2024 International Conference for Biomedical Robotics and Biomechatronics (BioRob), Heidelberg, Germany, Sep 2024.

Evaluating Mixed Reality Technology for Tracking Hand Motion for Shoulder Rehabilitation Assessment*

Sergio A. Salinas, *Student Member, IEEE,* Katarina Grolinger, *Member, IEEE,* Marie-Eve LeBel, and Ana Luisa Trejos, *Senior Member, IEEE*

***Abstract*— Shoulder injuries and conditions are common musculoskeletal complaints that can limit a patient's range of motion and daily activities. Recently, serious games and mixed reality technologies, such as the HoloLens, have been proposed for shoulder rehabilitation. However, it is unclear if this technology accurately tracks 3D hand movements for reporting therapy-related kinematic metrics. This paper presents accuracy and repeatability tests of the HoloLens 2 in tracking hand movements, and its potential for shoulder rehabilitation assessment. Comparisons were made between index fingertip, palm, and wrist movements captured by the HoloLens 2 and an Aurora electromagnetic system, which was used as the ground truth. A mixed-reality environment was developed to capture static hand positions, as well as dynamic hand movements performed during a shoulder physiotherapy-based exercise. The tracking data were used to calculate several kinematic metrics. The results show that the HoloLens 2 hand-tracking system is accurate to within a median of 10.2 mm and has repeatability comparable to the Aurora system, with the palm exhibiting the best results. The HoloLens 2 data are suitable for computing kinematic metrics for shoulder rehabilitation assessment, achieving accuracies above 86.9% for all of the tested metrics. Metrics such as time-to-speed peak and the log dimensionless jerk were found to have significant differences between dynamic hand movements. These findings support the mixed reality technology potential to assist shoulder rehabilitation through immersive and interactive environments.**

I. INTRODUCTION

Shoulder affections are one of the most frequent musculoskeletal complaints [1], [2], which usually worsen with age [1], [3], limiting the patient's range of motion, dexterity, and muscle strength [2], [4]. Since shoulder motion is essential for completing basic activities of daily living such as eating, bathing, and dressing, restoring their function is a critical goal in upper limb rehabilitation. Therefore, the intensity and precision of the rehabilitation exercises are important for the patient's recovery even in the presence of shoulder pain [4]–[6]. However, it is difficult for patients to endure the pain and remain motivated while doing repetitive exercises, affecting their performance and increasing recovery time [7]–[9]. Furthermore, clinicians often subjectively evaluate therapy progress, and patients lack information about their performance [7].

As a potential solution, the combination of serious games (SGs) and virtual reality (VR), augmented reality (AR), or mixed reality (MR), has been proposed in rehabilitation. SGs are video games made to educate or train rather than entertain. In VR, the users interact with a completely immersive virtual world, while AR shows virtual elements superimposed on top of a real environment, and MR combines virtual and physical content in an interactive system [10]. The disadvantage of VR is that it limits the patient's observation of their body movements and the real environment [8], [11]. Therefore, the person is not aware of their surroundings and cannot interact with physical objects [9]. The AR and MR approaches solve this issue, but MR can deliver more immersive experiences, track user movements, and provide more intuitive interactions [11].

When combined with therapy exercises, SGs and MR can increase treatment intensity and patient engagement while also enabling the recording of movement data [8], [12], [13]. Motion tracking systems included in videogame consoles and virtual or mixed reality devices have been used to record these data. The Kinect employs an optical system [14]–[16], while the Wii uses inertial sensors [16], [17]. Both devices are examples of console technologies. The High Tech Computer Vive [18] is a VR headset, while the HoloLens 2 (HL2) [13] falls into the MR category. The motion data captured by these types of systems are relevant in physiotherapy because they can be used to calculate kinematic metrics (KMs) related to the patient's performance [19]–[21].

In terms of motion tracking for shoulder rehabilitation exercises, the data recorded are usually coming from shoulder, elbow, wrist, and hand movements [9], [13], [15], [18]. The KMs computed from those data are used to assess patient's movement accuracy, efficacy, efficiency, planning, smoothness, and speed [19]–[21]. Therefore, the accuracy and repeatability of the device used for motion tracking are crucial for determining expected measurement errors.

Recently, the HL2 has been used for tracking hand movements in medical applications, such as training for obstetric sonography [22], upper-limb rehabilitation [9], [23], and robotics rehabilitation [13]. Nam et al. [13] compared this MR technology to a robotic system using a rehabilitation-

*This work is supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada under grant RGPIN-2020-05648, by a Bell-Western 5G grant, by the Canadian Foundation for Innovation (CFI), by the Ontario Research Fund (ORF), by the Canada Research Chairs Program under File #950-233069, by scholarships awarded to S. Salinas from the Transdisciplinary Training Award from the Bone and Joint Institute at Western University, the Danny Ho Software Engineering Scholarship, and the Colombian Ministry of Science, Technology, and Innovation.

S. A. Salinas and K. Grolinger are with the Department of Electrical and Computer Engineering (ECE), Western University, London, ON, Canada, ssalina@uwo.ca, kgroling@uwo.ca,

M. E. LeBel is with the Department of Surgery, Western University, and the McFarlane Hand & Upper Limb Centre, St. Joseph's Health Care, London, ON, Canada, mlebel4@uwo.ca,

A. L. Trejos is with the ECE and the School of Biomedical Engineering, Western University, London, ON, Canada, (phone: +1-519-661-2111 ext. 89281), atrejos@uwo.ca.

based reaching task. They estimated three KMs using the index finger position. The tracking data of the index finger were also used by Soares et al. [24] to test the accuracy and repeatability of the HL2 in a user-computer interaction application. However, the HL2 tracking system requires further research to validate its accuracy and repeatability in tracking different parts of the hand. Specifically, the wrist and palm movements are variables of interest due to their importance for performing activities of daily living [25]. Moreover, an evaluation of the usability of the tracking data to determine KMs for upper limb rehabilitation assessment is necessary.

