



# Robustness and static-positional accuracy of the SteamVR 1.0 virtual reality tracking system

Lucia Grazia Sansone<sup>1</sup> · Ronny Stanzani<sup>1</sup> · Mirko Job<sup>1</sup> · Simone Battista<sup>1</sup> · Alessio Signori<sup>2</sup> · Marco Testa<sup>1</sup>

Received: 9 March 2021 / Accepted: 18 September 2021  
© The Author(s) 2021

## Abstract

The use of low-cost immersive virtual reality systems is rapidly expanding. Several studies started to analyse the accuracy of virtual reality tracking systems, but they did not consider in depth the effects of external interferences in the working area. In line with that, this study aimed at exploring the static-positional accuracy and the robustness to occlusions inside the capture volume of the SteamVR (1.0) tracking system. To do so, we ran 3 different tests in which we acquired the position of HTC Vive PRO Trackers (2018 version) on specific points of a grid drawn on the floor, in regular tracking conditions and with partial and total occlusions. The tracking system showed a high inter- and intra-rater reliability and detected a tilted surface with respect to the floor plane. Every acquisition was characterised by an initial random offset. We estimated an average accuracy of  $0.5 \pm 0.2$  cm across the entire grid (XY-plane), noticing that the central points were more accurate ( $0.4 \pm 0.1$  cm) than the outer ones ( $0.6 \pm 0.1$  cm). For the Z-axis, the measurements showed greater variability and the accuracy was equal to  $1.7 \pm 1.2$  cm. Occlusion response was tested using nonparametric Bland–Altman statistics, which highlighted the robustness of the tracking system. In conclusion, our results promote the SteamVR system for static measures in the clinical field. The computed error can be considered clinically irrelevant for exercises aimed at the rehabilitation of functional movements, whose several motor outcomes are generally measured on the scale of metres.

**Keywords** Virtual reality · Accuracy testing · Tracker occlusion · Validation Study [Publication Type] · SteamVR tracking · HTC Vive PRO

## 1 Introduction

Recent advances in the sensor industry, together with the lowering of production costs, have significantly changed the perspective of virtual reality (VR) technology. Cipresso and colleagues (Cipresso et al. 2018) described the different types of VR based on the level of immersion perceived by the user, therefore distinguishing three main groups of systems. The non-immersive and semi-immersive VR systems represent the cheapest VR solution, reproducing a virtual image or a 3D scenario, respectively, using either monitors

or desktops. Conversely, immersive VR (IVR) systems provide the user with a full-body simulated experience with a stereoscopic view of the surrounding environment using a head-mounted display (HMD) and a positional tracking system integrated with audio and haptic devices.

IVR allows for experiencing a greater degree of realism by concealing the interface between the real and virtual world from the user and creating the illusion of presence in 3D scenarios (Lombard and Ditton 1997). One of the main opportunities provided by IVR systems is the possibility to create an ad hoc scenario to test hypotheses that are otherwise difficult to assess in the real world (Nabiyouni et al. 2017). For this reason, these systems are rapidly expanding into numerous research applications and scientific disciplines, such as research, rehabilitation, education, medical and surgical procedures (Cipresso et al. 2018; Izard et al. 2018; Campo-Prieto et al. 2021).

Nonetheless, the applicability of IVR for quantitative measures in research is still under discussion, with a major

✉ Lucia Grazia Sansone  
luciagrazia.sansone@medicina.unige.it

<sup>1</sup> Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINO GMI), University of Genova, Campus of Savona, Via Magliotto 2, 17100 Savona, SV, Italy

<sup>2</sup> Biostatistics Unit, Department of Health Sciences, University of Genova, Genova, Italy

limit in the existing trade-off between costs and precision in the tracking of the movements.

Expensive movement analysis systems reported in the literature, such as Vicon and Optitrack, embed IVR in their platforms, integrating precision motion capture with a 3D immersive virtual experience. However, in later years, the general development of gaming platforms has cut the costs of IVR systems and devices, from 100 k USD—or more—to less than 1 k USD. Among them, the HTC VIVE system represents the most widely used commercial solution for consumer VR applications, due to its better tracking performance (Ikbal et al. 2021).

In the literature, different studies explored the feasibility and reliability of such IVR system for precise experimental measurements. However, the variability in the overall results precludes any definitive conclusion. Some works reported high tracking performance with accuracy and precision values ranging from millimetre up to submillimetre scales (Spitzley and Karduna 2019a; Ameler et al. 2019; Veen et al. 2019a,b; Jost et al. 2019; Ikbal et al. 2021). Jost and colleagues (Jost et al. 2019) affirmed that the HTC VIVE system has the potential to accurately track human movement in biomechanical and physiotherapy research and in the clinical setting. In contrast, other studies highlighted that the same system presents clear tracking issues (Niehorster et al. 2017), with error values that can vary from centimetres (Hemphill et al. 2020) up to metres for dynamic movements (Borges et al. 2018). The variability in the existing results is related to the various methodologies, experimental setups and intended application of the different available works. As an example, some studies have analysed planar movements (Niehorster et al. 2017; Borges et al. 2018), while others have been focussed on 3D motions (van der Veen et al. 2019a,b), highlighting a lack of a consensus in the methodologies to compare performance across various applications. In this sense, the work of Ikbal and colleagues (2021) laid a solid foundation towards the identification of a standard procedure to evaluate the performance of the HTC VIVE tracking in static and dynamic conditions for a custom application. However, in their comprehensive study, they did not consider the effects of external interferences in the working area by conducting their evaluations in a controlled environment. Such conditions do not entirely reflect a standard-use scenario where occlusions in the acquisition area can affect the quality of tracking (Niehorster et al. 2017). Hence, our study aimed at systematically exploring the static-positional accuracy of the HTC VIVE tracking and its robustness against possible tracking occlusions occurring inside the capture volume.

## 2 Methods

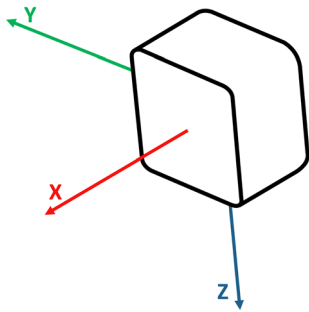
### 2.1 Overview of the HTC VIVE pro and the SteamVR tracking system

The Vive Pro System configuration included a head-mounted display (HMD), two controllers, two lighthouses BASE stations, and two VIVE trackers 2.0 (2018). The HMD is equipped with two AMOLED lenses with a resolution of  $1440 \times 1600$  pixels per eye with a refresh rate of 90 Hz, covering a nominal field of view of  $110^\circ$  (2021). The position and orientation of the trackable elements in the space (i.e. HMD, controllers or trackers) are obtained through the SteamVR tracking system, which operates on a combination of inertial and outside-in tracking principles (Borges et al. 2018).

The tracked position and orientation are updated primarily by inertial measurement units through dead reckoning (path integration) allowing high update rates (Niehorster et al. 2017). The BASE stations limit and correct the intrinsic “drift” error of the inertial measurements by providing additional kinematic data through the solution of the so-called perspective-n-point (PnP) problem (Maciejewski et al. 2020). The system extrapolates positional and orientation values from a set of sensors (IR photodetectors) located on the trackable device (i.e. HMD, controllers, or trackers) illuminated by the lighthouses. The latter emits a wide IR synchronisation blink followed by two wide range IR pulses which sweep the tracking area repeatedly within a 120 degree angle, one axis at a time from left to right and then from top to bottom. By knowing the angular velocity of the device, and the time between the synchronisation blink and the detection of the laser pulse, the system determines the directions in which each photodetectors are located. The directions of at least four non-coplanar photodetectors are the basis to solve the PnP mathematical problem and consequently improve the accuracy of the tracking.

### 2.2 Experimental setup

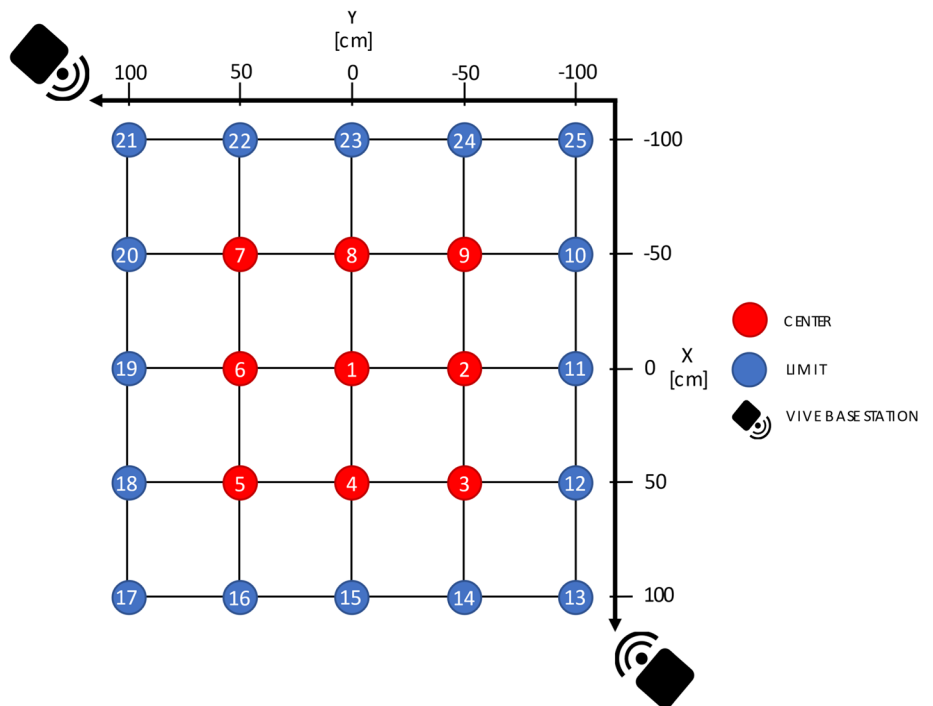
The HTC VIVE Pro system was used for the different experimental sessions. The experimental sessions were set in a  $7 \times 5$  m room, with neither reflective surfaces nor natural lighting exposure. The floor alignment was verified with a spirit level. According to the HTC Vive Headset User Manual (HTC Corp 2020), the two lighthouses were fixed to the ceiling at the height of  $h = 3$  m with a pitch inclination (Y-axis) of  $46 \pm 1$  deg and a roll inclination (X-axis) lower than  $1 \pm 1$  deg. The axes are shown in Fig. 1. Lighthouses were placed at a distance of  $d_c = 4$  m



**Fig. 1** Rotation axes of a VIVE lighthouse base station

from each other. Such angulation was fixed, following the maximal limits specified in the manual mentioned above, considering the installation height of the lighthouses, which were coupled in “Sync cable” mode. At the middle of the junction line that connects the floor projections of the two lighthouses, we identified the centre of a Cartesian grid. The grid was a square area of side  $L = 2$  m subdivided into sub-square regions of  $l = 0.5$  m side, drawn on the floor with a  $d = 2$  cm wide adhesive tape. By doing this,  $n_p = 25$  points were identified and categorised (relying on the specific position of the grid, Fig. 2) in: centre grid (CG) points, i.e. from 1 to 9, and limit grid (LG) points, i.e. from 10 to 25. Point 1 was considered as the origin of the XY plane. The grid was used as a reference to set the virtual gaming space using the SteamVR application.

