avaTTAR: Table Tennis Stroke Training with On-body and Detached Visualization in Augmented Reality

Dizhi Ma*
Elmore Family School of Electrical and Computer Engineering
Purdue University
West Lafayette, USA
ma742@purdue.edu

Xiyun Hu*
School of Mechanical Engineering
Purdue University
West Lafayette, USA
hu690@purdue.edu

Jingyu Shi
Elmore Family School of Electrical and Computer Engineering
Purdue University
West Lafayette, USA
shi537@purdue.edu

Mayank Patel
School of Mechanical Engineering
Purdue University
West Lafayette, USA
pate1421@purdue.edu

Rahul Jain
Elmore Family School of Electrical and Computer Engineering
Purdue University
West Lafayette, USA
jain348@purdue.edu

Ziyi Liu
School of Mechanical Engineering
Purdue University
West Lafayette, USA
liu1362@purdue.edu

Zhengzhe Zhu
Elmore Family School of Electrical and Computer Engineering
Purdue University
West Lafayette, USA
zhu714@purdue.edu

Karthik Ramani
School of Mechanical Engineering
Purdue University
West Lafayette, USA
ramani@purdue.edu

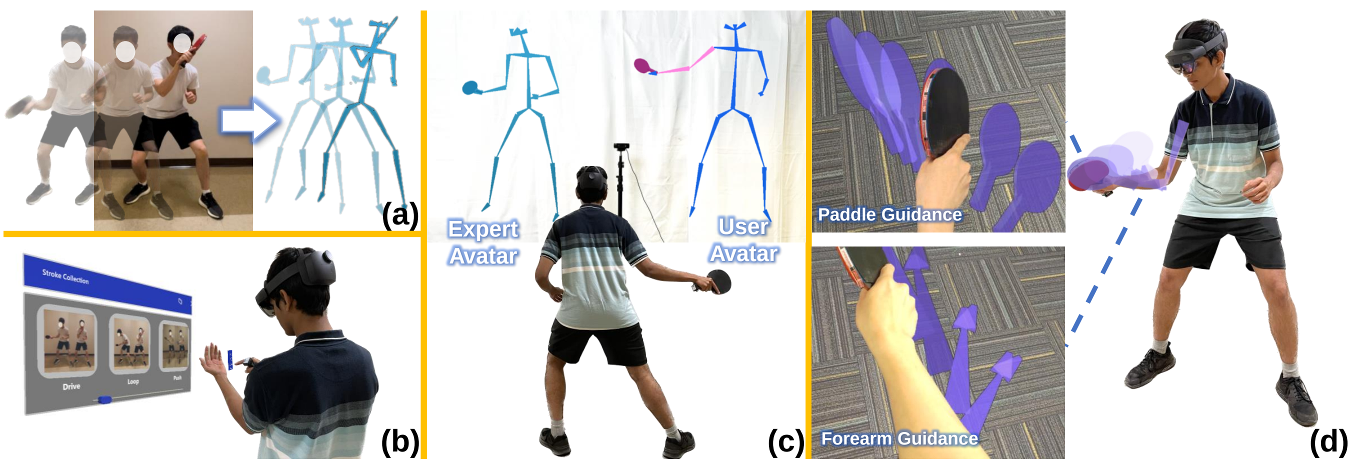

Figure 1: Overview of *avaTTAR*: An augmented reality system designed to assist table tennis stroke skills learning through *on-body* and *detached* visual cues. The expert records their body and paddle movement strokes (a) and saves them in our system. The user selects a specific table tennis stroke (b) during practice, the system provides a separate display of an Expert Avatar (in green) demonstrating the selected stroke alongside the User Avatar (in blue), accurately reproducing the user's real-time posture, and highlighting the error joints (in pink), (c) concurrently, it closely assesses the user's movements, offering on-body guidance (in purple) for (d) optimizing both body and paddle positioning.

*Both authors contributed equally to this research.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0628-8/24/10
https://doi.org/10.1145/3654777.3676400

ABSTRACT

Table tennis stroke training is a critical aspect of player development. We designed a new augmented reality (AR) system, avaTTAR, for table tennis stroke training. The system provides both "***on-body***" (first-person view) and "***detached***" (third-person view) visual cues, enabling users to visualize target strokes and correct their attempts effectively with this dual perspectives setup. By employing a combination of pose estimation algorithms and IMU sensors, avaTTAR captures and reconstructs the 3D body pose and paddle orientation of users during practice, allowing real-time comparison with expert strokes. Through a user study, we affirm avaTTAR 's capacity to amplify player experience and training results.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**.

