.781 (0.958)	2.083 (1.798)
average	0.437 (0.114)	0.927 (0.548)	0.881 (0.531)	2.278 (1.199)	average	0.682 (0.460)	1.579 (1.159)	0.659 (0.455)	1.602 (1.142)
vel. error (m/s)					**vel. error (m/s)**				
S_{vl}	0.118 (0.069)	0.246 (0.068)	0.210 (0.128)	0.210 (0.162)	S_{vl}	0.182 (0.093)	0.210 (0.142)	0.164 (0.111)	0.228 (0.122)
S_{vf}	0.080 (0.056)	0.089 (0.061)	0.163 (0.126)	0.149 (0.122)	S_{vf}	0.084 (0.057)	0.156 (0.121)	0.121 (0.104)	0.119 (0.099)
S_{vi}	0.272 (0.078)	0.308 (0.143)	0.247 (0.114)	0.347 (0.166)	S_{vi}	0.290 (0.114)	0.297 (0.148)	0.259 (0.096)	0.327 (0.152)
average	0.156 (0.039)	0.214 (0.073)	0.206 (0.095)	0.235 (0.135)	average	0.185 (0.064)	0.221 (0.114)	0.181 (0.075)	0.224 (0.106)
dist. error (m)					**dist. error (m)**				
S_{dl}	0.174 (0.055)	0.173 (0.069)	0.243 (0.174)	0.283 (0.213)	S_{dl}	0.173 (0.061)	0.263 (0.190)	0.208 (0.130)	0.228 (0.164)
S_{df}	0.207 (0.079)	0.208 (0.047)	0.296 (0.124)	0.301 (0.139)	S_{df}	0.207 (0.063)	0.298 (0.128)	0.254 (0.112)	0.252 (0.111)
S_{di}	0.216 (0.065)	0.228 (0.083)	0.371 (0.205)	0.294 (0.163)	S_{di}	0.222 (0.073)	0.332 (0.185)	0.293 (0.168)	0.261 (0.131)
average	0.197 (0.037)	0.203 (0.034)	0.305 (0.154)	0.292 (0.146)	average	0.200 (0.035)	0.298 (0.146)	0.251 (0.122)	0.247 (0.113)
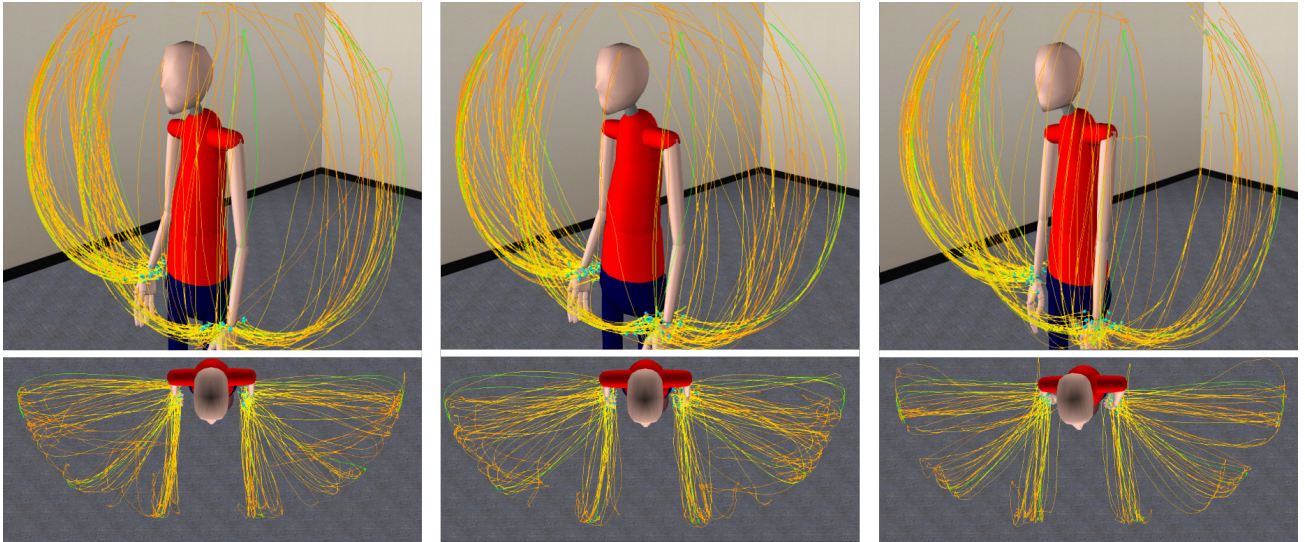
Fig. 7: Trajectories collected from the character's motions during experiment 2. The green trajectory shows the reference motion M_r. The small blue spheres represent points t_{p1} and t_{p2}. Left images: SLT shows fairly separated phases but the heights of the trajectories did not well correspond to the heights in M_r. Center images: MLA shows merged trajectories between phases M_i and M_l. Right images: SLA shows the best results, with the separation space and height traversed by each phase being closest to M_r.

