Supporting motion-capture acting with collaborative Mixed Reality

Computers & Graphics 124 (2024) 104090

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on RAGI 2024

Supporting motion-capture acting with collaborative Mixed Reality

Alberto Cannavò *, Francesco Bottino, Fabrizio Lamberti

Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, Turin, 10129, Italy

ARTICLE INFO

Keywords:
Collaborative virtual production
Acting rehearsal and performance
Motion capture
Mixed reality
Visual effects

ABSTRACT

Technologies such as chroma-key, LED walls, motion capture (mocap), 3D visual storyboards, and simulcams are revolutionizing how films featuring visual effects are produced. Despite their popularity, these technologies have introduced new challenges for actors. An increased workload is faced when digital characters are animated via mocap, since actors are requested to use their imagination to envision what characters see and do on set. This work investigates how Mixed Reality (MR) technology can support actors during mocap sessions by presenting a collaborative MR system named CoMR-MoCap, which allows actors to rehearse scenes by overlaying digital contents onto the real set. Using a Video See-Through Head Mounted Display (VST-HMD), actors can see digital representations of performers in mocap suits and digital scene contents in real time. The system supports collaboration, enabling multiple actors to wear both mocap suits to animate digital characters and VST-HMDs to visualize the digital contents. A user study involving 24 participants compared CoMR-MoCap to the traditional method using physical props and visual cues. The results showed that CoMR-MoCap significantly improved actors' ability to position themselves and direct their gaze, and it offered advantages in terms of usability, spatial and social presence, embodiment, and perceived effectiveness over the traditional method.

1. Introduction

Technological improvements are enabling new ways of producing films [1]. For instance, recent advancements regarding computer-generated imagery (CGI) and visual effects (VFX), have significantly attracted the attention of many researchers and practitioners [2]. These technologies are no longer limited to science-fiction films but are increasingly utilized across various genres [3]. Although they have started to be massively integrated into the traditional film production pipeline, these technologies are posing new challenges for both technical and acting crew [4]. As a matter of example, it is possible to consider the scenario of films containing motion capture (mocap) shoots. In this way of acting, the recorded movements of actors are leveraged to animate computer-generated characters that are integrated into the filmed scenes during post-production [5]. In this scenario, actors are requested to perform while imagining digital contents that will be added only at a later stage, potentially altering the actual surroundings in a significant way [2].

Difficulties for the actors increase when they are requested to interact with other people of the crew on the set who are controlling/animating digital characters, whose appearances do not align with the real bodies or are entirely virtual [4]. In fact, using mocap technology it is possible to animate characters who do not present anthropomorphic characteristics or proportions. This aspect adds further complexity for actors who have to portray these roles controlling a differently shaped character by using only their own bodies and movements as a reference. Many actors express frustration when rehearsing mocap scenes, since what they are asked to perform on stage varies greatly from traditional acting methods taught in drama schools [6].

The effects of these difficulties become more evident during the production phase, when the actual context of the scene is introduced, as differences between the performance and the environment can lead to severe inconsistencies. For instance, actors may struggle to align their movements with the desired appearance or react appropriately to their surroundings. The majority of these problems are addressed during the rehearsal and shooting, by asking the actors to perform the same scene several times.

In such scenarios, actors are typically aided by mechanisms aimed at guiding their focus and actions toward placeholder props. These props are used to indicate the location and shape of the corresponding digital elements within the scene. However, physical constraints may hinder or make ineffective the use of such solutions [4]. In addition, using these mechanisms may limit the capability of actors to keep eye contact with the other people present on the set, i.e., other actors and staff. This limitation arises from the fact that actors are requested to look simultaneously at the other people and their virtual counterparts to effectively control their character's actions.

* Corresponding author.
E-mail addresses: alberto.cannavo@polito.it (A. Cannavò), francesco.bottino@studenti.polito.it (F. Bottino), fabrizio.lamberti@polito.it (F. Lamberti).

https://doi.org/10.1016/j.cag.2024.104090
Received 13 June 2024; Received in revised form 26 August 2024; Accepted 16 September 2024
Available online 19 September 2024

In addition to physical props, another common practice is the use of a technique based on laser pointing, especially when actors are requested to follow the movements of an object or a character. However, ensuring precise synchronization with pre-computed or real-time animations may be challenging and can potentially lead to inaccuracies in the performance [2]. These inaccuracies do not only extend the time needed for shooting the scenes but also pose challenges during post-production. Mismatches in actors' gazes often need extensive post-processing efforts to adjust the animations or to re-align the recorded movements [2].

Mixed Reality (MR) and Virtual Reality (VR) technologies may offer potential solutions to the above challenges. In the last years, the use of such technologies increased rapidly and became prominent for many companies [7,8]. Successful examples of their application have been confirmed for supporting film-making [9,10], digital storytelling [11], set configuration and visualization [12], scene pre-visualization [13, 14], and more. In scenarios including mocap acting, MR/VR technologies can be leveraged to provide actors with a preview of the scene that closely resembles the final product. The enhanced visualization allows actors to be more aware of the virtual surroundings, since it removes the need for them to imagine the virtual contents while they are performing. This capability has the potential to enhance actors' performance [15].

Integrating MR/VR does not only supports actors in performing more intuitively and effectively [2] but also reduces the need for extensive post-production effort, since it enhances the authenticity of their performance [15]. For instance, the ability to perceive the actual size of digital characters allows the actors to interact with them more naturally [2]. Moreover, these technologies enhance actors' ability to empathize with their surroundings and facilitate emotional connections with the characters by providing a clearer understanding of what the scene actually contains. The improved awareness enabled by these technologies can also be beneficial for directors, who are allowed to better communicate their creative vision and plans for the shooting session on set [2].

Despite the numerous benefits brought by the use of MR/VR technologies, research work is still needed to make them become commonplace in the cinema industry. Most of the works in the literature proposed VR-based solutions that risk to fall short when the environment in which the scene takes place is not fully digital. In this case, MR could represent a valid alternative to support the actors during the rehearsal, since it simplifies the operations needed to virtually reconstruct the environment (as the real environment could be leveraged) or track the movement of the objects used in the scene. Moreover, some actions could be more straightforward to simulate (such as knotting a rope and receiving accurate haptic feedback) or execute (like climbing a staircase) in MR compared to VR. Notwithstanding, also the use of MR-based solutions remains quite limited, probably due to technological constraints [16]. In fact, to the best of the authors' knowledge, MR-based solutions supporting mocap acting with multiple actors have not been proposed yet.

Based on these considerations, the present paper proposes a system named CoMR-MoCap designed to explore the use of MR for helping actors when rehearsing or shooting scenes involving mocap and VFX. The proposed system could not only have beneficial impacts on the actors' performance, but could also reduce the time and effort required to correct wrong behaviors during the shooting. Experiments were conducted with 24 participants to compare the proposed approach with the traditional one, based on the use of physical props and visual cues. Results showed the benefits of the MR method both in objective and subjective terms.

2. Related work

As anticipated in Section 1, the combination of mocap with MR and VR technologies has already been explored in the literature. For instance, the work in [17] used VR and mocap within a training context. More specifically, the authors introduced a VR system that can be used as a tool for self-learning basketball-related technical gestures. By making use of an affordable mocap suit, the real-time skeleton data of the user's arm is reconstructed in a virtual space, thus making it possible for him or her to compare own movements with reference gestures presented through a ghost metaphor. Similarly, the works in [9,18] proposed VR-based systems for training in Tai Chi movements. Mocap suits are leveraged also in these systems to track the movements of both the trainer and the trainee in real time. Within the virtual environment, the trainees can receive feedback on their performance and visualize the correct movements to be mirrored, in this case performed by the trainee.

