Remote Extended Reality with Markerless Motion Tracking for Sitting Posture Training

Xupeng Ai\textsuperscript{1} and Sunil K Agrawal\textsuperscript{1}

\textsuperscript{1}Affiliation not available

May 17, 2024
Remote Extended Reality with Markerless Motion Tracking for Sitting Posture Training

Xupeng Ai, Graduate Member, IEEE, Sunil K. Agrawal*, Member, IEEE

Abstract—Dynamic postural control during sitting is essential for functional mobility and daily activities. Extended reality (XR) presents a promising solution for posture training in addressing conventional training limitations related to patient accessibility and ecological validity. We developed a remote XR rehabilitation system with markerless motion tracking for sitting posture training. Forty-two healthy subjects participated in this proof-of-concept pilot study. Each subject completed 24 rounds of multi-directional reach tasks using the system and 24 rounds without it. Motion data were collected via Zoom meetings using built-in cameras in laptops. Functional reach test scores were analyzed to assess the impact of the system on motor performance. Four standard questionnaires were used to assess the effects of this system on presence, simulator sickness, engagement, and enjoyment. Our results indicate that the remote XR training system significantly improved functional reach performance and proved highly effective for telerehabilitation. XR interaction also enhanced training engagement and enjoyment. By bridging the spatial gap between patients and therapists, this system enables personalized and engaging home-based intervention. Additionally, it facilitates more natural movements by eliminating body marker constraints and laboratory limitations. This study should serve as a stepping stone to advancing novel remote XR rehabilitation systems.

Index Terms—Remote Rehabilitation System, Virtual Reality and Interfaces, Markerless Motion Tracking, Postural Control

I. INTRODUCTION

Dynamic postural control during sitting requires maintaining postural stability and orientation during voluntary movements [1]. It is the cornerstone of functional mobility and is essential for performing daily activities [2]. However, this ability may be impaired in people with neuromotor disorders such as in spinal cord injury and cerebral palsy, significantly affecting their life quality [3].

Conventional training methods, including core muscle strengthening, stability ball training, and proprioceptive exercises, have been proven effective in postural control recovery [4]. However, providing sufficient intervention required for recovery in conventional training can be costly and labor-intensive [5], [6]. In addition, substantial repetitions of these exercises may demotivate patients and hinder their adherence to the training regimen [7], [8]. Therefore, researchers are exploring novel technologies to overcome these challenges.

Extended reality (XR), including virtual reality and augmented reality, is gaining popularity in postural training [9], [10]. By merging physical and virtual reals, XR could provide engaging interactive experiences to subjects, allowing them to interact with 3D-rendered virtual items in real time [11]. The gamification property inherent in XR can enhance enjoyment and encourage voluntary participation, leading to greater training intensity and improved rehabilitation outcomes [12], [13]. Further studies indicate that XR interaction could expedite motor learning by inducing neurophysiological adaptations and stimulating neural plasticity [14]–[16].

In our prior research, we have developed an XR-enhanced rehabilitation system where subjects practice multi-directional reach tasks within a virtual environment [17]. Our findings indicate that XR interaction significantly improves motor performance and balance. Despite promising results, our system relies on three flagship XR headsets to provide interaction experiences and employs nineteen infrared cameras and twenty-nine reflective markers for motion tracking. Substantial costs and limited portability restrict the system to specialized laboratory settings, hindering patient accessibility and adherence to treatment plans [18]. Additionally, marker placement before training could be time-consuming, and marker mislabeling may introduce noise that necessitates manual correction in post-processing [19], [20]. Furthermore, conducting functional movement tests and posture training solely in the laboratory restricts the ecological validity and generalizability of our findings to real-world scenarios [21].

Recent research underscores the potential of integrating XR into telerehabilitation to expand potential clinical applications and impact on patients [22]–[24]. In this study, we developed a remote XR sitting posture training system. We leveraged an open-source deep learning model to realize real-time markerless motion tracking. A custom game application was developed to deliver the XR intervention experience. We designed a training paradigm where functional movement ability testing and sitting posture training occurred during Zoom meetings through low-cost built-in cameras in mobile devices. Forty-two healthy subjects participated in usability testing at home. Each subject completed 24 rounds of multi-directional reach tasks using the remote XR training system and another 24 rounds without it. Our findings indicate that the system could significantly improve functional mobility and motor performance. In addition, remote XR interaction increased training intensity, enjoyment, and engagement. By eliminating the need for body markers and lab restrictions, our system offers high ecological validity and generalization potential. Further extension of our system to other posture training protocols may promote the development of innovative remote XR rehabilitation technologies.
II. METHODOLOGY

