AR-Enhanced Workouts: Exploring Visual Cues for At-Home Workout Videos in AR Environment

Yihong Wu
wuyihong@zju.edu.cn
State Key Lab of CAD&CG
Zhejiang University, China

Lingyun Yu
Lingyun.Yu@xjtlu.edu.cn
Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu, China

Jie Xu
elvinxu0405@126.com
State Key Lab of CAD&CG
Zhejiang University, China

Dazhen Deng
dengdazhen@zju.edu.cn
State Key Lab of CAD&CG
Zhejiang University, China

Jiachen Wang
wangjiachen@zju.edu.cn
State Key Lab of CAD&CG
Zhejiang University, China

Xiao Xie*
xxie@zju.edu.cn
Department of Sports Science
Zhejiang University, China

Hui Zhang
zhang_hui@zju.edu.cn
Department of Sports Science
Zhejiang University, China

Yingcai Wu
ycwu@zju.edu.cn
State Key Lab of CAD&CG
Zhejiang University, China

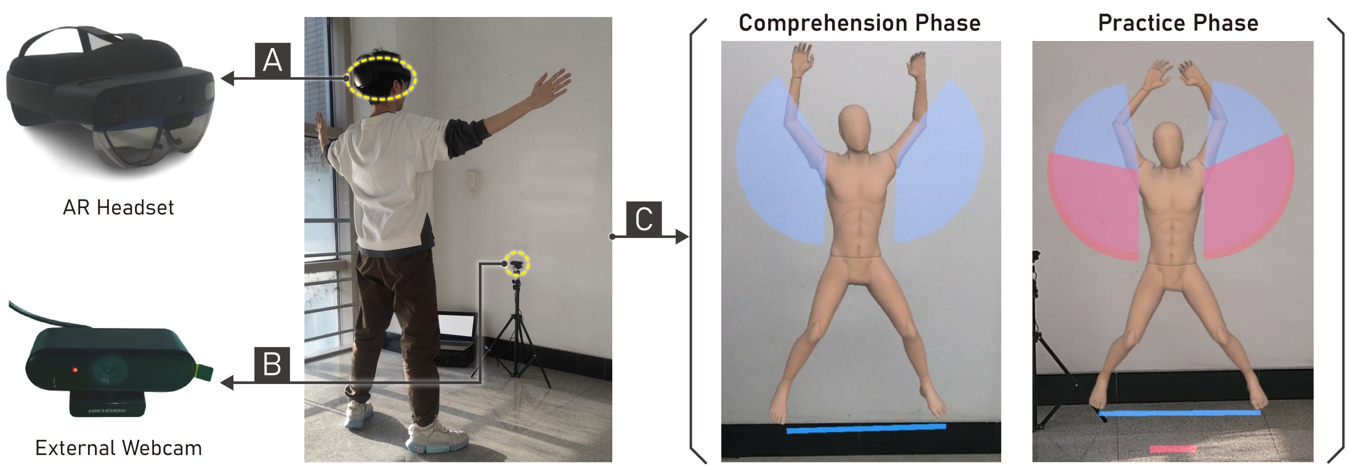

Figure 1: The example usage scenario of our study. (A) HoloLens 2, offering an augmented reality experience with workout videos and visual cues; (B) an external webcam, capturing user's movements in real-time; (C) visual cues for distinct workout learning stages (*Comprehension* and *Practice*), with blue signifying example exercises and red denoting user performance.

ABSTRACT

In recent years, with growing health consciousness, at-home workout has become increasingly popular for its convenience and safety. Most people choose to follow video guidance during exercising. However, our preliminary study revealed that fitness-minded people face challenges when watching exercise videos on handheld devices or fixed monitors, such as limited movement comprehension due to static camera angles and insufficient feedback. To address these issues, we reviewed popular workout videos, identified user requirements, and came up with an augmented reality (AR) solution. Following a user-centered iterative design process, we proposed a design space of AR visual cues for workouts and implemented an AR-based application. Specifically, we captured users' exercise performance with pose-tracking technology and provided feedback via AR visual cues. Two user experiments showed that incorporating AR visual cues could improve movement comprehension and enable users to adjust their movements based on real-time feedback. Finally, we presented several suggestions to inspire future design and apply AR visual cues to sports training.

*Corresponding author: Xiao Xie, xxie@zju.edu.cn

UIST '23, October 29–November 01, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0132-0/23/10…$15.00
https://doi.org/10.1145/3586183.3606796

CCS CONCEPTS

• **Human-centered computing** → **Visualization**; *Empirical studies in visualization*;

KEYWORDS

SportsXR, movement learning, argumented reality

ACM Reference Format:
Yihong Wu, Lingyun Yu, Jie Xu, Dazhen Deng, Jiachen Wang, Xiao Xie, Hui Zhang, and Yingcai Wu. 2023. AR-Enhanced Workouts: Exploring Visual Cues for At-Home Workout Videos in AR Environment. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), October 29–November 01, 2023, San Francisco, CA, USA*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3586183.3606796

1 INTRODUCTION

At-home workouts, such as yoga and high-intensity interval training (HIIT), have gained increasing popularity among fitness-minded people in recent years [2, 13]. The primary reasons for its high popularity include: **Convenience**, as many at-home workouts do not require additional equipment, allowing people to start exercising in a familiar and controllable environment at any time; **Safety**, as it maintains privacy and mitigates the risk of exposure to crowded public areas during the COVID-19 pandemic; **Accessibility**, as people with physical limitations can tailor workout programs to accommodate their own pace and requirements.

People usually search for instructional workout videos on fitness apps (e.g., 8Fit [1]) or video websites (e.g., YouTube [14]), then watch and follow them on mobile phones or televisions. However, our survey revealed several problems associated with this learning process: **First**, it is hard for users to quickly learn the correct movements from the 2D screen-based video. Workout videos typically present exercise movements using one or more fixed camera angles, which can create a gap in understanding for beginners due to the three-dimensional nature of human movements. Beginners require a more flexible observation of exercise movements. Moreover, many exercises involve supine/prone positions (such as planks) or head movement (such as sit-ups). As beginners often glance at the screen to check the exercise movements, this could result in compromised posture. **Second**, pre-recorded workout videos often fail to provide effective feedback. During exercise practice, beginners often require external assistance (e.g., full-body mirrors or others' guidance) to ensure they are performing the movements correctly, which raises the barrier for movement replication.

Figure 2: This example demonstrates how users can observe the exercise of virtual human from different angles.

Augmented reality (AR) technology can provide solutions to the aforementioned problems. First, with an AR Head Mounted Device (HMD), users can freely observe a 3D virtual human performing exercises (Fig. 1(C), Fig. 2). The AR headset can keep the workout video centered in the user's field of view, preventing irregular posture caused by glancing at the screen. Second, during practices, users can obtain real-time visual feedback about their movements from AR devices. For example, the visual cues can indicate the disparities between users' movements and standards (Fig. 1, Practice Phase), which can aid users in refining their movements [48].

Nevertheless, there are two primary challenges when utilizing AR technology for workout training. First, existing AR devices contain powerful toolkits for modeling the surrounding environments, but lack support for monitoring the full-body status of the users. In this study, we introduce an additional camera to complement the information on user bodies and propose a pipeline for AR training. Leveraging 3D pose tracking technology, we capture users' skeleton joint data. The data is used to provide AR visual cues (Fig. 1, Practice Phase) for users to adjust their movements immediately.

The second challenge pertains to the lack of clarity regarding the design space of visual cues and their effectiveness in the context of AR training. To begin, we conducted a preliminary study with fitness-minded people, using interviews and questionnaires to identify the limitations and potential improvements in workout video learning. After obtaining the results, we worked closely with fitness enthusiasts to progressively refine the design space for AR visual cues in workout videos. Based on our pipeline and the design space, we developed an AR-enhanced workout video application. To assess the effectiveness of AR visual cues in enhancing exercise comprehension and practice, we designed two user experiments. From the experiment, we learned that fitness beginners would benefit from workouts enhanced by 3D immersion, and the use of visual metaphors can make the workouts more engaging and memorable.

The main contributions of this paper are as follows:

(1) A pipeline integrating 3D pose tracking technology for real-time visual feedback on exercise performance.
(2) A user-centered exploration of the design space for AR visual cues in workout videos.
(3) Two user experiments to evaluate AR visual cues, along with design suggestions and lessons learned from user interviews.

2 RELATED WORK

In this section, we present prior studies related to our work, divided into two categories: visual cues for physical training and immersive methods for physical training.

2.1 Visual Cues for Physical Training

In traditional physical training, trainees often rely on external guidance, such as coaches, to understand and adjust their movements. However, this approach can be expensive and inconvenient, leading to an increase in self-training research in recent years [20, 36]. Visual cues offer a user-friendly and lightweight option, serving as feedback to help users recognize discrepancies in their movements.

As an intuitive approach to pose assessment, the visualization of **skeleton joints** is widely used in self-training systems. With pose tracking technology, many studies display a series of joint points via video playback [11]. For example, Wang et al. [59] detected human movements in skiing and overlaid skeleton joints on the athlete. Deng et al. [26] recognized table tennis player skeletons for action analysis. Fieraru et al. [28] reconstructed human 3D poses using video frames, but the absence of focused guidance led to insufficient feedback for users. Other research emphasizes real-time feedback for in-exercise adjustments [23, 58, 64]. Chen et al. [20] developed a yoga training system that extracts trainees' body contours and key joints, aiding users in refining movements through visual comparisons. Dittakavi et al. [27] introduced an explainable machine learning model for users to understand how to make effective adjustments. While beneficial, directly displaying full-body joints introduces excessive visual elements that might hinder users from recognizing the key aspects of the movement.

In addition to skeleton joints, **trajectories** are widely employed in human movement visualization. Trajectories are usually displayed in two-dimensional [5, 17] or three-dimensional [47, 53] forms, used for demonstrating movement direction or quality evaluation. A representative work, MotionPro [8], illustrates baseball and golf swing trajectories in videos. Clarke et al. [22] presented Reactive Video, using trajectories to emphasize differences by overlaying standard motions. For professional athletic training, Liu et al. [22] proposed PoseCoach, examining common pose attributes and offering visual cues across various dimensions, such as positional and angular data.

Apart from human body movements, there are **external cues** to show useful information for exercising. For example, Woźniak et al. [60] attached sensors to shoes, capturing foot strikes while running and highlighting relevant areas with color. Vidal et al. [56] instructed yoga learners to wear laser lights, using changes in environmental light to evaluate the movement quality. External cues can also be presented in a metaphorical way [15, 46]. Stergiou et al. [52] drew sharp strikes on a virtual body, marking out the key areas to be stretched while dancing. Tsampounaris et al. [55] enhanced user comprehension by coupling dance movements with virtual objects like wooden boxes.