KEYWORDS

Augmented Reality, Motor Learning, Table Tennis

ACM Reference Format:

Dizhi Ma, Xiyun Hu, Jingyu Shi, Mayank Patel, Rahul Jain, Ziyi Liu, Zhengzhe Zhu, and Karthik Ramani. 2024. avaTTAR: Table Tennis Stroke Training with On-body and Detached Visualization in Augmented Reality. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3654777.3676400

1 INTRODUCTION

Stroke training can be considered the most fundamental for all players among the wide range of table tennis techniques [13, 27, 46, 65]), as it forms the basis for executing various shots effectively. Traditional table tennis stroke training involves coaches who demonstrate specific techniques, players who imitate the strokes, and receive feedback and corrections from the coach [47, 55, 72]. Alternatively, players often try to learn strokes by watching online video clips and mimicking the techniques demonstrated. Some players record their own footage and review it for self-improvement. Recently, advancements in commercial Virtual Reality (VR) and Augmented Reality (AR) equipment have opened up new opportunities to practice table tennis skills in immersive AR/VR games [39, 67]. Regardless of which methods to learn strokes, players often struggle with perceiving the correct motion trajectories spatially, either from an expert's demonstration or their own attempts, leading to a vague understanding of the correct movements and how to rectify errors.

In an attempt to help the user understand their motion and feedback on errors, prior research has explored techniques for analyzing video data and displaying the visualization on screen. Researchers have compared user performance with that of experts and provided corresponding feedback using video clips for general motor learning [8, 10, 15, 22] and sports [44, 69] with computer vision algorithms. Such methods also empowered AR and VR-based sports training systems, which provide more immersive and real-time visual experiences that traditional methods cannot replicate. Noteworthy examples span a spectrum of sports and physical practices, including basketball [43], dancing [1], and beyond [38, 53, 73]. These works typically present visual cues in a third-person view, "detached" from the user. However, mirroring skills through detached cues may not always be intuitive compared to first-person view [74]. Researchers have addressed this challenge from a first-person perspective in VR [26], aiming to increase learner posture accuracy by rendering virtual avatars of both the learner and the expert together.

To this end, we introduce avaTTAR, an AR system designed for table tennis stroke movement learning. Inspired by our preliminary interview with 11 table tennis players, avaTTAR aims to provide a dual-view perspective visualization mechanism by combining on-body visualizations with detached ones, facilitating a better understanding of both the expert's and the user's execution. The "***on-body***" visuals overlay virtual cues directly onto the user's physical body, enabling first-person view guidance in AR. The "***detached***" visuals separate cues from the user, facilitating third-person observation of movements. By showing the user and expert execution with avatars and providing ***on-body*** and ***detached*** visualization, avaTTAR could be used in common table tennis training sessions such as shadow practice [2, 16] and multi-ball practice [21].

Our system can capture the motion of both experts and users by employing a 3D pose estimation algorithm [49] and an Inertial Measurement Unit (IMU) sensor to track the paddle orientation. The system provides recording software to record experts' strokes using this visual-sensor solution, storing them in a database. Meanwhile, the system also provides an AR sub-system for training, when using the AR system for training, users can then choose specific strokes to practice, with two detached avatars displayed during training: one representing the expert's recorded demonstration and the other reflecting the user's real-time execution. Users could change the viewpoint of the detached avatars in whatever way they prefer. The system continuously compares the user's motion with the expert's, offering detached feedback that highlights incorrect joints or the paddle and corresponding on-body trajectory guidance to help correct their strokes.

In summary, we contribute:

- A design rationale for table tennis stroke training, informed by formative interviews with experienced table tennis players.
- A table tennis stroke training system in AR, incorporates a dual-view perspective visualization mechanism, featuring on-body and detached visual cues, enabling an intuitive experience of table tennis skills acquisition.
- A user study showcasing the educational effectiveness of the system over a traditional video-based learning method, together with a separate study evaluating the usability of the system.

2 RELATED WORK

2.1 Sports Data Visualization

With the advent of online video-sharing platforms, social media, and portable devices equipped with cameras, motion analysis methods for sports have become increasingly prominent due to their accessibility and convenience. In team sports, the tactical dimension has also been addressed through visual analytics systems that unveil winning strategies [62, 63]. The evolution of deep convolution networks and algorithms in computer vision has dramatically expanded the potential for analyzing sports data within videos and delivering feedback, encompassing the tracking of objects like balls and subjects like players. Video-based visual analytics systems have been used to enhance game videos [9] and to annotate videos [12]. These techniques also facilitate posture comparison among different players. Guo *et al.* presented DanceVis [22], which employs a deep learning pose estimation model to analyze videos, thus enhancing analysis efficiency and automatically yielding individual global performance curves. Similarly, Wang *et al.* [69] developed a coaching system for skiing. This system tracks pose trajectories and identifies anomalous poses to provide corresponding "good examples."