A. Subjects

Forty-two healthy subjects participated in this study (age = 23 ~ 47; females = 17; left-handed = 6; mean height = 173.1 ± SE = 4.7 cm and mean weight = 65.5 ± SE = 8.7 kg). Approval for all ethical and experimental procedures in this paper was sought and granted by the Institutional Review Board (IRB) of Columbia University under Protocol No. AAAR7781. Informed consent was received from all human subjects. The IRB approval date was 10/11/2023.

B. Markerless Motion Tracking

MediaPipe is an open-source deep learning framework developed by Google for human pose estimation [25], [26]. By feeding webcam frames to the BlazePose module, MediaPipe accurately predicts 33 body landmark positions in real time, without relying on physical markers or specialized lab equipment [27]. Previous studies have demonstrated the reliability and validity of the MediaPipe for studying movements in telerehabilitation [27], [28]. For this study, which centers around seated movement analysis, we deliberately selected the fine-tuned MediaPipe version (0.7.9) that provides more precise predictions for upper body landmarks.

C. Remote XR System Setup

Schematic of the remote XR rehabilitation system setup is shown in Fig. 1. A researcher sat in front of a laptop (Dell XPS 9730, Intel Core i7-13700H CPU, 16 GB RAM, NVIDIA RTX 4060 GPU) with two extended monitors (InnoView PM808) in the lab (Fig. 1a). The researcher invites a subject to join a remote Zoom meeting at home through a device with a webcam (e.g., desktop, laptop, tablet, smartphone with an extended monitor). The subject sat on a bench without foot support, with the entire body visible to the webcam (Fig. 1b). Customized Python scripts built upon the MediaPipe framework ran on the researcher’s laptop. These scripts captured frames from the subject Zoom view, extracted body landmark positions and executed XR interaction programs. The program execution results were rendered on the frames and then displayed on another screen, shared with the subject in real time (see Fig. 1a schematic diagram). Movement data were stored locally. The subject sat upright, extended the arm to ninety degrees of shoulder flexion. Two virtual horizontal bars were rendered at the shoulder-level height, with the edges aligned to the wrist point (Fig. 1c).

D. Experiment Protocol

The schematic diagram of the experiment design is shown in Fig. 2. A postural star sitting test (PSST) [29] was performed before training (Fig. 2a). Subjects were instructed to reach in four primary directions: dominant (D), nondominant (ND), front (F), and back (B); using their dominant arm for D, F, and B directions, and the nondominant arm for ND. Subjects sat upright and fully extended the arm, then reached as far as possible along a virtual horizontal bar without losing balance. A color change from purple to blue provided feedback when
the wrist touched the virtual bar, assisting subjects in following the correct reaching direction. The wrist movement trajectory during the PSST was recorded. The distance between the initial and farthest wrist positions along the virtual horizontal bar was referred to as the functional reach test score (FRTS) in that specific direction [30].

Each subject completed both an XR training session and a no-XR training session. The session orders were counterbalanced. Specifically, forty-two subjects were randomly assigned to two equal groups: one group underwent the XR training session first and conducted the experiment as a-b-c-d-e-f in Fig. 2, while the other group began with the no-XR training session and followed the sequence a-e-f-d-b-c. Training environment was virtual in the XR session but real in the no-XR session. Reaching targets were classified into three difficulty levels: basic, medium, and hard. During the XR training session, targets were represented as virtual bronze, silver, and gold coins. In contrast, the no-XR training session utilized real small items at subjects' home. For each subject, basic level targets were positioned at the farthest reaching positions measured in the pre-training PSST, while medium and hard level targets were placed 10% and 20% FRTS farther away from the basic level targets, respectively. Subjects completed 96 bouts (24 rounds × 4 directions) of reach task in one training session. Each reach was associated with a score: basic, medium, and hard level targets corresponded to scores of 1, 2, and 3, respectively. During training, subjects were motivated to achieve the highest possible score. A color change to red provided visual feedback when the wrist touched the virtual coins. A ten-minute intermission ensued after the first training session, followed by a PSST prior to the second training session (see Fig. 2d).