Regarding the use of MR or Augmented Reality (AR) technologies with mocap, the work in [19] presented a self-learning tool for refining golf movements. The proposed system allows a trainee to compare his or her own movements tracked by an inertial mocap suit with pre-recorded movements performed by a trainer, visualizing them on a Microsoft HoloLens device. The authors of [20] combined an Optical See-through Head-mounted Display (OST-HMD) with an inertial mocap suit to reconstruct a virtual avatar that can be used to facilitate real-time and full-body interactions in an augmented environment.

Although the works reviewed so far confirmed the benefits of combining mocap with MR and VR, they are mainly focused on training and sports, or have general application. Moving to works in the literature which specifically targeted the cinema industry, it is worth mentioning the contribution reported in [12]. In this work, the authors introduced an MR system designed to support directors in validating the setup of film sets. In particular, the system allows directors to visualize and manipulate computer-generated assets within the real environment using an OST-HMD (for viewing) and a tablet device (for manipulation). The system showed numerous benefits for the directors, who can efficiently explore various configurations of a virtual scene before physically arranging objects within the physical environment. The work reported in [15] represents another example of combining the considered technologies to support tasks concerning the cinema industry. In this case, an Android application was proposed for low-budget film production, enabling actors to seamlessly transition between viewing the real environment, which includes green screen areas, and a synthetic environment generated by overlaying digital contents onto the green areas. A system supporting previsualization is proposed in [14]. The system enables the members of the staff operating with moving cameras (i.e., videographers) to test the trajectories to be followed during the shooting without the need for the real actors' physical presence on the set. To this aim, the system leverages the tracked position and orientation of the camera to overlay virtual avatars.

Considering works aimed at supporting the actors' performance in scene rehearsal or shooting, several solutions based on MR and VR have been recently proposed. As a matter of example, the work in [2] describes a VR system designed to assist actors in rehearsing scenes that include VFX. The VR technology was leveraged not only to immerse the actors within the virtual environment but also to make them experience interactions with the so-called "dynamic scenario" features (e.g., picking up objects, moving furniture, or adjusting lighting) all at their own pace. The system can be configured to make an actor rehearse a scene both individually or in a collaborative way, supporting both on-set and off-set scenarios. Another example is presented in [4]. In this work, the authors proposed a way to cope with the difficulties that actors may face when scenes to be performed include virtual characters presenting sizes different than human ones. In particular, the authors developed an immersive VR system that lets actors visualize their virtual avatar's body while simultaneously seeing other virtual characters, each with distinct sizes, from their own perspective.

Despite the possibility to achieve high-fidelity simulations and visualizations, the use of VR technology for scene rehearsal also comes with some issues. For instance, using VR for rehearsing scenes in

which actors have to interact with elements of the real environment would generally entail significant modeling and reconstruction efforts to prepare the required assets, since all the elements belonging to the environment (both the digital and the real one) have to be created. Moreover, to support actors' interactions with the objects, sophisticated tracking techniques should be implemented for maintaining the coherence between the digital and the real worlds. Furthermore, specific interactions, such as touching a fluid or manipulating a rope, may prove challenging to replicate accurately in VR. Similarly, actions that rely on actors utilizing the physical set, such as climbing stairs, may present difficulties in virtual environments [16].

To cope with the above limitations, works like [16,21] proposed to adopt MR technology for scene rehearsal. More specifically, the authors of [21] presented a rehearsal system that allows actors to visualize virtual characters through an OST-HMD using a first-person perspective. To demonstrate the capabilities of the system, a use case was proposed that represents a battle between two samurai. The first samurai was portrayed by a real actor wearing the OST-HMD, whereas the second one was a purely virtual character. The system also featured the accurate tracking of the real actor's sword and vibrotactile feedback (e.g., provided when the swords clash) to enhance the level of immersion and provide realistic sensory cues. The authors of [16], in turn, proposed a solution that combines MR and mocap. By wearing an OST-HMD (a Microsoft HoloLens device), an actor can visualize digital contents overlapped to the real world, thus facilitating interactions with virtual objects and characters (controlled with mocap) that could present variable sizes. The authors envisioned two possible use cases for their system, which differ based on who actually portrays the virtual character through mocap and who wears the OST-HMD. In the first use case, these are two distinct subjects; hence, the actor can see in MR someone else portraying the virtual character with mocap. In the second use case, the actor can see himself or herself augmented while portraying the virtual character. Although experiments demonstrated the validity of applying the devised rehearsal method in the first use case, the authors claimed they could not investigate the second use case due to technological constraints concerning the limited field of view of the Microsoft HoloLens device, which was cutting off from the augmented view digital elements close to the actor's point of view. Moreover, like in the previous work, the architecture of this system does not support collaborative interactions among multiple actors controlling virtual characters and visualizing the virtual environment at the same time.

The present paper explores the combination of mocap and MR technology, like in [9,17–20] but, with respect to these works, it focuses on the cinema industry. In particular, it proposes a method that can be leveraged by the actors for rehearsing scenes. Like in [2,4], the devised CoMR-MoCap system envisages a collaborative approach that allows multiple actors to interact within the same scene. However, differently than in these two works, the system relies on MR. Moreover, with respect to previous works which also leveraged MR for the considered task [16,21], the proposed system features a collaborative approach and overcomes technological limitations related to the field of view by adopting a VST-HMD.

3. CoMR-MoCap system

As said above, the present paper proposes a MR system named CoMR-MoCap designed to allow multiple actors wearing mocap suits and VST-HMDs (later also abbreviated HMDs) to visualize in real time the digital contents that are supposed to be integrated into the shooted scenes during post-production. Digital contents could potentially encompass virtual characters or elements animated either through prerecording or in real time utilizing mocap. This section aims to provide some details on the system architecture and its usage workflow.

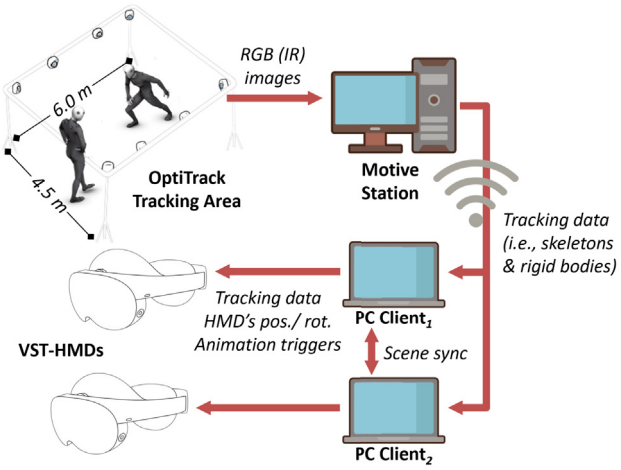


Fig. 1. Architecture of the CoMR-MoCap system.

3.1. Architecture

The overall architecture of the CoMR-MoCap system is shown in Fig. 1.

As stated in Section 2, technological constraints related to the limited field of view of the Microsoft HoloLens device prevented the authors of [16] from applying their MR-based rehearsal system in all the use cases involving mocap they originally envisaged. To overcome this limitation, the present paper proposes to use VST-HMDs as MR devices for delivering the digital contents. More specifically, the pass-through capability of the Meta Quest Pro[1] was exploited.