These prior works primarily rely on 2D pose estimation. While 2D pose estimation is effective in many cases, it may lead to limitations [45, 56, 70] in accurately capturing the full-depth and three-dimensional aspects of complex movements. By generating a comprehensive dataset featuring 3D human poses for fitness training, Fieraru *et al.* [15] introduced AIFit. This system leverages the dataset for training purposes, offering valuable posture comparison natural language feedback base on the 3D body pose analysis. Liu *et al.* [44] introduced PoseCoach, which offers customized visualization and feedback based on the specified pose attributes from running videos. PoseCoach also relies on 3D pose analysis. However, this customization lacks real-time interaction and immediate feedback, restricting skill refinement. Both AIFit and PoseCoach utilize vision-based methods to acquire 3D information. Motion sensing devices could also be employed to capture 3D body poses. Clarke's ReactiveVideo [10] aligns experts' poses from a Microsoft Kinect device with novices' poses in videos, adjusting playback based on user proficiency.

Recognizing existing gaps, such as the absence of prompt feedback in current methods and the increasing inclination towards 3D pose analysis, we are motivated to build an AR system that presents a more intuitive and effective solution for sports motor skill training, especially in the context of table tennis.

2.2 XR-based Sports Training

Using XR for sports training taps into the power of interactive visuals and simulated environments, offering athletes opportunities to learn, practice, and refine their skills. In non-traditional sports, XR-based training systems have shown incredible potential. Nozawa *et al.* [53] innovated an indoor VR ski training system that showcases the movements of professional players while facilitating comparisons between users and these seasoned athletes. Kajastila *et al.* [38] devised an augmented climbing wall, incorporating it with three interactive applications. Ikeda *et al.* [30] project users' postures as virtual shadows on the ground, providing the corresponding feedback.

In the area of general physical training and conventional sports, Han *et al.* [24] developed a VR system for self-practicing Tai Chi Chuan, employing an optical see-through head-mounted display (HMD). This setup allows users to be surrounded by multiple virtual coaches, with the added flexibility to adjust perspectives. Prior research suggests that while demonstrating the coach's movements is essential for skill transfer to the trainee, displaying the trainee's own movements is equally crucial in enhancing their self-awareness. Chan *et al.* [7] utilized a wearable sensor-based motion capture suit to document users' movements and subsequently rendered them within a virtual environment. Analogously, Takahashi *et al.* [64] employed wearable sensors to capture and reconstruct three-dimensional body poses, then trained baseball batters. However, these wearable sensor approaches, while potentially yielding precise estimations, inevitably introduce inconveniences and complexities since the wearable suits and sensors can be cumbersome or uncomfortable to wear, potentially affecting the natural movement of the user. Additionally, RGBD sensors, such as Microsoft Kinect, were used by Anderson *et al.* [1] in their YouMove system. This system empowers experts to record sequences of physical movements, and novices to practice and learn them using an AR mirror. In a similar process, Zhou *et al.* [77] delved into the effectiveness of movement acquisition through a mixed reality mirror.

These prior studies displayed feedback adjacent to or encircling the user, adopting a third-person perspective. While this perspective can offer valuable insights, a first-person viewpoint (FPV) has the potential for more intuitive visualization [23, 26, 43, 74]. In recent research on basketball, an in-situ display approach [43] was explored, in which basketball trajectories were visualized through an AR head-mounted display, offering trajectory visualization directly within the user's field of view. For motor skill learning, Han *et al.* [23] investigated FPV for arm movement learning, and [26, 74] utilized FPV for movement learning in Virtual Reality (VR), enabling users to observe both their own virtual skeleton and the teacher's virtual skeleton. In this paper, we introduce an embodied approach in AR, in which the virtual body is superimposed on the physical one, allowing the user to actively conform to the template.

Our objective is to offer users a comprehensive visualization experience by juxtaposing both ***detached*** expert and user avatars while incorporating ***on-body*** cues for movement correction.

2.3 Training Systems for Table Tennis

In table tennis, computer-aided methods have advanced to make gameplay analysis easier. These methods include an annotation tool [12] that integrates with computer vision algorithms, thus enhancing the analysis of the dynamics of table tennis. Additionally, a data visualization tool [9] has emerged to provide insight into game video analysis from the comfort of one's home. Another distinct example is Wang's work [68], which adapted the Internet of Things (IoT) devices to track arm and paddle movements. The resulting training system interprets technical attributes and the corresponding indicator values, presenting these data through an intuitive software interface.