**Fig. 2** Top view of the grid setup. The central points (CG) are identified by the red numbers on the grid nodes, the limit points (LG) by the blue ones



### 2.3 Experimental protocol

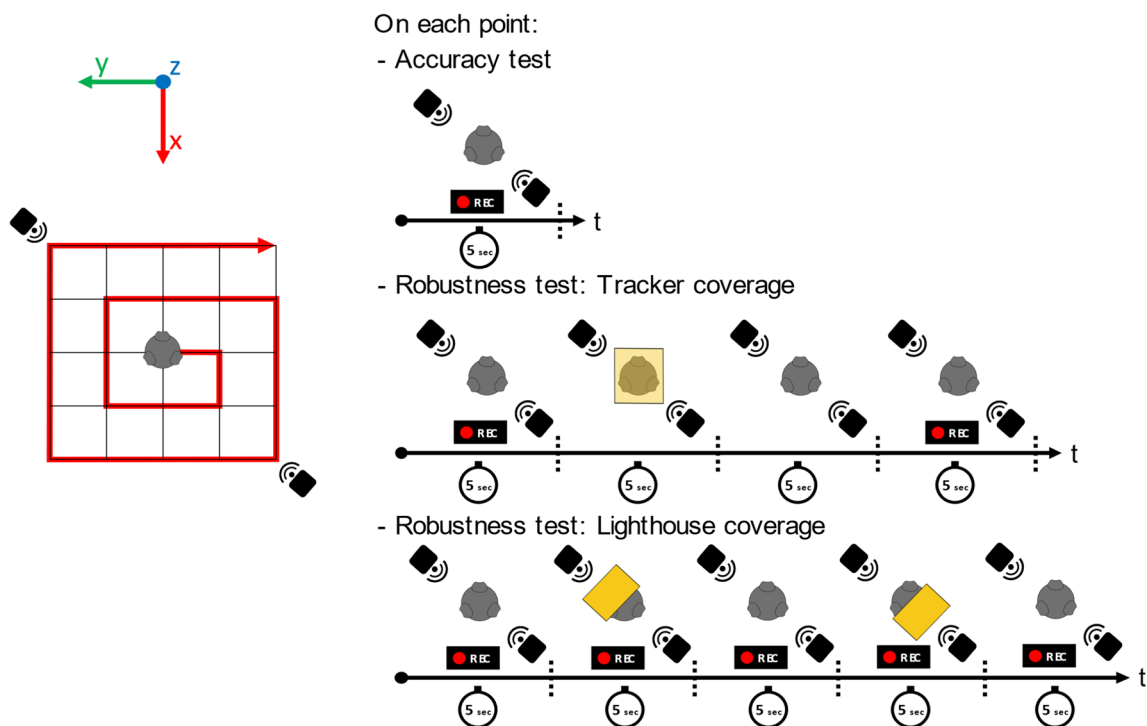
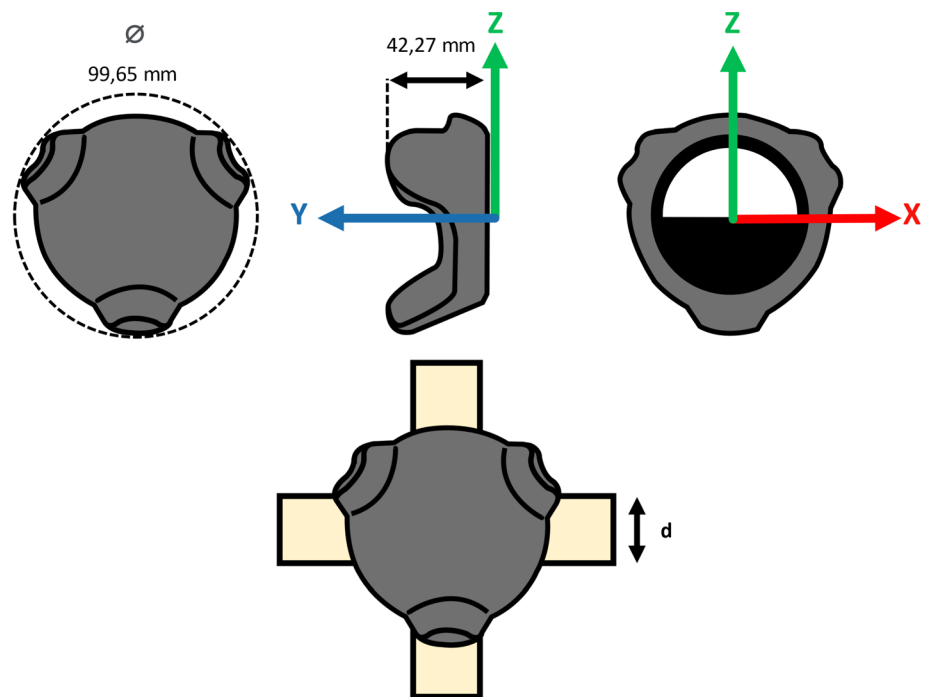
To validate the positional accuracy of the tracking system, we evaluated the  $x, y, z$  coordinates of the VIVE tracker (2018 version) registered in all the points of the grid. For each measurement session, the tracker was moved clockwise along the grid, from the centre to the outer borders. The choice to use the tracker instead of any other trackable support (headset or controller) was led by the optimal position of its reference frame, located at its base in direct contact to the floor, as shown in Fig. 3.

Each point measure was acquired for 5 s with a sampling rate of  $f_s = 40$  Hz, using custom software developed within the framework of Unreal Engine (v4.24).

To evaluate the positional accuracy of the Lighthouse technology and its robustness against possible tracking occlusion inside the capture volume, we ran three different tests:

- *Accuracy (AC) test*: this test aimed at evaluating the intra- and inter-trackers measurement variability. We conducted  $n_r = 4$  acquisitions with two different trackers, identified as T1 and T2. In each measurement repetition, the trackers' positions were acquired at each point of the grid according to the configuration shown in Fig. 4 with no obstruction in the capture volume.
- *Robustness to total occlusion (RTO)*: this test aimed at evaluating the system's robustness against total occlusions of a fixed duration,  $t_o = 5s$ . The tracker's position was acquired two times on each point of the grid, before

**Fig. 3** Dimensions and coordinate system of the 2018 Vive marker (a); the schematic representation of the cardboard base support under the tracker (b) (HTC Corp 2017)



**Fig. 4** Measurement session configurations and different test protocols shown from the top view. Note that the yellow squares represent the boxes used for the tracker occlusions

and after the tracker coverage, following the protocol displayed in Fig. 4. Before the second acquisition, a time of  $t_w = 5\text{ s}$  was waited, in order to allow the tracker to reacquire the signal. A time of 5 s is enough to both lose and

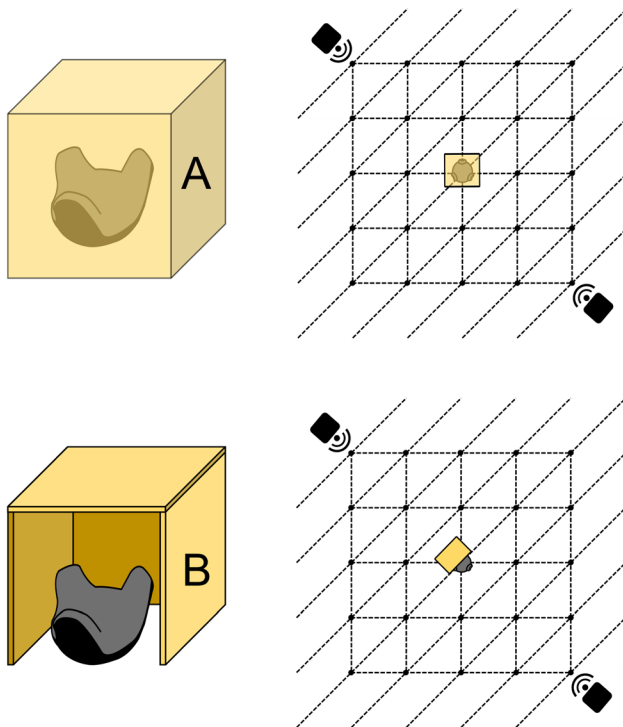
regain the sensor tracking, verified by observing in the SteamVR interface the real-time information about the connection status of the trackers. Measurement sessions were repeated four times on a single tracker.

- **Robustness to partial occlusion (RPO):** this test aimed at evaluating the robustness of the system against partial occlusions. The sight of the tracker was alternatively occluded to one of the lighthouses. The tracker's position was acquired five consecutive times at each point of the grid, following the protocol displayed in Fig. 4. Measurement sessions were repeated four times on a single tracker.

To ensure repeatability of marker positioning, a cardboard base support was placed underneath it. This support consists of two rectangular elements positioned perpendicular to each other, with the smaller side equal to  $d$ , as shown in Fig. 3. This cross-like shape allows at the same time for centring the marker on the nodal point of the grid and for aligning precisely its orientation as needed. For each test, the position of the tracker was standardised, with the status led facing towards the positive direction of the X-axis.

Two simple cardboard boxes, Box A and Box B in Fig. 5, were used to occlude the tracker visibility, respectively, for the RTO and the RPO tests. Box A covered entirely the tracker. Box B had an open side to cover the tracker visibility from one lighthouse at a time. For the RPO test, diagonal lines were added to the grid to allow repeatable directionality of Box B (dashed lines in Fig. 5).

Before each measurement session, the lighthouses, the SteamVR and the Unreal Engine software were restarted.



**Fig. 5** Schematic representation of Boxes A and B used to occlude the tracker, respectively, for the RTO and RPO tests

## 2.4 Data analysis

The  $(x, y, z)$  coordinates of the tracker were registered for 5 s at each of the  $n_p$  points on the grid, obtaining about  $n = 200$  samples for each point and per axis. This was repeated  $n_r$  times. The single sample in our dataset was identified with:

$$P_{ji,a}^{(k)} \tag{1}$$

where  $k = x, y, z$  identifies the axis,  $i = 1, \dots, n$  identified the  $i$ -th sample in the  $j$ -th point of the grid ( $j = 1, \dots, n_p$ ) and  $a = 1, \dots, n_r$  identified the  $a$ -th repetition. Therefore, the total number of samples is  $N_{tot} = N_a \cdot n_r$ , where  $N_a$  is the number of samples from the  $a$ -th acquisition.

All the signal processing and statistical data analysis were carried out using MATLAB (Matlab R2019b).

### 2.4.1 Pre-processing

A first-order low-pass Butterworth filter was applied for each acquired grid point to denoise the signal. Based on the FFT analysis, the cut-off frequency was set to 1.81 Hz to have 85% of the signal power spectral density.

Subsequently, a first graphical evaluation of the dataset highlighted that each repetition had a random offset  $P_a^{(k)}$ , as shown in Fig. 6. We therefore subtracted  $P_a^{(k)}$  from the dataset as follows:

$$P_{ji,a}^{(k)*} = P_{ji,a}^{(k)} - \bar{P}_a^{(k)} \tag{2}$$

$$\bar{P}_a^{(k)} = \frac{1}{N_a} \cdot \sum_1^{N_a} P_{ji,a}^{(k)} \tag{3}$$

where  $\bar{P}_a^{(k)}$  is the average value of the  $a$ -th acquisition and  $P_{ji,a}^{(k)*}$  is the corrected sample.

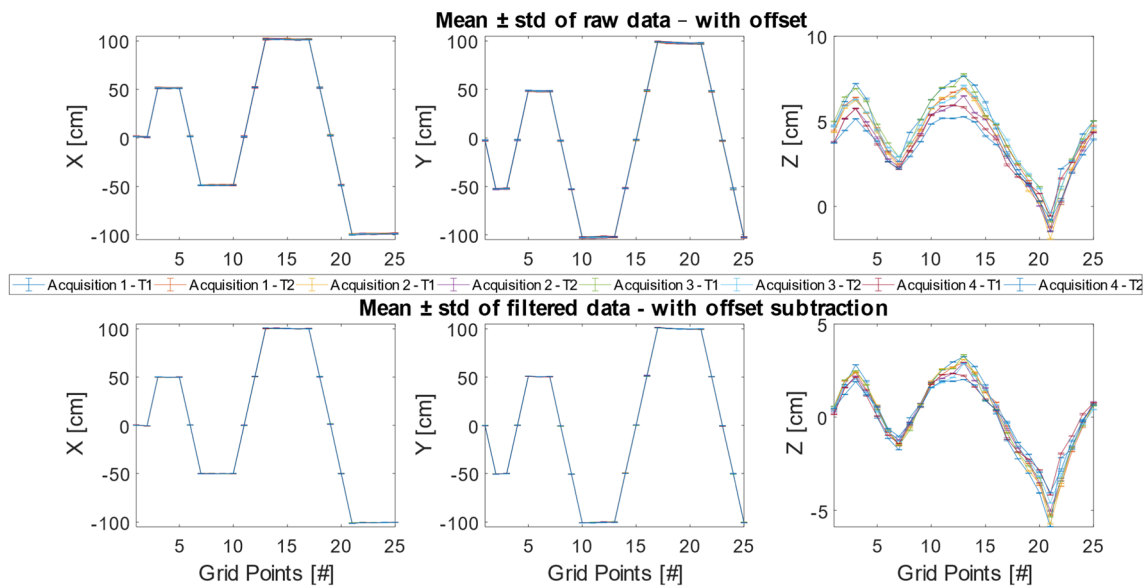
All three test datasets were filtered and adjusted by applying the offset subtraction.