Therefore, the main objective of this research is to perform a preliminary evaluation of the accuracy and repeatability of an MR system, in tracking hand movements, including index fingertip, palm, and wrist, for its application in shoulder rehabilitation. The secondary goal is to investigate the usability of the tracking data to compute KMs for shoulder rehabilitation assessment.

II. Related Work

The Microsoft MR system, the HoloLens (HL), has found applications in diverse healthcare areas such as gait analysis, medical training, rehabilitation, and surgical navigation [26]. In the area of upper limb rehabilitation, Codino et al. [7] used the first version of the HL headset to develop and test the usability of an SG for shoulder rehabilitation exercises. Pillai et al. [9] made several SGs for shoulder, elbow, wrist, and finger therapy for the HL2. In addition to a usability test, they compared estimations of shoulder and elbow angles made with the HL2 hand tracking algorithm to the same angles obtained with a Kinect device, finding insignificant differences.

Franzo et al. [23] also compared an HL2 to a Kinect, but they tracked the three-dimensional (3D) movements of a participant's right hand while performing a reaching task exercise designed for ataxic patients. They found that the subjects achieved the goals in less time with the MR system than with the Kinect-based prototype, and the MR immersion made it easier to understand the depth and position of the targets. Furthermore, they noted that the HL2 has a higher sample rate than the Kinect, and concluded that the HL2 can follow unexpected patient movements better. A similar reaching exercise for upper limb rehabilitation, but in two dimensions, was considered by Nam et al. [13] to compare the performance of the HL2 to an end-effector-based rehabilitation arm. They tracked the hand movements and computed three KMs: average speed, total movement time, and curvilinearity ratio (a relation between the minimal distance to complete a movement and the actual distance performed). The differences found in the subjects' performance were produced by specific characteristics of the robotic system, such as arm support and haptic feedback. The authors concluded that a combination of MR and the robot could be suitable for upper-limb exercises for stroke survivors.

In general, MR systems have a promising future in advanced upper-limb therapy applications, alone or joined with a complementary device such as a Kinect or a robot [9], [13]. In particular, the capability of the HL2 to track the user's hand motion can be used to compute more KMs than those reported by Nam et al. [13]. Therefore, a comprehensive analysis of the accuracy and repeatability of the data that can be extracted from these MR devices is crucial to establish its usability in reporting these metrics. Soares et al. [24] tested accuracy and repeatability on the HL2 index fingertip tracking using several experiments and an OptiTrack system as ground truth. They registered an accuracy of around 20 mm and a repeatability between 5.8 mm and 11 mm. However, the tasks performed during the experiments were not based on therapy exercises; therefore, the effect on rehabilitation performance parameters was not measured. In terms of orientation accuracy and precision, Costa et al. [27] evaluated the ability of the HL2 to detect the orientation of a static QR code. They used a spherical structure with hexagonal holes as ground truth and reported HL2 average accuracy of $0.755°$ and average precision of $0.018°$.

As a contribution, this paper presents accuracy and repeatability results for 3D motion tracking obtained from the HL2, not only for the index fingertip but also for the hand palm and wrist. Moreover, the accuracy test is extended to analyze the results of six KMs used in upper limb rehabilitation assessment.

III. Materials and Methods

In order to evaluate the accuracy and repeatability of an HL2 for tracking the 3D positions of the index fingertip, palm, and wrist, two tasks were designed. Both tasks were adapted for an MR interaction. The first task was used for the accuracy and repeatability tests. The second task was used for accuracy tests, as well as for computing upper limb rehabilitation performance KMs. Two of the authors performed both tasks several times while the mentioned 3D positions were simultaneously tracked by the HL2 and an Aurora electromagnetic system, which was used as ground truth. The participants were right-handed, between 40 and 55 years of age, and of both genders, without any visible or diagnosed upper body impairment. Fig. 1 depicts the methodology followed in this study. An expanded explanation of the devices, tasks, materials, and methods to record and process data, and calculate accuracy, repeatability, and rehabilitation KMs, are presented in the following subsections.

A. Hardware

The Aurora system consisted of a Planar 20-20 electromagnetic field generator, a control unit, a sensor interface unit, and three 5-degree-of-freedom sensors to capture the 3D location of the index fingertip, palm, and wrist. This Aurora provided 3D location signals with a root-mean-square error (RMSE) of 0.7 mm at a sampling frequency of 40 Hz, in a cubic workspace with a side length of 500 mm [28]. The sensors were attached to the hand using 3M medical tape and positioned on the palm-side of the hand, so as not to interfere with the HL2 tracking algorithm.

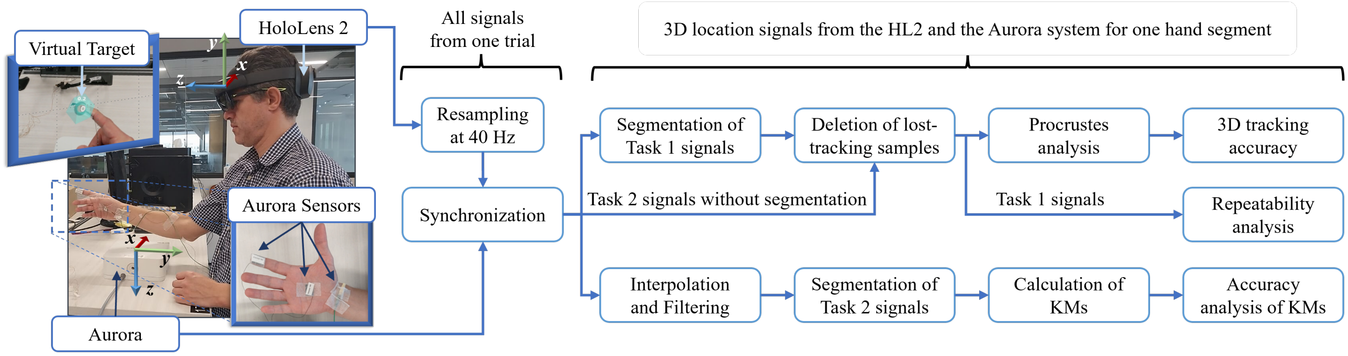

Figure 1. Methodology for processing hand motion data captured during a trial. Data from index fingertip, palm, and wrist 3D locations are captured with the Aurora system and the HoloLens 2 while performing two tasks based on an upper–limb rehabilitation exercise. The HoloLens and Aurora reference frames are represented using a red arrow for the x axis, a green arrow for the y axis, and a blue arrow for the z axis.