To examine the usefulness of different visual cues, Semeraro et al. [48] conducted user studies with three visual cues in 2D instructional workout videos. However, videos with 2D visual cues encounter limitations when representing 3D data (e.g., depth ambiguity in lines). A systematic design guideline for 3D visual cues in physical training is still lacking. Thus, we aim to employ a user-centered approach to explore and validate the design space of 3D visual cues for workout videos.

2.2 Immersive Methods for Physical Training

In recent years, advances in AR and VR technologies have demonstrated great potential in improving learning experiences [29, 34]. As sports naturally involve extensive 3D data, AR/VR-related methods are increasingly investigated in physical training [16, 38, 50].

Wearing head-mounted displays (HMDs), learners can receive augmented visual feedback in various sports, such as skiing [41, 61], baseball [33, 66], and basketball [18, 37, 54]. Additional devices (e.g., sensors, controllers) are usually used to deliver precise, sport-specific training experiences based on the characteristics of each sport. Racket sports are a typical example [40, 63], where controllers are utilized in virtual worlds as substitutes for rackets. SPinPong [62] equips a Feedback Racket with motion sensors to capture physical data like spin and impact during play, and then enhances ball trajectories with different visual cues. Motion capture devices are also common for equipment-free exercises [19, 21]. For instance, Hülsmann et al. [30] proposed a VR application that presents real-time Tai Chi postures, highlighting virtual body parts to indicate movement errors. However, incorporating these devices increases the threshold, making them less accessible for home use.

As computer vision technology matures, movement data can be collected via regular cameras, thereby lowering the barrier to entry immersive methods [32, 49]. For example, Kyan et al.[35] used the Microsoft Kinect camera system to recreate ballet learners' performances and provide visual guidance. Lin et al. [37] employed external cameras to capture real-time basketball players' free-throw trajectories. This approach allows players to intuitively observe and compare three-dimensional trajectories, which leads to an improvement in their shooting skills. Focusing on at-home scenarios, our work aims to reduce device requirements for better convenience and accessibility. Inspired by related studies, we plan to introduce an additional camera to track 3D poses and incorporate immersive techniques to enrich users' at-home workout experiences.

3 PRELIMINARY STUDY

In this section, we present our interview and an online survey with individuals who are interested in fitness. Our goal is to gain insight into their routines to learn from workout videos, and the challenges they encounter (Fig. 3).

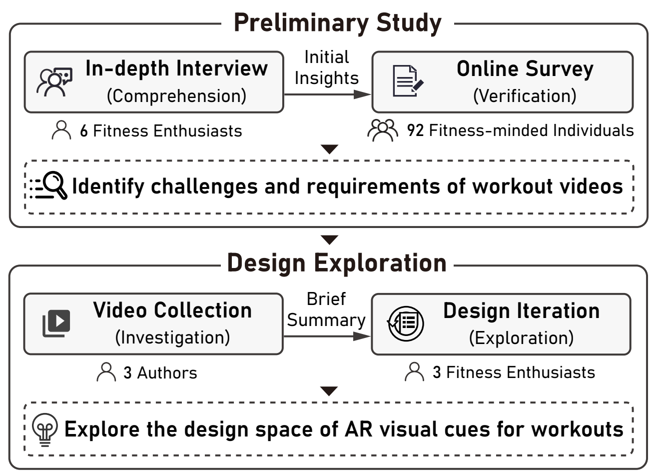

Figure 3: Our study follows a user-centered design process: in the preliminary stage, we gathered design requirements through user interviews and online surveys; in the design exploration phase, we worked closely with users and iteratively refined the design space of AR visual cues for 3D workouts.

3.1 Procedures

Interview. We conducted semi-structured interviews with six fitness enthusiasts (I1 - I6; Male=4, Female=2; Age:22–35 years) who had engaged in regular physical activities for at least two years and had experience with at-home workout videos. Each interview lasted approximately 30 minutes. We inquired about their general exercising process and how videos are involved in the process. We

Table 1: Representative questions from our questionnaire.

No.	Question Content	Question Format
1.	Have you ever learned to exercise through a workout video?	Single-choice (Yes/No)
2.	Do you think visual cues are frequently used in these workout videos?	Single-choice
3.	What kinds of visual cues have you observed in workout videos?	Multiple-choice (With extra input)
4.	Are the visual cues in these videos helpful for you when learning to exercise?	Single-choice (Yes/No)
5.	Briefly describe the aspects you pay attention to while watching workout videos.	Fill-in-the-blank
6.	Do fixed camera angles in workout videos impact your movement learning?	Single-choice
7.	What factors may reduce the quality of your exercise when following a workout video?	Multiple-choice (With extra input)
8.	Share your suggestions or opinions on current workout videos.	Fill-in-the-blank

then presented them with three sets of video clips to understand their preference. Each set showcased the same exercise in four variations (with/without camera angle switching and with/without visual cues). The clips were around 25 seconds long and covered common workout categories, such as aerobic training and core exercises. The visual cues' designs followed the guidelines proposed by Semeraro et al. [48]. After watching a set of clips, interviewees picked the best one and explained their reasoning. Finally, we encouraged them to discuss potential improvements to their existing routines. From the interview transcripts, we identified and outlined a range of challenges that individuals may encounter while learning to exercise through at-home workout videos.

Online Survey. To further ground the findings from interviews, we created a questionnaire based on these initial insights. The questionnaire was distributed within the university, covering topics such as exercise frequencies, experience with workout videos, awareness of visual cues in workout videos, and more (Table. 1). In order for respondents to understand what a "visual cue" is, the questionnaire provided sample images clearly pointing out internal visual elements such as arrows and body highlighting. Moreover, if a respondent chooses "*Never seen*" for Q2, then Q3 and Q4 will not be presented to her/him. In the end, we collected a total of 189 responses, of which 92 respondents had prior experience learning from workout videos (Male=31, Female=61; Average Age=23.16 years, SD=2.45 years). Among these respondents, 55 participated in occasional workouts, while the remaining 37 had a regular weekly workout routine. Their responses supported our analysis of design requirements to enhance existing at-home workout videos.

3.2 Findings and Discussions

The preliminary study revealed the routines and requirements of fitness-minded individuals when learning from workout videos.

3.2.1 General Process of Learning from Workout Videos.

In the user interviews, all interviewees reported that they would sometimes learn unfamiliar exercises through workout videos from YouTube [14] or fitness apps (e.g., Keep [4]) to adapt their workout programs. This primarily depends on the evolution of their fitness goals (e.g., building muscle or reducing body fat percentage) or the need to protect their bodies. For instance, I2 mentioned, *"I tried various knee-friendly exercises last year due to my leg injury."*

When faced with an unfamiliar exercise, five interviewees stated that they would first watch the video to gain a general understanding of the movements before attempting to perform them. They emphasized the importance of watching the movements as a foundation for their subsequent exercise performance. The last interviewee reported that he was accustomed to following the exercises directly and *"adjusted the movements in real time as I imitated them"* (I4).

Based on their responses, the workout learning process can be divided into two main stages: the **Comprehension Phase** and the **Practice Phase**. We will summarize the responses and indicate the stages they belong to.

3.2.2 Challenges Faced in Learning from Workout Videos.

While using videos to learn exercises is common for fitness enthusiasts, several challenges have emerged in their learning process. Notable concerns include *"fixed camera angles make it difficult to grasp a comprehensive view of the workout"* (I2, Comprehension Phase) and *"insufficient feedback to indicate whether I'm performing the exercise correctly"* (I1, Practice Phase). Questions Q6 and Q7 in the questionnaires further provide evidence for the prevalence of issues identified, and the suggestions and feedback from the respondents (Q8). After analyzing 92 valid questionnaires, we identified four major challenges:

C1 **Fixed camera angles limit the complete view of a workout.** When we asked interviewees to choose their preferred workout videos, all opted for those with camera angle switching. This led us to hypothesize: "Limited camera angles in videos provide insufficient exercise information, hindering users' movement comprehension." The online survey (Q6) validated our hypothesis. Among 92 respondents, 8 selected *"No, a single camera angle is sufficient for understanding,"* 22 chose *"No, multiple camera angles are adequate,"* while the remaining 63 opted for *"Yes, I want to be more flexible in my observations."* Hence, overcoming the constraints of fixed camera angles and allowing users to flexibly observe movements is a significant challenge for current workout videos.

C2 **Existing workout videos offer inadequate visual guidance.** Visual cues have proven their effectiveness in learning movements [48, 52, 65]. Our preliminary results also corroborate this: all six interviewees preferred workout videos with visual cues, and 92.31% (Q4, 72 out of 78) of survey respondents who had experienced visual cues found them helpful for learning workouts. However, visual cues are not widely available in current workout videos (Q2): out of 92 respondents, 14 chose *"Never seen,"* 46 chose *"Occasionally seen,"* and the remaining 32 chose *"Often seen."* We suspected that the observed discrepancy might stem from the high cost of creating visual-enhanced videos and the absence of prevalent design guidelines for such cues.

C3 **Users lack feedback on their exercise performance.** Three out of six interviewees (I1, I4, I6) reported difficulty in assessing the performance of their movements without external support (e.g., full-body mirrors or guidance from others) owing to a lack of immediate feedback. We included this concern as an option in Q7, and 74 out of 92 respondents selected it. Addressing this deficiency is crucial, as improperly executed movements may not only compromise the workout's effectiveness but also pose a potential risk of physical injuries.

C4 **Concentrating on the screen can lead to improper posture.** I4 described the problems encountered: *"When I am following workouts that require prone or supine positions, frequently turning my head to see the screen leads to incorrect posture."* Out of 92 respondents, 74 shared the same opinion (Q7). After gathering suggestions for improving workout videos (Q8), we inferred that users desire videos to remain within their field of view. However, commonly used devices such as mobile phones and television struggle to meet this requirement.

3.2.3 Potential Improvements for Workout Videos.

After compiling the suggestions for enhancing workout videos, we conclude two requirements to tackle the observed challenges:

Displaying workouts in an AR environment. Currently, workout videos are mostly displayed on 2D screens, leading to unavoidable issues with fixed camera angles and screen dependency. We propose to employ head-mounted displays (HMDs) to tackle the loss of workout details during the Comprehension Phase (C1). HMDs allow users to freely observe a virtual human's workout movements in an AR environment. Nevertheless, users may lose sight of the virtual human while doing certain exercises. To mitigate this, we can keep the movements displayed within the user's field of view during the Practice Phase (C4), highlighting the unique benefits of the AR environment.