After the XR training session, subjects completed four standard questionnaires: the Simulator Sickness Questionnaire (SSQ) [31], the Game Engagement Questionnaire (GEQ) [32], the Universal Enjoyment Questionnaire (UEQ) [33], and the Witmer & Singer Presence Questionnaire (PQ) [34]. After the no-XR training session, subjects filled out the same questionnaires, excluding the presence questionnaire.

E. Data Collection

Whole-body landmarks were identified using MediaPipe, and movement data were recorded during experimental sessions (Fig. 2a, b, d, e). To standardize the representation of the movement distance, we normalized the pixel distances by the actual length of the upper arm, converting them into centimeter units.

Self-report data from questionnaires (Fig. 2c, f) were collected using a seven-point Likert scale [34]. To standardize the scores, we normalized each answer sheet’s total score to the range 0 ~ 1 by dividing the sum of all question scores by the maximum possible score.

F. Measures

Training outcome measures are summarized in Fig. 2. Our measures fall into two categories: objective measures (e.g.,
FRTS, intensity score) and subjective measures (including SSQ, GEQ, UEQ, and PQ scores). Notably, the training sequence has a substantial impact on the objective measures [35]. Consequently, the post-training PSST was conducted before the second training session. Additionally, for all subjects, only the intensity score collected from the first training session was recorded.

1) **FRTS**: FRTS refers to as the maximum distance a subject can reach while maintaining balance. Previous research has demonstrated its reliability and validity in evaluating upper limb function and seated postural control [30]. In this study, we compared pre- and post-training FRTSs in four directions to assess the impact of the remote XR intervention on motor performance.

2) **Intensity Score**: Goal-oriented posture training promotes motor skill exploration, and the count of challenging and meaningful repetitions can serve as an indicator of training dosage [12]. In this study, reaching targets at basic, medium, and hard levels were assigned scores of 1, 2, and 3, respectively. We then calculated the cumulative score across all reaches in the first training session for each subject to assess the training intensity.

3) **SSQ Score**: Simulator sickness, also known as motion sickness in virtual environments, occurs when interacting with simulated settings. This discomfort arises due to a mismatch between visual cues suggesting self-motion and the vestibular system’s inability to detect corresponding inertial forces [31]. Common symptoms of simulator sickness include nausea, vomiting, eye fatigue, dizziness, and ataxia. In this study, we employed the SSQ as a self-report checklist to evaluate the severity of 16 symptoms related to simulator sickness in subjects after their training sessions.

4) **GEQ Score**: Engagement pertains to the level of involvement and attention exhibited by subjects during game play. It is a subjective metric influenced by factors like the state of mind, dissociation, and psychological absorption [32]. In this study, the engagement level of each subject during training was assessed by the GEQ score. An example question from the GEQ is: “Did you feel like time stands still or stops during training?”

5) **UEQ Score**: Enjoyment refers to the positive experience and satisfaction derived from playing a game. It is a subjective construct influenced by elements such as pleasure, task challenge level, and competence [33]. The enjoyment level of each subject during training was assessed by the UEQ score. An example question from the UEQ is: “Did you feel pleasurable in interacting with virtual items?”

6) **Immersion Score**: XR encompasses a spectrum of reality technologies such as augmented reality, augmented virtuality, and virtual reality. These technologies exhibit varying degrees of merging physical and virtual reals, forming a reality-virtuality (RV) continuum [36]. At one extreme of this continuum lies the physical real world, while the opposite end represents the absolute virtual world [37].

The immersion level of an XR system reflects its ability to deliver simulated virtual interaction experiences to users. This objective measure is influenced by both hardware characteristics (e.g., rendering quality, refresh rate, and field of view) and software design elements (e.g., virtual item texture, lighting, color, and interaction logic). In this study, we employed the approach proposed by Selzer et al. [38] to compute the immersion score for our XR system. Subsequently, we normalized the score within a range of 0 to 1, representing the relative position of our XR system along the RV continuum (with the no-XR training setup defaulting to 0).

7) **PQ score**: Presence describes the sensation of existing physically within a virtual environment, where subjects perceive themselves as an integral part of the digital realm. This subjective measure is influenced by factors such as emotional variation, psychological state, and the quality of XR intervention [34]. We employed the PQ score to assessed the presence level during remote XR training, with a baseline of 0 for no-XR training. An example question in PQ is “Were you able to anticipate what would happen next in response to the actions that you performed within the virtual environment?”