The HL2 is a wireless head-mounted-display device [29] that has see-through holographic lenses with a resolution of 2k and 3:2 ratio, visual projection with a focal length of 1.08 mm and a field of view of 96.1°, a camera of 1 Megapixel resolution for time-of-flight depth sensing, and a visual camera of 8 Megapixels. This device contains an algorithm for hand tracking based on a two-handed fully articulated model.

B. *Experimental Tasks*

The Aurora and the MR systems recorded 3D location data while the following experimental tasks were performed:

Task 1: This task was used for accuracy and repeatability tests of 3D position tracking using static hand positions. Ten virtual objects (named targets) were touched with the right-hand index fingertip for 3 seconds each. Each of the ten 3D positions was referred to as a static position. The targets had a 3D icosahedron shape delimited by a sphere with a radius of 35 mm. They were shown sequentially on the vertexes and center of a rectangular cuboid. The size of this cuboid was selected as 350 mm in height and length and 250 mm in width, so that the targets and the hand were inside the workspace of the Aurora system.

Task 2: This task was used to calculate KMs and evaluate the HL2 accuracy with respect to the Aurora, during dynamic hand movements. It was based on a therapeutic exercise for dynamic shoulder rehabilitation, which requires the patient to touch, as quickly as possible, physical objects randomly appearing in a reachable space in front of them with a finger or hand. The exercise includes shoulder horizontal abduction or adduction and flexion or extension movements. An MR-based SG was developed for this experimental task. The SG had a duration of 60 seconds and showed targets randomly inside the same cuboid used for Task 1. In contrast to Task 1, the interaction between the target and the user's index fingertip required 0.2 seconds for the target to change its location. Each hand movement from one target to the next one was considered as a basic repetition. Task 2 was performed under four conditions: 1) high speed, by moving from one to the next target as quickly as possible using a natural arm speed, 2) medium and 3) low speeds, with guidance to reduce the arm motion speed, and 4) a jerky motion, by vibrating the arm.

Each participant performed each task eight times. The execution of one Task 1 followed by one Task 2 was considered as a trial. Task 1 was the same during all trials, while Task 2 was sequentially modified by the mentioned conditions in each trial. For example, Trial 1 included one Task 1 and one Task 2 at high speed; Trial 2 consisted of one Task 1 and one Task 2 at medium speed, and so on. After going through the four conditions of Task 2, the sequence was repeated. The SG, running in the HL2, and the data recording from the Aurora system were restarted between trials.

C. *Data Recording*

The game engine Unity v2020.3.42f1, the Mixed Reality Toolkit v2.8.3, and the MR OpenXR plugin v1.8.0 were used to create the MR environment, define user–MR interactions, and record hand motion data. Using the HL2 hand-tracking capabilities, 3D locations of the right-hand index fingertip, palm, and wrist were recorded in addition to a timestamp, 3D locations of the virtual targets, and the number of targets touched. The sampling frequency of the HL2 was around 55 Hz. The measurements coming from the Aurora system were recorded using proprietary software from Northern Digital Inc. The number of frames, the status of the sensors, and 3D locations were recorded. A representation of the data recording and processing, as well as, the used devices are shown in Fig. 1.

D. *Synchronization of Signals*

Since the HL2 and the Aurora system recorded the tracking data in different devices and at different starting times and sample frequencies, the following procedure for synchronizing signals from each trial was applied: 1) invert the y and z axes from the HL2 signals to have the same orientation as the z and y axes from the Aurora signals, respectively, according to the orientation of the HL2 and Aurora reference frames during the experiments (the frames are shown in Fig. 1), 2) resample HL2 signals at 40 Hz using linear interpolation, 3) calculate the lag between the Aurora signals and the HL2 signals according to their maximum cross-correlation, 4) apply the lag to the Aurora signals, and 5) split the signals into Task 1 and Task 2 using the HL2 timestamp.

E. Accuracy of 3D Location Data

For the accuracy analysis of the MR system, data related to the 3D location of the hand from both tasks were considered. The index fingertip, palm, and wrist data were processed independently using the same procedure. In order to obtain data related to only static positions, without the transition movements between targets, the intervals of time for which the finger was in contact with the targets were used for segmenting the signals from Task 1 (See Fig 1). In contrast, signals from Task 2 were not segmented to maintain dynamic hand movements and compare the HL2 accuracy between static and dynamic hand motions.

After the segmentation, the signals were cleaned from lost–tracking samples. The Aurora system lost tracking when its sensors were outside of its workspace, while the HL2 lost hand tracking when the hand was outside of the camera field of view. Since the comparison of the signals would be made point to point for each time instant, the lost tracking data needed to be removed from the data set. Then, the HL2 3D location points P_{HL} were transformed into the Aurora reference frame P_{HL}^{Au} by following the Procrustes analysis, since the HL2 and the Aurora tracking data were measured with respect to different world frames. The Procrustes analysis searches for the optimal transformation matrix between two reference frames by minimizing the Euclidean distance between shared data points [30]. MATLAB v2022b was used to perform the analysis and obtain the rotation matrix R and the translation vector c between the HL2 and Aurora reference frames. Although it is possible to modify the scale of the HL2 data when transforming them to the Aurora frame, this modification was not made in order to preserve the HL2 data as measurements of physical distances. To complete the transformation, the Equation 1 was applied to the HL2 signals.

$$P_{HL}^{Au} = P_{HL} * R + c \, [mm] \tag{1}$$

The accuracy of the HL2 was computed by taking the 3D location data from the Aurora system P_{Au} as the ground truth. The accuracy of each trial and task was estimated using RMSE, calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left|P_{iAu} - P_{iHL}^{Au}\right|^2} \, [mm]. \tag{2}$$

In this equation, n is the number of 3D location samples from a specific trial and task, and $|.|$ represents the norm of a vector.