In VR applications, earlier work like that of Brunnett *et al.* [6] built immersive table tennis simulation systems. However, their focus leaned towards constructing realistic virtual environments rather than optimizing study efficiency. Subsequent research [50, 54, 71] has shed light on the effectiveness of training in VR landscapes. Michalski *et al.* [50] conducted tests with a VR game, while Wu *et al.* [71] focused on providing multimodal cues to improve the successful return rate. Oagaz *et al.* [54] indicated players' incorrect postures by highlighting the joints of the user's reconstructed skeleton in the virtual environment. These collective efforts paint a multifaceted picture of training possibilities within VR-based table tennis interventions. However, current VR simulations suffer limitations [39] including the absence of genuine physical and precise movement feedback, even in the latest table tennis VR games [67]. Shifting the focus to augmented reality initiatives, the origins trace back to [35], which introduced a reactive AR table. Subsequent work, including those by Mayer [48], also embraces AR tables, offering enhanced visualization. These advances have excelled in effectively showing the results of various techniques, including ball trajectory, ball placement, and return rate. However, it is important to highlight that these often overlooked the crucial elements of motor skill learning, specifically focusing on the player's body and paddle, and lacked essential guidance.

In this work, our main objective is to assist the stroke training of table tennis by integrating motion guidance through virtual avatars in AR.

3 DESIGN RATIONALE

We conducted interviews with expert players and table tennis coaches (Section 3.1), to understand their training experience, challenges and pain points, training methods, and coaching experience (if any). We use these data and the results of these interviews to derive the design requirements for our system (Section 3.2). A summary of the insights, requirements, and subsequent components is shown in Figure 2.

3.1 Formative Interview

Our methodology to obtain the insights began with formative interviews involving a diverse group of 11 table tennis players. These interviews were divided into several sections, each focusing on critical aspects of their table tennis experience.

We conducted in-person interviews with each interviewee individually. The interviews were semi-structured with open-ended questions. The entire interview procedure lasted between 15 to 30 minutes, depending on the interviewees. We recorded the audio of all conversations and converted them into text for analysis.

The demographics of the interviewees are as follows: the average age was 38.6 years (SD = 19.2), and their cumulative experience in table tennis averaged 20.0 years (SD = 14.7). Of the participants, seven had prior experience practicing under the guidance of a coach, with an average coaching duration of 4.7 years (SD = 4.5). Only two participants were familiar with AR, and none had used AR for table tennis practice.

3.1.1 Interview Topics. During the formative interviews, we explored several key topics, including:

- Table Tennis Training Experience. Participants were asked to share their personal experiences with table tennis stroke training, including any formal coaching or self-guided practice.
- Challenges and Pain Points. We inquired about the specific challenges and difficulties they encountered during stroke training, seeking to identify common pain points.
- Coaching and Mentorship. For those with coaching or mentoring experience, we delved into their coaching methodologies and the impact of coaching on trainees' performance.

3.1.2 Findings and Insights. In our conversations with the interviewees, we gained valuable insights. These insights helped us understand the details of stroke training and the difficulties that players encounter when training alone.

I1: Learning Stroke with Execution Cycle. One of the fundamental insights that emerged from our interviews relates to the essence of stroke training in table tennis. Most strokes can be deconstructed into three distinct stages: the "prepare - back swing - fore swing - recovery" cycle [13, 65]. The key to executing a correct stroke lies in mastering the precise positions and transitions within these stages and translating them into a smooth motion. Our interviewees emphasized that the path to improvement involves rigorous repetition to form muscle memory.

I2: Lack of Comparison when Practicing without Coach. A significant portion of table tennis training occurs without the presence of a coach. Although self-practicing is essential due to the sheer volume of repetition required to master the sport, it comes with its own set of challenges. Specifically, without a coach's guidance, it becomes difficult to discern when a stroke is being executed incorrectly. Even more concerning, continued practice without correction can lead to the formation of muscle memory for incorrect motions, making subsequent corrections more challenging.

I3: Missing Effective Input for Stroke Correction. Coaching and mentorship were highlighted as invaluable resources for stroke training. Coaches can provide multifaceted assistance, including: ***Demonstration***: Coaches can effectively demonstrate the correct movement and stroke execution, offering a visual reference for players. ***Feedback***: Coaches have the unique capacity to provide immediate, tailored, and constructive feedback during practice sessions. However, training with a coach may still present challenges in transferring skills, as bridging the gap from observing third-person movements to adopting an egocentric perspective can be complex.