The MR application to be run on the HMDs in order to visualize digital contents overlapped to the video see-through feed coming from the onboard cameras has been developed with the Unity[2] game engine (v2022.3.9f1). The application includes all the necessary assets, i.e., 3D models, prerecorded animations, and spatialized sound effects required by the particular screenplay. The animations and sound effects can be triggered by the actors by pressing buttons available on the hand controllers. Alternatively, in those scenarios in which the actors need to keep their hands free, the other members of the crew or the director can activate them by using a keyboard.

As shown in the figure, an 8 RGB(IR)-camera (Prime13W[3]) OptiTrack system with tracked area of 4.5 m × 6.0 m is leveraged for mocap. The optical tracking data referring to both the actors' skeletons and rigid bodies are streamed from the workstation running the OptiTrack software (Motive v2.2[4]) to some PC clients over a 2.4 GHz Wi-Fi connection. Each PC client is assigned to a single actor and is responsible for managing/configuring the mapping between the tracked data and the virtual character's joints. The NatNet Software Development Kit (SDK)[5] is leveraged to this purpose.

In order to track the point of view of the actor in the virtual scene, three different alternatives were evaluated: (i) using the integrated tracking capabilities of the HMD, (ii) using the OptiTrack markers to track the head of the skeleton, or (iii) a rigid body mounted over the HMD. After measuring the transmission delays and the usability of each approach, the best solution found was actually to merge the tracking data gathered by the HMD and the OptiTrack system. More specifically, the 3D position of the actor's point of view relies on the

[1] Meta Quest Pro: https://www.meta.com/quest/quest-pro/.
[2] Unity: https://unity.com/.
[3] Primex 13W: https://optitrack.com/cameras/primex-13w/.
[4] Motive: https://optitrack.com/software/motive/.
[5] NatNet SDK: https://optitrack.com/software/natnet-sdk/.

positional tracking data of the OptiTrack skeleton's head, whereas its orientation is managed through the HMD. This approach reduces the delays related to the transmission of tracking data, by transferring a small amount of data over the network per frame. More specifically, it avoids sending the orientation data gathered by the OptiTrack system, as this information is reconstructed using the HMD's integrated tracking. Furthermore, this approach does not transmit data related to additional markers (required, e.g., in the third alternative), since the positional data from the OptiTrack system are already used by the MR application to reconstruct the actor's skeleton. Finally, the high accuracy of the OptiTrack system's positional tracking makes this approach more effective compared to the first alternative.

To register the two reference systems, i.e., the tracking spaces of the Meta Quest Pro and the OptiTrack system, a calibration procedure has been devised to be performed at the beginning of the rehearsal session, asking the users to remain still in T-pose for a few seconds. In this way, the offsets to convert the coordinates gathered by the HMD tracking system to the OptiTrack one are computed. It was decided to perform calibration only once (when a new user activates the system), as it was observed that the results were sufficiently accurate to support the entire rehearsal session. Nevertheless, using the system for a long time may introduce unacceptable tracking inaccuracies. Future work may consider implementing a novel calibration procedure optimized to run in background during the rehearsal session without significantly impacting user experience. For instance, the stage could be populated with reflective markers for the OptiTrack system and AR markers for the HMDs. By simultaneously capturing the positional data of these markers and using techniques such as Singular Value Decomposition (SVD), it would be possible to compute the transformation matrix that best aligns the two reference systems. This approach would allow for the continuous adjustment of tracking data from both the systems, ensuring proper registration even for long sessions.

To keep the PC clients synchronized, the Unity Netcode for Game Objects[6] library is used. This library offers utilities for developing networking and multiplayer applications with Unity. More specifically, one of the PC clients is requested to start the MR application as the host to which the other clients will connect. In this way, the session state of the applications running on the other HMDs and connected to the host results synchronized. This choice allows all the users to share and visualize the same session state (e.g., playback of prerecorded animations or sound effects) in real time. Moreover, the library offers functionalities to manage network connections and matchmaking, as well as the creation of lobbies.

The motion-to-photon latency, i.e., the time between the movement of the users and the rendering of their movements in the virtual scene, was measured to be approximately 50 ms. This latency is the result of a series of cumulating factors. First, the OptiTrack tracking based on visual markers introduces an initial delay related to the time needed for converting 2D images captured by each camera to 3D positional data. These data are then processed and converted into an appropriate format for transmission using the NatNet streaming protocol. Further delay are thus due to the processing with Motive and the NatNet SDK, as well as to the transmission of resulting information to Unity. Finally, once received, data should be unpacked and assigned to the corresponding elements in the scene for controlling virtual character animations; these last operations are subject to the limited computational power of the HMDs.

To reduce the impact of latency, a number of countermeasures were taken during the experimental evaluation described in the following section. In particular, the system was configured to work within a local network (even though the NatNet SDK and Unity Netcode library support remote connections), thus reducing transmission delay and loss of information. Moreover, the HMDs were connected to PC clients using the tethering mode to increase their computational power (the current implementation used desktop PCs, though backpacks could also be employed to improve mobility).

[6] NetCode: https://docs-multiplayer.unity3d.com/netcode/current/about/.

3.2. Workflow

Once the application has been populated with the necessary assets, it has to be deployed on the HMDs. It is worth observing that the preparation of these assets is not expected to entail particular efforts, as they are generally created for previsualization purposes before the actual rehearsal with the actors. Integrating them in the CoMR-MoCap system requires only the definition of the scene logic, which can be easily accomplished with the Unity game engine.

After the user has worn the HMD, a menu is shown to make him or her configure the connection with the other users. More specifically, the user can choose whether to host the session or connect to another PC client. In the first case, a lobby room is created to wait until the other users are connected. This lobby is identified by a code, which is automatically generated by the system when the user starts hosting. The same code can be used by the other users to connect (second case). In the lobby, the user can select the character to portray (in the experiment, the users had the possibility to choose between two characters). Once the users have selected their character, the rehearsal can be started by interacting with the Graphical User Interface (GUI) of the application.

Through the application, digital contents are added to the Meta Quest Pro pass-through to visualize assets intended to be added in the post-processing stage. The HMDs can be worn not only by the actors but also by the other components of the staff, e.g., directors or camera operators, to have a better visualization of the scene being shot.

4. Experiment

A user study was conducted with the aim of analyzing to what extent MR can be a valid support to rehearse collaborative mocap scenes. This section first introduces the design of the script and the scene adopted in the experiment. Afterwards, it describes the rehearsal methods contrasted in the study and the experimental procedure adopted. Finally, it provides details on the metrics used and the participants involved.

4.1. Script and scene design

In order to carry out the study, a custom script was created for a mocap scene rehearsal as already proposed in [4,16]. The script was designed to contain a number of challenges that actors may face when rehearsing or shooting scenes including mocap. More specifically, the aspects that were chosen to be particularly stressed during the experiment include: (i) directing the gaze of the actors on specific virtual elements, (ii) positioning the body (or specific body parts like the hands) of the actors to properly react to received stimuli or cope with script requirements, and (iii) making the actors have the correct emotional reactions to the events happening in the scene.

By moving from these considerations, a scene (and the corresponding script) was designed to satisfy the following requirements: (i) the script should envisage multiple actors controlling virtual characters using mocap; (ii) the virtual characters should present deformable and/or extendable body parts as well as (iii) non-anthropomorphic body parts such as wings, horns, and tails that may modify the actors' perception of space and proprioception; (iv) the scene should require the actors to interact with virtual objects and characters generated with either CGI or controlled with mocap; (v) the script should contain events that require the actors to simulate emotional reactions.

After having defined the script requirements, a number of recent films containing mocap scenes were analyzed to get inspiration, as done in [16]. Ultimately, "The Hobbit"[7] and "The Lord of the Rings"[8]

[7] The Hobbit: https://www.youtube.com/watch?v=Wu9XPEdBelY.