F. Repeatability of 3D Location Data

A repeatability test to evaluate HL2 ability to estimate the 3D position of the hand in every measurement repetition was also performed. A comparison between the dispersion of the 3D position measurements obtained during Task 1 from the HL2 and the Aurora was done. Usually, a repeatability test is only made over the evaluated system [24], [31], [32], the HL2 in this case. However, the dispersion of the Aurora measurements was included in this study because the natural hand motion generates changes in the reference 3D position over time. The test was applied to both systems following the ISO 9283 [33] standard, which has been used to test the repeatability of other tracking systems [31], [32] as well as the HL2 [24]. The standard deviation of the measurements was computed from each system (HL2 or Aurora), static 3D position, and trial, using the following equation:

$$\sigma_{3D} = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m} (l_i - \bar{l})^2} \, [mm], \tag{3}$$

where m is the number of 3D location samples from a specific 3D static position, $\bar{.}$ represents the mean over the m samples, and l_i is defined by (4), where x_i, y_i, and z_i, are the components of the i^{th} 3D location sample.

$$l_i = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2} \, [mm] \tag{4}$$

The repeatability, Rep_s, evaluated for each system and each static position was then computed using the following equation:

$$Rep_s = \bar{l} + 3\sigma_{3D} \, [mm]. \tag{5}$$

G. Kinematic Metrics

In addition to the accuracy and repeatability tests, an analysis of HL2 capability to produce suitable information for rehabilitation performance assessment was performed. KMs obtained from the HL2 data to the same KMs computed from the Aurora data collected during Task 2 were compared. This task was selected because the movements were more similar to an actual rehabilitation exercise than those performed during Task 1.

The preprocessing of the HL and Aurora signals is shown in Fig. 1. After the synchronization, position signals were filtered using a fourth-order low-pass Butterworth filter with a 6 Hz cutoff frequency without taking out the lost tracking samples but replacing them through linear interpolations. This filtering approach was applied in order to maintain the dynamic of the movements over time and obtain dynamic KM values. Later, a segmentation of the signals into basic repetitions of hand movements between one and the next target was made. Then, the following KMs were computed for each basic repetition: path length (PL), mean speed (MS), normalized mean speed (NMS), number of speed peaks (NSP), log dimensionless jerk (LDJ), time to speed peak (tSp), and movement accuracy (MAcc). Table I shows their definitions and equations. The selection of the KMs was made considering their evidence of quality and reliability for shoulder kinematic assessment [19]–[21]. For example, PL is used to assess the efficiency of a patient's movement during therapeutic exercises; MS, NMS, NSP, and LDJ measure movement smoothness; movement planning is evaluated with tSp; and MAcc assesses the patient's ability to reach a 3D location with their arm.

TABLE I

KINEMATIC METRICS FOR SHOULDER ASSESSMENT, THEIR DEFINITIONS AND EQUATIONS

KM: Definition	Equation
Path Length (PL): Length of travelled trajectory	$PL = \sum_{i=1}^{k-1} \lvert P_{i+1} - P_i \rvert \; [mm]$ (6)
Mean Speed (MS) The arithmetic mean of speed samples	$MS = \frac{1}{k-1} \sum_{i=1}^{k-1} \left\lvert \frac{dP}{dt} \right\rvert_i \; [mm/s]$ (7)
Normalized Mean Speed (NMS) MS divided by the maximum speed (S_{max})	$NMS = \frac{MS}{S_{max}}$ (8)
Number of Speed Peaks (NSP) The count of speed peaks greater than 60% of S_{max}	$NSP = \sum_{i=1}^{p-1} \left\lvert \frac{dP}{dt} \right\rvert_i > 0.6 S_{max}$ (9)
Log Dimensionless Jerk (LDJ) Measure of movement smoothness	$LDJ = -ln \left(\frac{(t_2 - t_1)^3}{S_{max}^2} \int_{t_1}^{t_2} \left\lvert \frac{d^3 P}{dt^3} \right\rvert^2 dt \right)$ (10)
Time to S_{peak} (tSp) The time between the movement starts and when S_{peak} happens	$tSp = t \left(\left\lvert \frac{dP}{dt} \right\rvert == S_{peak} \right) \; [s]$ (11)
Movement Accuracy ($MAcc$) Distance between the fingertip (P_f) and the center of a target (P_t)	$MAcc = \lvert P_f - P_t \rvert \; [mm]$ (12)

P: 3D–location vector, k: number of elements of the P vector, p: number of peaks, t_1: time when the finger touches a target, and t_2: time when the finger touches the next target.

Since the KMs have different units and ranges of values, the normalized metric presented in Equation 13 was applied to estimate the accuracy of the HL2 (KM_{HL}), using the Aurora KMs as ground truth (KM_{Au}). Similar to the normalized mean absolute error and the normalized root mean square error, applied in other studies [34]–[36], this equation normalizes the error with the range of the ground truth measurements. This approach facilitates the comparison across different variables that can include small or zero values for the ground truth KMs.

$$Acc = 100 \left(1 - \left\lvert \frac{KM_{Au} - KM_{HL}}{max(KM_{Au}) - min(KM_{Au})} \right\rvert \right) [\%] \quad (13)$$

In order to identify differences in accuracies, repeatability, and KMs, statistical tests were performed (a p <0.05 was considered significantly different). Normality and non-parametric tests were made using MATLAB v2022b. The Wilcoxon Rank Sum test was employed to compare two variables and the Kruskal-Wallis test for more than two variables with a Bonferroni correction to adjust the significance level.

Type text here

IV. RESULTS

No data were deleted from the HL2 measurements, in the time intervals of interest. However, the Aurora lost track of the wrist sensor during one trial, affecting the data recorded. Therefore, those particular data were deleted from the corresponding HL2 data stream, affecting 2.1% of the dataset.