### G. Statistical Analysis

Statistical analysis was conducted using SPSS (IBM, v29) with a significance level set at 0.05. Normality was assessed through the Shapiro-Wilk test and visual inspection of Q-Q plots.

To evaluate the impact of remote XR interaction on FRTSs in pre- and post-training PSST, we employed the two-way mixed between-within group analysis of variance (ANOVA). Levene’s and Mauchly’s tests were used to examine homoscedasticity and multicollinearity, respectively.

An independent t-test was conducted to detect the effect of XR on intensity scores. Paired t-tests were used to compare SSQ, GEQ, and UEQ scores between XR training sessions and no-XR training sessions.

### III. RESULTS

#### A. Immersion and Presence

The normalized immersion score for our remote XR training system is 0.34. Fig. 3 shows its relative position on the RV continuum. The subjects’ self-reported PQ scores (mean ± standard deviation) for the remote XR training experience are 0.31 ± 0.06.

#### B. Functional Reach Performance

Average FRTSs in PSSTs before and after the XR and the no-XR interventions were visually depicted in Fig. 4. Across all directions, the training condition (XR or no-XR) significantly interacted with the test time (pre- or post-training): \( p = 0.002, 0.032, 0.045 \) for D, F, B, and \( p < 0.001 \) for ND.

For the front and nondominant directions, only the XR group showed a significant increase in FRTS after training.
C. Training Intensity

As shown in Fig. 5a, there was a significant difference in intensity scores for no-XR intervention ($M = 138.52$, $SD = 26.43$) and XR intervention ($M = 176.62$, $SD = 38.20$; $t(40) = -3.76$, $p < 0.001$, two-tailed). Mean difference = -38.10, 95% CI = [-58.58, -17.61] was very large ($\eta^2 = 0.26$).

D. Simulation Sickness Level, Engagement, and Enjoyment

As shown in Fig. 5b, no statistical difference was found in SSQ scores between no-XR ($M = 0.14$, $SD = 0.03$) and XR ($M = 0.14$, $SD = 0.03$), $t(41) = -0.98$, $p = 0.33$, two-tailed). The mean difference was -0.01 (95% CI = [-0.02, 0.01]). $\eta^2 = 0.02$ indicated a small effect size.

As shown in Fig. 5c, a significant difference exhibited in GEQ scores between no-XR ($M = 0.30$, $SD = 0.04$) and XR ($M = 0.66$, $SD = 0.05$), $t(41) = -37.95$, $p < 0.001$, two-tailed). The mean difference was -0.36 (95% CI = [-0.38, -0.34]). The $\eta^2 = 0.97$ indicated a very large effect size.

As shown in Fig. 5d, there was a significant difference in UEQ scores between no-XR ($M = 0.43$, $SD = 0.07$) and XR ($M = 0.81$, $SD = 0.05$), $t(41) = -39.56$, $p < 0.001$, two-tailed). The mean difference was -0.38 (95% CI = [-0.40, -0.37]). The $\eta^2 = 0.97$ indicated a very large effect size.

IV. DISCUSSION

Our results show that the remote XR rehabilitation system could significantly improve functional reach performance. Multi-directional reach movements necessitate intricate neuromuscular control for motor planning while maintaining postural equilibrium [39]. Prior studies report that XR interaction could effectively induce neurophysiological adaptations, including improved interhemispheric balance, enhanced cortical connectivity, and increased muscle cortical representation [14]. These adaptations could stimulate neural plasticity, positively impacting motor function. Repetitive and goal-oriented active practice is a key factor for promoting neural plasticity [12]. In this study, despite both groups performing an equal number of reach movements during training, the XR group achieved a higher intensity score than the no-XR group. This suggests that subjects in the XR training session actively explored and honed motor skills beyond their pre-test stability limits, leading to more pronounced improvements in the FRTS compared to the no-XR training condition.

Previous research has highlighted the potential of the gamification characteristic within XR to enhance cognitive and emotional engagement, thereby promoting voluntary participation and optimizing intervention outcomes [12], [15]. Our findings are consistent with existing literature. Subjects achieved significantly higher GEQ and UEQ scores in the XR training session, indicating greater engagement and enjoyment during sitting posture training. Some subjects even became so absorbed in earning virtual coins during the XR intervention session that they momentarily forgot that they were undertaking a posture training task. In contrast, subjects perceived the no-XR training task as monotonous and aimed to complete it swiftly. Prior studies emphasize that adherence to training protocols and active emotional engagement significantly impact
treatment effectiveness [40], [41]. In this study, the positive attitude and high motivation towards training likely played a pivotal role in the observed longer training duration and improved rehabilitation outcomes for the XR condition.