ity assumptions for performing a two-way ANOVA the data were transformed using a standard natural power transformation (p-values reported below resulted from not significant tests for homogeneity of variance). The reported estimated means have been back-transformed to reflect the original data and the standard deviation has been reported as interval (SDI) due to the non-linear back transformation. An alpha level of .05 was used for the initial analysis.

Considering the distance compliance with the reference motion M_r, the results for the two-way ANOVA indicated a significant main effect for the visualization type, per phase (S_{dl}: $F(1, 36) = 5.755, p = .022$; S_{di}: $F(1, 36) = 7.360, p = .009$) and overall ($S_d$: $F(1, 36) = 10.588, p = .002$). A review of the group means for

the averaged distance factor (S_d) indicated that the error of the group using user-perspective stereo vision ($M = 0.197$ $SDI = [0.175, 0.221]$) had a significantly lower level of error than the group interacting without stereo ($M = 0.258$ $SDI = [0.229, 0.291]$). The analysis confirmed our initial conclusions and we can state that user-perspective stereo vision resulted in motions with significantly higher compliance to the reference motion.

In addition, we examined the participants capability of respecting key poses in time with the reference motion M_r. Both visualization type and the presence of the avatar (disjointly) showed main effects. The visualization type produced a main effect per phase (t_{d1}: $F(1, 36) = 5.755, p = .022$; t_{de}: $F(1, 36) = 9.280, p = $

.004) and overall (t_d: $F(1, 36) = 21.016, p < .001$). Considering the averaged estimated means we can infer that participants in user-perspective condition better respected key times ($M = .654 \ SDI = [.345, .936]$) in respect to the other groups without ($M = 1.475 \ SDI = [1.184, 1.767]$). Similarly, the presence of avatar, also produced main effects per phases (t_{d1}: $F(1, 36) = 6.870, p < .013$), t_{de}: $F(1, 36) = 16.416, p < .001$) and overall ($t_d$: $F(1, 36) = 24.942, p < .001$)). The presence of avatar ($M = .555 \ SDI = [.264, .847]$) helped the user to better respect the original motion key poses with respect to time in respect to the groups without the avatar ($M = 1.565 \ SDI = [1.274, 1.856]$).

Considering the joint factor analysis, in phase t_{d1} and t_{de}, the effect trended toward significance (t_{d1}: $F(1, 36) = 3.225, p < .080$ and t_{df}: $F(1, 36) = 3.710, p < .062$). Similarly S_{dl} and S_{df} also trended toward significance (S_{dl}: $F(1, 36) = 3.449, p < .071$ and S_{df}: $F(1, 36) = 3.527, p < .068$). Looking at the estimated mean comparisons it seems that avatar use (both in time and distance) improved the reproduction of motions when only mono vision was used. However, since there is no statistical significance and the interaction appeared only on subsets of the data, the main estimates have not been dropped.