A. Accuracy of 3D Location Data

Fig. 2 shows a box plot of the RMSE obtained from Task 1 (static positions) and Task 2 (dynamic motion). The median RMSE was below 10.2 mm for all of the hand segments and the two tasks, which is lower than the mean error of around 20 mm reported by Soares et al. [24]. This discrepancy lies in the differences in the experiments, ground-truth equipment, and signal processing. Moreover, the irregular sampling frequency and delays of the HL2 data are other critical factors. The methodology shown in Fig. 1 is proposed as a solution to reduce the effects of these factors on the HL2 tracking accuracy.

The minimum median error was obtained by tracking the palm, with 7.06 mm for static positions and 6.71 mm for dynamic motion. There were no significant differences in accuracies between tasks (p=0.64). However, a significant difference was found between the palm accuracy compared to the index and wrist accuracies (p <0.01 and p <0.001, respectively) during Task 1. The outperformance of the palm, compared to the wrist and the index fingertip, can be caused by the tracking algorithm included in the HL2 to capture the hand position. This algorithm continuously adjusts a hand model to the actual user's hand. Therefore, the adjustment produces greater variance in the fingertip and wrist measurements compared to the palm. These results suggest that tracking palm motion with the HL2 should be considered as an alternative to the motion tracking of index fingertip typically made [13], [23], [24].

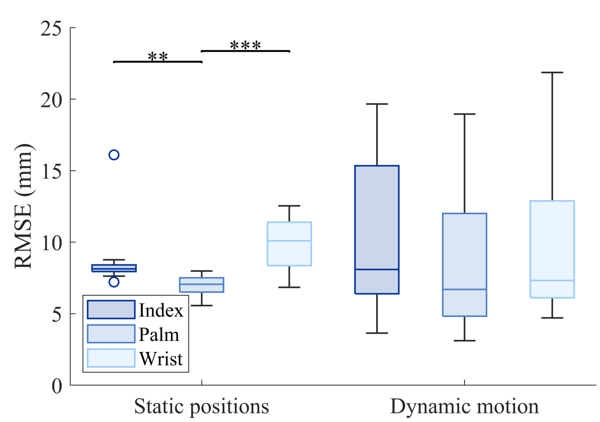

Figure 2. RMSE obtained from comparing the HoloLens 2 tracking data to the Aurora system during a static positions task and a dynamic motion task, related to the 3D location of the index fingertip, palm, and wrist. Here, ** indicates p <0.01 and *** p <0.001.

B. Repeatability of 3D Location Data

In terms of repeatability, Fig. 3 shows that the dispersion was similar for both systems, ranging from 5.74 mm to 15.08 mm for the HL2 and from 5.85 mm and 13.56 mm for the Aurora. The minimum medians were obtained from the palm measurements for both systems, 8.44 mm for the HL2 and 7.87 mm for the Aurora. Measurements from the three hand segments between systems showed insignificant differences (p=0.16 for the index, p=0.17 palm, and p=0.17 wrist). Furthermore, there were insignificant differences between hand segments in the HL2 (p=0.57 index vs. palm, p=0.3 index vs. wrist, and p=1 palm vs. wrist). These results imply that the tested MR system has a similar repeatability to the Aurora system, confirming, with the accuracy results, the HL2 reliability in tracking hand positions.

C. Kinematic Metrics

For the KM computation, 812 basic repetitions from Task 2 were considered. The accuracies obtained are shown in Fig. 4 and Fig. 5. MAcc has accuracy only from the finger data because the palm and wrist did not have contact with the target during the experiments. For easy visualization the outliers have been removed. The percentages of outliers for each KM were 10%, 13.7%, 5.2%, 29.3%, 16.1%, 8.1%, and 5.3%, for PL, MS, NMS, NSP, tSp, LDJ, and MAcc, respectively.

To investigate the accuracy of the majority of KM estimations, the 10^{th} percentile was calculated for each KM, obtaining 93.1%, 90%, 86.9%, 90.9%, 95%, 89.6%, and 88.4%, for PL, MS, NMS, NSP, tSp, LDJ, and MAcc, respectively. In general, there were insignificant differences between the KMs accuracies obtained from the analyzed hand segments (See Fig. 4). However, differences in accuracy were found depending on the Task 2 experimental condition (See Fig. 5). The PL and MS obtained better estimations during low-speed conditions (medians of 99.6% and 99.8%, respectively) and worse estimations for high-speed conditions (medians of 95.4% and 93.3%). The NMS and LDJ were better estimated for medium and low-speed conditions (medians above 96.5%). The accuracy of NSP estimations presented less variability for medium-speed conditions and

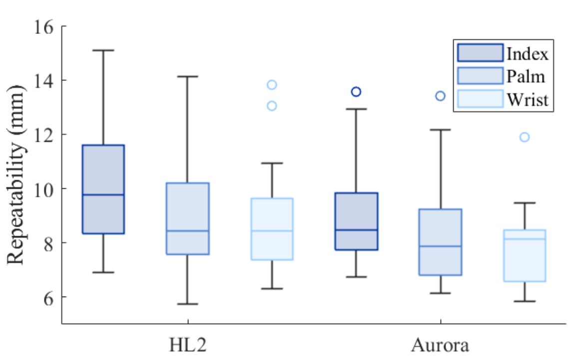

Figure 3. Repeatability obtained from the HoloLens 2 and the Aurora measurements while tracking 3D locations of the index fingertip, palm, and wrist. Note that the y axis does not start at zero.