Combining visual cues with workouts in an AR environment. Visualization is useful in understanding and comparing information [44, 57], which has proven to be an effective cue for workouts [48]. Our goal is to explore the design space of visual cues for 3D workout movements, analyze the pros and cons of various designs, and summarize relevant design principles and insights (C2, Comprehension Phase). Moreover, with the aid of 3D posture tracking technology [6, 12], users' exercise postures can be accurately captured. By providing feedback through visual cues (e.g., key skeletal joint motion trajectories), users can assess whether their movements are correct (C3, Practice Phase).

These requirements guided our system design: 1. integrating AR for observing workouts from various angles (C1) and centering them in the user's field of view (C4); 2. proposing a visual cue design space for movement guidance (C2); 3. utilizing pose tracking technology to capture user movements and provide feedback (C3).

4 DESIGN EXPLORATION

To come up with a design space for 3D visual cues, we start with popular workout videos in order to understand the designs and workout types. Then we initiate and iteratively fine-tune the design space in collaboration with three dedicated fitness enthusiasts.

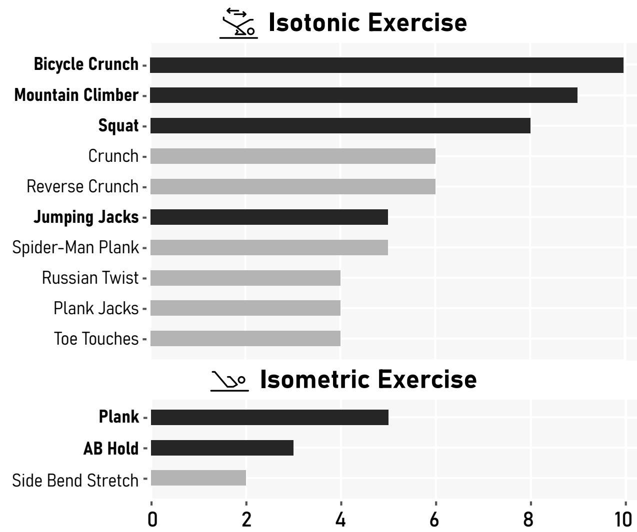

Figure 4: Our statistics on the frequency of both *isotonic* and *isometric exercises* in popular workout videos. The exercises highlighted in bold text were selected for our experiments.

4.1 Video Data Analysis

We collected the top 20 English at-home workout videos on YouTube (each with over 30 million views), to understand how current exercise videos guide beginners. Accompanied by images, most videos usually contain voice guidance to tell the audience how to perform the movement correctly. Only one video incorporates simple visual elements (arrows) to emphasize key body parts during exercises. Considering the advantages of visual cues for motor learning [39, 48], we assumed that the current prevalent video presentations could be considerably improved, especially for people with hearing impairments or different language backgrounds. Hence, our goal is to strengthen the guiding role of visual information in AR exercise scenarios, as a complement or replacement to voice guidance. After the analysis of the audio, we identified the key elements that visual cues could cover, such as postures, rhythms, workout tricks, and crucial body parts. The aforementioned insights have been integrated into the design space. Details can be found in Section 4.2.

In addition, workouts can be categorized into various types, each requiring distinct visual designs to enhance training. To facilitate the creation of a design space and guide our selection of workouts for user experiments, three authors reviewed and cross-validated these 20 videos, cataloging 144 distinct workouts. Among them, 131 are *isotonic exercises*, which involve changes in muscle length. The remaining 13 are *isometric exercises*, requiring muscle engagement in a stationary position [3]. The most prevalent workout types are both showcased in Fig. 4 and serve as an essential corpus for designing our AR visual cues.

4.2 Design Space

The construction of the design space is composed of two stages. First, we proposed a series of visual cue designs based on the suggestions in the preliminary study (questionnaire Q5), the insights

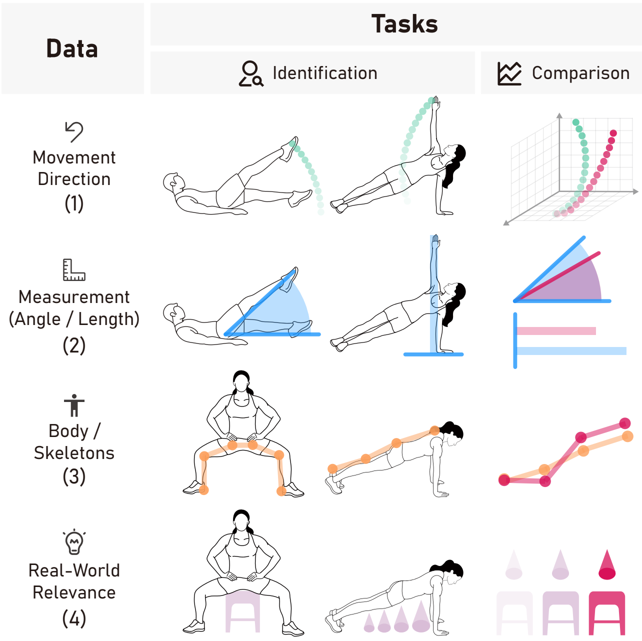

Figure 5: Our design space is based on data and task partitioning, featuring four visual cue designs: green for *directional cues*, blue for *measurement cues*, orange for *skeleton joints*, and pink for *visual metaphors*. In the comparison task, we use red to uniformly represent the user's own exercise performance for feedback purposes.

from video analysis, and relevant studies in 2D [43, 48] and 3D [42] visual designs for movement learning. We also referred to instructional workout videos with visual cues from popular fitness apps (such as Keep [4], 8Fit [1]).

Next, we categorized visual cues according to the dimensions that users concern (e.g., displayed information, purpose). We proposed initial AR visual cue designs for each dimension and invited three interviewees (I2, I3, I5) to evaluate if these designs addressed their needs and challenges. After each discussion, we collected feedback and revised our categories and designs. Following several iterations, we determined three essential dimensions of visual cues required for workouts in AR environments: **Data**, **Tasks**, and **Situations**.

4.2.1 Dimension I: Data.

We refer to Semeraro et al.'s [48] classification of different movement properties in 2D workout videos and provide a systematic description tailored to AR scenarios. Since workout videos inherently showcase motion-related information, we divide the information to be presented into **motion** and **non-motion data**.

Motion data refers to the movements that users need to pay attention to rather than just the goals they need to achieve. We further decompose it into *direction* and *distance* according to the law of motion. Fig. 5 (1) illustrates a typical *movement direction* visual cue example (green curve). To portray the movement direction of a specific body part, we use a sequence of 3D spheres. The spheres' transparency diminishes over time, clarifying the starting and ending points for users. Notably, we avoid a common line design for representing directions, as 3D lines at varying depths could have identical 2D projections when viewed from certain angles [31, 39], leading to ambiguity. In contrast, the sphere design enables users to perceive depth differences through changes in sphere size, thereby reducing ambiguity. Fig. 5 (2) presents examples of *movement distance* visual cues (blue line or angle). As human limb movements involve both rotation and translation, we describe them with angle and length, respectively. Specifically, we use pies and bars to display measurements in a widely acceptable way. However, isometric exercises are usually static muscle works without body movement. Using motion data fails to convey information about the muscles or joints that are engaged during the workouts.

Non-motion data refers to the goals that users only need to aim to achieve. We decompose it into *in-body* and *out-of-body* information, as users generally focus on body parts related to their workout. Fig. 5 (3) illustrates a typical *body skeleton* visual cue example (orange lines) that aids users in identifying significant workout-related joints. In addition to the body skeleton, highlighting relevant body parts is effective. As for out-of-body information, we employ *visual metaphors* to reflect the connections between workouts and the real world [52]. For example, in Fig. 5 (4), a pink chair is used to show where to squat, while cones indicate the appropriate body height for a plank exercise. Although such visual metaphors are vivid and intuitive, creating suitable real-world metaphors for complex workouts can be challenging.

4.2.2 Dimension II: Tasks.

During different stages of workout learning, users require different information from visual cues. The Comprehension Phase requires a thorough understanding of movements, while the Practice Phase demands immediate feedback. Given that existing research demonstrates the applicability of visualization for various levels of tasks [25, 44, 51], we classify visual cues based on the tasks to be completed, namely, **identification** and **comparison** (Fig. 5).

For **identification** tasks, visual cues present key movement information to facilitate understanding, as detailed in Section 4.2.1. For **comparison** tasks, we employ comparison-based visualization taxonomy [24, 57] to design methods. We primarily apply *superposition* (Fig. 5(1), (2, Angle), (3)), overlaying users' exercise performance onto standard visual cues. This direct overlaying method is intuitive and avoids potentially overwhelming visual elements in *juxtaposition*, which may exceed users' cognitive capacity. Exceptionally, we apply *juxtaposition* (Fig. 5 (2, Length)), arranging two bars side by side as overlapping two differently colored bars hinders discernment. Lastly, for visual metaphor cues, we apply *explicit encoding* (Fig. 5(4)), which involves additional visual channels for comparison. As metaphor-based visualization typically involves transparency and color channels, we offer real-time feedback through color/transparency changes. For example, a transparent pink chair indicates a squat far above the chair's surface, while a non-transparent red chair denotes a squat below it.

4.2.3 Dimension III: Situations.

The situation dimension, stemming from diverse scenarios during the Practice Phase, requires a discussion of the specific workout movements. Many workouts involve frequent viewpoint shifts (e.g., sit-ups) or require supine or prone positions (e.g., planks). As a result, the 3D virtual human and visual cues (Fig. 6, **within sight**)

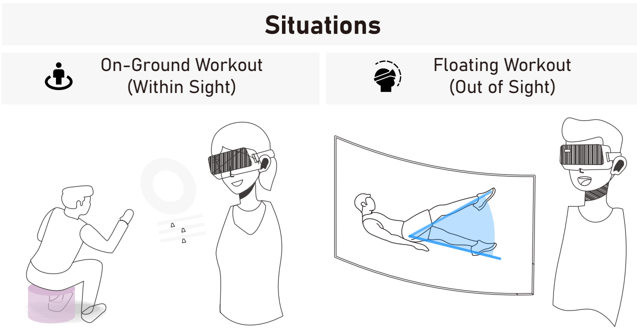

Figure 6: Our design space is also based on situation partitioning. We present workout movements in two distinct ways: *on-ground* and *floating*. This depends on whether the virtual human remains within the user's sight during exercises.

may not consistently remain in the user's sight during exercises. To tackle this issue, we employ a floating screen to maintain example movements and visual feedback in the user's central field of view (Fig. 6, **out of sight**).

5 IMPLEMENTATION

In this section, we introduce our method of displaying 3D virtual human motion in an AR environment, binding visual cues, and providing real-time feedback through these visual cues.