### 2.4.2 Accuracy test: intra- and inter-tracker analysis

To evaluate the single tracker repeatability, an intra-tracker analysis was conducted. We calculated the difference between the position measurement in each grid point and the ideal grid point.

Then, we focussed on a detailed inter-tracker analysis to explore eventual differences between the two trackers, since many applications require tracking more than one point at a time.

Thus, we conducted the Bland–Altman analysis (described in Sect. 3.1.1) and we computed the average



**Fig. 6** AC test dataset processing: plots of the mean and standard deviations calculated on each grid point for each axis of acquisition (X, Y and Z), before (upper plots) and after (lower plots) the offset subtraction for both tracker T1 and T2

deviation between the positional measurements obtained by T1 and T2 for the XYZ volume (Eq. 4), for the XY plane (Eq. 5), and for the Z-axis (Eq. 6).

$$\Delta_{xyz}^i = \sqrt{(\delta_{x1x2}^i)^2 + (\delta_{y1y2}^i)^2 + (\delta_{z1z2}^i)^2} \quad (4)$$

$$\Delta_{xy}^i = \sqrt{(\delta_{x1x2}^i)^2 + (\delta_{y1y2}^i)^2} \quad (5)$$

$$\Delta_z^i = \delta_{z1z2}^i \quad (6)$$

where  $i = 1, \dots, 25$  and  $\delta_{x1x2}^i$ ,  $\delta_{y1y2}^i$ ,  $\delta_{z1z2}^i$  are the absolute differences between the measurement of the two tracker, for each axis (Eq. 7 for the X-axis, the same for Y and Z).

$$\delta_{x1x2}^i = \text{abs}(x_{T1}^i - x_{T2}^i) \quad (7)$$

Similarly, we computed the deviation between the positional measurements and the gold standard, represented by the ideal grid points coordinates.

### 2.4.3 Statistical analysis

The Bland–Altman (BA) analysis (Giavarina 2015) was conducted for the AC test to evaluate the measurement agreement between two trackers, T1 and T2. Acquisitions performed with T1 and T2 were identified as  $P_{ji,a}^{(k)}$  and  $Q_{ji,a}^{(k)}$ , respectively (following the convention of Eq. 1).

Before performing the BA analysis (Fig. 13), it was verified that the difference vector  $D_{ji,a}^{(k)}$  (Eq. 8), followed a normal distribution by observing histogram plots and skewness–kurtosis values.

$$D_{ji,a}^{(k)} = P_{ji,a}^{(k)} - Q_{ji,a}^{(k)} \quad (8)$$

Moreover, to investigate if there was a dependency between the difference T1-T2 and the trackers' position on the grid, a linear regression was conducted on the BA data for the X- and Y-axis. The resulting line and the corresponding slope were displayed on the BA plots.

The RTO test aimed at evaluating the behaviour of the system before and after the tracker coverage, by quantifying the difference in the grid points acquisitions before and after the total occlusion.

Pre- and post-coverage data were saved in two matrices  $P1_{ji,a}^{(k)}$  and  $P2_{ji,a}^{(k)}$ , respectively, defined as Eq. 1. Then the difference matrices  $D12_{ji,a}^{(k)}$  were defined as reported in Eq. 9.

$$D12_{ji,a}^{(k)} = P1_{ji,a}^{(k)} - P2_{ji,a}^{(k)} \quad (9)$$

Regarding the partial occlusion robustness test, the tracker position was recorded five times for each grid point, according to the different conditions described in Fig. 4. These five recordings were collected in different matrices  $P1, P2, P3, P4, P5$  and were compared two-by-two by analysing the difference matrices  $D12, D23, D34, D45, D51$ , defined as in Eq. 10.

$$\begin{aligned}
 D12_{ji,a}^{(k)} &= P1_{ji,a}^{(k)} - P2_{ji,a}^{(k)} \\
 D23_{ji,a}^{(k)} &= P2_{ji,a}^{(k)} - P3_{ji,a}^{(k)} \\
 D34_{ji,a}^{(k)} &= P3_{ji,a}^{(k)} - P4_{ji,a}^{(k)} \\
 D45_{ji,a}^{(k)} &= P4_{ji,a}^{(k)} - P5_{ji,a}^{(k)} \\
 D51_{ji,a}^{(k)} &= P5_{ji,a}^{(k)} - P1_{ji,a}^{(k)}
 \end{aligned}
 \tag{10}$$

Since the difference matrices for both the RTO and RPO tests (Eqs. 9 and 10) were not normally distributed, the nonparametric BA (NPBA) (Bland and Altman 2010) was applied. For the NPBA, the upper limits of agreement (ULoA) and the lower limits of agreement (LLoA) were identified by the 97.5th and the 2.5th percentiles (Bland and Altman 1999), each of them characterised by a 95% confidence interval (CI) calculated using a percentile bootstrap method based on 10 k samples (Davison and Hinkley 1997). The percentile method was chosen for its conservative nature, as it tends to produce wider CI less sensitive to population value and sample size (Jung et al. 2019). The ULoA and LLoA and the respective (CI) identified the limits of agreement (LoA) range. In all the BA and NPBA plots, we reported the corresponding grid point numbers near the circles that are located above the lower end of ULoA’s CI and/or under the upper end of the LLoA’s CI. This helped us understanding which grid points were more responsible for greater LoAs’ ranges.

### 2.4.4 Accuracy computation

The positional accuracy over the XY plane was computed as defined in Eq. 11:

$$Acc = mean(abs(\hat{d} - D))
 \tag{11}$$

$\hat{d}_i$  is the average geometric distance between the average origin measurement  $\hat{x}_1$  and the average positional measurement  $\hat{x}_i$  on the  $i$ -th over all grid points, as defined in Eq. 12.

$$\hat{d}_i = \sqrt{(\hat{x}_1 - \hat{x}_i)^2 + (\hat{y}_1 - \hat{y}_i)^2}
 \tag{12}$$

Likewise, D is the geometric distance vector obtained from Eq. 12 by considering the distance between the ideal origin, i.e. (0,0), and all the ideal grid points positions on the plane.

Then, to evaluate whether the accuracy varies depending on the position on the grid, we performed the same calculation, this time by differentiating the 8 CG (from 2 to 9) from the 8 LG (11,13,15,17,19,21,23,25) points defined above.

$$Acc_z = mean(abs(\hat{z} - Z))
 \tag{13}$$

For the Z-axis, we computed the accuracy as the mean value of the absolute difference between the ideal plane

points ( $z_i = 0$ ) and the average positional measurement  $\hat{z}_i$  (Eq. 13). Similarly, to the XY plane accuracy computation, the z-accuracy was computed for all grid points and differentiating the selected 8 CG from 8 LG points, as described above.

## 3 Results

### 3.1 Accuracy test: intra- and inter-tracker analysis

Both trackers identified a tilted surface for the floor plane on which the test was conducted, as shown in Fig. 9. The two surfaces were obtained by using, for each grid point, the average value of the processed data reported in Fig. 6. The surfaces identified with both trackers are comparable. The maximum Z-variation was lower than 8 cm and observed along with the  $2 \times 2$  m grid plan diagonal, resulting in a slope lower than 0.025.

The average values of the processed data, acquired with both trackers on each grid point, appear superimposed in comparison to the dimensions of the grid.

For the intra-tracker analysis, we reported the difference distribution for each axis separately in the boxplots of Fig. 10, where XT1 is the X-axis measurement for tracker T1 and so on. The boxplots showed a similar trend for all the acquisitions. The outliers in XT1 and XT2 are associated with the 19th and 21st grid points. The outliers in YT1 and YT2 are associated with the 17th, 24th and 16th, 17th points, respectively. Therefore, the outliers are located on the limit of the grid and more precisely to its outer corners (Fig. 2).

Then, we calculated the average deviation between the positional measurements obtained by the two trackers T1 and T2, by varying the position on the XY plane. Differences on the XY plane and in the volume XYZ are reported in Table 1.

Similarly, we computed the deviation between the positional measurements and the gold standard, represented by the ideal grid points coordinates, by merging both trackers acquisitions (Table 2).

**Table 1** The deviation between T1 and T2 positional measurements (mean ± SD)

T1 and T2 measurement deviation [cm]			
	XY plane	Z	XYZ
All points	0.37 ± 0.11	0.31 ± 0.15	0.49 ± 0.15
CG points	0.30 ± 0.07	0.26 ± 0.08	0.4 ± 0.1
LG points	0.41 ± 0.12	0.34 ± 0.18	0.54 ± 0.16

**Table 2** Deviation between positional measurements and the gold standard

Deviation from the gold standard [cm]			
	XY plane	Z	XYZ
All points	$0.65 \pm 0.31$	$1.67 \pm 1.15$	$1.87 \pm 1.06$
CG points	$0.48 \pm 0.24$	$1.06 \pm 0.69$	$1.24 \pm 0.59$
LG points	$0.74 \pm 0.32$	$2.01 \pm 1.23$	$2.22 \pm 1.11$

Values were differentiated by the number of points and for the XY plane, the Z-axis and XYZ volume

### 3.1.1 Bland–Altman analysis

In some RTO and RPO tests, the tracker lost its visibility on the LG points closest to the cameras, i.e. 12, 13, 14, 20, 21 and 22. This issue was probably due to the proximity to the vertical visibility limit of one of the two cameras (Fig. 2), but it was not present in all the acquisitions. In Fig. 7, the processed dataset of the RTO test is displayed. The grey bands on the plots indicated the data related to the points closest to the cameras. Even when the tracker was not completely lost in these grey areas, the values were characterised by visible artefacts. The same was observed for the RPO test (shown in Fig. 8). Thus, two BA analyses for the inter-trackers evaluation for the AC test were conducted: the first with all the grid points and the latter by excluding the ones closest to the camera field of view (FoW) limit. By removing these points, all the LoA ranges decreased by about 1 mm.

The BA plots, obtained by excluding the grid points closest to the camera FoW, are shown in Fig. 9, 10, 11, while those with the complete dataset are reported in "Appendix 1".

### 3.1.2 Accuracy computation

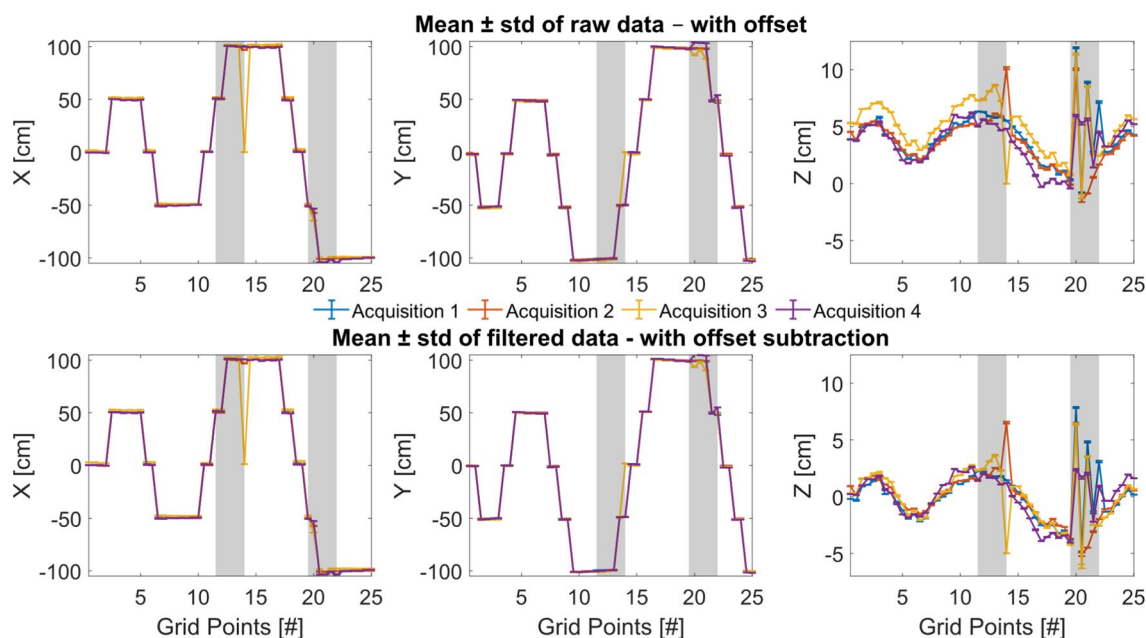
The computed accuracy on the XY plane, considering all the grid points, was  $0.5 \pm 0.2$  cm. An accuracy of  $0.4 \pm 0.1$  cm and  $0.6 \pm 0.1$  cm was obtained by differentiating the CG and the LG points, respectively. In line with that, the maximum deviation from the ideal position was observed in the grid points closest to the cameras. The deviation between the tracker's measurements and the ideal grid positions was due to both an intrinsic uncertainty ( $\sim 1$  mm) characterising the manual grid construction method and the systematic error occurring during the tracker repositioning on the grid points.