[8] The Lord of the Rings: https://www.youtube.com/watch?v=zpLpkfFFsnI.

(a) (b) (c)

Fig. 2. Virtual characters included in the script used in the experiment: (a) dragon, (b) ent, and (c) spider.

were selected, since they include characters and narrative contexts that satisfy the requirements listed above. Moreover, they also represent good examples of films characterized by massive adoption of mocap and VFX.

The designed scene involves two characters: a dragon (i.e., an anthropomorphic version of the Smaug character, shown in Fig. 2(a)), and an ent (depicted in Fig. 2(b)). Both the characters are controlled through mocap by two participants, each wearing also an HMD for visualizing the digital contents. An additional terrifying character, i.e., the spider named Shelob (illustrated in Fig. 2(c)), was included in the script to elicit fear in Smaug and make the latter respond properly to its aggressive behavior. Using the approach adopted also in [4,16], the animations of the spider were triggered programmatically by one of the operators involved in the experiment.

In the scene, the participant who plays the role of the dragon has to rescue the ent. When the dragon gets close to the cage in which the ent is imprisoned, it is scared by the spider sleeping just right outside (Fig. 3(a)). To make the cage magically disappear, the dragon uses a rope (hanging from the ceiling) to tie the artifact and swing it toward the ent.

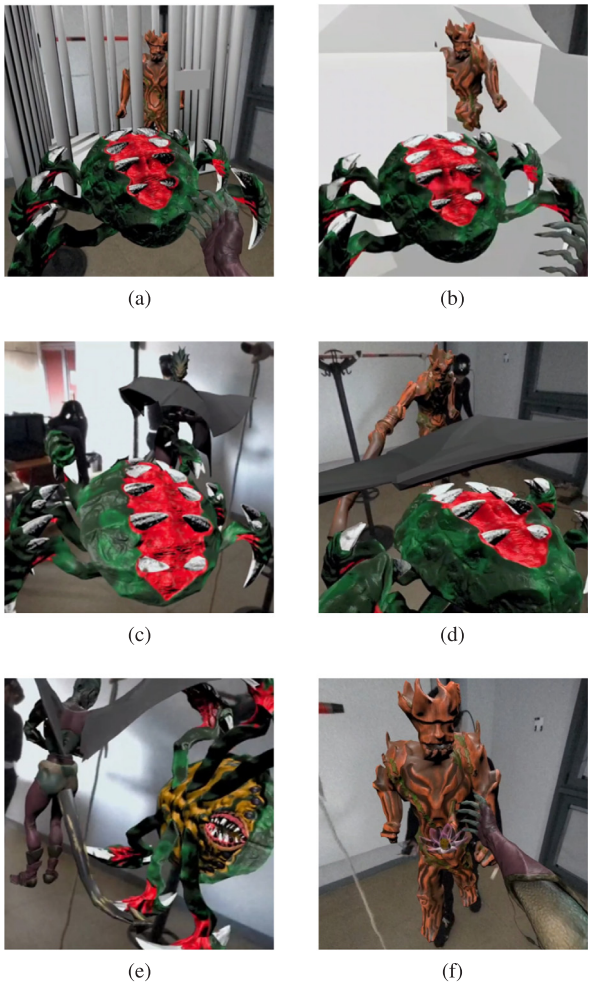
Once the ent catches the artifact and activates it by positioning the hands in the correct position, the cage explodes (Fig. 3(b)). The sound of the explosion wakes up the spider, who starts to attack the dragon with its claws. The dragon has to protect itself with its wings but, during the fight, it is injured (Fig. 3(c)). The ent extends one of the arms to strike the spider from behind and attract its attention (Fig. 3(d)). The dragon hits the spider with its tail knocking it out (Fig. 3(e)). To thank the dragon for its support, the ent grows a branch with a flower from its chest that the dragon can use to recover from injuries (Fig. 3(f)). The dragon takes the flower in its hand and places it on the wings to heal them.

During the performance of the two actors, several actions have to be carried out involving interactions with both digital elements (e.g., fighting against the spider or using the flower to heal the dragon's wings) and real elements (e.g., knotting a rope around the artifact). In addition, the participants were asked to animate virtual characters that have unusual parts, i.e., extensible arms, or non-anthropomorphic parts, such as wings and tails. Finally, emotional reactions (e.g., being scared by the spider) are envisaged. It is worth observing that actions such as knotting a rope around an object have been included in the script also to confirm the relevance of MR against fully virtually reconstructed scenes, as this kind of actions could be difficult to reproduce in VR.

The full script is available for download at http://tiny.cc/u2sjzz

4.2. Rehearsal methods

The design of the experiment was inspired by the works done in [2,16]. More specifically, the main objective of the experiment was to compare the proposed MR method for scene rehearsal against the traditional method based on physical props and visual cues (in the following referred to as TR).



(a) (b) (c) (d) (e) (f)

Fig. 3. Salient moments of the devised script from the actors' perspectives: (a) the dragon looks at the spider in front of the cage; (b) the ent uses the artifact to make the cage disappear; (c) the dragon defends from the attack of the spider; (d) the ent extends the arm to touch the back of the spider; (e) the dragon uses the tail to knock out the spider; (f) a flower grows from the chest of the ent.

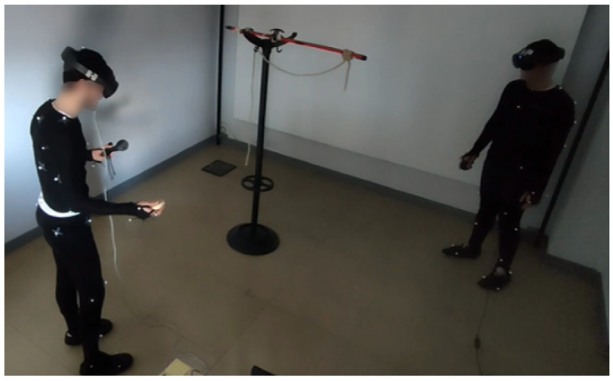
The two methods adopted in the experiment are shown in Fig. 4. Videos of the experiments are also available for download at http://tiny.cc/z2sjzz.

In the TR method (Fig. 4(a)), the participants were requested to rehearse the scene by using their imagination, physical props, and visual cues in place of the digital contents added during post-production. This approach is commonplace in the cinema industry, as reported, e.g., in [22,23].

More specifically, the position of the spider was represented through laser pointing. Adhesive tape on the floor was used to indicate the boundaries of the cage in which the ent is imprisoned. The tail and wings of the dragon were represented using physical props worn by the participant during the experiment. The extension of the ent's arm for touching the spider, the magic flower extracted from the chest, and the claws of the spider attacking the dragon had to be imagined by the actors. The other stage props such as the artifact to be used for making the cage disappear and the rope were available for both the methods, as they were physical scene objects included in the script.

With the MR method (Fig. 4(b)), the participants were allowed to directly visualize the digital contents, i.e., the actual body shapes of both the dragon and the ent characters, the spider and its animations, the cage, and the flower, overlapped with the real-world scene.

(a)



(b)

Fig. 4. Rehearsal methods considered in the experiment: (a) the TR method uses a red laser to point the position of the spider on the floor, adhesive tape near the feet of the participant on the right to indicate the boundaries of the cage, and physical props (i.e., the wings and the tail) worn by the other participant; (b) in the MR method, the participants wear VST-HMDs to visualize digital contents. Both the methods envisage physical scene objects, i.e., the artifact held by the participant on the left, the rope, and its support.

