An Initial Study on Enhancing AR Physical Coaching: Multimodal Posture Feedback Using Local ML and LLMs

J. Nivala[1], V. Månsson[2], A. Geris[3,4] & G. Alce[5]

[1,2,3,5]*Department of Design Sciences, LTH, Lund University. Formstråket, Lund, Sweden* | [4]*Department of Instructional Technologies, Manisa Celal Bayar University. Manisa, Turkey*

[1]jo5580no-s@student.lu.se, [2]viktor.mansson0316@gmail.com, [5]gunter.alce@design.lth.se, [3,4]ali.geris@design.lth.se / ali.geris@cbu.edu.tr

DOI: 10.54337/aau.icdvrat26.08

Abstract

This preliminary study presents an Augmented Reality (AR) coaching application for physical rehabilitation. Using a Meta Quest 3 and an iPhone, the system offers a low-strain environment for home exercises. In a user study (N=11), participants performed shoulder presses guided by a 3D avatar. A local machine learning model analyzed their form, triggering personalized audio feedback via a Large Language Model (LLM). NASA-TLX evaluations showed minimal user exertion. Despite minor tracking and API limits, the system demonstrates strong potential for integrating immersive, multimodal AR into accessible physical therapy.

1. Introduction

The integration of immersive technologies and artificial intelligence (AI) is opening new frontiers in healthcare, particularly in physical rehabilitation and preventive medicine. For patients recovering from injuries or individuals training for medical reasons, maintaining proper form during physical exercises is crucial to build strength and prevent secondary injuries. However, home-based rehabilitation often lacks the real-time, expert supervision required to ensure movements are executed correctly, and adherence to home exercise programs remains a well-documented challenge (Argent et al. 2018). Augmented Reality (AR) headsets combined with mobile tracking devices, offer a promising solution by providing immersive, real-time visual guidance, and prior review work suggests that AR can support engagement and rehabilitation-related outcomes in physical therapy contexts (Gil et al. 2021).

When designing AR applications for physical exertion, the choice of interaction methods is a critical factor, as it directly affects user strain, occlusion, and overall usability. Research by Goh et al. highlights the advantages and limitations of various interaction types (touch-based, mid-air gesture-based, and device-based) in mobile AR environments (Goh et al. 2019). Given that wearable headsets can cause fatigue during prolonged usage, interactions must be designed to minimize physical burden. To address these limitations, Nizam et al. emphasize the importance of multimodal interaction techniques, which combine different input methods to make applications more intuitive and less straining (Nizam et al. 2018). Since physical rehabilitation requires users to move their bodies freely, relying solely on physical or manual interactions can be restrictive and counterproductive.

Despite these advancements, there remains a need for accessible, hands-free AR systems that provide intelligent, multimodal feedback without interrupting the user's physical flow. This paper addresses this gap by introducing an AR application designed to help users optimize their form during exercises, such as the shoulder press, for medical rehabilitation or personal health.

The aim of this preliminary study is twofold: first, to develop an AR-based virtual coach that tracks user form via an iPhone and provides multimodal (visual and auditory) feedback through a headset; and second, to compare the accuracy and latency of a local machine learning (ML) model against a Large Language Model (LLM) in generating personalized feedback. To guide this preliminary study, we formulated the following research questions:

1. How suitable is AR as a platform for providing usable and low-strain guidance during physical rehabilitation exercises?
2. How can a local machine learning model and a cloud-based Large Language Model (LLM) be practically integrated to provide multimodal feedback in an AR coaching system?

2. Methodology

2.1. System Overview and Setup

The proposed AR coaching system comprises an iPhone application for body tracking and a Meta Quest 3 headset for the main AR interface. As shown in Figure 1a, users are guided by a 3D avatar (downloaded from Mixamo[1] and animated through a prerecorded motion capture animation) that visually demonstrates a standing shoulder press. Concurrently, the iPhone's camera tracks the user, overlaying a red skeleton on their body using Unity's AR Human Body Manager, and streams this video to the headset via a peer-to-peer WebRTC connection using SimpleWebRTC[2]. The user interface incorporates a movable panel displaying this video stream (Figure 1b), ensuring users can monitor their movements without experiencing severe visual occlusion.

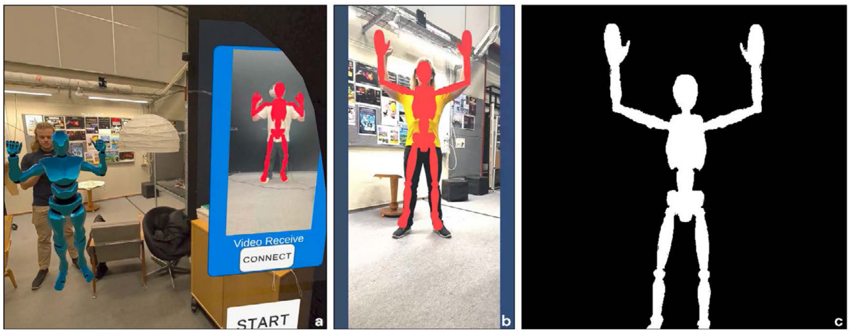

Figure 1: System architecture: AR interface, tracking, and ML pipeline.