Meanwhile, for the Z-axis, by considering the acquisitions in all the grid points, the accuracy value was equal to  $1.7 \pm 1.2$  cm. The accuracy was equal to  $1.1 \pm 0.7$  cm and  $2.1 \pm 1.4$  cm for the CG and LG points, respectively.

### 3.2 Robustness to occlusions test

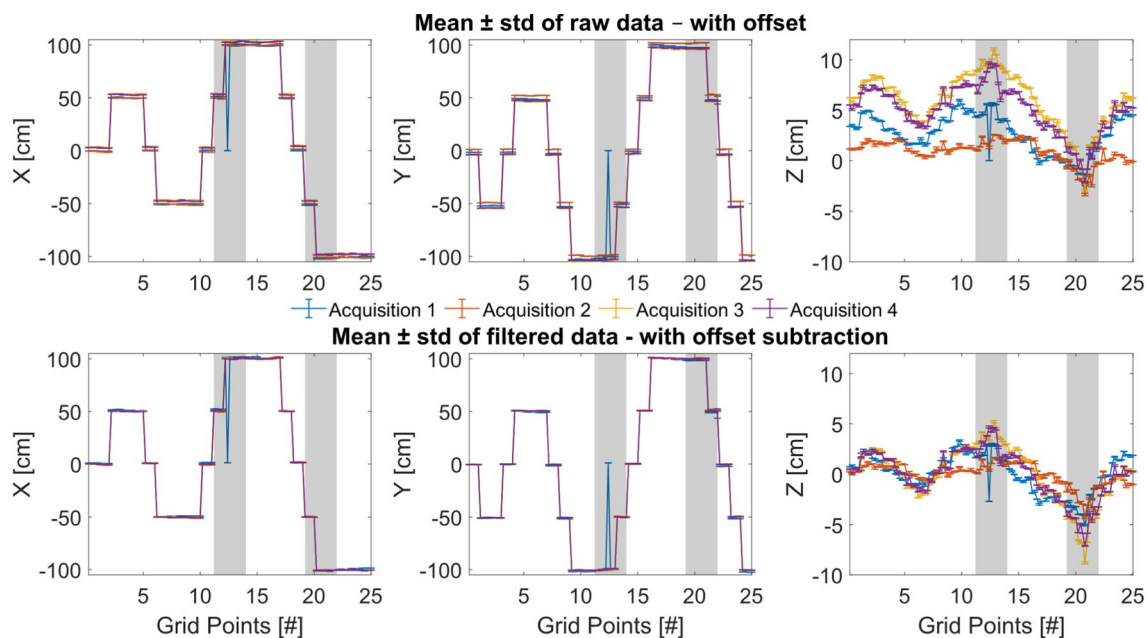
The NPBA analysis, for both the RTO and RPO tests, was conducted by excluding the points close to the cameras' field of view, as described for the AC test.

- *Total occlusion*: for the RTO test, the Bland–Altman plots and the values of the respective parameters (ULOAs,



**Fig. 7** Plot of the mean and standard deviations of both raw data (upper plots) and the processed data (filtered and without offset—lower plots) for the RTO test for X-, Y- and Z-axis





**Fig. 8** Plot of the mean and standard deviations of both raw data (upper plots) and the processed data (filtered and without offset—lower plots) for the RPO test for X-, Y- and Z-axis

LLOA, BIAS) are reported in Fig. 12. To summarise, the calculated LoAs were lower than 1.5 cm for the X-axis, 1 cm for the Y-axis and 2 cm for the Z-axis.

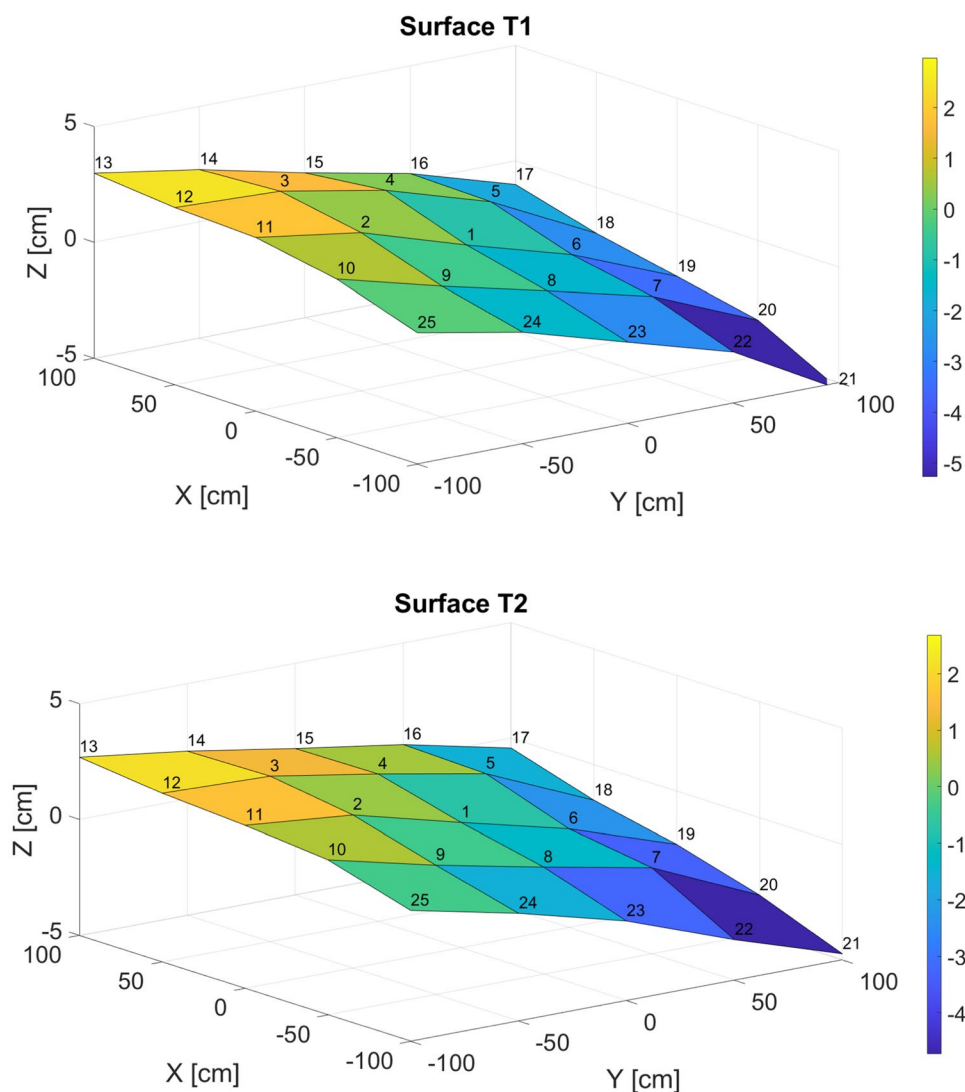
- *Partial occlusions*: Regarding the RPO test instead, the NPBA plots are reported in "Appendix 2", whereas values of the parameters (LoA, median and slope) are reported in Table 3.

## 4 Discussion

This study aimed at increasing the available evidence about the performance of the SteamVR tracking system, in terms of static-positional accuracy and its robustness in suboptimal operating conditions. Specifically, our main efforts were directed towards the systematic evaluation of the effects of partial and total occlusions in the recording volume. This issue represents a pivotal point in the field of motion tracking, and it has been highlighted as one of the main sources of measurement variability in various applications (Jiménez Bascones et al. 2019; Conconi et al. 2021). Thanks to the inherent characteristics of the SteamVR tracking system, that set it apart from the competitors in consumer-grade VR applications (Ikbal et al. 2021), several studies have already explored the capabilities of this technology for research and scientific applications. In particular, most of the literature underlines how systems based on SteamVR tracking provide excellent measurement performance with accuracy values that vary in millimetre or even sub-millimetre scales

(Spitzley and Karduna 2019b; Ameler et al. 2019; Veen et al. 2019a,b; Jost et al. 2019; Ikbal et al. 2021). For instance, Jost and colleagues (Jost et al. 2019) validated the positional and rotational performance of such technology in tracking the controller inside a room-scale (i.e. lighthouses positioned at 5.6 m apart) and standing configurations (i.e. lighthouses positioned at 2.6 m apart) against an optoelectronic motion-capture system. They obtained a sub-millimetric measurement difference ( $0.74 \pm 0.42$  mm under room-scale calibration and  $0.63 \pm 0.27$  mm for standing calibration trials) for robot-driven motions, which increased up to  $3.97 \pm 3.37$  mm for human movements. Similar evidence was found by Spitzley and Karduna (2019b), who compared the performance of the SteamVR tracking with a gold-standard magnetic tracking system, obtaining average error values below 0.35 mm. Other works highlighted the limitations of the system, reporting error values that vary in centimetres up to metre scales for dynamic movements (Niehorster et al. 2017; Borges et al. 2018; Hemphill et al. 2020). By implementing an HTC Vive system as a mobile-unit for room-to-room therapy, a recent work estimated the mean translational errors for the motion tracker, the controllers, and the HMD as  $(2.43 \pm 1.57)$ cm,  $(3.63 \pm 1.27)$ cm, and  $(2.10 \pm 0.61)$ cm, respectively (Hemphill et al. 2020). Among the variability of reported results in the literature, our study supported the evidence of a millimetric precision of the tracking system across the XY plane, identifying an average accuracy of  $0.5 \pm 0.2$  cm. As for the Z-axis, measurements showed a slope among the direction linking the two lighthouses,

**Fig. 9** Colour map of the recorded Z-axis positions on the grid, both for T1 and T2 (upper plots) and 3D representations of the colour map (lower plots). The Z-axis has a different scale than the other two axes, to better show the slope

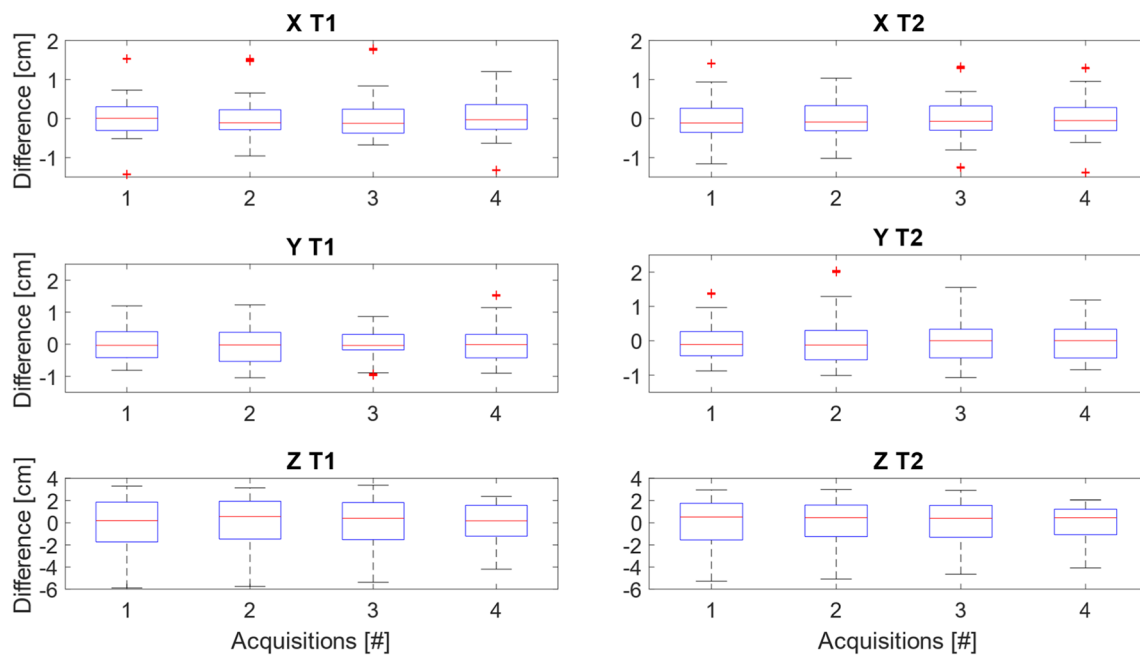


which increased the variability with respect to the XY plane, i.e.  $(1.7 \pm 1.2)$  cm. More interestingly, the further analysis conducted on the different points of the recording area indicated that the central part should be considered as the ideal registration position, as pointed out by the higher level of accuracy for both the XY plane and the Z-axis.