Finally, considering the velocity profile maintenance ANOVA reported only a trend on the presence of the avatar main effect on the phase S_{vl} ($F(1, 36) = 3.449, p < .071$). Even though this conclusion could not be supported statistically, by analyzing the motions visually it is possible to notice that in the presence of the avatar participants reached more accurately (in terms of distance and timing) peaks and key points, resulting in small partial velocity errors. Changes in acceleration could be however noticed in a few trials, due to users assessing the virtual character's motion and trying to catch up with the tutor. The motions appeared less fluid in these cases.
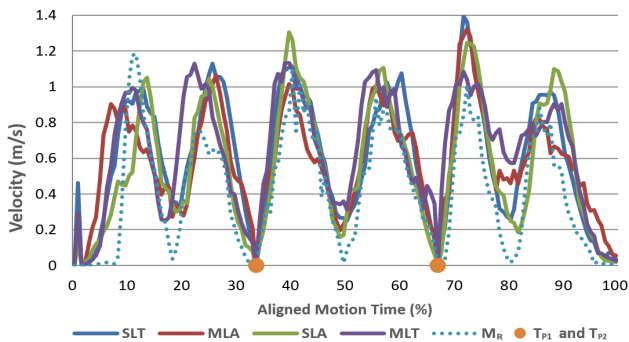

Fig. 8: Aligned velocity profiles of reference and replicated motions in each variation (orange dots depict t_{d1} and t_{d2}).

5.3 Discussion

Our study provides evidence that both the use of avatars and stereo vision positively affect motion reproduction tasks.

The presence of avatar and stereo vision both improve the ability to preserve spatial alignments when reproducing motions. Training applications represent a typical scenario where both these conditions would be useful. Accurate motion perception is particularly important when demonstrated motions have relationships with the environment or objects, in which cases key poses will have to be well perceived in order to achieve effective training.

We can also observe that the avatar helps driving the improvement of the motion reproduction when stereo vision is not employed. Considering a scenario where motion reproduction is performed without stereo vision, displaying an avatar is a key recommendation. This is for instance the case of several applications related to delivery of physical exercises at home, where it is difficult to implement stereo vision.

6 EXPERIMENT 3: MOTION MODELING

The third experiment investigated the usability of a complete interface for immersive motion modeling by direct demonstration (see Figure 9). In this experiment participants were asked to perform motions to be performed by a virtual character.
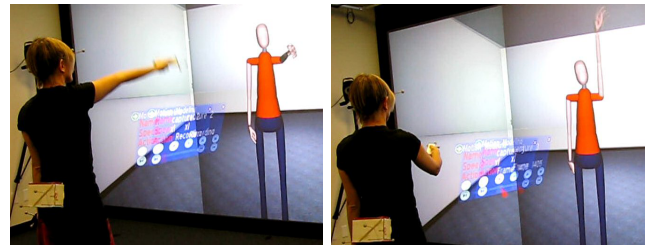

Fig. 9: Experiment 3 evaluated an immersive motion modeling by demonstration interface. Motion recording with or without avatar (left image) and playback (right image) were evaluated.

Two variations of the system were considered by varying the use of the avatar, as summarized in Table 6. Given the focus on investigating the usability of the system, and that detailed trajectory analysis was already explored in the previous experiments, only questionnaires were used for evaluation.

TABLE 6: Configurations of experiment 3 (G = Group, M = Male, F = Female).

Label	G	M	F	Description
SLN	5	3	7	*Recording phase without the avatar*
SLA	4	5	5	*Recording phase with the avatar*

Similarly to experiment 2 the scenario included a blue virtual tutor character placed in front of the user. The tutor was used only at the beginning of the experiment to demonstrate to the user an upper-body motion composed by a sequence of simple gestures

(arm raises, elbow bends, arm pointings, etc.). At this stage the participant was required to memorize the overall motion until satisfaction, being allowed to move and replicate the motion physically to help with the memorization. The sequence of motions included both the left and the right arm.