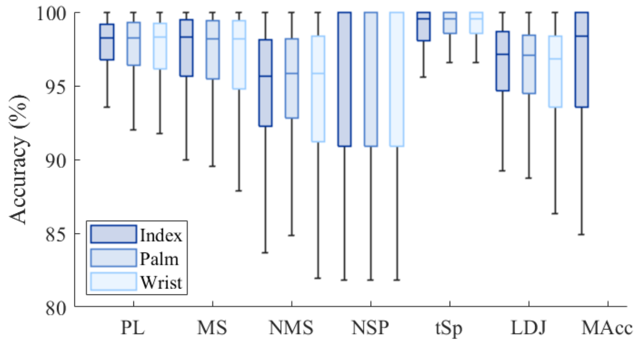

Figure 4. Accuracy obtained by comparing KMs calculated with HoloLens 2 tracking data to the Aurora data, related to the index fingertip, palm, and wrist. See Table I for KM definitions. Note that the y axis starts at a number different than zero.

significant differences with low-speed and jerky movements, but an insignificant difference with high-speed conditions (p=0.32). Estimations of tSp had lower accuracy during low-speed movements. For MAcc estimations, no experimental condition produced a significantly higher accuracy.

These accuracy results demonstrate the usability of an MR device, such as the HL2, in upper–limb rehabilitation assessment, without a preference for any segment of the hand studied. Moreover, the high accuracy in detecting jerky movements makes this MR system suitable for assessing stroke [15], [19]–[21], [37], ataxic [23], [38], and injured patients [9], for example. In contrast, the reduced accuracy during high-speed movements of some metrics such as PL, MS, and NMS can be generated by its dependency on the computation of movement speed and the speed peak that each system can measure (HL2 and Aurora). The sampling frequency of the HL2 that makes it able to follow unexpected movements [23] can also produce this reduction in accuracy.

An analysis of each KM was also performed to investigate the usability of the HL2 data to distinguish between low,

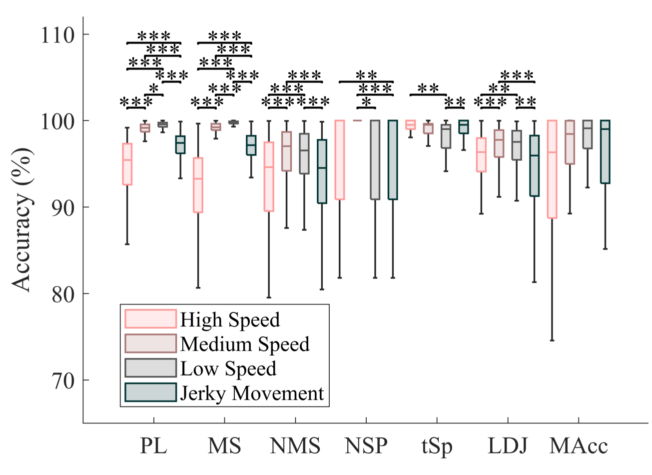

Figure 5. Accuracy obtained by comparing KMs calculated with HoloLens 2 tracking data to the Aurora data, related to four different experimental conditions during a shoulder rehabilitation exercise. See Table I for KM definitions. Here, * indicates p <0.05, ** p <0.01, and *** p <0.001. Note that the y axis starts at a number different than zero.

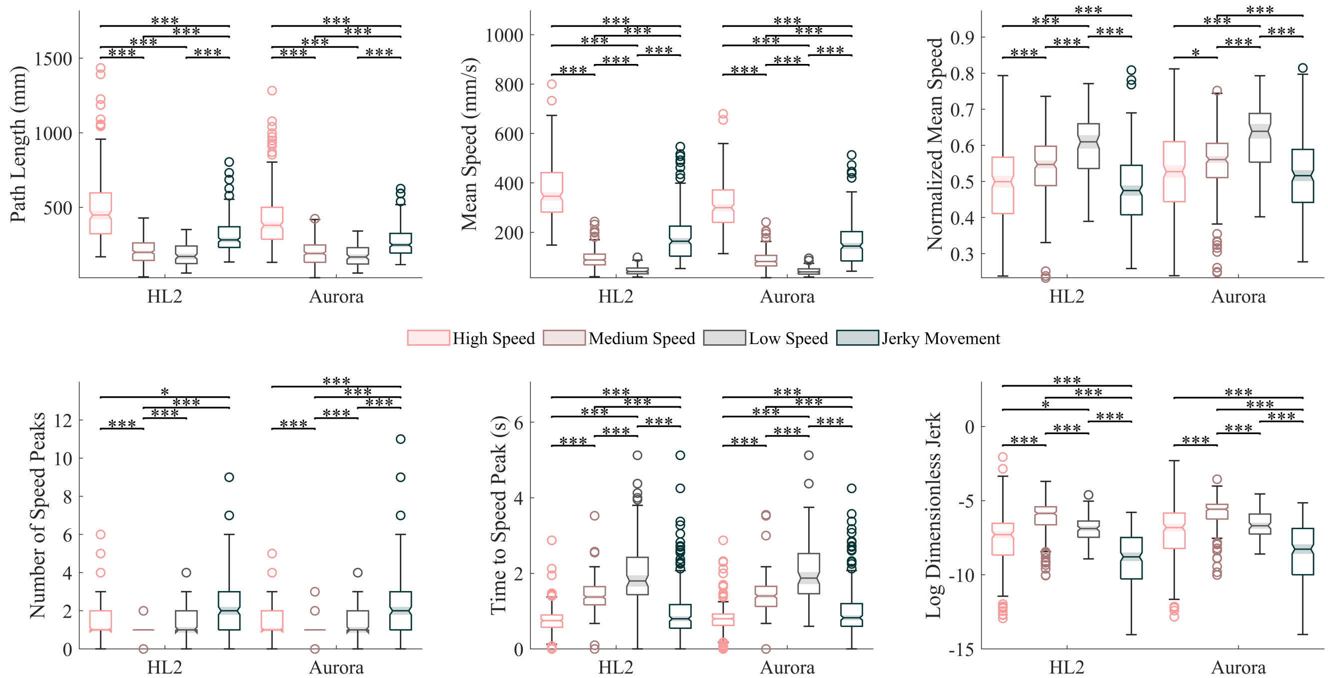

Figure 6. Comparison of the kinematic metrics obtained with the HoloLens 2 and the Aurora data from four different experimental conditions during a shoulder rehabilitation exercise. See Table I for KM definitions. Here, * indicates p <0.05, ** p <0.01, and *** p <0.001.

medium, and high speeds and jerky movements. Fig. 6 shows box plots for each KM and system under the four different conditions. Since MAcc presented insignificant differences, it was not included. The notch of each box has been highlighted in order to represent the region of possible statistical significance. The results of MS confirmed that the experiments were performed at different speeds (p <0.001), the obtained interquartile ranges were: 31.8 mm/s to 55.4 mm/s for low speed, 68.3 mm/s to 111.6 mm/s for medium speed, 281.8 mm/s to 442 mm/s for high speed, and 103.4 mm/s to 224.9 mm/s for jerky movements. PL was significantly larger for the high-speed condition (p <0.001) and distinguished jerky movement (p <0.001) from the others, whereas PL showed insignificant differences between medium and low-speed conditions (p=0.38).