According to our findings, the analysis of Niehorster and colleagues (2017) also reported how the accuracy of the HTC Vive system decreased in the identification of the HMD position along the vertical axes. They found good performance on planar measurements but a variability in the recorded height ranging from  $\sim 40$  cm to  $\sim 4$  cm across space, following a tilted surface. Conversely, by placing a robotic arm in the centre of the acquisition volume at a height of 0.90 m and tracing its motion with a VIVE tracker, Amler and colleagues (2019) reported sub-millimetric accuracies along all the three coordinate axes. These discrepancies could be due to the strict dependence of the SteamVR

tracking system's performance on the experimental setup and the application in which it is used (Ikbal et al. 2021). In line with that, our results could be influenced by the vertical distance of 3 m, which places the measurement grid at the lower limit of the acquisition volume. A different setup, with a grid positioned at 1 m height, may have led to better accuracy results. However, the correct identification of the floor level plays a fundamental role for the game development, since it represents the origin of the SteamVR tracking system, defined during the calibration procedure.

The use of two different trackers did not significantly affect the overall quality of the measures and the trajectories of the  $(x,y,z)$  measures across the grid followed similar trends independently from the specific tracker or acquisition. This was even clearer after the initial offset removal in the data pre-processing, indicating a high inter-rater and test-retest reliability. Niehorster (2017) highlighted the presence of a random offset between different acquisitions and



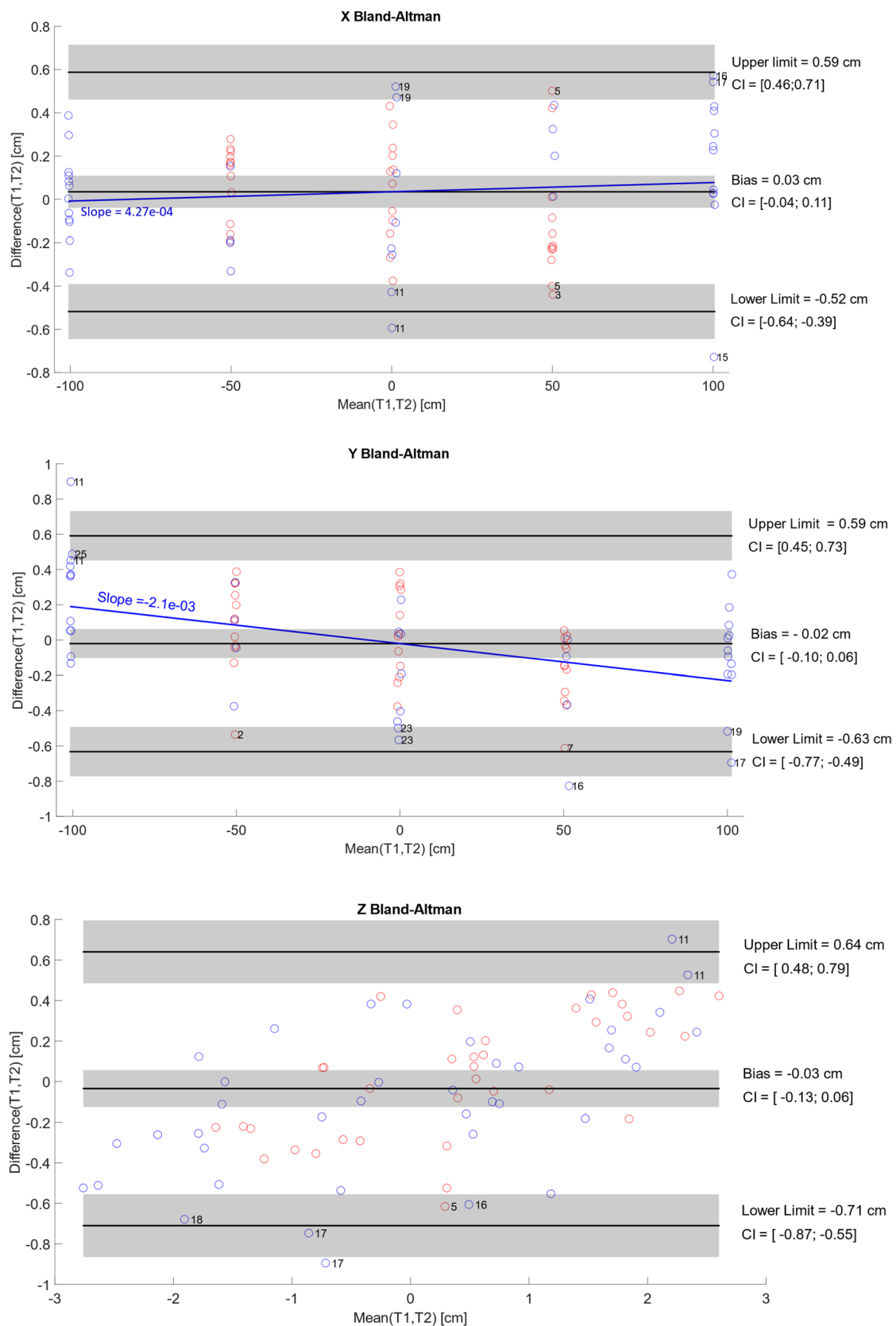
**Fig. 10** Distribution of the difference between the position measurement in each grid point and the ideal grid point. Boxplots were grouped by tracker, axis and acquisition

after brief signal losses, stating that this may continuously influence the measurement, and even a calibration procedure could be ineffective to correct it. Our robustness to occlusion tests, however, support only partially these conclusions: tracking artefacts occurred specifically on the points closest to the lighthouses' field of view limits. Limited visibility from the SteamVR lighthouses resulted in variability and loss of accuracy, with a consequent spatial-dependant behaviour of the measures. This was evidenced in the Bland–Altman plots since most of the points falling close to the LoA correspond to the outermost points of the grid. Therefore, particular attention must be paid to setting up the system by selecting the recording volume that benefits from the greater visibility from both the tracking stations. Then, we demonstrated the robustness of the SteamVR tracking system for short transient occlusions occurring in the capture volume since the nonparametric Bland–Altman analysis showed small LoA ranges comparable to the one of the accuracy test ( $\sim 2$  cm).

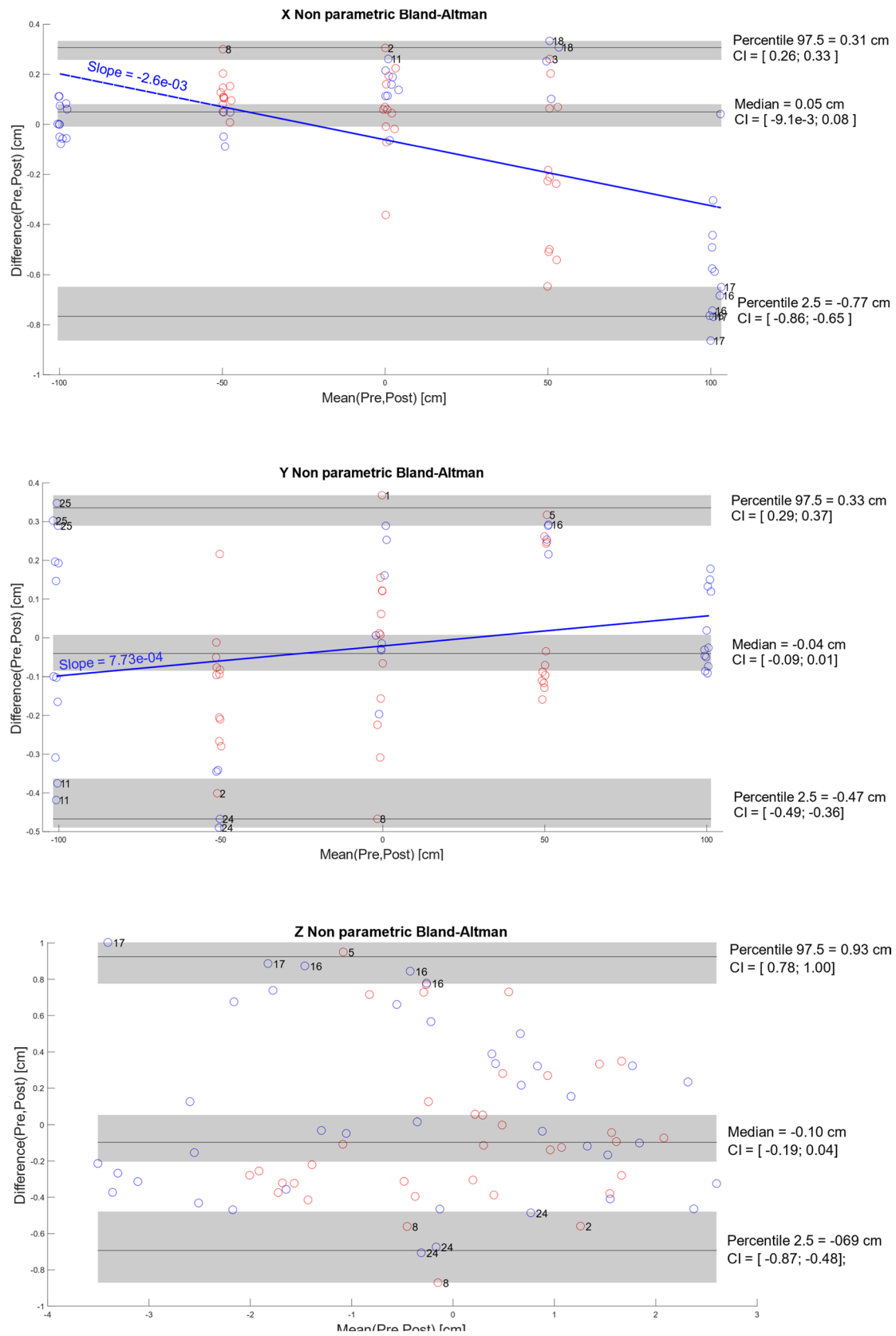
The empirical findings of the present study should be evaluated in light of two main limitations. Our systematic approach was majorly focussed on the in-depth evaluation of the tracking occlusion issue across the recording area, limiting the analysis to static-positional acquisitions on a single plane. Further efforts should aim at characterising completely the SteamVR tracking, exploiting the presented methodology for volumetric and dynamic measures for position and orientation. Moreover, the intrinsic nature of SteamVR tracking considers the combined contribution of

the Inertial Measurement Units (IMUs) and optical sensors. In the presence of occlusion, IMUs can be used for position and orientation estimation. In the first case, the position is estimated by double integration on the accelerometer data, even if this measurement, therefore, presents a considerable drift (Sitole et al. 2020). The gyroscopes of the IMUs provide the orientation, and the inevitable integration drift is partially corrected for pitch and roll by the accelerometer data. In contrast, the yaw correction is more difficult, having to rely on the magnetometers (Stanzani et al. 2020). The drift generally depends on the IMUs used, their calibration and environmental factors, e.g. temperature (Paternain et al. 2013). To evaluate the drift magnitude on all the IMU-derived measurements, a thorough analysis of the effect of time under-recording occlusion conditions would be needed. Important to notice, the new version of the SteamVR tracking system 2.0 has been recently commercialised and, by allowing the simultaneous connection of 4 lighthouses, it could have better performances and the potential to overcome the issues highlighted in the present work. In line with that, future studies could investigate the robustness to occlusions of the new technologies.