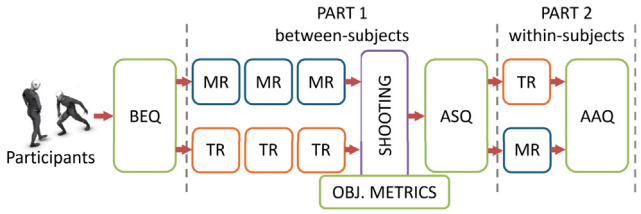


Fig. 5. Study design.

4.3. Procedure

The design of the experiment followed a mixed design, as proposed in [2,16]. More specifically, the procedure considered in the experiment is reported in Fig. 5.

The first part was arranged with a between-subjects design. The participants were randomly assigned to two equal-sized groups. At each group, a different rehearsal method, i.e., MR or TR, was assigned. Then, the participants were introduced to the objective of the experiment and to the script that they were asked to perform (presented in Section 4.1). The participants were requested to fill in a before-experience questionnaire (BEQ) aimed at collecting demographics and information regarding their previous experience with acting and MR applications. Afterwards, they were given free time (around 15/20 min) to familiarize with the actions and dialogs contained in the script.

When the participants considered themselves ready to act, two of them (one playing the role of the dragon, the other portraying the ent) underwent three rehearsals of the scene by using only the method assigned at the beginning of the experiment (i.e., either MR or TR). During the rehearsal, the participants were allowed to look at the script displayed on a large screen and ask for clarifications about it. The average duration of one rehearsal using either MR or TR was about 2 min. After all the rehearsals, the participants performed the actual shooting of the scene, without leveraging any visual aid (neither MR nor physical props). Like in [16], this approach was adopted to make the comparison of the two methods as fair as possible. During the shooting, objective metrics detailed in Section 4.4 were collected. After the shooting, an after-shooting questionnaire (ASQ) was administered to the participants in order to collect subjective feedback.

To complete the experiment, the pair of participants was requested to rehearse the scene again portraying the same character but using the alternative method. In this way, a direct comparison of the two methods and the corresponding participants' feedback were collected through an after-alternative questionnaire (AAQ). In this respect, the second part of the experiment can be regarded as following a within-subjects design.

4.4. Evaluation criteria and metrics

During the experiment, both subjective and objective metrics were collected through standard questionnaires and by logging tracking data, respectively.

4.4.1. Subjective measurements

As anticipated in Section 4.3, the participants were asked to fill in three questionnaires, i.e., BEQ, ASQ, and AAQ throughout the experiment. All the questionnaires are available for download at http://tiny.cc/43sjzz.

The BEQ asked the participants to indicate their previous experience with acting and MR applications as well as reporting demographic information.

After the shooting of the scene using the assigned method, the ASQ was administered. This questionnaire was aimed at collecting subjective feedback related to the usability of the experimented method, as well as the perceived effectiveness, embodiment, spatial and social presence.

Usability was evaluated through the System Usability Scale (SUS) [24]. This choice was made after observing that SUS is designed to be broadly applicable across various experiences [25]. In fact, it has been used for many task-based usability assessments even with extreme rewording [26]. In particular, it has been considered as versatile enough to support the evaluation of different technologies (even not mediated by a GUI), such as interactive voice response systems [27], wearable interfaces [28], AR systems [29], hand-tracking methods [30], etc.

Aspects concerning spatial and social presence, not considered in [4], were investigated by leveraging the questionnaire proposed in [31]. The participants were asked to provide scores on a 1-to-5 scale (from not at all to very much).

To evaluate the perceived effectiveness, the statements adapted from [2] were used. In this case, the participants had to rate each statement on a 1-to-5 Likert scale (from strongly disagree to strongly agree).

Differently than in [16], the sense of embodiment was also measured using the questionnaire proposed in [32], which requested the participants to express a score on a 1-to-7 scale (from not at all to very much).

Once the participants had also experimented with the alternative rehearsal method, they were invited to fill in the AAQ. This questionnaire was aimed at collecting the overall preferences of the participants for the two rehearsal methods, as done in [2,16]. To this aim, the participants were requested to rank the MR and TR methods according to a number of characteristics. Finally, a more in-depth picture of the proposed system was reconstructed through the questionnaire proposed

in [33], with the aim to investigate the suitability of MR for supporting rehearsal activities. The participants expressed their score on a 1-to-5 scale (from strongly disagree to strongly agree).

While filling in the questionnaires, the participants were provided explanations and examples to help them align, e.g., on what can be regarded as a usable or unusable system, characterize levels of embodiment, or distinguish aspects related to spatial and social presence.

4.4.2. Objective measurements

As anticipated, the objective measurements were collected for both the groups only during the shooting of the scene. In this phase, the participants were not aided by any physical prop or visual hint, and solely relied on the knowledge acquired during the rehearsal. The objective measurements included the two metrics defined in [16], i.e., the *eye* and *body* distance, designed to evaluate the eye gaze and spatial positioning.

More specifically, the first metric estimated the distance between the points of interest that the participant was expected to look at and his or her actual gaze. The distance was computed by projecting the participant's gaze to a plane that is perpendicular to the gaze and contains the point of interest as proposed in [16]. Similarly to [16], values of this metric were computed only at specific moments in time by averaging the measured distances in a time interval centered at the event occurrence. For this metric, the following events and corresponding points of interest were considered: (i) the dragon has to look at the spider placed in front of the cage (in the following the metric computed for this event is referred to as $EyeDist_1$); (ii) the ent is requested to look at the spider while it is attacking the dragon $EyeDist_2$); (iii) the dragon ($EyeDist_3$) and (iv) the ent ($EyeDist_4$) have to point their gaze to the flower growing on the chest of the ent.

The second metric evaluated the distance between parts of the participant's body and the position these parts were expected to assume. Similarly to the eye distance, this metric was computed at specific moments in time. More specifically, the following events and corresponding body parts were considered: (i) the position of the ent's hands when the cage is unlocked using the artifact (in the following the metric computed for this event is referred to as $BodyDist_1$); (ii) the position of the dragon's hip when it is attacked by the spider ($BodyDist_2$); (iii) the position of the ent's hand when it is requested to distract the spider by touching its back $BodyDist_3$; (iv) the position of the dragon's hand when the flower is to be collected from the ent ($BodyDist_4$).

To compute the values of these metrics, the tracking data obtained through the HMD and the OptiTrack system were leveraged. Hence, during the shooting, the participants were requested to wear the mocap suit. Sound effects useful for helping them to synchronize with the animations in the script were played using external speakers for both the rehearsal methods.

4.5. Participants

Like in [16], the experiment was carried out by considering 24 participants (14 males and 10 females). Participants were aged between 20 and 31 ($\overline{x} = 25.29$, $SD = 2.56$).

According to the information collected at the beginning of the experiment, the majority of the participants (i.e., 95.83%) had "minimal" or "none" experience with acting. Only 4.17% of the participants reported "good" expertise. None of them had acted in scenes including characters controlled with mocap. Regarding their familiarity with MR technology, 62.50% of the participants had "never" or "sometimes" used this kind of devices, whereas the remaining stated to use MR "regularly".