The results of NMS and tSp showed higher values for low-speed followed by medium-speed conditions (p <0.001). The tSp metric was also significantly different for high and jerky movements (p <0.001), whereas NMS was not different (p=0.13). Similar to tSp, LDJ was different for each of the conditions (p <0.05) when calculated with the HL2 data. LDJ was higher for medium-speed conditions, followed by low-speed, high-speed, and jerky movements. From NSP, the low-speed condition showed lower results (p <0.001) in both systems.

These analyses reveal that the MS, LDJ, and tSp calculated with the HL2 data can be used to estimate and differentiate under the four studied conditions: low, medium, high and jerky movements. Additionally, with PL is possible to distinguish between an exercise performed by a healthy subject (high-speed condition) and a patient with jerky movements, while a patient with low-speed movements can be assessed using NMS and NSP. Results of MS, NMS, NSP, and LDJ are comparable with those reported for stroke patients [15], [21], [37]. However, they are not the same because of the difference in the rehabilitation exercise performed.

V. CONCLUSION

The HL2 has a hand-tracking system accurate and precise for applications that require static and dynamic movements. This MR system exhibits a median RMSE below 10.2 mm in tracking 3D locations of the index fingertip, palm, and wrist. Moreover, the repeatability of the HL2 measurements is comparable with the Aurora system repeatability. From the studied hand segments, the palm showed the best accuracy and repeatability results.

Data coming from an MR system, such as the HL2, are suitable for computing KMs for upper–limb rehabilitation assessment. Specifically, we obtained accuracies above 90% for PL, MS, NSP, and tSp, and above 86.9% for NMS, LDJ, and MAcc, for the HL2 when compared to the ground-truth system. Higher accuracies were reached for movement speeds below 111.6 mm/s. The KMs were computed with the index fingertip, palm, or wrist tracking data without a significant difference. These KMs can be used to detect low, medium, and high-speed motion as well as jerky upper-limb movements.

The results of this study open the possibility that serious gaming in combination with MR technology can be used appropriately and accurately to assist and support the rehabilitation of patients with a range of shoulder conditions. This research is part of a series of projects on serious gaming for shoulder rehabilitation. Other MR environments will be developed to include additional physiotherapy exercises and kinematic metrics.

ACKNOWLEDGMENT

The authors thank Michael Taran, Marcus Johnson, Megan Ginham, Elliot Lam, and Alexander Nelson for their contribution to the serious game development, and Harshu Kannappan for assisting during experimentation.

REFERENCES

[1] C. D. Djade, T. V. Porgo, *et al.*, "Incidence of shoulder pain in 40 years old and over and associated factors: A systematic review," *European Journal of Pain*, vol. 24, no. 1, pp. 39–50, 2020.

[2] R. Teasell, N. Hussein, *et al.*, *Stroke Rehabilitation Clinician Handbook*. Canadian Partnership for Stroke Recovery, 2020. [Online]. Available: www.ebrsr.com

[3] C. J. Hodgetts, C. Leboeuf-Yde, *et al.*, "Shoulder pain prevalence by age and within occupational groups: A systematic review," *Archives of Physiotherapy*, vol. 11, pp. 1–12, 2021.

[4] C. Payne, A. Jaggi, *et al.*, "Rehabilitation for shoulder arthroplasty," *Orthopaedics and Trauma*, vol. 29, no. 5, pp. 313–323, 2015.

[5] C. Leftley, J. Walton, *et al.*, *Shoulder Rehabilitation: A Comprehensive Guide to Shoulder Exercise Theraphy*, 2nd ed. Manchester: Shoulderdoc.co.uk, 2018.

[6] U. G. Longo, L. Risi Ambrogioni, *et al.*, "Physical therapy and precision rehabilitation in shoulder rotator cuff disease," *International Orthopaedics*, vol. 44, no. 5, pp. 893–903, 2020.

[7] S. Condino, G. Turini, *et al.*, "Wearable augmented reality application for shoulder rehabilitation," *Electronics*, vol. 8, no. 10, pp. 1178-1–16, 2019.

[8] H. L. Phan, T. H. Le, *et al.*, "Effectiveness of augmented reality in stroke rehabilitation: A meta-analysis," *MDPI Applied Sciences*, vol. 12, no. 4, pp. 1848-1–17, 2022.

[9] A. Pillai, M. S. H. Sunny, *et al.*, "Gamification of upper limb rehabilitation in mixed-reality environment," *MDPI Applied Sciences*, vol. 12, no. 23, pp. 12 260-1–18, 2022.

[10] M. C. Howard and M. M. Davis, "A meta-analysis and systematic literature review of mixed reality rehabilitation programs: Investigating design characteristics of augmented reality and augmented virtuality," *Computers in Human Behavior*, vol. 130, pp. 107 197-1–17, 2022.

[11] R. M. Viglialoro, S. Condino, *et al.*, "Review of the augmented reality systems for shoulder rehabilitation," *MDPI Information*, vol. 10, no. 5, pp. 154-1–14, 2019.

[12] R. Karamians, R. Proffitt, *et al.*, "Effectiveness of virtual reality- and gaming-based interventions for upper extremity rehabilitation poststroke: A meta-analysis," *Archives of Physical Medicine and Rehabilitation*, vol. 101, no. 5, pp. 885–896, 2020.