Overall, the presented evidence, together with the proven robustness of the tracking to both partial and total occlusions, promotes the SteamVR system for static measures in the clinical field. The estimated error can be considered clinically irrelevant for exercises aimed at the rehabilitation of functional movements of multiple joints simultaneously whose several motor outcomes



**Fig. 11** BA plots of the AC test for T1 and T2 for the X-axis (upper plot), Y-axis (middle plot) and Z-axis (lower plot). The blue line is the linear regression of the data. The red and blue circles indicate the means of the  $n$  samples corresponding to each of the CG and LG points, respectively



**Fig. 12** NPBA plots of the robustness test pre and post the total occlusion, for the X-axis (upper plot), Y-axis (middle plot) and Z-axis (lower plot). Each plot point is the mean of n samples. The red and

blue circles indicate the means of the n samples corresponding to each of the CG and LG points, respectively

**Table 3** NPBA analysis results for the five different conditions (D12, D23, D34, D45, D51) of the RPO test

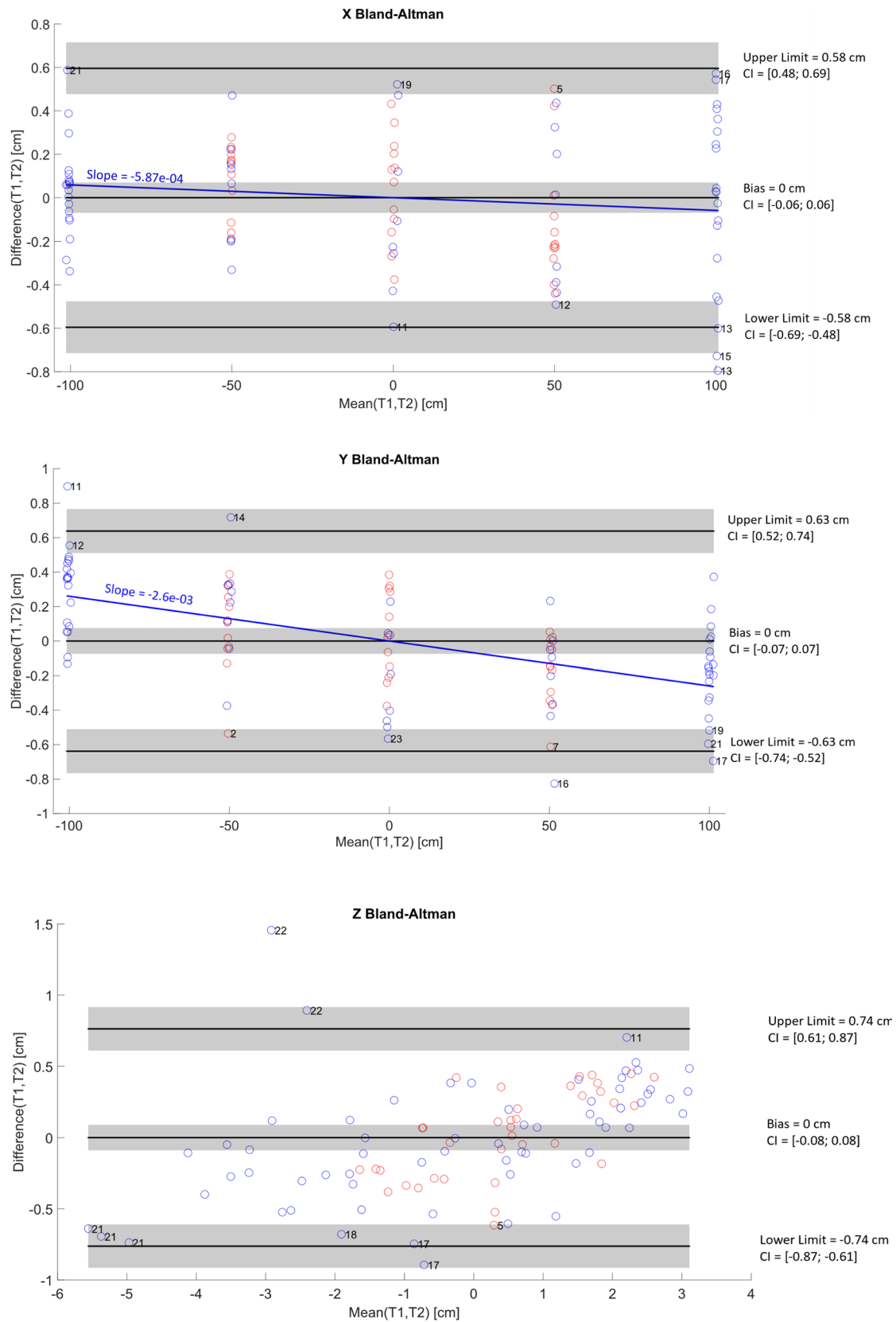
	Percentile 2.5 [cm]	Median [cm]	Percentile 97.5 [cm]	Slope
<b>X</b>				
D12	-0.94 CI=[-1.07; -0.39]	-0.06 CI=[-0.08; -2.9e-03]	0.51 CI=[0.17; 0.61]	2.1e-04
D23	-0.45 CI=[-0.67; -0.35]	0.04 CI=[2.3e-03; 0.08]	0.99 CI=[0.43; 1.12]	-1.8e-03
D34	-0.56 CI=[-0.73; -0.22]	-0.02 CI=[-0.03; -4e-03]	0.22 CI=[0.16; 0.23]	-7.4e-04
D45	-0.17 CI=[-0.18; -0.13]	5.5e-03 CI=[-6.1e-03; 0.02]	0.26 CI=[0.11; 0.32]	2.1e-04
D51	-0.30 CI=[-0.35; -0.26]	-3.9e-03 CI=[-0.05; 0.03]	0.94 CI=[0.47; 1.18]	2.1e-03
<b>Y</b>				
D12	-0.31 CI=[-0.48; -0.11]	0.03; CI=[-5.1e-03; 0.07]	0.55 CI=[0.45; 0.65]	1.2e-03
D23	-0.60 CI=[-0.71; -0.45]	-0.04 CI=[-0.11; -6.8e-3]	0.29 CI=[0.20; 0.33]	-1.2e-03
D34	-0.36 CI=[-0.50; -0.17]	-0.01 CI=[-2.2e-03; 0.05]	0.28 CI=[0.13; 0.31]	4.0e-04
D45	-0.15 CI=[-0.18; -0.09]	-4.1e-03 CI=[-0.02; 9.5e-04]	0.27 CI=[0.15; 0.36]	-3.4e-04
D51	-0.43 CI=[-0.46; -0.34]	2.1e-03 CI=[-0.03; 0.04]	0.39 CI=[0.31; 0.47]	-8.1e-05
<b>Z</b>				
	Percentile 2.5 [cm]	Median [cm]	Percentile 97.5 [cm]	
D12	-1.44 CI=[-1.61; -0.84]	0 CI=[-0.19; -1.1 e-03]	0.61 CI=[0.44; 0.91]	
D23	-0.48 CI=[-0.87; -0.30]	0.10 CI=[-0.02; 0.18]	1.11 CI=[0.84; 1.55]	
D34	-0.44 CI=[-0.57; -0.27]	0.03 CI=[1.1e-03; 0.08]	0.64 CI=[0.35; 0.80]	
D45	-0.29 CI=[-0.35; -0.20]	-0.01 CI=[-0.06; 0.02]	0.30 CI=[0.24; 0.41]	
D51	-1.08 CI=[-1.33; -0.65]	-1.3e.03 CI=[-0.07; 0.10]	0.64 CI=[0.50; 0.71]	

are generally measured on the scale of metres (e.g. back forward flexion, reaching exercises, etc.) and where previous VR systems were studied and deployed (Hemphill et al. 2020). The person immersed in the IVR is free to move and to get in contact with trackers which can be represented by real-world objects in the virtual environment (Maciejewski et al. 2018). Real-world objects can be chosen according to the patient needs and preferences making the application tailored and more attractive. Trackers could be considered as real external focus of attention (EFA). Research reported that EFAs are more effective than internal focus of attention (IFA) in terms of motor performance and learning improvements (Rossettini et al. 2017) (Piccoli et al. 2018). Indeed, EFAs provide a higher level of movement efficacy and kinematics across

different types of tasks, skill levels and age groups (Wulf 2013). In line with that, performing an IVR exercise can increase patient engagement, leading to better rehabilitative outcomes (Laut et al. 2015). Simple examples of some possible rehabilitative applications could be referred to reaching exercises with objects located on the ground for patients with low back pain, weightlifting exercises and certainly, the y balance test (Powden et al. 2019).

## Appendix 1: Bland–Altman plots for the AC test with all grid points

See Fig. 13.



**Fig. 13** Bland–Altman plots of the AC test for T1 and T2 for X-axis (upper plot), Y-axis (middle plot) and Z-axis (lower plot). Blue line represents the linear regression of the data. The red and blue circles

indicate the means of the n samples corresponding to each of the CG and LG points, respectively

## Appendix 2: Bland–Altman plots to RPO test

See Fig. 14.

**Fig. 14** NPBA for the 3 axes (x, y, z) for the different combinations of the 5 conditions of the RPO test. For the X- and Y- axis, the blue line represents the linear regression of the data. The red and blue circles indicate the means of the n samples corresponding to each of the CG and LG points, respectively