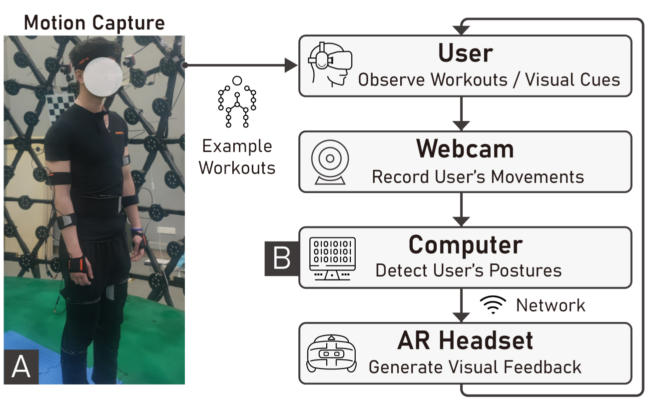

Figure 7: The workflow of AR-based physical training.

5.1 Hardware and Software

We utilized a HoloLens 2 to implement the AR-Enhanced workouts. The HoloLens 2 weighs 566g and supports a 52-degree diagonal field of view with 2048×1080 resolution and a 75 Hz refresh rate. To minimize the equipment thresholds, we chose a standard 1280x720 external webcam (Hikvision DS-E11) for capturing users' movements. The webcam is connected to a laptop (Intel Core i5-8625U CPU, NVIDIA MX250 GPU, 8GB memory) via a USB 2.0 cable.

In terms of software, we used MRTK [9], a mixed reality toolkit in Unity, to develop the AR application for HoloLens. To track the user's 3D pose from live streaming data, we employed the well-established Python library Mediapipe [6], which captures 3D skeletal data comprising 33 articulation points in real time. Finally, we enabled data exchange with HoloLens via the wireless network and developed a backend server using the Python Flask framework.

5.2 Workflow

Data Preparation. We first prepared 3D workout animation examples. Our 3D virtual mannequin originated from Mixamo[7]. Additionally, we utilized some of Mixamo's workout animations, such as bicycle crunch. For missing workouts, we employed MVN Link [10], a professional motion capture device, to obtain accurate 3D workout animations (Fig. 7(A)). The device is used only during the data preparation phase to create sample motions of the virtual human. All human motion animations produced are stored in our application in .fbx file format.

Presenting AR-Enhanced Workouts. We imported the skeletal animation and bound it to the 3D virtual human, enabling users to view exemplary workouts in an AR environment via HoloLens. To associate visual cues with the 3D virtual human, we extracted a sequence of 22 key skeleton joint positions from the animation. Based on the skeleton data, we computed the 3D coordinate sequences necessary to render the four types of visual cues. In the initial state, we positioned visual cues near corresponding body parts, thereby creating an AR-enhanced workout animation (as shown in Fig. 8).

Generating Visual Feedback. Once the user began the exercise, her/his full-body movements were captured by a webcam and transmitted to a laptop in real time. This laptop functioned as the backend (Fig. 7(B)), employing the computer vision library Mediapipe to extract the user's real-time 3D skeleton joint data. However, due to individual skeletal variations, directly mapping user joints to the virtual human skeleton could introduce errors. Hence, calibration was required prior to exercise. During calibration, we asked users to adopt a T-Pose, calculated the distance between joints, and normalized the data. Next, we performed a proportional conversion between the user and virtual human skeleton information to minimize mapping errors. Finally, the skeleton joint data needed for visual cue rendering was transmitted wirelessly to the HoloLens, which displayed real-time visual feedback (Fig. 9).

5.3 Interaction

Our system offers the following interactions to improve usability:

Workout/Visual Cue Selection. Users can select different workout movements or types of visual cues by tapping a virtual panel.

Playback Control. Users can play, pause, and adjust the speed of workout movements (0.5x to 2x) using virtual buttons. However, during the user experiment, participants were restricted from switching workouts/visual cues or using playback controls, which ensured their focus remained on the visual presentation and control experimental variables.

Observation Angle Control. At present, users can change their observation angle of the virtual human by walking on the ground. Other interaction modalities, such as voice or gaze control, may also be natural to adjust observation angles during exercises. Future studies could be conducted to evaluate the effectiveness of different control methods.

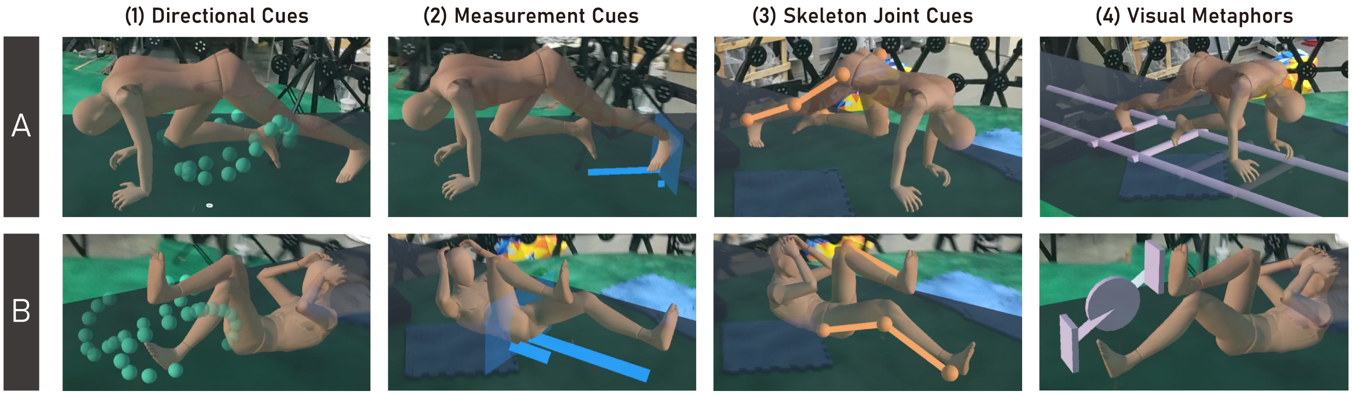

Figure 8: Examples of our AR-Enhanced workout animations. (A) *Mountain Climber.* (B) *Bicycle Crunch.*

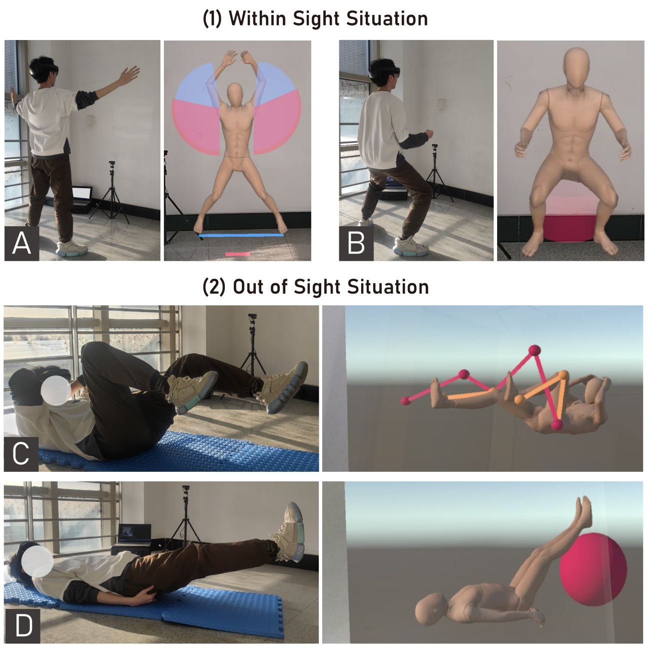

Figure 9: Examples of our AR-Enhanced workout feedback. (A) The arm angles and distance between feet guide users to perform *Jumping Jacks.* (B) The transparent chair indicates that the user's hips are far above the standard *Squat.* (C) The floating workout displays the joint points of the user's legs. (D) The non-transparent yoga ball indicates that the user's legs are far below the standard *AB Hold.*

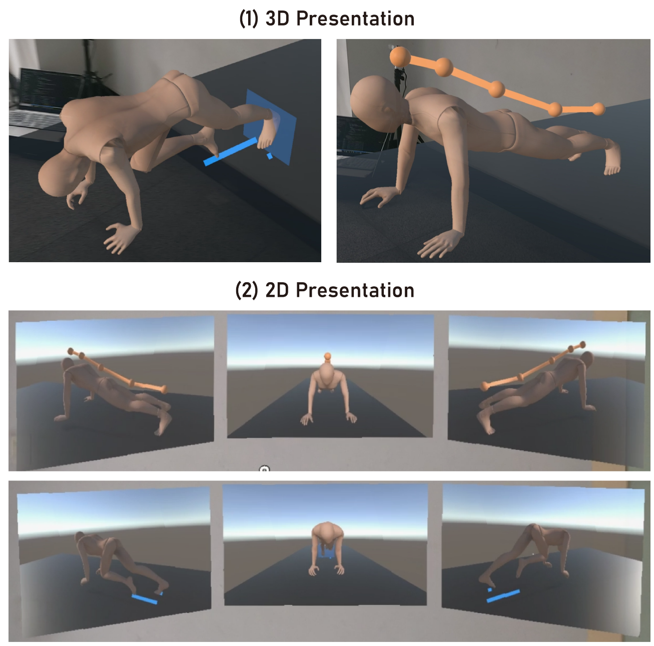

Figure 10: Examples of our 2D and 3D presentations. (1) shows lengths for *Bicycle Crunch* and skeleton joints for *Plank.* (2) shows the same workout from three fixed camera angles.

6 STUDY OF COMPREHENSION PHASE

In the first phase of our experiment, we aimed to investigate whether AR can enhance workout comprehension for fitness beginners. To this end, we conducted the user study under controlled conditions, collecting subjective ratings and qualitative feedback from users.

6.1 Experimental Variables

Our application introduces two types of visual enhancements over the existing 2D at-home workout videos: a transition from 2D to 3D and the incorporation of various types of visual cues. Therefore, the primary goal of our user experiment is to validate whether using 3D presentation enhances the understanding of workouts and to assess the appropriateness of different visual cues for various workout movements. To achieve this, we concentrate on two experimental variables: **dimensions (2D/3D presentation)** and **different visual cues** for the same workout.

2D/3D Presentation of Workouts. We want to verify that fitness beginners would benefit more from workout-related information presented in 3D immersion than on a 2D screen. To validate this, we provided both 2D and 3D presentations for each workout and visual cue (see Fig. 10) and instructed users to compare them. As a single camera angle in a 2D presentation might cause information bias for movements, we offered three viewpoints (the maximum number of fixed camera angles in the workout videos we surveyed): front-facing, 45 degrees to the left back, and 45 degrees to the right back (Fig. 10(2)). To control the interference of HMDs, we asked users to view both the 2D and 3D presentations wearing an HMD. The 2D presentations were fixed in real-world locations to simulate monitors, while the 3D virtual human was placed on the ground.