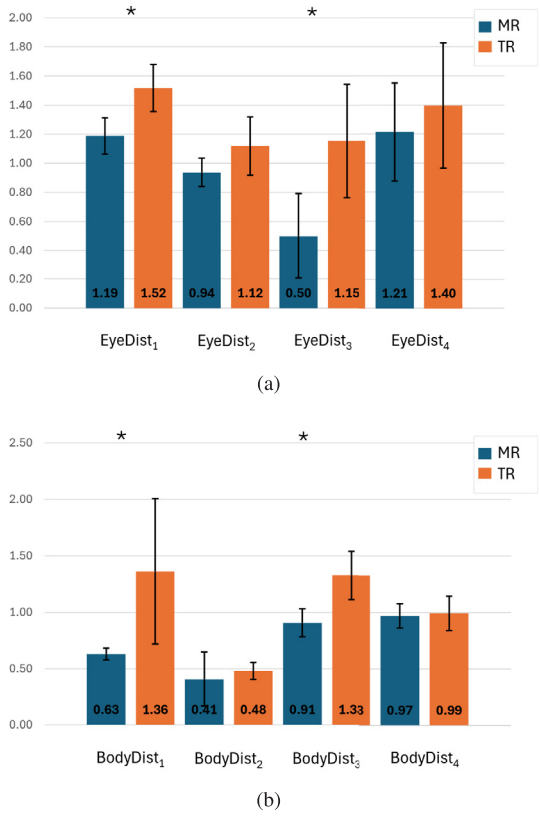


Fig. 6. Objective results based on the (a) eye and (b) body distance metrics. Significant differences are marked with *.

5. Results

Data collected during the experiment were statistically analyzed by using MS Excel with the Real Statistics add-on. After having studied the normality of data using the Shapiro–Wilk test and found that prerequisites for parametric tests were not met, it was decided to use non-parametric tests. More specifically, the Mann–Whitney and Wilcoxon Signed-Rank tests were used depending on whether samples could be considered as independent (for the objective metrics and the first sections of the questionnaire) or paired (for the last section of the questionnaire). Correlations between the profile of the users and scores assigned to the items of the questionnaires were studied using Spearman's rank correlation coefficient (ρ), since data did not meet the assumptions for using Pearson's correlation analysis.

5.1. Objective results

The objective results based on the eye and body-part distance collected during the shooting are shown in Fig. 6. It can be observed that, generally, the values referring to the MR method were smaller than the TR one. Starting from the eye distance, it can be noticed that, differently than in [2] where no statistically significant differences were found, in the present study significant differences were spotted for several parts of the script: the dragon looks at the spider at the beginning of the scene (MR: 1.18 vs. TR:1.51, $p = .020$), and the ent looks at the flower (MR: 0.50 vs. TR: 1.15, $p = .031$). It is worth noticing that significant differences were found for actions that involve both the characters (i.e., the dragon and the ent).

For what it concerns the body-part distance, significant differences were observed for two events, i.e., (i) when the dragon defends itself from the attack of the spider (MR: 0.91 vs. TR: 1.33, $p = .013$), and (ii) collects the flower from the ent (MR: 0.63 vs. TR: 1.36, $p = .005$).

Table 1
Subjective results concerning usability based on SUS [24]. Bold font indicates the best value (significant differences) for the two rehearsal methods.

Statement	MR	TR	*p-value*
I think that I would like to use this system frequently	**4.17**	2.75	< .001
I found the system unnecessarily complex	1.83	2.08	.733
I thought the system was easy to use	4.08	3.50	.184
I think that I would need the support of a technical person to be able to use this system	3.00	2.33	.131
I found the various functions in this system were well integrated	**4.50**	3.67	.045
I thought there was too much inconsistency in this system	1.75	1.92	.710
I would imagine that most people would learn to use this system very quickly	4.42	4.25	.549
I found the system very cumbersome to use	**1.33**	2.42	.048
I felt very confident using the system	4.42	4.17	.705
I needed to learn a lot of things before I could get going with this system	1.58	1.83	.393
SUS Score	**80.21**	69.38	.024
Grade	B	D	
Adjective rating	Excell.	Ok	

5.2. Subjective results

In the following, the subjective results collected through the ASQ and AAQ are reported.

5.2.1. ASQ

For what it concerns the subjective measurements, the participants rated the MR method as characterized by a higher usability than the TR (MR: 80.21 vs. TR: 69.38, $p = .024$). Based on the categorization reported in [34], the overall SUS scores obtained by the two methods correspond to the grades B ("Excellent") and D ("Ok"), respectively.

The higher scores obtained by the MR method can be explained by analyzing the individual statements of the SUS questionnaire (reported in Table 1). In particular, the participants expressed their interest in using the MR method more frequently than the TR one (MR: 4.17 vs. TR: 2.75, $p < .001$), and the functionalities provided for the rehearsal were found to be more integrated in the MR method than in the TR one (MR: 4.50 vs. TR: 3.67, $p = .045$). Finally, they rated the TR method as more cumbersome than the MR one (TR: 2.42 vs. MR: 1.33, $p = .048$).

Regarding the other dimensions already considered in [16] and analyzed through the ASQ, Fig. 7 reports the average scores. Statistically significant differences can be spotted in favor of the MR method for both spatial (MR: 4.15 vs. TR: 3.02, $p = .015$) and social presence (MR: vs. TR: 3.52, $p = .002$) as well for effectiveness (MR 4.56 vs. TR: 3.66, $p = .003$).

As done for the SUS, more insights can be obtained by investigating the single statements of each dimension. Starting from spatial presence, the participants felt the objects to be part of the environment they were experiencing more with the MR method than the TR one (MR: 4.33 vs. TR: 3.33, $p = .020$). Moreover, they were more confident in interacting with digital contents using the MR method than the TR one (MR: 4.58 vs. TR: 3.33, $p = .002$). Finally, the MR method made the participants have more the instinct to interact with the digital contents they were seeing/imagining than the TR one, even though these interactions were not explicitly reported in the script (MR: 4.42 vs. TR: 2.67, $p = .005$).

Moving to social presence, when using the MR method the participants had a better feeling that the other virtual characters (i.e., the ent, the dragon, or the spider depending on the character they were portraying) were also able to see and hear them (MR: 4.33 vs. TR: 3.17, $p = .022$). The improvement brought by the use of the MR method allowed the participants to interact better with the other characters than the TR one (MR: 4.50 vs. TR: 3.25, $p.009$). Furthermore, when rehearsing the scene with the MR method, the participants felt more in the same place as the other characters (MR: 4.83 vs. TR: 3.83, $p = .001$) and they could speak more directly to them (MR: 4.83 vs. TR: 3.67, $p < .001$). Finally, the participants felt to be more present when interacting with the other characters they were seeing/hearing when using the MR method than the TR one (MR: 4.67 vs. TR: 3.67, $p = .005$).

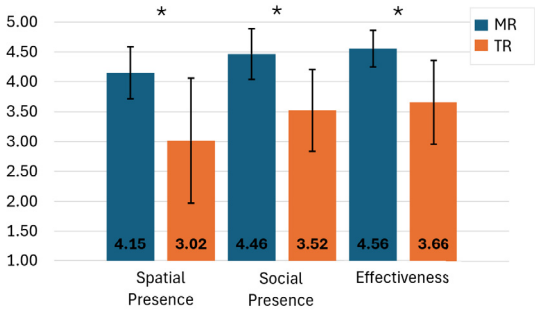


Fig. 7. Subjective results concerning spatial presence, social presence, and effectiveness as investigated in [2,31]. Significant differences are marked with *.