After this preliminary step the user was asked to use the virtual user interface to model the motion previously observed. The motion was modeled by direct recording via the motion capture markers. The tutor character was hidden in all variations. In variation SLA, the red user's avatar was displayed during recording. In variation SLN the avatar was hidden during motion recording, and the participant was required to record motions without any visual feedback.

The interaction with the graphical interface consisted of a virtual pointer floating in front of the user's hand (in front of the Wii-mote controller). Two buttons from the controller were used to select actions from virtual panels and to show/hide the interface at will. Since the implementation used user-perspective stereo vision the interface was perceived as floating in front of the user and it could be repositioned in space at any time to avoid occlusions with the scene or to be better approachable. The interface provided tools for recording, playing, trimming, and inspecting specific portions of the recorded motions. The user could also model and store more than one motion until satisfaction, with the possibility to re-select or discard motions.

6.1 Data Collected

At the end of each trial, a questionnaire was administered about the usability, preferences, user experience, and also asking for suggestions for improving the interface and the overall approach. Except for a few open questions requesting feedback, the questionnaire consisted of seven-point Likert-scale items. See Appendix A for an excerpt of the questionnaire. The full motions saved by the participants were also stored but they were only used to validate if the users performed all the motions required to be modeled.

6.2 Results and Discussion

Since the task of modeling motions by demonstration required to handle a more complex interface and implied more steps and familiarization with the system, as expected, the task was rated by the participants to be more difficult than the tasks in the previous experiments. Looking at the control questions, when asked about confidence on completing the task correctly and confidence on being able to coach the task to someone else, we observed a 15% decrease of confidence with respect to experiments 1 and 2. In experiments 1 and 2 the average level of confidence was 6.4 out of 7, while in Experiment 3 it was 5.7 out of 7.

Considering the targeted questions and the open comments the users positively rated the overall experience. The average rating for the question "Rate how comfortable you felt during the performance in all the aspects of the activity (1=extremely uncomfortable, 7=extremely comfortable)" was 5.5.

Figure 10 summarizes results of selected questions from the post-activity questionnaire. Q6 highlights the fact that users in SLN wanted additional training before engaging in the motion modeling activity. Comments and suggestions also noted that they did not know how to behave when the recording phase started. Before the beginning of each recording phase the system warned the user with a timer countdown with textual messages, but this was not sufficient for them to grasp the activity in their first attempt. After a few trials they were then able to record motions correctly. This fact was also reflected by question Q14 of the SLN questionnaire, where users expressed the hypothesis that having an avatar during the recording session would have improved the usability of the system (6.29 out of 7). Users performing in SLA felt that the avatar helped them to better understand their motions (Q15-SLA 6.57 out of 7) and they did not feel distracted by it (Q14-SLA 6.57 out of 7). These results are consistent with experiments 1 and 2.
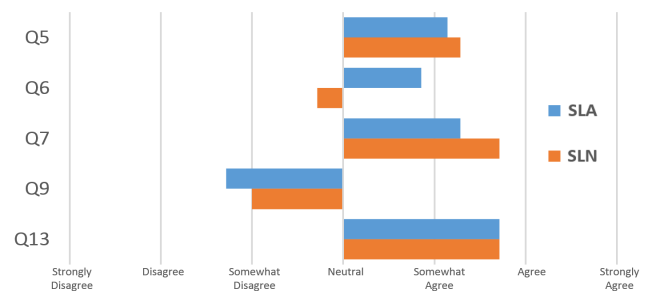

Fig. 10: Results from selected usability questions for Experiment 3. The corresponding questions are available in Appendix A.

7 CONCLUSIONS

The presented experiments have uniquely studied the effectiveness of avatars and user-perspective stereo vision during task performance. Our results have shown the viability of the approach of direct motion demonstration for modeling and reproducing motions. The correct use of avatars has showed to have great potential to improve performances in a number of situations; however, we have also observed that there are critical design choices that highly influence the suitability of the configurations to different types of interaction needs.