Different Visual Cues. We also want to verify that each type of visual cue has unique advantages in presentation depending on the workout being performed. To validate this, we consulted with fitness education experts and designed four different visual cues for each exercise movement (as shown in Fig. 8), including *directional cues*, *measurement cues*, *skeleton joints*, and *visual metaphors*. It is worth mentioning that for isometric exercises, which involve merely static postures, we did not design *directional* and *measurement cues*, as these cues are used to describe motion data. All the workouts and visual cues are available in our supplementary material.

6.2 Experiment Settings

Workout Selection. To enhance the generalizability of our experimental findings, we selected common workout movements as the basis for our evaluation. Specifically, we chose the four most common isotonic exercises (*bicycle crunch*, *mountain climber*, *squat*, and *jumping jacks*) and the two most common isometric exercises (*plank* and *AB hold*) from popular at-home workout videos (highlighted in Fig. 4). Notably, we excluded some similar workouts to increase the diversity of workout movements. For instance, *crunch* and *reverse crunch* occurred more frequently than *jumping jacks*, but were not selected due to their similarity to *bicycle crunch*.

Experimental Environment. To ensure a realistic simulation of diverse home environments, we conducted the experiment in various locations, including a laboratory room and open dormitory areas. Participants were provided with an adequate area that allows free movement and observation (at least a cubic space larger than 2.5m × 2.5m × 2.5m) with appropriate lighting.

Evaluation Metrics. Our evaluation metrics are derived from the work of O'Brien and Toms [45] to measure user engagement:

(1) *Flexibility*: How free do you feel to observe the workout movements in the current condition?
(2) *Comprehensibility*: How easy is it to understand the workout movements in the current condition?
(3) *Fun*: How much fun is it to learn about the workout movements in the current condition?
(4) *Memorability*: How easy is it to remember the workout movements in the current condition?

6.3 Participants and Pre-Survey

Our target scenario is for people who are learning unfamiliar workouts in an AR-enhanced environment. Therefore, we recruited 24 fitness beginners (male=16, female=8; average age=23.09 years, SD=2.17 years) from a university. All participants reported that they had not received systematic fitness training and did not engage in any regular physical activity. They had no physical limitations and had never experienced symptoms of motion sickness. Six participants had previous experience using VR headsets, and three had experience using AR headsets, but none had used immersive methods to learn workouts before.

Considering that each participant had different levels of familiarity with various workouts, we conducted a pre-survey prior to the formal experiment. Participants were asked to rate their familiarity with six selected workouts on a scale of 1 (low) to 5 (high). Then in the formal experiment, we would only present each participant with unfamiliar workouts they rated as 1 or 2 (1 - no prior exposure, 2 - occasional exposure). Among all participants, 23 were unfamiliar with *bicycle crunch*, 4 with *squat*, 16 with *jumping jacks*, 20 with *mountain climber*, 14 with *plank*, and 18 with *AB hold*. As *squat* is a commonly known exercise with inadequate valid participants, it was not included in our subsequent discussion.

6.4 Procedures

For clarity of exposition, in the following sections, we refer to the workout animation based on a specific visual cue in a particular dimension as **a condition of the workout**. For example, Fig. 8(A) shows four conditions of Mountain Climber. Hence, each isotonic exercise has eight conditions (2 dimensions × 4 cues), while each isometric exercise has four conditions (2 dimensions × 2 cues).

Before the experiment began, participants were told to wear the HoloLens and watch tutorial *squat* animations to familiarise themselves with our AR application. We ensured their comfort while wearing the HoloLens and understanding of the basic workflow.

During the experiment, participants were instructed to observe the workouts that they had expressed unfamiliarity with in the pre-survey. However, due to the varying number of workouts that each participant was unfamiliar with, counterbalancing the workout order was difficult. As a solution, we randomly allocated the workout order to reduce its potential influence on the experiment results. According to the results of the pre-survey, 24 participants should complete a total of 91 workout observations.

For each workout observation, we presented participants with all conditions. We counterbalanced the order of 2D and 3D presentations, with 45 showing 2D cues first and 46 showing 3D cues first. We also counterbalanced the order of the different visual cues. For a total of 59 isotonic exercise observations, we rotated through all 24 possible permutations of the four visual cues (randomized) to balance the order in which users observed the visual cues. For the remaining 32 isometric exercise observations, 16 showed the skeleton joints first, and 16 showed the visual metaphors first. When participants were observing each condition of a workout, they were required to roughly understand the meaning of the current visual cue (usually between 5 and 30 seconds) before requesting to switch to the next condition. After observing all conditions of a workout, we asked participants to rate each condition on a 7-point Likert scale based on the four evaluation metrics. To assist them in recalling these visual cues, a thumbnail of each visual cue was provided.

After participants viewed all the unfamiliar workouts, we would conduct a brief semi-structured interview to collect their opinions on the AR-enhanced workouts. On average, each participant spent approximately 50 minutes throughout the experiment. We compensated these participants in accordance with our school's policy.

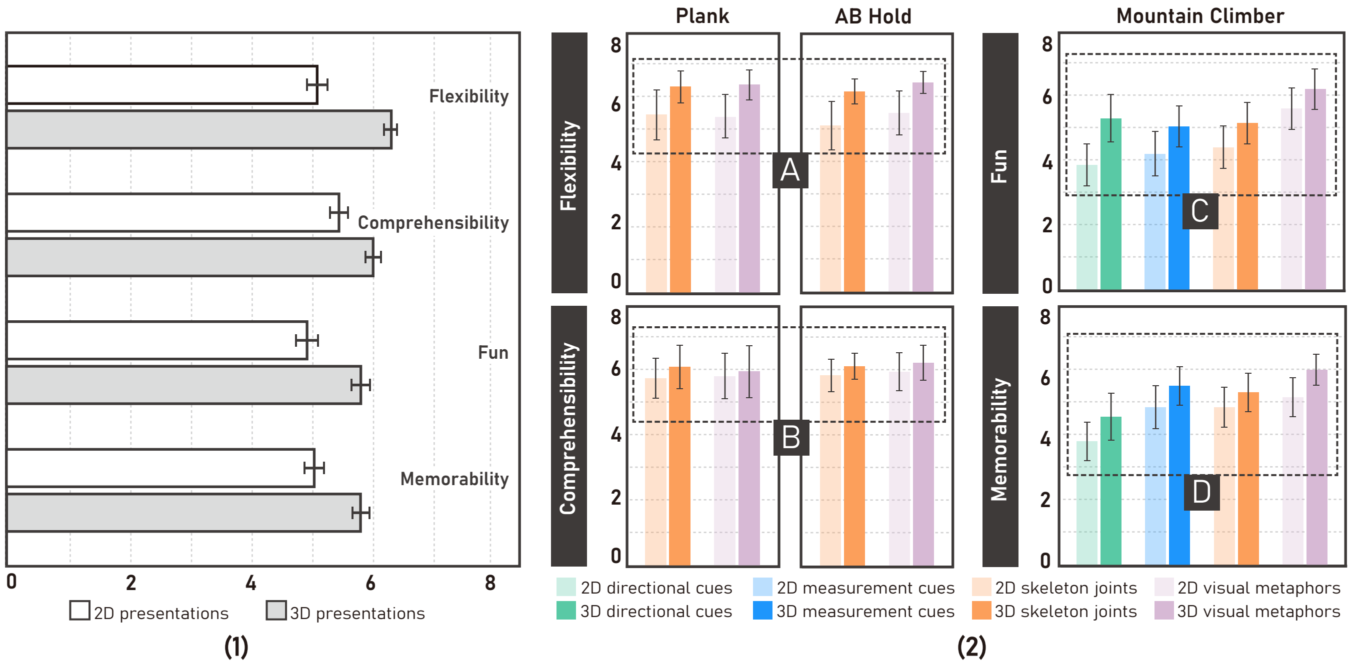

Figure 11: (1) User ratings for all the 2D/3D presentations. (2) Typical user ratings from several workouts, where each color indicates the type of visual cue, and transparency is used to distinguish 2D/3D. Error bars denote 95% confidence intervals, and the complete figure is available in the supplementary material.

6.5 Results and Discussions

After the experiments, we conducted a comparative analysis of descriptive data, including user ratings and qualitative feedback, to derive essential findings.

Insight 1: 3D presentation improves beginners' understanding of the workout movement. For the five workouts, we discovered that the 3D presentations consistently scored higher on all four metrics than their 2D counterparts (see Fig. 11(1)). Thus, we concluded that for fitness beginners, 3D presentations are more effective at facilitating the understanding and memory of workout movements. This conclusion is further supported by the subjective feedback we gathered from participants. One participant (P12) shared that the 3D virtual human performing the exercises felt *" very impressive as if a real coach was guiding me through the movements."* Another participant (P3) noted that the unrestricted observation provided by the 3D presentation helped notice details that might have been missed otherwise: *"Initially, on the 2D screen, I failed to recognize that bicycle crunches require coordinating the upper body and legs. However, when I observed the virtual person from behind, I realized this through the rhythm of their head and knee movements"*. Overall, the majority of participants provided positive feedback on the immersive 3D approach, indicating its effectiveness in increasing realism and enhancing understanding.

Insight 2: The advantage of 3D presentations in enhancing understanding diminishes for simpler exercises. As shown in Fig. 11(B), the average comprehensibility scores for Plank and AB Hold are slightly higher in the 3D presentations than in 2D. However, the error bars reveal that this improvement is not significant, given the considerable overlap in the 95% confidence intervals between the two presentations. In comparison to the flexibility metric in Fig.11(A), we deduce that the primary benefit of 3D presentations for isometric exercises lies in allowing users to observe the exercises without restrictions, rather than enhancing their understanding of the movements. During the interviews, a few participants also expressed this view. For example, P22 mentioned: *"The 2D display of the plank exercise is enough for me to understand this simple movement; I even think that the three viewpoints are unnecessary"*.

This phenomenon is also somewhat reflected in the user ratings. Participants' ratings were more diverse for 2D presentations than for 3D. In all the isotonic exercise ratings (3 workouts × 4 metrics × 4 visual cues), 40 out of 48 scores had a higher standard deviation for the 2D presentation. Similarly, in all the isometric exercise ratings (2 workouts × 4 metrics × 2 visual cues), 10 out of 16 scores had a higher standard deviation for the 2D presentation. We speculate that this phenomenon is partly due to the differences in participants' ability to imitate movements. While some participants found the 3D presentation apparently more flexible and comprehensive, others felt that the 2D video had already provided sufficient information.

Insight 3: Visual metaphors are engaging for their fun and memorability. To verify our second hypothesis, we compared the performance of the four 3D visual cues across different workouts. We used rankings as an indirect measure since the scores of different metrics cannot be directly compared. For each workout, we calculated the average ranking of each visual cue based on the four evaluation metrics. Table 2 presents these average rankings.