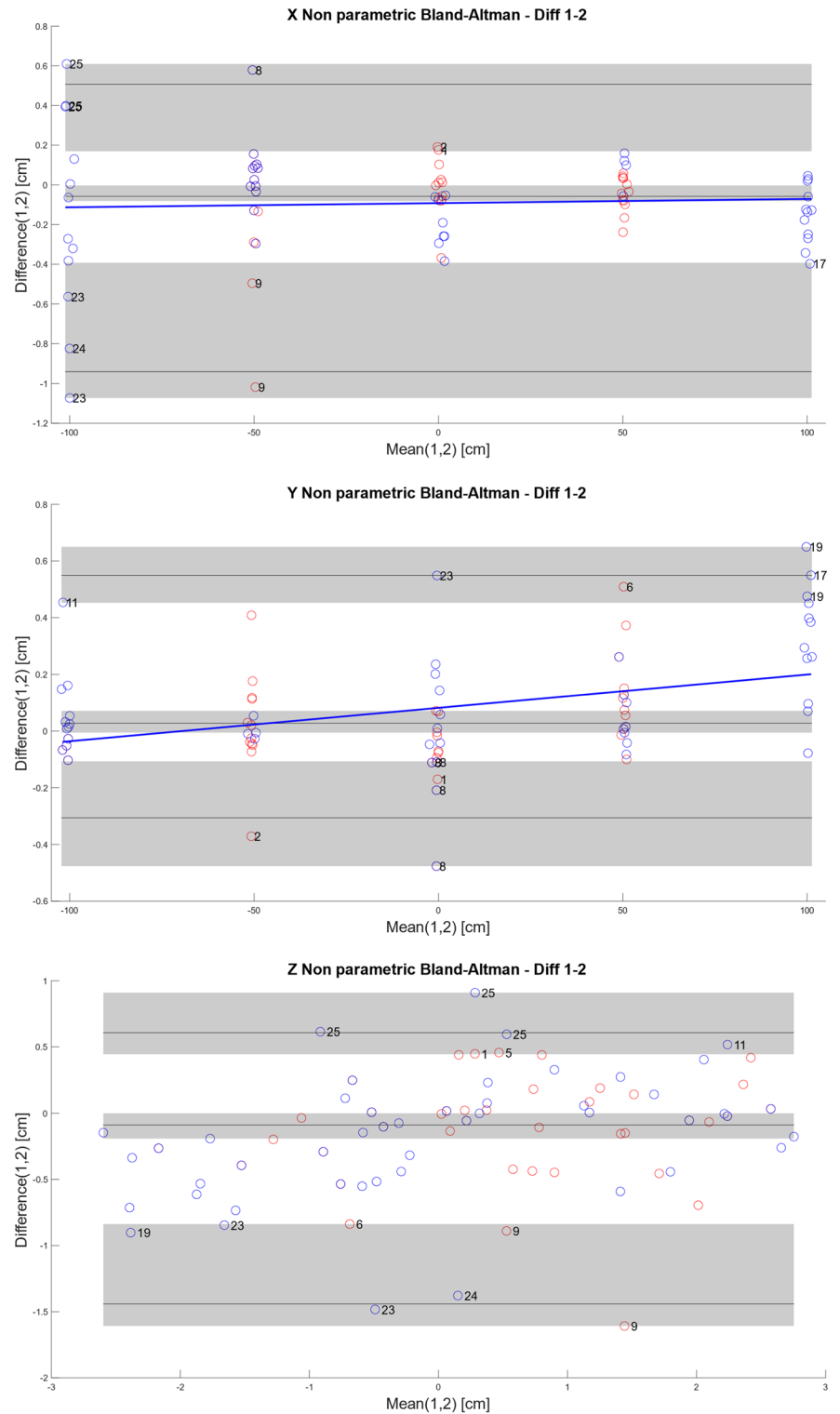




Fig. 14 (continued)

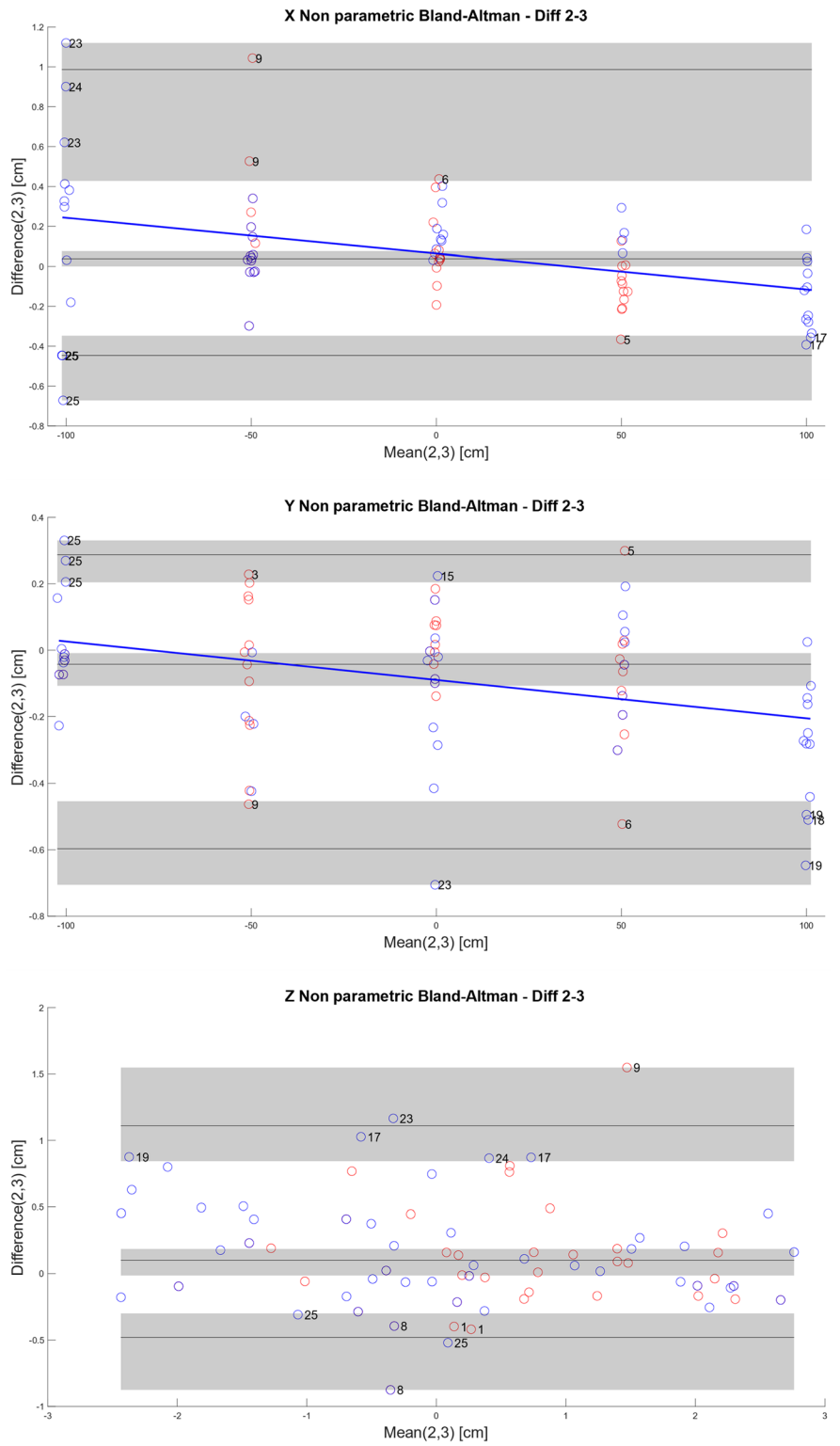


Fig. 14 (continued)

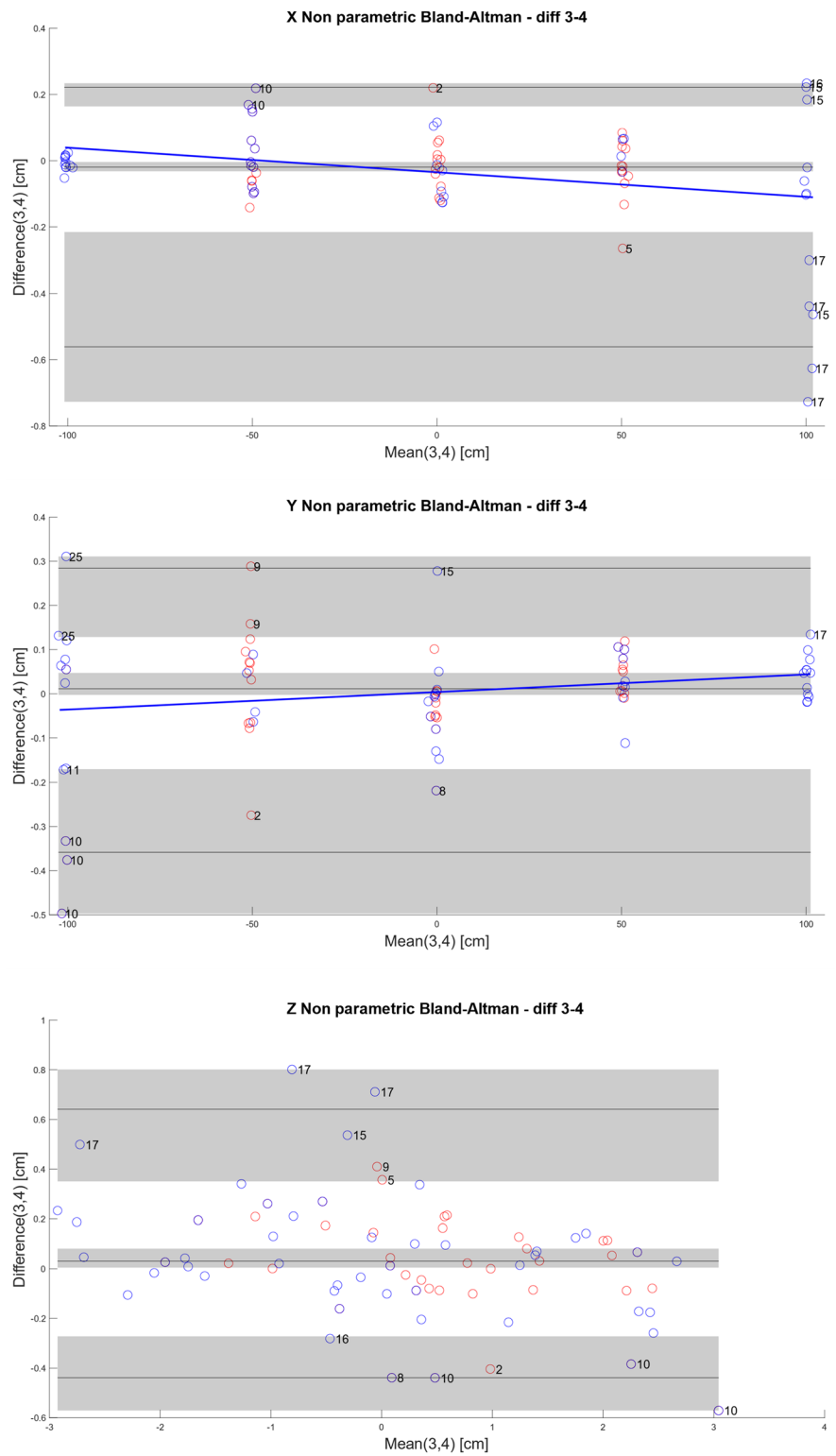


Fig. 14 (continued)