Our experiments confirm that the use of user-perspective stereo vision with direct interaction is the optimal choice in terms of task accuracy and completion time, when precision tasks are involved. Direct interaction made users to be 6 times more

accurate and 2 times faster than in other conditions. For example, in the widely adopted scenario of stroke rehabilitation, tasks often involve repeatedly reaching regions in space. In such cases user-perspective stereo vision will lead to better accuracy when measuring the rehabilitation progress.

User-perspective stereo vision also improved replication of spatial relationships even when the task was transferred to the avatar's space (by a factor of 3). When the task involved motion reproduction, stereo vision showed improvements both in terms of synchronization and compliance with the reference motion. The use of avatars produced increased attention to the avatar space, allowing users to better observe and address motion constraints and qualities. Coupling avatar use with stereo vision resulted in users paying more attention to the motions within the virtual environment, improving the realism and correctness of the motions. These findings represent key factors to consider when designing applications for distant training in collaborative spaces where the kinesthetic component is fundamental with respect to the environment.

In addition, the avatar use showed a trend towards the possibility to improve motion reproduction in the cases where stereo vision was not present. This factor suggests that in applications where stereo vision is not practical to be used (such as in homes or clinics), the use of avatars can improve the understanding of motions displayed by virtual tutors.

In summary, if the involved tasks require generation or reproduction of motions with desired qualities, such as in training applications where the environment or objects and tools are key factors, the use of avatars and stereo vision will improve that ability. However, if the goal is to accomplish tasks no matter the type of motions used, direct interaction in user-perspective will be more efficient. Besides the additional instrumentation (stereo glasses), stereo vision has showed to be always beneficial.

The above conclusions were found to be statistically significant and additional important trends and observations were also made. We have noticed that the small display induced users to not overestimate their capabilities during precision tasks. However, the small display increased their execution time and frustration. Although further investigation should be performed to fully support this theory, a new guideline on designing gesture-based game controllers can be drawn. Given that the dimensions of the target visualization system was observed to affect user expectations, game controller precision and other difficulty settings could be dynamically adjusted with respect to display size in order to manage user frustration.

Overall, the presented results provide important new quantification and observations in each of the performed experiments, leading to a new understanding of the tradeoffs involved when designing avatar-based training systems.

APPENDIX A
QUESTIONNAIRE OF EXPERIMENT 3

Extract of the questionnaire administered in Experiment 3 (likert-scale rated between 1=strongly disagree, 7=strongly agree):

Q5) The interface was simple and easy to understand.
Q6) The interface could be used without training.
Q7) 3D vision is important to model motions.
Q9) I would have preferred to use a standard computer.
Q13) The approach to model motions was effective.
Q14*) The avatar was distracting while recording.
Q14**) Seeing my motions while recording would have helped me.
Q15*) The avatar helped me to be more precise.
(*SLA only, **SLN only.)

ACKNOWLEDGMENTS

REFERENCES

[1] P. L. Alfano and G. F. Michel. Restricting the field of view: perceptual and performance effects. *Percept Mot Skills*, 70(1):35–45, Feb 1990.

[2] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice. YouMove: enhancing movement training with an augmented reality mirror. In *Proceedings of User Interface Software and Technology (UIST)*, pages 311–320. ACM, 2013.

[3] K. W. Arthur. Effects of field of view on performance with head-mounted displays. Doctor of Philosophy, Computer Science. The University of North Carolina at Chapel Hill, 2000.

[4] A. P. Atkinson, W. H. Dittrich, A. J. Gemmell, and A. W. Young. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33(6):717–746, 2004.

[5] R. Ball and C. North. Effects of tiled high-resolution display on basic visualization and navigation tasks. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1196–1199, New York, NY, USA, 2005. ACM.

[6] R. Blake and M. Shiffrar. Perception of Human Motion. *Annual Review of Psychology*, 58(1):47–73, 2007.