Table 2: The average ranking of each 3D visual cue on the four evaluation metric scores. The visual cues with the highest average ranking for each workout are highlighted in bold.

	Directional Cues	Measurement Cues	Skeleton Joints	Visual Metaphors
Bicycle Crunch	2.5	4	**1.25**	2.25
Jumping Jacks	3.5	**1.25**	2.25	2.75
Mountain Climber	3.5	3	2	**1.5**
Plank	/	/	1.75	**1.25**
AB Hold	/	/	2	**1**

Visual metaphors were generally well-received, primarily due to their strong performance in fun (ranking first in 3 out of 5 workouts) and memorability (ranking first in 4 out of 5 workouts). Taking Mountain Climber as an example, its 3D visual metaphor (Fig. 8(A4)) significantly outperformed the other three cues in fun (Fig. 11(C)) and memorability (Fig. 11(D)). Many participants shared their opinions, such as P4, who noted that this approach was closely related to real life. *"Seeing the metaphor of stepping on bike pedals, I can already imagine how to exert myself in this kind of exercise."*

However, not all metaphor designs were universally appreciated. For instance, P11 (who scored the plank metaphor 2 for fun) said, *"In the plank exercise, placing pink nails beneath the body creates an overly aggressive and oppressive feeling. Although I can understand the intent behind the method, I would prefer to watch such instructional videos with a more relaxed mindset"*. Consequently, crafting intuitive and suitable visual metaphor cues presents a challenge, particularly for exercises that are harder to imaginatively connect to real-world situations.

Insight 4: Measurement cues may struggle with information loss in certain workouts. As shown in Table 2, directional cues and measurement cues, although used exclusively for isotonic exercises, haven't received high ratings in these workouts. A typical response comes from P15, *"The green ball path is clear and interesting, especially the heart shape it forms during the jumping jacks. But trying to remember the exact angles and curves from the path seems hard, and it sort of lowers my confidence in learning the moves."*

Exceptionally, measurement cues achieved the highest average ranking (1.25) in presenting jumping jacks. To investigate the reasons, we focused on user feedback regarding measurement cues. Then we speculate that the nature of the jumping jack movement contributes to this result. In jumping jacks, users' hands and feet remain mainly on the same plane (Fig. 1, Comprehension Phase), and the displayed angle and length information aligns with the movement, making it intuitive and easy to understand. In contrast, bicycle crunches and mountain climbers both involve three-dimensional trajectories, which leads to information loss when only displaying the vertical distance from the virtual person's foot node to a certain plane (Fig. 8(2)). As P4 commented, *"The blue bar chart initially caught my eye, but after taking a closer look, I realized it wasn't as accurate as just drawing the trajectory because I couldn't tell when the foot was being lifted."* We thus concluded that the main reason for the poor performance of measurement cues is the information loss resulting from the dimensionality reduction of motion data. This highlights the importance of avoiding using measurement cues for complex 3D trajectories when designing 3D workout scenes.

7 STUDY OF PRACTICE PHASE

Based on the findings from our initial experiment, we plan to expand our research in the second phase of the study to explore the potential impact of real-time visual feedback on the ability of fitness beginners to adapt and refine their movements during practice.

7.1 Experimental Variables

Another enhancement of our application is the real-time visual feedback designed to assist users in adjusting their movements. Hence, we seek to determine whether such feedback can assist fitness beginners in identifying flaws in their movements and enable them to make timely adjustments. For this phase, the experimental variables are as follows:

With/Without Visual Feedback Provided. We established two conditions for the same workout movement: one with feedback provided and the other without feedback (as depicted in Fig. 12). Additionally, we ensured that in both conditions, the example movement was consistently presented in the same format at the center of the user's field of view (both in on-ground form or floating form).

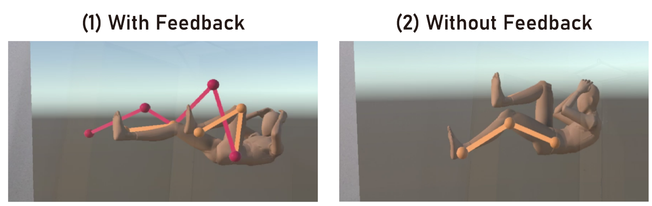

Figure 12: Two conditions of our second experiment: (1) with and (2) without visual feedback provided to the user.

7.2 Experiment Settings

Visual Cue Selection. To ensure consistency across our user experiments, we utilized the same five workouts from the previous study and incorporated the highest-ranking visual cues to provide visual feedback (see Table 2). These cues included using measurement cues for the *jumping jacks*, skeleton joints for the *bicycle crunch*, and visual metaphors for the *mountain climber*, *plank*, and *AB Hold*. The specific feedback methods are described as follows:

(1) *Jumping jacks*: Users can compare their arm angle and distance between feet with the example movement using the red fan and bar (as shown in Fig. 9(A)), respectively.
(2) *Bicycle Crunch*: Users can see the red skeleton joints of their legs overlapping the orange joints of the example movement (as shown in Fig. 9(C)).

(3) *Mountain Climber*: Users can see a red ladder beneath the virtual human. When one of their feet is in the correct position, the corresponding step will be highlighted in non-transparent red.
(4) *Plank*: Users can see several cones beneath the virtual human. When a body part of the user falls below a standard height, the corresponding cone turns red.
(5) *AB Hold*: Users can see a yoga ball under the virtual human's leg. When the user lifts their legs below/at/above the standard height, the yoga ball will turn red/pink/almost transparent (as shown in Fig. 9(D)).

Experimental Environment. Similar to our previous experiment, we did not restrict the study to a fixed location. We further ensured that the experimental site offered sufficient space for participants to exercise without constraints, while also accommodating necessary equipment, such as laptops and tripods.
Evaluation Metrics. We collected qualitative feedback on the effectiveness from both cognitive and practical perspectives. The specific evaluation questions are shown in Fig. 13.

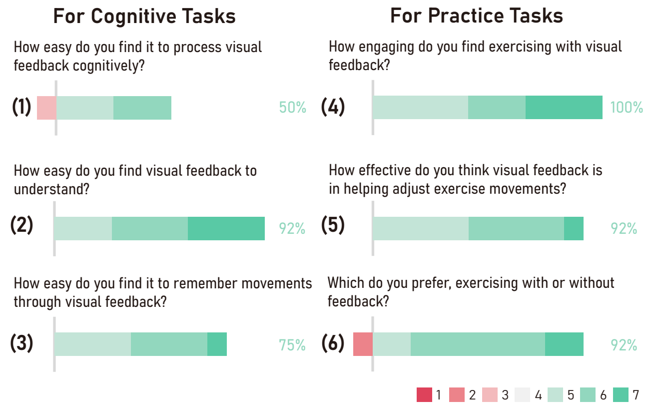

Figure 13: User ratings for the second experiment. The percentages indicate the proportion of positive feedback (≥5 points). Neutral attitudes (=4 points) are not shown in bars.

7.3 Participants

Considering our target scenario, we recruited a total of 12 participants from a university (male=4, female=8; average age=21.75 years, SD=1.83 years). Among the 12 participants, 4 had previously taken part in our previous experiment and were unfamiliar with all five workouts. The other 8 participants were additionally recruited, and similarly, we conducted a pre-survey to ensure they had no fitness experience and had not mastered any of the five workouts. In addition, we signed an informed consent form with each participant to ensure their voluntary participation in the study.

7.4 Procedures

Before the experiment began, we instructed participants to stand in front of the camera, assume a T-Pose for calibration, and wear the HoloLens to avoid any physical discomfort. Subsequently, we sequentially demonstrated five workout movements, explained the meanings of different visual feedback, and guided participants in investigating the connection between their body movements and the provided visual feedback.

After users gained a basic understanding, we presented five workouts in random order, requesting them to perform corresponding exercises (12 continuous repetitions for isotonic exercises and a 15-second hold for isometric exercises). Among these workouts, 2-3 provided visual feedback, while the rest did not. We also ensured a counterbalanced distribution of visual feedback for each workout among the 12 participants, with 6 of them receiving feedback and the other 6 not receiving it. Furthermore, we recorded each participant's movements as they carried out a set of workouts. Between different workouts, we allowed participants to rest for 2 to 5 minutes, ensuring they did not feel overly fatigued before moving on to the next set of workouts. This is to minimize the impact of fatigue stemming from continuous exercise.

In the post-study survey, we investigated both subjective and objective aspects. For the subjective perspective, we asked participants to answer various questions using a 7-point Likert scale (see Fig. 13) and gathered their subjective opinions through interviews. As for the objective perspective, we invited two experts with over five years of fitness experience to review the recorded videos of participants completing each set of workouts (12 participants × 5 workouts in total). The experts were not informed whether the users received feedback or not, and they could only observe the users' motions to identify instances where the initial movements were not correct but improved after adjustments during the workout. On average, each participant's experiment lasted approximately 30 minutes, and suitable compensation was provided.

7.5 Results and Discussions

Based on the subjective ratings from users and instances of real-time adjustments found by fitness experts, we summarized and discussed the following findings:

Insight 1: Visual feedback aids fitness beginners in recognizing and adjusting deviations in their movements. From a subjective perspective, our participant survey revealed that all respondents found the real-time feedback exercise approach engaging (Fig. 13(4), 100% rated 5 or higher). The vast majority expressed positive opinions on the effectiveness of visual feedback for movement adjustment (Fig. 13(5), 92% rated 5 or higher), and 92% of participants preferred an exercise method incorporating feedback.

From an objective perspective, expert E1 found 8 instances of self-adjustment and movement optimization in 70 exercise videos, 6 of which involved feedback. Expert E2 detected 7 such instances, 6 with feedback, and 6 overlapping examples between E1 and E2 (5 with feedback, 1 without). Given the higher number of instances with feedback than without, we inferred that visual feedback has a positive impact on adjusting movements, in addition to participants' abilities to imitate movements.

Insight 2: Visual feedback may impose an additional cognitive load on fitness beginners. In our survey, visual feedback received lower ratings for cognitive tasks compared to practice tasks. As shown in Fig. 13(1), only 50% of participants found it easy to handle visual feedback when performing exercises. Some participants expressed concerns; for instance, P12 (who rated question (1) with a score of 3) stated, *"The red fan shape showing my arm angle during jumping jacks felt weird, because the angle of my raised right* *arm was shown on the virtual person's left arm. If the exercise has an asymmetrical movement, this kind of feedback actually messes with my intuition"*. Furthermore, P4 commented, *"Viewing visual feedback to align with standard movements made me feel more tired, especially when looking at all those skeleton points in bicycle crunches; I just gave up. I'd rather not spend a lot of time and energy trying to figure out discrepancies in my movements during exercise"*. These findings suggest that we should try to minimize users' cognitive load when providing feedback, such as by reducing the number of visual elements in skeleton joints.