Open-source commercial games are commonly used in XR posture training [5], [15]. However, prior research indicates that excessive recreational features in these games may lead to cognitive overload and distract subjects from posture training goals [42]. To balance the entertainment and rehabilitation requirements, we adopted suggestions from the literature [12], [22], [23], [43], [44] and developed a custom XR game application. The application featured clear functional reach targets with varying difficulty levels. When subjects successfully reached a virtual coin, its color changed to red as positive feedback. Besides, leveraging MediaPipe for markerless motion tracking, our remote XR interaction application eliminated body marker constraints and allowed direct hand interaction with virtual objects. This may promote more natural movement and facilitate the transfer of motor skills acquired during training to everyday activities. Our application is available to the research community upon request following the IRB guidelines (access link: https://roar.me.columbia.edu/content/trust). Extending our application to other posture training protocols may promote the development of innovative systems for remote movement analysis and telerehabilitation.

Simulator sickness describes the discomfort experienced in simulated environments. This discomfort is directly linked to the immersion level of the XR system [42]. In this study, remote XR interaction occurred on a distant screen, resulting in significantly lower immersion and presence scores compared to the fully-immersive virtual reality headset used in our previous research [17]. Our findings align with the immersion trends reported in existing literature for reality technologies [45]. The low simulator sickness reported by subjects using our remote XR training system could be attributed to this low immersion level. Additionally, we objectively measured and normalized the immersion score for our system and positioned it on the RV continuum. Future studies working on different remote XR systems can adapt our approach to position their systems accordingly. If their system closely resembles ours, the findings of this study may offer valuable insights for developing their XR-enhanced telerehabilitation systems. Furthermore, our approach of integrating XR with remote sitting posture intervention could potentially enhance postural control recovery in patients with neuromotor disorders. However, rigorous clinical testing is necessary in future studies to address potential exacerbation of XR side effects in specific patient populations.

Our system relies on remote Zoom meetings through low-cost built-in cameras on mobile devices, making it well-suited for home-based rehabilitation. Despite promising results, our system has some limitations. During the experiment, we observed that MediaPipe occasionally struggled to accurately recognize the lower limb during forward and backward reach trials due to picture occlusion. Given our primary focus on upper body movement, this occlusion issue likely has minimal impact in this study. Future studies can opt for an alternative version of MediaPipe specifically tailored for precise whole-body tracking. This choice involves a trade-off: while accuracy is enhanced, running speed may be compromised. Additionally, all data collection and remote XR training were driven by codes running on the researcher’s laptop. This approach eliminates the need for subjects to download additional applications but simply join a Zoom meeting. Although this setup ensures data privacy protection, it does impose a heavy load on the researcher’s hardware. Our study employed an XPS 9730, achieving a tested refresh rate of 11–26 Hz based on calculation and rendering fluctuation. Future studies could explore fine-tuning the MediaPipe model to enhance prediction speed through model compression. Furthermore, instead of directly extracting frames from the Zoom application, we chose to capture frames from an external screen in real time. This approach traded the entire system’s running speed for the generalization ability. Researchers extending our system to their studies can select their familiar remote meeting software such as Microsoft Teams, Google Meet, and GoToMeeting. Future optimization efforts could explore direct frame extraction from the chosen software.

V. CONCLUSION

In this study, we developed a remote XR rehabilitation system for sitting posture training. By leveraging markerless motion tracking and a custom game application, the system was highly usable for telerehabilitation and significantly efficient in improving motor performance. The XR interaction led to greater engagement, enjoyment, and improved functional reach performance compared to the no-XR interaction without introducing significant simulator sickness issue. Our approach bridges the spacial gap between subjects and therapists, enabling personalized home-based intervention. This system should serve as a step stone toward advancing novel remote XR rehabilitation systems.

ACKNOWLEDGEMENTS

The authors want to thank all participants in this study. All work reported in this paper were performed in the Columbia Robotics and Rehabilitation (ROAR) Laboratory. We gratefully acknowledge support from the grant NIH R01HD10190.

REFERENCES