1 https://www.mixamo.com/, accessed 2026-01-05
2 SimpleWebRTC, Unity Asset Store, https://assetstore.unity.com/packages/tools/network/ simplewebrtc-309727, accessed 2026-01-05.

2.2. Data Pipeline and Classification

To analyze the user's form without imposing a heavy computational load on the headset, frames from the video stream are extracted twice per second (yielding 84 images per set) and sent as Base64 strings to a local Python server via WebSockets. The server processes these images using color thresholding to isolate the red skeleton, yielding a normalized, centered black-and-white image (Figure 1c). This segmented image is then split into grids to extract feature vectors. A Support Vector Machine (SVM) classifier, trained on 1,013 collected images and categorized into five distinct posture classes (e.g., lower good, left bad, upper good), evaluates the vectors to predict the user's real-time form.

To verify the classifier's reliability, a subsequent evaluation was conducted using a newly collected dataset of 152 images. The SVM model achieved an overall classification accuracy of 76.87%. While the accuracy for correct posture classes was highly reliable (exceeding 92%), performance dropped for specific incorrect postures (e.g., 41.9% for "right bad"). This discrepancy is primarily attributed to class imbalance within the training data; since participants naturally performed well during the initial data collection, the model lacked sufficient examples of erroneous postures to train on.

2.3. Multimodal Feedback Generation

To ensure accurate and non-intrusive guidance, the system aggregates the posture classifications over each 8-repetition set. If the count of a specific error (e.g., the left hand being lower than the right) exceeds a predefined threshold of 5, the system triggers corrective feedback. This classified error is fed into a prompt for Google Gemini via its API, generating concise, personalized, and encouraging text instructions. This text is subsequently converted to audio using Google Cloud Text-to-Speech (TTS) and played to the user between sets, providing a seamless, multimodal coaching experience without interrupting the physical flow.

2.4. User Testing Procedure and Demographics

To evaluate the system, a preliminary user study was conducted with 11 participants aged 23–35 (8 male, 2 female, 1 preferred not to say). The sample primarily consisted of students with generally low prior AR experience (9 out of 11 participants rated their experience below 5 on a 0–10 scale). Participants performed two sets of eight shoulder presses using only their body weight to minimize injury risk. The study aimed to assess the system's usability, the classifier's accuracy, and overall user exertion. Data were collected through direct observation and an anonymous questionnaire incorporating the NASA Task Load Index (TLX) (NASA Human Systems Integration Division n.d.) to evaluate cognitive and physical workload demands.

3. Results

3.1. Workload, Usability, and Feedback

As summarized in Table 1, the perceived workload measured via the NASA-TLX form was notably low, with all average scores falling below 4 out of 10. This indicates that the overall load and demand of using the application were quite manageable. Participants generally felt successful in accomplishing the exercise tasks, rating their performance an average of 7.00. However, responses for performance were somewhat divided, with four participants rating their performance a 5 or below.

Metric	**Score**	**Metric**	**Score**
Usability & Feedback		*Workload (NASA-TLX)*	
Using Intuition	8.55	Temporal Demand	3.63
Coach Helpfulness	8.55	Physical Demand	3.45
Setup Intuition	6.91	Effort	3.36
Stream Helpfulness	5.18	Mental Demand	2.27
Audible Helpfulness	5.00*	Frustration	2.18
Audible Accuracy	4.91*		

Table 1: Average Scores for Workload, Performance, and Usability (Scale 0-10).
**Affected by API technical errors. Adjusted averages excluding outliers: 6.88 and 6.75.*

3.2. Usability and Feedback

The intuitiveness of using the application was rated highly (8.55), though the setup process scored slightly lower (6.91). Regarding the feedback modalities, the 3D coach avatar was the most preferred and helpful feature (8.55), with 9 out of 11 participants favoring it over the video stream and audible feedback.

Initially, the audible feedback received low scores for helpfulness (5.00) and accuracy (4.91). However, this was skewed by a technical failure where the LLM (Gemini) ceased to provide feedback during three test runs due to API limits. Excluding these outliers, the adjusted scores for audible helpfulness and accuracy rose significantly to 6.88 and 6.75, respectively.

The video stream was perceived as less helpful (5.18). Four participants found the red skeleton overlay intrusive or annoying, with one noting that the large overlay obscured the presented feedback. Additionally, several users found it difficult to look at the video stream and the avatar simultaneously. Interaction with the stream panel also proved challenging; although the panel was movable, the pinch-to-move mechanism was hard to grasp for novices, and some participants required assistance to interact with the virtual buttons.