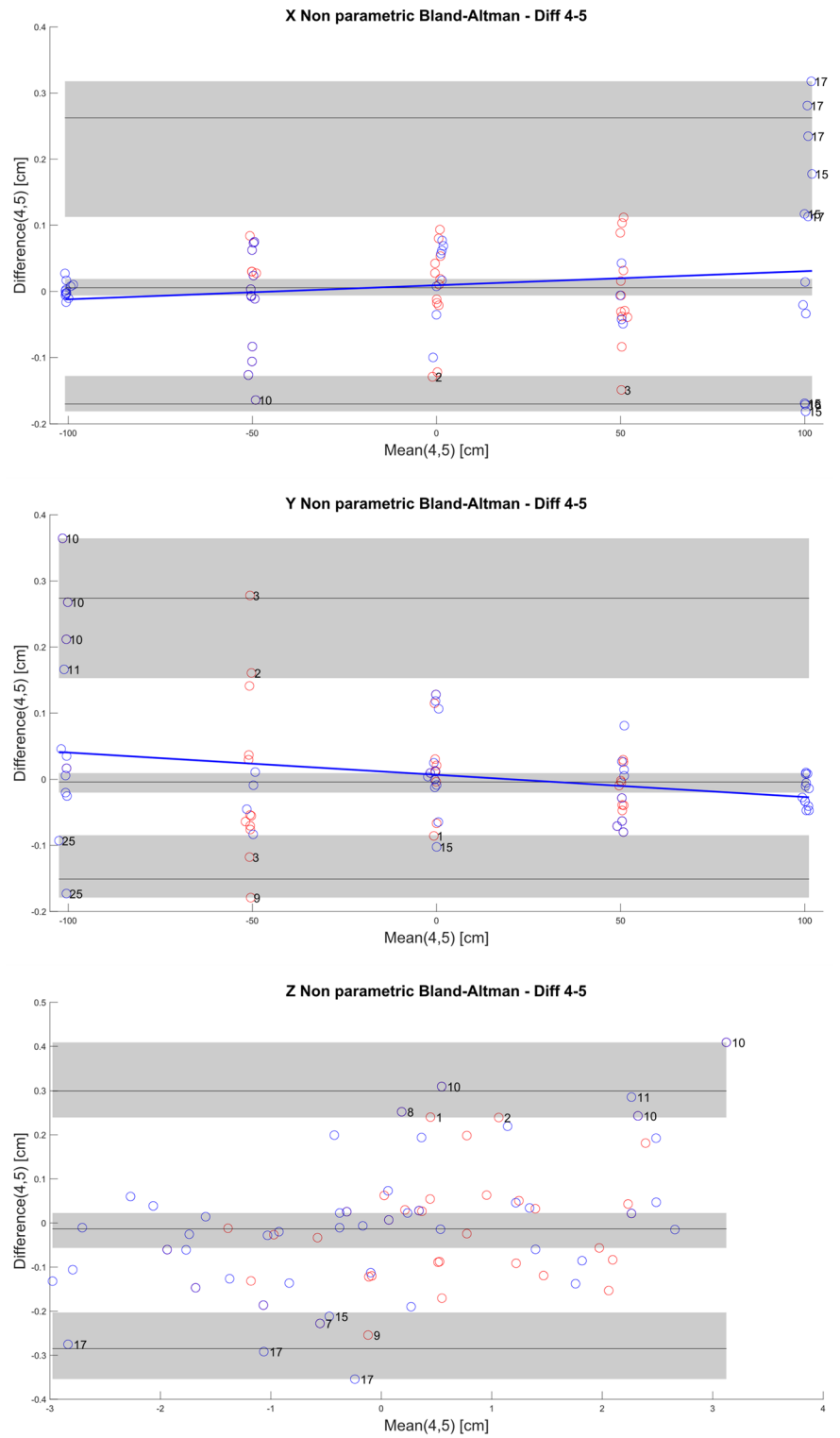
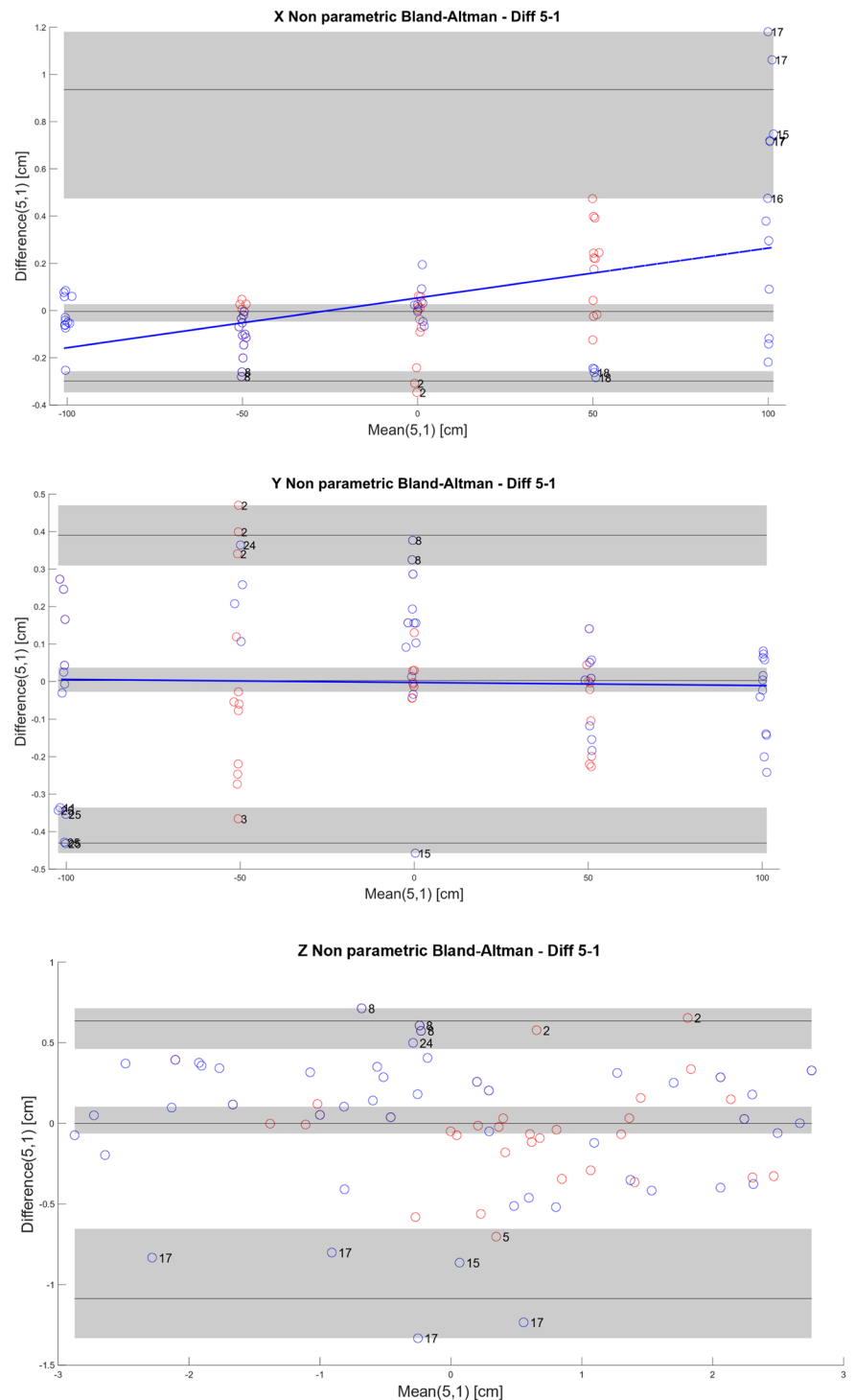


Fig. 14 (continued)



**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10055-021-00584-5>.

**Acknowledgements** This work was developed within the framework of the DINOGMI Department of Excellence of MIUR 2018-2022 (232-2016).

**Authors' contributions** All authors made substantial contributions to the conception and design, data acquisition, or analysis and interpretation of data. All authors participated in drafting the article or revising it critically for important intellectual content. All authors gave final approval of the version to be published. All authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Funding** This research did not receive any external funding.

**Data availability** The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ameler T, Blohme K, Brandt L, et al (2019) A comparative evaluation of SteamVR Tracking and the OptiTrack system for medical device tracking. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). pp 1465–1470
- Bland JM, Altman DG (2010) Statistical methods for assessing agreement between two methods of clinical measurement. 9
- Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160. <https://doi.org/10.1177/096228029900800204>
- Borges M, Symington A, Coltin B, et al (2018) HTC Vive: analysis and accuracy improvement. In: 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS). pp 2610–2615
- Campo-Prieto P, Cancela JM, Rodríguez-Fuentes G (2021) Immersive virtual reality as physical therapy in older adults: present or future (systematic review). *Virtual Real*. <https://doi.org/10.1007/s10055-020-00495-x>
- Cipresso P, Giglioli IAC, Raya MA, Riva G (2018) The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2018.02086>
- Conconi M, Pompili A, Sancisi N, Parenti-Castelli V (2021) Quantification of the errors associated with marker occlusion in stereophotogrammetric systems and implications on gait analysis. *J Biomech* 114:110162. <https://doi.org/10.1016/j.jbiomech.2020.110162>
- Davison AC, Hinkley DV (1997) *Bootstrap Methods and Their Application*. Cambridge University Press
- Giavarina D (2015) Understanding Bland Altman analysis. *Biochem Med* 25:141–151
- Hemphill S, Nguyen A, Rodriguez ST et al (2020) Mobilization and calibration of the HTC VIVE for virtual reality physical therapy. *Digit Health* 6:2055207620950929. <https://doi.org/10.1177/2055207620950929>
- HTC Corp. (2020) 2Q29100 HTC Vive headset User Manual
- HTC Corp. (2017) HTC Vive Tracker Developer Guidelines (v1.3)
- Ikkal MS, Ramadoss V, Zoppi M (2021) Dynamic pose tracking performance evaluation of htc vive virtual reality system. *IEEE Access* 9:3798–3815. <https://doi.org/10.1109/ACCESS.2020.3047698>
- Izard SG, Juanes JA, García Peñalvo FJ et al (2018) Virtual reality as an educational and training tool for medicine. *J Med Syst* 42:50. <https://doi.org/10.1007/s10916-018-0900-2>
- Jiménez Bascones JL, Graña M, Lopez-Guede JM (2019) Robust labeling of human motion markers in the presence of occlusions. *Neurocomputing* 353:96–105. <https://doi.org/10.1016/j.neucom.2018.05.132>
- Jost TA, Drewelow G, Koziol S, Rylander J (2019) A quantitative method for evaluation of 6 degree of freedom virtual reality systems. *J Biomech* 97:109379. <https://doi.org/10.1016/j.jbiomech.2019.109379>
- Jung K, Lee J, Gupta V, Cho G (2019) Comparison of bootstrap confidence interval methods for GSCA using a monte carlo simulation. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2019.02215>
- Laut J, Cappa F, Nov O, Porfiri M (2015) Increasing patient engagement in rehabilitation exercises using computer-based citizen science. *PLoS ONE* 10:e0117013. <https://doi.org/10.1371/journal.pone.0117013>
- Lombard M, Ditton T (1997) At the heart of it all: the concept of presence. *J Comput-Mediat Commun* 3:34. <https://doi.org/10.1111/j.1083-6101.1997.tb00072.x>
- Maciejewski M, Piszczek M, Pomianek M (2018) Testing the SteamVR trackers operation correctness with the OptiTrack system
- Maciejewski M, Piszczek M, Pomianek M, Pałka N (2020) Design and evaluation of a steamvr tracker for training applications – simulations and measurements. *Metrol Measur Syst* 27:601–614
- Nabiyouni M, Scerbo S, Bowman DA, Höllerer T (2017) Relative effects of real-world and virtual-world latency on an augmented reality training task: an AR simulation experiment. *Front ICT*. <https://doi.org/10.3389/fict.2016.00034>
- Niehorster DC, Li L, Lappe M (2017) The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. *i-Perception*. <https://doi.org/10.1177/2041669517708205>
- Paternain S, Tailanian M, Canetti R (2013) Calibration of an inertial measurement unit. In: 2013 16th international conference on advanced robotics (ICAR). IEEE, Montevideo, Uruguay, pp 1–6
- Piccoli A, Rossetini G, Cecchetto S et al (2018) Effect of attentional focus instructions on motor learning and performance of patients with central nervous system and musculoskeletal disorders: a systematic review. *J Funct Morphol Kinesiol* 3:40. <https://doi.org/10.3390/jfmk3030040>
- Powden CJ, Dodds TK, Gabriel EH (2019) The reliability of the star excursion balance test and lower quarter y-balance test in healthy adults: a systematic review. *Int J Sports Phys Ther* 14:683–694
- Rossetini G, Testa M, Vicentini M, Manganotti P (2017) The effect of different attentional focus instructions during finger movement tasks in healthy subjects: an exploratory study. *Biomed Res Int* 2017:e2946465. <https://doi.org/10.1155/2017/2946465>
- Sitole SP, LaPre AK, Sup FC (2020) Application and evaluation of lighthouse technology for precision motion capture. *IEEE Sens J* 20:8576–8585. <https://doi.org/10.1109/JSEN.2020.2983933>
- Spitzley KA, Karduna AR (2019a) Feasibility of using a fully immersive virtual reality system for kinematic data collection. *J Biomech* 87:172–176. <https://doi.org/10.1016/j.jbiomech.2019.02.015>
- Stanzani R, Dondero P, Mantero A, Testa M (2020) Measurement accuracy of an upper limb tracking system based on two Hillcrest Labs BNO080 IMU sensors. An environmental assessment. *IEEE Sens J*
- van der Veen SM, Bordeleau M, Pidcoo PE et al (2019a) Agreement analysis between vive and vicon systems to monitor lumbar postural changes. *Sensors (basel)*. <https://doi.org/10.3390/s19173632>

van der Veen SM, Bordeleau M, Pidcoe PE et al (2019b) Agreement analysis between vive and vicon systems to monitor lumbar postural changes. *Sensors* (basel, Switzerland). <https://doi.org/10.3390/s19173632>

Wulf G (2013) Attentional focus and motor learning: a review of 15 years. *Int Rev Sport Exerc Psychol* 6:77–104. <https://doi.org/10.1080/1750984X.2012.723728>

VIVE Pro European Union (2021) <https://www.vive.com/eu/product/vive-pro/>. Accessed 13 Jul 2021

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.