Insight 3: The current AR headset design negatively impacts the execution of specific exercise movements. Out of the 12 participants, 5 reported discomfort in their exercising with the AR headset. In addition to its weight and challenges in maintaining its position during physical activities, the HoloLens incorporates multiple hardware modules in the rear head area. This design aspect causes difficulties when carrying out exercises requiring a supine position (e.g., bicycle crunches and ab holds). As a result, there is a substantial increase in the risk of hardware damage, prompting participants to alter their posture to prevent collisions with the ground while performing these workouts.

8 DISCUSSION

In this section, we present the implications that we have discovered through our research on visual cue design and AR-enhanced physical training.

8.1 In-situ Visual Cue vs. Co-located Visual Cue

First of all, visual cues are essential for guiding movements during self-training. Our experiments were designed to explore effective and convenient ways to display these visual cues. It is important to note that the most effective way to guild users' movements is through in-situ visual cues that fully match the user's body skeleton and display it from a first-person perspective. Nevertheless, we also noticed that presenting visual cues in either a first-person (in-situ) or third-person (co-located) perspective can yield specific benefits and drawbacks in an immersive environment.

In-situ visual cues presented in first-person perspective are the simplest and most direct way to reproduce or guide movement, as seen in successful applications in basketball [37] and skiing [61]. These cues present 3D trajectories within the user's field of view, allowing for quick understanding and adjustment of shooting angles or skiing directions. Additionally, the in-situ visual cue has a natural cognitive advantage in presenting 3D directions, as users only need to understand the meaning behind the 3D trajectory to follow its guidance. In comparison, when presented in a co-located way, users have to transfer the presented information to their own perspective, potentially leading to discordance or additional cognitive costs. In our study, we observed a mismatch in feedback when users were facing a virtual human doing jumping jacks (e.g., the real-time position of their right arm was drawn on the virtual human's left arm). While turning around the virtual human can be a solution, it would result in the virtual human's back facing the user, potentially causing issues with obstructed motion information.

However, sometimes the limited viewing angle of head-mounted devices can restrict the availability of in-situ visual cues from a first-person perspective. For instance, in our investigation of at-home workout scenarios, the in-situ visual cues were attached to the user's legs and feet during the "mountain climber" exercise. This means that, while maintaining the posture of a "mountain climber," the user may not be able to directly see the visual cues due to the user's view direction and the HoloLens' inherent viewing angle limitation of 52 degrees diagonal. This shows the challenge of applying visual cues in first-person perspective, as their effectiveness is dependent on movements occurring within the user's field of view or the ease with which these visual cues can be easily perceived. Given the physical limitations, co-located visual cues displayed around a virtual human and a third-person viewing approach would be the optimal choices in this scenario. Additionally, our experiments showed that users could partially alleviate the visual cue mismatch by becoming more skilled and establishing simple mapping relationships.

8.2 Design Implications

Based on our experimental data and findings, we have derived some design implications to inspire future AR-enhanced home workout scenarios. Firstly, designing visual cues around hit rates may be suitable for AR-enhanced workouts. During the study of the comprehension phase, user ratings indicated a less favorable response to directional cues, suggesting a preference for assessing exercise quality based on hit rate rather than movement direction/trajectory. Secondly, there are disparities in body forms between users and the virtual human. To improve user comprehension of workouts and provide a more personalized experience, tailoring the virtual human's shape to match the user's body shape and refining the visual cue association is essential. Lastly, considering that users have diverse preferences for visual cues, especially visual metaphors, it is essential to utilize a universal data format. This approach will provide users with a broader selection of options to choose from.

8.3 Limitations

8.3.1 Latency. Throughout the experiments, we observed an inherent delay between the user's actual action and the receipt of visual feedback. Analyzing the recorded video, we can estimate the delay time, which is approximately 0.4 seconds. For isometric exercises, the latency does not have a significant effect due to less intensive movements. For isotonic exercises, however, users may experience an increased cognitive burden as they need to offset their movement rhythm with the virtual human. Further quantitative experiments are required to evaluate the influence of latency on motor learning outcomes and user acceptance. This may involve comparing users' perceptions of different levels of latency and validating its impact on movement accuracy across varying latency conditions.

Additionally, we attempted to trace the sources of these physical delays, including GPU computation, wireless network transmission, and AR rendering. To lessen this, upgrading to hardware with higher performance, like advanced graphics processing units and low-latency network devices, can be a plausible solution.

8.3.2 Physical Limitations of AR Headsets. The HoloLens 2 in our study fulfills basic workflow requirements, but it also has some limitations that might influence user performance. First, its weight can put stress on the neck, demanding caution with exercises involving extensive shoulder-neck movement. Second, its size, such as the large battery pack at the back, may pose a collision risk with the ground, especially in supine positions. Lastly, its limited FOV constrains visual cue design, necessitating specific distancing for users to observe visual cues or the virtual human's full-body movements. Hence, future AR headsets that are lighter, more compact, and offer a larger FOV will be better suited for home exercise scenarios.

8.3.3 Design Considerations. Our visual cue design primarily focuses on presenting movement data, with limited discussion on AR-specific factors. For example, to maintain the visibility of virtual body movements within a restricted FOV, we kept it floating and smaller. However, the effectiveness of these design considerations, such as the placement and the scale of the virtual body, requires further experimental research and evidence. Moreover, investigating how to adjust visual cues according to changes in the physical environment (e.g., color variations, distance from the virtual entity) and incorporating diverse interaction modalities (e.g., touch, voice, gaze control) during exercises can provide valuable insights for designing AR fitness scenarios.

8.3.4 System Errors. We used a 3D pose tracking algorithm that is suitable for at-home scenarios while minimizing equipment requirements. It relies on a single camera instead of using multiple cameras, which can lead to less accurate motion capture performance. The errors might accumulate and affect the visual feedback provided by the visual cues. The performance of the algorithm also diminishes in complex situations like occlusion during exercises. Furthermore, we have not conducted a quantitative analysis of the participants' exercise quality as it depends on various factors, including physical strength and imitation ability. To explore methods for improving exercise quality through visual cues, a deeper understanding of kinesiology and additional engineering work would be necessary, which might go beyond the scope of this paper.

9 CONCLUSION

In conclusion, our study addressed the challenges associated with at-home workout videos, including limited movement comprehension and insufficient feedback, through the implementation of augmented reality (AR) technology. By utilizing a user-centered iterative design process and 3D pose tracking technology, we developed an AR-enhanced workout video application that incorporated visual cues to improve movement comprehension and enable real-time feedback. Our two user studies demonstrated the effectiveness of AR visual cues in improving exercise experience and provided design implications for future applications. In the future, we plan to allow the use of personalized feedback based on individual body characteristics and fitness levels to further enhance the effectiveness of AR visual cues.

ACKNOWLEDGMENTS

The work was supported by NSFC (62202424, U22A2032), and the Collaborative Innovation Center of Artificial Intelligence by MOE and Zhejiang Provincial Government (ZJU).

REFERENCES

[1] 2023. 8Fit. Retrieved January 22, 2023 from https://8fit.com/
[2] 2023. *Covid-19 has forced us to exercise indoors. But home fitness has been shaping our lives for decades.* Retrieved January 22, 2023 from https://www.bbc.com/worklife/article/20200504-covid-19-update-quarantine-home-workouts-during-coronavirus
[3] 2023. Forma Gym. Retrieved January 22, 2023 from https://formagym.com/isotonic-or-isometric-exercise-which-will-help-you-most/
[4] 2023. *Keep.* Retrieved January 22, 2023 from https://www.gotokeep.com/
[5] 2023. Kinovea. Retrieved January 22, 2023 from https://www.kinovea.org/
[6] 2023. MediaPipe. Retrieved January 22, 2023 from https://google.github.io/mediapipe/
[7] 2023. Mixamo. Retrieved January 22, 2023 from https://www.mixamo.com/
[8] 2023. MotionPro. Retrieved January 22, 2023 from https://www.motionprosoftware.com/index.htm
[9] 2023. MRTK. Retrieved January 22, 2023 from https://github.com/microsoft/MixedRealityToolkit-Unity
[10] 2023. MVN Link. Retrieved January 22, 2023 from https://www.movella.com/products/motion-capture/xsens-mvn-link
[11] 2023. OnForm. Retrieved January 22, 2023 from https://www.getonform.com/sports
[12] 2023. OpenPose. Retrieved January 22, 2023 from https://github.com/CMU-Perceptual-Computing-Lab/openpose
[13] 2023. *The pandemic's home-workout revolution may be here to stay.* Retrieved January 22, 2023 from https://www.washingtonpost.com/road-to-recovery/2021/01/07/home-fitness-boom/
[14] 2023. Youtube. Retrieved January 22, 2023 from https://www.youtube.com/
[15] Sarah Fdili Alaoui, Frederic Bevilacqua, and Christian Jacquemin. 2015. Interactive visuals as metaphors for dance movement qualities. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 5, 3 (2015), 1–24.
[16] Hashem A Almusawi, Christopher M Durugbo, and Afaf M Bugawa. 2021. Innovation in physical education: Teachers' perspectives on readiness for wearable technology integration. *Computers & Education* 167 (2021), 104185.
[17] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: enhancing movement training with an augmented reality mirror. In *Proceedings of the ACM symposium on User Interface Software and Technology.* 311–320.
[18] Alberto Cannavò, Filippo Gabriele Pratticò, Giuseppe Ministeri, and Fabrizio Lamberti. 2018. A movement analysis system based on immersive virtual reality and wearable technology for sport training. In *Proceedings of the International Conference on Virtual Reality.* 26–31.
[19] Jacky CP Chan, Howard Leung, Jeff KT Tang, and Taku Komura. 2010. A virtual reality dance training system using motion capture technology. *IEEE Transactions on Learning Technologies* 4, 2 (2010), 187–195.
[20] Hua-Tsung Chen, Yu-Zhen He, and Chun-Chieh Hsu. 2018. Computer-assisted yoga training system. *Multimedia Tools and Applications* 77 (2018), 23969–23991.
[21] Philo Tan Chua, Rebecca Crivella, Bo Daly, Ning Hu, Russ Schaaf, David Ventura, Todd Camill, Jessica Hodgins, and Randy Pausch. 2003. Training for physical tasks in virtual environments: Tai Chi. In *Proceedings of the International Conference on Virtual Reality (VR).* 87–94.
[22] Christopher Clarke, Doga Cavdir, Patrick Chiu, Laurent Denoue, and Don Kimber. 2020. Reactive video: adaptive video playback based on user motion for supporting physical activity. In *Proceedings of the ACM Symposium on User Interface Software and Technology.* 196–208.
[23] Caleb Conner and Gene Michael Poor. 2016. Correcting exercise form using body tracking. In *Proceedings of the SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems.* 3028–3034.
[24] Dazhen Deng, Weiwei Cui, Xiyu Meng, Mengye Xu, Yu Liao, Haidong Zhang, and Yingcai Wu. 2022. Revisiting the Design Patterns of Composite Visualizations. *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–16. https://doi.org/10.1109/TVCG.2022.3213565
[25] Dazhen Deng, Aoyu Wu, Haotian Li, Ji Lan, Yong Wang, Huamin Qu, and Yingcai Wu. 2022. KB4VA: A Knowledge Base of Visualization Designs for Visual Analytics. arXiv:2211.02567 [cs.HC]
[26] Dazhen Deng, Jiang Wu, Jiachen Wang, Yihong Wu, Xiao Xie, Zheng Zhou, Hui Zhang, Xiaolong (Luke) Zhang, and Yingcai Wu. 2021. EventAnchor: Reducing Human Interactions in Event Annotation of Racket Sports Videos. In *Proceedings of CHI Conference on Human Factors in Computing Systems.* 1–13.
[27] Bhat Dittakavi, Divyagna Bavikadi, Sai Vikas Desai, Soumi Chakraborty, Nishant Reddy, Vineeth N Balasubramanian, Bharathi Callepalli, and Ayon Sharma. 2022. Pose Tutor: An explainable system for pose correction in the wild. In *Proceedings of the CVF Conference on Computer Vision and Pattern Recognition.* 3540–3549.
[28] Mihai Fieraru, Mihai Zanfir, Silviu Cristian Pirlea, Vlad Olaru, and Cristian Sminchisescu. 2021. AIFit: Automatic 3d human-interpretable feedback models for fitness training. In *Proceedings of the CVF Conference on Computer Vision and Pattern Recognition.* 9919–9928.