Concerning effectiveness, the MR method aided the participants in better positioning themselves within the environment than the TR one (MR: 4.58 vs. TR: 4.00, $p = .050$) and feeling more comfortable with the gestures they were requested to perform (MR: 4.58 vs. TR: 4.00, $p = .050$). Additionally, the MR method helped the participants to express the emotional state of the played characters through facial expressions (MR: 4.17 vs. TR: 3.00, $p = .012$) and gestures (MR: 4.25 vs. TR: 3.17 $p = .027$) more effectively than the TR one. Furthermore, the participants using the MR method demonstrated higher levels of emotional engagement (MR: 4.33 vs. TR: 3.17, $p = .015$) compared to those using the TR one. The participants also reported that with the MR method they were able to use the staging space more effectively than with the TR one (MR: 4.83 vs. TR: 3.75, $p < .001$). Finally, the MR method made the participants more confident than the TR one regarding their performance during the rehearsal (MR: 4.83 vs. TR: 4.00, $p = .010$), thus enhancing their readiness for shooting (MR: 4.83 vs. TR: 3.83, $p = .001$).

As mentioned above, differently than in [16], the dimension related to embodiment was also investigated in the study performed in the present paper. Overall, the participants expressed a higher sense of embodiment with the MR method than the TR one (MR: 5.27 vs. TR: 3.96, $p = .050$). More specifically, with the MR method the participants had a higher feeling that their real body was drifting to the virtual one (MR: 6.08 vs. TR: 4.25, $p = .023$) and their appearances were turning into that of the virtual character (MR: 5.42 vs. TR: 3.58, $p = .037$). Moreover, the MR method made the participants have a higher feeling that they were wearing different clothes from what they were actually using in the experiment (MR: 5.25 vs. TR: 2.75, $p = .013$). In addition, the participants felt more sensations in their bodies when they saw digital contents with the MR method than the TR one (MR: 5.83 vs. TR: 3.50, $p = .023$), thus making them perceive that their bodies could be more affected by the virtual elements (MR: 6.17 vs. TR: 3.92, $p = .004$). Finally, the MR method improved the sense of touch of virtual objects that were interacted with the virtual body (MR: 5.00 vs. TR: 2.67, $p = .008$), thus making the participants recognize that the touch was caused by the virtual contents (MR: 4.40 vs. TR: 2.67, $p = .032$) and perceive a higher sensation when their bodies touched digital contents (MR: 5.42 vs. TR: 3.17, $p = .008$).

Table 2
Subjective results concerning the suitability of the MR method for rehearsing mocap scenes based on the analysis tool proposed in [33].

Statement	$\bar{x}$ (SD)
Watching the virtual objects was as natural as watching real-world objects	4.29 (0.55)
I had the impression that virtual and real objects belonged to the same world	3.83 (0.92)
I had the impression that I could touch and grasp the virtual objects	4.21 (0.88)
I had the impression that the virtual objects were in the real world rather than simply projected on a screen	3.96 (0.75)
I had the impression of seeing virtual objects as three-dimensional and not as mere flat images	4.63 (0.49)
I do not notice differences between real and virtual objects	3.00 (1.06)
I had not to make an effort to recognize virtual objects as being three-dimensional	4.75 (0.68)

The analysis of correlations between the profile of the participants and the scores assigned to the items of the questionnaires did not show sufficient statistical evidence to conclude that significant relationships between the variables exist, as ρ coefficients were in the range -0.13 and 0.31 with $p > .05$.

Regarding the direct comparison of the two methods, the results confirmed the higher appreciation of the participants for the MR one ($p < .001$). More specifically, the latter was ranked as the best method for what it concerns positioning in the scene ($p < .001$), controlling gaze direction ($p < .001$), and eliciting emotional involvement ($p < .001$). A clear ranking supported by a statistically significant difference between the two methods was not spotted for what it concerns synchronization with other characters.

5.2.2. AAQ

Results regarding the suitability of the MR method for rehearsing mocap scenes are reported in Table 2. The relatively high scores assigned by the participants confirmed the ability of the MR to support the rehearsal activity. A possible weakness of the proposed system could be related to the differences between the real and virtual objects noticed by the participants. The intermediate score (3.00) can be explained by the fact that no photorealistic assets were used in the experiment. Moreover, short delays or tracking inaccuracies experienced during the rehearsal could have potentially occurred, making more evident the digital contents with respect to the real ones.

The analysis of correlations did not show again significant relationships between variables ($-0.10 < \rho < 0.20$, $p > .05$).

6. Conclusion and future work

The work reported in this paper was aimed at exploring the use of MR for supporting multiple actors during the rehearsal of scenes involving mocap. To this purpose, the paper presented the design and development of a system named CoMR-MoCap, which allows actors to collaboratively visualize digital contents (supposed to be added in the post-production phase) overlapped with the real scene by wearing VST-HMDs. Digital contents can include both real and computer-generated elements, such as virtual objects, animations, VFX, and digital characters animated with mocap. The system was developed for the Meta Quest Pro as VST-HMD and OptiTrack as marker-based tracking system for mocap. It is worth noticing that the hardware setup considered in this work, especially the mocap system, may be regarded as suitable for high-budget productions. Even though these productions may benefit from higher-end props compared to those used in the experimental evaluation, it is reasonable to expect that many of the benefits offered by the use of MR would be preserved. In such contexts, MR could be also considered as an alternative tool, aimed not at fully replacing the use of props but rather offering a complementary solution for reducing reliance on the most costly or logistically challenging props (which could be easily simulated with minimal creation efforts). Nevertheless, the architecture of the devised system could be easily adapted/scaled down to support alternative technologies aimed at low-budget productions. In such application scenarios, the MR-based method may prove even more effective due to the possible absence of sophisticated props or the use of simpler mocap setups.

Experiments were conducted to compare the effectiveness of the proposed MR-based rehearsal method with the traditional approach relying on physical props and visual cues, considering both objective and subjective metrics. The obtained results indicated that the MR-based method outperformed the traditional method in terms of usability, spatial presence, social presence, perceived effectiveness, and embodiment. Furthermore, the MR-based method was found to be effective in assisting the users in directing their gaze toward virtual elements and positioning their bodies, especially when interacting with moving virtual objects or reacting emotionally in response to virtual events. These findings extend the current state of the art in the field, since previous works (e.g., [4,16]) did not consider interactions among multiple users.

It is worth observing that technological issues related to limited field of view of the HMD experienced in [16] were overcome, proving the capability of the proposed system to support also use cases in which the actor who wears the HMD is also controlling a virtual character through mocap. Moreover, differently than the existing literature, the proposed system enables collaboration into a MR scenario, allowing multiple users (i.e., the actors, the director, or members of the staff) to visualize simultaneously the same scene during rehearsal.

Considering the experimental evaluation performed in this work, it is possible to mention several ways to improve it. First, even though positive outcomes have been obtained, it is worth recalling that the number and background of participants involved in the user study (24 subjects with limited experience in acting with mocap) may not be representative of all possible production situations. Therefore, future experiments shall consider engaging skilled actors to have a more in-depth picture of the effectiveness of the proposed system. For instance, involving skilled actors would make it possible to compare their performance with that of the sample engaged in this study to determine whether the approach is more or less effective, as well as to investigate possible relationships between participants' profiles and the dimensions being explored. Moreover, experiments could be extended to stress other factors or consider different production situations and use cases in which, e.g., multiple actors/performers control the same virtual character (like in "Godzilla: King of the Monsters"[9]).

Second, the results regarding the suitability of the MR-based method for rehearsing mocap scenes have shown some issues that could be related to delays and tracking inaccuracies. Future studies may be devoted to further investigating their effects on actors' performance, with the final aim of further optimizing the devised system.

Third, other technologies used for virtual production, such as VR or LED walls, could be included in the comparison, in order to possibly identify further strengths and weaknesses of MR in supporting scene rehearsal.