[7] J. Blascovich and J. Bailenson. *Infinite Reality - Avatars, Eternal Life, New Worlds, and the Dawn of the Virtual Revolution*. William Morrow, New York, 2011.

[8] H. Brenton, M. Gillies, D. Ballin, and D. Chatting. The uncanny valley: does it exist. In *19th British HCI Group Annual Conference: workshop on human-animated character interaction*, 2005.

[9] J. Broeren, M. Rydmark, and K. S. Sunnerhagen. Virtual reality and haptics as a training device for movement rehabilitation after stroke: a single-case study. *Archives of physical medicine and rehabilitation*, 85(8):1247–1250, 2004.

[10] C. Camporesi, Y. Huang, and M. Kallmann. Interactive motion modeling and parameterization by direct demonstration. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA)*, 2010.

[11] T. Chaminade, J. Hodgins, and M. Kawato. Anthropomorphism influences perception of computer-animated characters's actions. In *Social Cognitive and Affective Neuroscience*. Books (MIT Press, 2007.

[12] J. E. Cutting and L. T. Kozlowski. Recognizing Friends By Their Walk - Gait Perception Without Familiarity Cues. *Bulletin Of The Psychonomic Society*, 9(5):353–356, 1977.

[13] P. Dukes, A. Hayes, L. Hodges, and M. Woodbury. Punching ducks for post-stroke neurorehabilitation: System design and initial exploratory feasibility study. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 47–54, March 2013.

[14] J. Han, G. Kurillo, R. Abresch, E. de Bie, A. Nicorici Lewis, and R. Bajcsy. Upper extremity 3d reachable workspace analysis in dystrophinopathy using kinect. *Muscle & nerve*, 2015.

[15] W. L. Johnson, J. W. Rickel, and J. C. Lester. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial intelligence in education*, 11(1):47–78, 2000.

[16] J. M. Knapp and J. M. Loomis. Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoperators and Virtual Environments*, 13(5):572–577, Oct. 2004.

[17] L. Kozlowski and J. Cutting. Recognizing the gender of walkers from point-lights mounted on ankles: Some second thoughts. *Perception & Psychophysics*, 23(5):459–459, 1978.

[18] G. Kurillo, T. Koritnik, T. Bajd, and R. Bajcsy. Real-time 3d avatars for tele-rehabilitation in virtual reality. *Studies in health technology and informatics*, 163:290296, 2011.

[19] J.-W. Lin, H.-L. Duh, D. Parker, H. Abi-Rached, and T. Furness. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Virtual Reality, 2002. Proceedings. IEEE*, pages 164–171, 2002.

[20] R. McDonnell, M. Breidt, and H. H. Bülthoff. Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.*, 31(4):91:1–91:11, July 2012.

[21] R. McDonnell, S. Jörg, J. McHugh, F. N. Newell, and C. O'Sullivan. Investigating the role of body shape on the perception of emotion. *ACM Transations on Applied Perception*, 6(3):14:1–14:11, Sept. 2009.

[22] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):626–633, Apr. 2012.

[23] R. P. McMahan, D. Gorton, J. Gresock, W. McConnell, and D. A. Bowman. Separating the effects of level of immersion and 3d interaction techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '06, pages 108–111, New York, NY, USA, 2006. ACM.

[24] A. Naceri, R. Chellali, F. Dionnet, and S. Toma. Depth perception within virtual environments: A comparative study between wide screen stereoscopic displays and head mounted devices. In *Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, 2009. COMPUTATIONWORLD '09. Computation World:*, pages 460–466, 2009.

[25] M. Narayan, L. Waugh, X. Zhang, P. Bafna, and D. Bowman. Quantifying the benefits of immersion for collaboration in virtual environments. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '05, pages 78–81, New York, NY, USA, 2005. ACM.

[26] T. Ni, D. A. Bowman, and J. Chen. Increased display size and resolution improve task performance in information-rich virtual environments. In *Proceedings of Graphics Interface 2006*, GI '06, pages 139–146, Toronto, Ont., Canada, Canada, 2006. Canadian Information Processing Society.