[13] J.-Y. Nam, S.-H. Lee, and W.-K. Song, "Comparison of reaching motion in mixed reality headset and end-effector-based robotic arm with flat panel display," in *2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. Seoul, Republic of Korea: IEEE, Aug. 21–24, 2022, pp. 1–6.

[14] A. Vakanski, J. M. Ferguson, and S. Lee, "Metrics for performance evaluation of patient exercises during physical therapy," *International Journal of Physical Medicine & Rehabilitation*, vol. 5, no. 3, pp. 403-1–6, 2017.

[15] A. Mobini, S. Behzadipour, and M. Saadat, "Test–retest reliability of Kinect's measurements for the evaluation of upper body recovery of stroke patients," *Biomedical Engineering Online*, vol. 14, pp. 75-1–13, 2015.

[16] Y. Liao, A. Vakanski, *et al.*, "A review of computational approaches for evaluation of rehabilitation exercises," *Computers in Biology and Medicine*, vol. 119, pp. 103 687-1–15, 2020.

[17] J. P. Proença, C. Quaresma, and P. Vieira, "Serious games for upper limb rehabilitation: A systematic review," *Disability and Rehabilitation: Assistive Technology*, vol. 13, no. 1, pp. 95–100, 2018.

[18] M. O. Powell, A. Elor, *et al.*, "Predictive shoulder kinematics of rehabilitation exercises through immersive virtual reality," *IEEE Access*, vol. 10, pp. 25 621–25 632, 2022.

[19] A. Schwarz, C. M. Kanzler, *et al.*, "Systematic review on kinematic assessments of upper limb movements after stroke," *Stroke*, vol. 50, no. 3, pp. 718–727, 2019.

[20] R. M. Maura, S. Rueda Parra, *et al.*, "Literature review of stroke assessment for upper-extremity physical function via EEG, EMG, kinematic, and kinetic measurements and their reliability," *Journal of NeuroEngineering and Rehabilitation*, vol. 20, no. 1, pp. 21-1–32, 2023.

[21] M. Hajihosseinali, S. Behzadipour, *et al.*, "Direction-dependency of the kinematic indices in upper extremities motor assessment of stroke patients," *Medical Engineering & Physics*, vol. 108, pp. 103 880-1–9, 2022.

[22] M. Birlo, P. J. E. Edwards, *et al.*, "CAL-Tutor: A HoloLens 2 application for training in obstetric sonography and user motion data recording," *Journal of Imaging*, vol. 9, no. 1, pp. 6-1–13, 2022.

[23] M. Franzo, S. Pascucci, *et al.*, "Exergaming in mixed reality for the rehabilitation of ataxic patients," in *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. Messina, Italy: IEEE, Jun. 22–24, 2022, pp. 1–5.

[24] I. Soares, R. B. Sousa, *et al.*, "Accuracy and repeatability tests on HoloLens 2 and HTC Vive," *Multimodal Technologies and Interaction*, vol. 5, no. 8, pp. 47-1–14, 2021.

[25] L. Reissner, G. Fischer, *et al.*, "Assessment of hand function during activities of daily living using motion tracking cameras: A systematic review," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 233, no. 8, pp. 764–783, 2019.

[26] A. Palumbo, "Microsoft HoloLens 2 in medical and healthcare context: State of the art and future prospects," *Sensors*, vol. 22, no. 20, pp. 7709-1–30, 2022.

[27] J. N. Costa, J. Gomes-Fonseca, *et al.*, "Ultrasound training simulator using augmented reality glasses: An accuracy and precision assessment study," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Glasgow, UK: IEEE, Jul. 11–15, 2022, pp. 4461–4464.

[28] N. D. Inc. (2023) Minimally invasive procedures with electromagnetic tracking technology. [Online]. Available: https://www.ndigital.com/electromagnetic-tracking-technology/aurora/

[29] M. Inc. (2023) About HoloLens 2. [Online]. Available: https://learn.microsoft.com/en-us/hololens/hololens2-hardware

[30] E. C. Chen, D. Allen, *et al.*, "Unified calibration technique for augmented-reality ultrasound-guided interventions," in *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*. Rome, Italy: IEEE, Oct. 26–28, 2022, pp. 495–500.

[31] F. Weichert, D. Bachmann, *et al.*, "Analysis of the accuracy and robustness of the leap motion controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, 2013.

[32] S. Blankemeyer, R. Wiemann, *et al.*, "Intuitive robot programming using augmented reality," *Procedia CIRP*, vol. 76, pp. 155–160, 2018.

[33] I. O. for Standardization, *ISO 9283:1998: Manipulating Industrial Robots: Performance Criteria and Related Test Methods*, 2nd ed. International Organization for Standardization, 1998.

[34] I. Ogashawara, M. P. Curtarelli, *et al.*, "Interactive correlation environment (ICE): A statistical web tool for data collinearity analysis," *Remote Sensing*, vol. 6, no. 4, pp. 3059–3074, 2014.

[35] K. Goldberg, T. Roeder, *et al.*, "Eigentaste: A constant time collaborative filtering algorithm," *Information Retrieval*, vol. 4, pp. 133–151, 2001.

[36] N. Grimmelsmann, M. Mechtenberg, *et al.*, "sEMG-based prediction of human forearm movements utilizing a biomechanical model based on individual anatomical/physiological measures and a reduced set of optimization parameters," *PLoS ONE*, vol. 18, no. 8, pp. e0 289 549-1–28, 2023.

[37] R. Colombo, I. Cusmano, *et al.*, "Test–retest reliability of robotic assessment measures for the evaluation of upper limb recovery," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 5, pp. 1020–1029, 2014.

[38] M. Franzo, S. Pascucci, *et al.*, "Breakthrough in occupational therapy with mixed-reality exergaming for cerebellar ataxia patients," in *2023 IEEE International Symposium on Medical Measurements and Applications, MeMeA 2023-Conference Proceedings*. Jeju, Republic of Korea: IEEE, Jun. 14–16, 2023, pp. 1–4.