Finally, further experiments could aim to compare the proposed system based on VST-HMDs with the previous work based on OST-HMDs [16], thus assessing the impact that the different visualization technology can have on the actors' performance.

The current capabilities of the CoMR-MoCap system could also be enhanced. For instance, the adoption of a marker-less tracking system would allow actors to move in larger spaces without relying on potentially intrusive equipment like the reflective markers of the OptiTrack system. However, such tracking technologies may introduce new challenges such as additional latency, tracking inaccuracies, and

[9] Godzilla — King of the Monsters: https://www.youtube.com/watch?v=g8PhummZur0.

drift that should be addressed before making a fair comparison. Integration of face tracking could also be explored, enabling actors to visualize not only the articulated body but also the facial expressions of virtual characters. To this aim, it would be possible to either leverage the face tracking capabilities possibly provided by the HMD (which is the case of the considered one) or rely on external hardware. However, these solutions would not come without challenges to be faced like, e.g., the limited facial expressiveness recognized by tracking technologies integrated in current HMDs or the difficulty for external systems to track the actor's face, since a large portion of it would be covered by the HMD.

CRediT authorship contribution statement

Alberto Cannavò: Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Francesco Bottino:** Software, Resources, Methodology, Investigation, Conceptualization. **Fabrizio Lamberti:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work has been developed in the frame of the VR@POLITO initiative. The research was supported by PON "Ricerca e Innovazione" 2014–2020 – DM 1062/2021 funds.

References

[1] Pires F, Silva R, Raposo R. A survey on virtual production and the future of compositing technologies. Avanca Cine J 2022;21(1):692–9.
[2] Bouville R, Gouranton V, Arnaldi B. Virtual reality rehearsals for acting with visual effects. In: Proc. international conference on computer graphics & interactive techniques. 2016, p. 1–8.
[3] Kumar A. Introduction to visual effects (VFX). In: Beginning VFX with autodesk Maya. Springer; 2022, p. 1–10.
[4] Kammerlander RK, Pereira A, Alexanderson S. Using virtual reality to support acting in motion capture with differently scaled characters. In: Proc. IEEE virtual reality and 3D user interfaces. 2021, p. 402–10.
[5] Allison T. More than a man in a monkey suit: Andy serkis, motion capture, and digital realism. Q Rev Film Video 2011;28(4):325–41.
[6] Soghomonian T. Ian mckellen: Filming "the hobbit" made me cry with frustration. 2012, https://www.nme.com/news/film/ian-mckellen-filming-the-hobbit-made-me-cry-with-f-877575. [Accessed 13 June 2024].
[7] Kang C-Y, Li T-Y. One-man movie: A system to assist actor recording in a virtual studio. In: Proc. IEEE international conference on artificial intelligence and virtual reality. 2021, p. 84–91.
[8] Ge R, Hsiao T-C. A summary of virtual reality, augmented reality and mixed reality technologies in film and television creative industries. In: Proc. IEEE eurasia conference on biomedical engineering, healthcare and sustainability. 2020, p. 108–11.
[9] Liu J, Zheng Y, Wang K, Bian Y, Gai W, Gao D. A real-time interactive Tai Chi learning system based on VR and motion capture technology. Procedia Comput Sci 2020;174:712–9.
[10] Nassar SGM. Engaging by design: Utilization of VR interactive design tool in mise-en-scène design in filmmaking. Int Des J 2021;11(6):65–71.
[11] Rizvic S, Okanovic V, Boskovic D. Digital storytelling. In: Visual computing for cultural heritage. Springer; 2020, p. 347–67.
[12] Sanna A, Lamberti F, De Pace F, Iacoviello R, Sunna P. ARSSET: Augmented reality support on SET. In: Proc. international conference on augmented reality, virtual reality and computer graphics. 2017, p. 356–76.
[13] Tamura H, Matsuyama T, Yokoya N, Ichikari R, Nobuhara S, Sato T. Computer vision technology applied to MR-based pre-visualization in filmmaking. In: Proc. Asian conference on computer vision. 2010, p. 1–10.
[14] Ikeda S, Taketomi T, Okumura B, Sato T, Kanbara M, Yokoya N, Chihara K. Real-time outdoor pre-visualization method for videographers—real-time geometric registration using point-based model. In: Proc. IEEE international conference on multimedia and expo. 2008, p. 949–52.
[15] Stamm A, Teall P, Benedicto GB. Augmented virtuality in real time for pre-visualization in film. In: Proc. IEEE symposium on 3D user interfaces. 2016, p. 183–6.
[16] Cannavò A, Pratticò FG, Bruno A, Lamberti F. AR-mocap: Using augmented reality to support motion capture acting. In: Proc. IEEE conference virtual reality and 3D user interfaces. 2023, p. 318–27.
[17] Cannavò A, Pratticò FG, Ministeri G, Lamberti F. A movement analysis system based on immersive virtual reality and wearable technology for sport training. In: Proc. international conference on virtual reality. 2018, p. 26–31.
[18] Chen X, Chen Z, Li Y, He T, Hou J, Liu S, He Y. ImmerTai: Immersive motion learning in VR environments. J Vis Commun Image Represent 2019;58:416–27.
[19] Ikeda A, Hwang D-H, Koike H, Bruder G, Yoshimoto S, Cobb S. AR based self-sports learning system using decayed dynamic TimeWarping algorithm. In: Proc. international conference on artificial reality and telexistence and eurographics symposium on virtual environments. 2018, p. 171–4.
[20] Damian I, Obaid M, Kistler F, André E. Augmented reality using a 3D motion capturing suit. In: Proc. augmented human international conference. 2013, p. 233–4.
[21] Ichikari R, Tenmoku R, Shibata F, Ohshima T, Tamura H. Mixed reality pre-visualization for filmmaking: On-set camera-work authoring and action rehearsal. Int J Virtual Real 2008;7(4):25–32.
[22] Okun JA, Zwerman S. The VES handbook of visual effects: industry standard VFX practices and procedures. Routledge; 2020.
[23] Dower J, Langdale P. Performing for motion capture: a guide for practitioners. Bloomsbury Publishing; 2022.
[24] Brooke J, et al. SUS-A quick and dirty usability scale. Usability Eval Ind 1996;189(194):4–7.
[25] Brooke J. SUS: A retrospective.. J Usability Stud 2013;8(2):29–40.
[26] Lewis JR. The system usability scale: Past, present, and future. Int J Hum–Comput Interact 2018;34(7):577–90.
[27] Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum–Comput Interact 2008;24(6):574–94.
[28] Kostov V, Ozawa J, Matsuura S. Analysis of wearable interface factors for appropriate information notification. In: Proc. IEEE international symposium on wearable computers. 2004, p. 102–9.
[29] Santos MEC, Taketomi T, Sandor C, Polvi J, Yamamoto G, Kato H. A usability scale for handheld augmented reality. In: Proc. ACM symposium on virtual reality software and technology. 2014, p. 167–76.
[30] Khundam C, Vorachart V, Preeyawongsakul P, Hosap W, Noël F. A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training. Informatics 2021;8(3):60–73.
[31] Lombard M, Ditton TB, Weinstein L. Measuring presence: The temple presence inventory. In: Proc. annual international workshop on presence. 2009, p. 1–15.
[32] Gonzalez-Franco M, Peck TC. Avatar embodiment: Towards a standardized questionnaire. Front Robotics AI 2018;5:74.
[33] Regenbrecht H, Schubert T. Measuring presence in augmented reality environments: Design and a first test of a questionnaire. 2021, p. 1:1–7, arXiv preprint arXiv:2103.02831.
[34] Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud 2009;4(3):114–23.