[29] David Hamilton, Jim McKechnie, Edward Edgerton, and Claire Wilson. 2021. Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education* 8, 1 (2021), 1–32.
[30] Felix Hülsmann, Jan Philip Göpfert, Barbara Hammer, Stefan Kopp, and Mario Botsch. 2018. Classification of motor errors to provide real-time feedback for sports coaching in virtual reality—A case study in squats and Tai Chi pushes. *Computers & Graphics* 76 (2018), 47–59.
[31] Eakta Jain, Yaser Sheikh, Moshe Mahler, and Jessica Hodgins. 2010. Augmenting Hand Animation with Three-dimensional Secondary Motion. In *Symposium on Computer Animation*. 93–102.
[32] Raine Kajastila, Leo Holsti, and Perttu Hämäläinen. 2016. The augmented climbing wall: High-exertion proximity interaction on a wall-sized interactive surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 758–769.
[33] Taihei Kojima, Atsushi Hiyama, Takahiro Miura, and Michitaka Hirose. 2014. Training archived physical skill through immersive virtual environment. In *Human Interface and the Management of Information*. 51–58.
[34] Mohammad Amin Kuhail, Areej ElSayary, Shahbano Farooq, and Ahlam Alghamdi. 2022. Exploring Immersive Learning Experiences: A Survey. In *Informatics*, Vol. 9. 75.
[35] Matthew Kyan, Guoyu Sun, Haiyan Li, Ling Zhong, Paisarn Muneesawang, Nan Dong, Bruce Elder, and Ling Guan. 2015. An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 2 (2015), 1–37.
[36] Qing Lei, Ji-Xiang Du, Hong-Bo Zhang, Shuang Ye, and Duan-Sheng Chen. 2019. A survey of vision-based human action evaluation methods. *Sensors* 19, 19 (2019), 4129.
[37] Tica Lin, Rishi Singh, Yalong Yang, Carolina Nobre, Johanna Beyer, Maurice A Smith, and Hanspeter Pfister. 2021. Towards an understanding of situated ar visualization for basketball free-throw training. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1–13.
[38] Tica Lin, Yalong Yang, Johanna Beyer, and Hanspeter Pfister. 2020. SportsXR–Immersive Analytics in Sports. *arXiv preprint arXiv:2004.08010* (2020).
[39] Jingyuan Liu, Nazmus Saquib, Zhutian Chen, Rubaiat Habib Kazi, Li-Yi Wei, Hongbo Fu, and Chiew-Lan Tai. 2022. PoseCoach: A Customizable Analysis and Visualization System for Video-based Running Coaching. *IEEE Transactions on Visualization and Computer Graphics* (2022), To appear.
[40] Laura Marchal-Crespo, Mark van Raai, Georg Rauter, Peter Wolf, and Robert Riener. 2013. The effect of haptic guidance and visual feedback on learning a complex tennis task. *Experimental Brain Research* 231 (2013), 277–291.
[41] Takashi Matsumoto, Erwin Wu, and Hideki Koike. 2022. Skiing, Fast and Slow: Evaluation of Time Distortion for VR Ski Training. In *Augmented Humans*. 142–151.
[42] Stefan C Michalski, Ancret Szpak, and Tobias Loetscher. 2019. Using virtual environments to improve real-world motor skills in sports: a systematic review. *Frontiers in psychology* 10 (2019), 2159.
[43] Moritz Mödinger, Alexander Woll, and Ingo Wagner. 2022. Video-based visual feedback to enhance motor learning in physical education—a systematic review. *German Journal of Exercise and Sport Research* 52, 3 (2022), 447–460.
[44] Tamara Munzner. 2014. *Visualization analysis and design*.
[45] Heather L O'Brien and Elaine G Toms. 2010. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (2010), 50–69.
[46] Katerina El Raheb, Marina Stergiou, Akrivi Katifori, and Yannis Ioannidis. 2019. Dance interactive learning systems: A study on interaction workflow and teaching approaches. *ACM Computing Surveys (CSUR)* 52, 3 (2019), 1–37.
[47] Katerina El Raheb, George Tsampounaris, Akrivi Katifori, and Yannis Ioannidis. 2018. Choreomorphy: A whole-body interaction experience for dance improvisation and visual experimentation. In *Proceedings of the International Conference on Advanced Visual Interfaces*. 1–9.
[48] Alessandra Semeraro and Laia Turmo Vidal. 2022. Visualizing instructions for physical training: Exploring visual cues to support movement learning from instructional videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1–16.
[49] Cezary Sieluzycki, Patryk Kaczmarczyk, Janusz Sobecki, Kazimierz Witkowski, Jarosław Maśliński, and Wojciech Cieśliński. 2016. Microsoft Kinect as a tool to support training in professional sports: augmented reality application to Tachi-Waza techniques in judo. In *Third European Network Intelligence Conference (ENIC)*. 153–158.
[50] Pooya Soltani and Antoine HP Morice. 2020. Augmented reality tools for sports education and training. *Computers & Education* 155 (2020), 103923.
[51] Robert Spence. 2001. *Information visualization*. Vol. 1.
[52] Marina Stergiou, Katerina El Raheb, and Yannis Ioannidis. 2019. Imagery and metaphors: From movement practices to digital and immersive environments. In *Proceedings of the International Conference on Movement and Computing*. 1–8.
[53] Richard Tang, Xing-Dong Yang, Scott Bateman, Joaquim Jorge, and Anthony Tang. 2015. Physio@Home: Exploring visual guidance and feedback techniques for physiotherapy exercises. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 4123–4132.
[54] Wan-Lun Tsai, Li-wen Su, Tsai-Yen Ko, Cheng-Ta Yang, and Min-Chun Hu. 2019. Improve the decision-making skill of basketball players by an action-aware VR training system. In *Proceedings of the International Conference on Virtual Reality and 3D User Interfaces (VR)*. 1193–1194.
[55] Georgios Tsampounaris, Katerina El Raheb, Vivi Katifori, and Yannis Ioannidis. 2016. Exploring visualizations in real-time motion capture for dance education. In *Proceedings of the Pan-Hellenic Conference on Informatics*. 1–6.
[56] Laia Turmo Vidal, Elena Márquez Segura, Christopher Boyer, and Annika Waern. 2019. Enlightened yoga: Designing an augmented class with wearable lights to support instruction. In *Proceedings of the Designing Interactive Systems Conference*. 1017–1031.
[57] Jarke J Van Wijk. 2005. The value of visualization. In *Proceedings of the IEEE International Conference on Visualization*. 79–86.
[58] Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2013. MotionMA: motion modelling and analysis by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1309–1318.
[59] Jianbo Wang, Kai Qiu, Houwen Peng, Jianlong Fu, and Jianke Zhu. 2019. AI coach: Deep human pose estimation and analysis for personalized athletic training assistance. In *Proceedings of the ACM International Conference on Multimedia*. 374–382.
[60] Paweł W Woźniak, Monika Zbytniewska, Francisco Kiss, and Jasmin Niess. 2021. Making Sense of Complex Running Metrics Using a Modified Running Shoe. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1–11.
[61] Erwin Wu, Takayuki Nozawa, Florian Perteneder, and Hideki Koike. 2020. VR alpine ski training augmentation using visual cues of leading skier. In *Proceedings of the CVF Conference on Computer Vision and Pattern Recognition Workshops*. 878–879.
[62] Erwin Wu, Mitski Piekenbrock, Takuto Nakumura, and Hideki Koike. 2021. SPinPong-virtual reality table tennis skill acquisition using visual, haptic and temporal cues. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2566–2576.
[63] Shuainan Ye, Zhutian Chen, Xiangtong Chu, Yifan Wang, Siwei Fu, Lejun Shen, Kun Zhou, and Yingcai Wu. 2020. ShuttleSpace: Exploring and analyzing movement trajectory in immersive visualization. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 860–869.
[64] Ziyi Zhao, Sena Kiciroglu, Hugues Vinzant, Yuan Cheng, Isinsu Katircioglu, Mathieu Salzmann, and Pascal Fua. 2022. 3D Pose Based Feedback For Physical Exercises. In *Proceedings of the Asian Conference on Computer Vision*. 1316–1332.
[65] Qiushi Zhou, Andrew Irlitti, Difeng Yu, Jorge Goncalves, and Eduardo Velloso. 2022. Movement guidance using a mixed reality mirror. In *Proceedings of the Designing Interactive Systems Conference*. 821–834.
[66] Liyuan Zou, Takatoshi Higuchi, Haruo Noma, Lopez-Gulliver Roberto, and Tadao Isaka. 2019. Evaluation of a virtual reality-based baseball batting training system using instantaneous bat swing information. In *Proceedings of the International Conference on Virtual Reality and 3D User Interfaces (VR)*. 1289–1290.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009