4. Discussion

Overall, the usability of the application and the users' performance in relation to their exertion suggest that the AR coaching concept works effectively. However, qualitative feedback highlighted specific interaction and technical issues that affect the overall experience.

Given that the application is designed for physical workouts, it was expected that physical demand would be a notable part of the overall exertion. Differences in user fitness levels and their perception of the headset's weight likely contributed to the variance in physical demand scores. Mental demand was generally low, indicating that the instructional design was not overwhelming. However, one outlier reported a mental demand of 6/10, which may indicate that technical glitches or the specific test scenario induced stress. Frustration levels were also low on average, but qualitative feedback suggested that the pace of the tests and the positioning of the video stream which required users to constantly shift their gaze were the primary sources of annoyance for the outliers.

Interactions with virtual buttons and panels proved unintuitive for some users, directly contributing to the mental demand and setup frustration. For instance, one user with very little AR experience attempted to press virtual buttons using their entire hand rather than a finger, failing to activate them. Furthermore, the pinch gestures (using the thumb and index/middle finger) required to reposition the stream panels were highly unreliable. Even when users performed the instructed gesture, the system registered it inconsistently. Improving the reliability of these interactions or automating the panel placement is crucial for users to comfortably view both the avatar and themselves without manual adjustments.

Several significant technical problems were discovered during testing. The temporary failure of the Gemini API severely hindered the audible feedback loop. More importantly, the 2D-to-3D body tracking was less accurate than expected, struggling with diversity in body disposition, clothing, and background noise. Since multiple real-life postures could translate to the same overlaying skeleton pose, this inaccuracy cascaded into the local classification model, occasionally confusing users with incorrect visual feedback. Finally, users experienced application rotation issues, where the Quest 3 headset would unexpectedly change its orientation when removed and put back on between test runs.

5. Conclusion and future work

This preliminary study suggests that AR can be a promising platform for guiding users through physical rehabilitation and exercises. The system successfully provides a low-strain, functional environment where users receive real-time visual guidance and post-set audible feedback.

Despite its functional state, several technical challenges were identified and partially addressed. During development, the audible feedback system encountered API rate limits with Gemini 2.5 Flash, which caused the feedback to halt. This was resolved by transitioning to a more lightweight model (*gemma-3-4b*); while this restored uninterrupted functionality, it presented a trade-off with a slight risk of reduced feedback quality. Additionally, to combat the 3D body tracking inaccuracies, the local ML classification model was updated with a broader dataset of user images, yielding a slight but notable improvement in accuracy.

To further optimize the system for clinical-grade reliability and seamless user experience, future work will focus on three main architectural improvements. First, regarding **Tracking and Overlay**, the 3D skeleton overlay will be replaced with a 2D skeleton, allowing for more precise, transformable joint tracking that better matches user disposition, which will require new data collection to rebuild a fitting classification model. Second, to improve **UI Integration** and eliminate the need for users to look back and forth, the video stream panel and the avatar will be designed to overlap. By making them semi-transparent, users will see their own movements in direct correlation to the dynamic size of the avatar. Finally, for **System Stability and Evaluation:** A ”reset view” functionality (similar to Meta's native controls) will be implemented to fix application rotation issues upon removing the headset.

Acknowledgment

The authors gratefully acknowledge the Faculty of Engineering(LTH) for providing the Meta Quest 3 headset used in this project. We also extend our gratitude to Stefan Lindgren and Carolina Larsson at the Humanities Lab in Lund for granting us access to their Motion Capture lab, which was essential for creating our avatar's animations. Finally, we are deeply grateful to all the participants who volunteered to test the application during its development.

References

Argent, R., Daly, A. & Caulfield, B. (2018), 'Patient involvement with home-based exercise programs: can connected health interventions influence adherence?', *JMIR mHealth and uHealth* 6(3), e8518.

Gil, M. J. V., Gonzalez-Medina, G., Lucena-Anton, D., Perez-Cabezas, V., Ruiz-Molinero, M. D. C. & Martín-Valero, R. (2021), 'Augmented reality in physical therapy: systematic review and meta-analysis', *JMIR Serious Games* 9(4), e30985.

Goh, E. S., Sunar, M. S. & Ismail, A. W. (2019), '3d object manipulation techniques in handheld mobile augmented reality interface: A review', *IEEE Access* 7, 40581–40601.

NASA Human Systems Integration Division (n.d.), 'NASA Task Load Index (TLX) Rating Scale PDF'. Accessed: 2026-01-05. https://ntrs.nasa.gov/api/citations/20000021488/downloads/20000021488.pdf

Nizam, S. S. M., Abidin, R. Z., Hashim, N. C., Lam, M. C., Arshad, H. & Majid, N. A. A. (2018), 'A review of multimodal interaction technique in augmented reality environment', *International Journal on Advanced Science, Engineering and Information Technology* 8(4-2), 1459–1466.