[27] D. Perez-Marcos, M. Solazzi, W. Steptoe, O. Oyekoya, A. Frisoli, T. Weyrich, A. Steed, F. Tecchia, M. Slater, and M. V. Sanchez-Vives. A fully immersive set-up for remote interaction and neurorehabilitation based on virtual body ownership. *Frontiers in Neurology*, 3:110, 2012.

[28] E. D. Ragan, R. Kopper, P. Schuchardt, and D. A. Bowman. Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE Transactions on Visualization and Computer Graphics*, 19(5):886–896, May 2013.

[29] P. S. A. Reitsma and N. S. Pollard. Perceptual metrics for character animation: sensitivity to errors in ballistic motion. *ACM Transactions on Graphics*, 22(3):537–542, July 2003.

[30] W. Steptoe, A. Steed, and M. Slater. Human tails: Ownership and control of extended humanoid avatars. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):583–590, 2013.

[31] R. Sternberg. *Practical Intelligence in Everyday Life*. Cambridge University Press, 2000.

[32] D. S. Tan, D. Gergle, P. Scupelli, and R. Pausch. With similar visual angles, larger displays improve spatial performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 217–224, New York, NY, USA, 2003. ACM.

[33] G. T. Thielman, C. M. Dean, and A. Gentile. Rehabilitation of reaching after stroke: task-related training versus progressive resistive exercise. *Archives of physical medicine and rehabilitation*, 85(10):1613–1618, 2004.

[34] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the quality of the computer graphics matter when judging distances in visually immersive environments. *Presence: Teleoperators and Virtual Environments*, 13(5):560–571, Oct. 2004.

[35] E. Velloso, A. Bulling, and H. Gellersen. Motionma: Motion modelling and analysis by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1309–1318, New York, NY, USA, 2013. ACM.

[36] C. Ware and P. Mitchell. Reevaluating stereo and motion cues for visualizing graphs in three dimensions. In *Proceedings of the 2Nd Symposium on Applied Perception in Graphics and Visualization*, APGV '05, pages 51–58. ACM, 2005.

[37] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *ACM Transactions on Applied Perception*, 6(2):8:1–8:14, Mar. 2009.

[38] B. G. Witmer and P. B. Kline. Judging perceived and traversed distance in virtual environments. *Presence: Teleoperators and Virtual Environments*, 7(2):144–167, Apr. 1998.

[39] Y. Wu, S. V. Babu, R. Armstrong, J. W. Bertrand, J. Luo, T. Roy, S. B. Daily, L. C. Dukes, L. F. Hodges, and T. Fasolino. Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *Visualization and Computer Graphics, IEEE Transactions on*, 20(4):626–635, 2014.

Carlo Camporesi is an R&D engineer at *Avatire* developing cloud compute solutions for distributed simulation and rendering. He obtained his PhD in computer graphics and animation from the University of California, Merced (UCM). He joined UCM in 2008 and during his research work established the UCM virtual reality facility. Before joining UCM he was a research fellow at the National Research Council of Italy (ITABC) in Rome, where he developed several virtual heritage projects. In 2010 he was a visiting research associate at the Institute of Creative Media at the City University of Hong Kong. His research interests are in virtual reality, computer animation, motion capture, distributed systems and human computer interaction.

Marcelo Kallmann is founding faculty and associate professor of computer science at the University of California, Merced. He holds a PhD from the Swiss Federal Institute of Technology in Lausanne (EPFL) and before joining UC Merced in 2005 he was a research faculty member at the University of Southern California (USC), and a research scientist at the USC Institute of Creative Technologies. His areas of research include computer animation, virtual reality and motion planning. In 2012 he was program co-chair for the 5th International Conference on Motion in Games, in 2013 he was a guest co-editor of the Computer Animation and Virtual Worlds Journal, and in 2014 he was associate editor for ICRA 2015.