I4: Requiring Learner-centric education. During training, individuals have diverse preferences for the content they wish to observe. Different strokes may require distinct viewing angles, and the specific focus of observation varies from person to person. For example, some may prioritize analyzing arm and footwork, while others may concentrate on monitoring paddle positioning and rotation.

In the following section, we outline the specific features and functions of our system that directly address the identified needs and challenges voiced by the interview participants.

3.2 Design Requirements

From the initial interviews with the potential user with table tennis training and coach experience, as well as the limitations of existing works, we identify the following design requirements:

3.2.1 R1: Enable body-paddle motion reconstruction. Based on the insight from **I1**, which emphasizes the significance of mastering precise stroke stages, our first design requirement, is aimed at facilitating stroke training. It comprises two essential components:

- C1: Reconstruction of body movements to assess posture and positioning during strokes. According to the interviewee's feedback: *"... I would like to see the movement for the **body and the paddle** ..."*

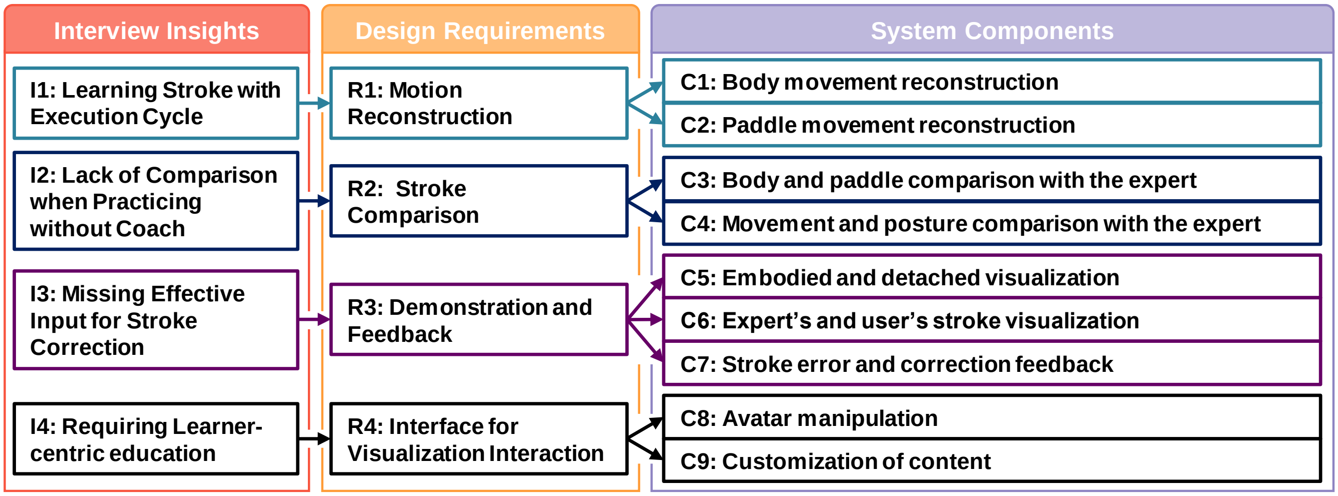

Figure 2: Illustration of design requirements and system components. We have extracted four key design requirements from our formative interviews with individuals experienced in table tennis training. In response, we have elaborated on these requirements, resulting in the development of nine detailed modules for implementation.

- C2: Thorough reconstruction of paddle movements to scrutinize the subtleties of stroke execution.

We followed YouMove [1] and used the skeleton for explicit visualization [75] of the body since YouMove also includes full-body movement learning.

3.2.2 R2: Integrate stroke comparison. Derived from insights **I2**, which highlights the challenges of practicing without a coach, our second design requirement, introduces a core feature to address these challenges. This requirement aligns with the need to bridge the gap between coaching and individual practice. We've identified two core components:

- C3: A side-by-side comparison of the user's body and paddle movements with those of an expert, allowing the user to observe the errors for improvement. As expressed by an interviewee: *"...So when you have a coach, I mean, a coach can constantly give you feedback ... someone can **tell you that you did this wrong right at that moment**"*
- C4: Comparative analysis of overall movement and important posture about established standards and best practices, helping users strive for excellence. In the words of an interviewee: *"...like suppose I play a stroke if it tells me exactly how I should correct it and **what my angle is, what the angle should be of an expert**"*

3.2.3 R3: Demonstration and feedback visualization. In response to insights gathered from **I3**, which emphasises the role of coaching and mentorship in stroke training, our third design requirement focuses on addressing the challenges faced when practicing without a coach. This multifaceted requirement is designed to add to awareness during movement execution and encompasses three key elements:

- C5: A two-view perspective, presenting information in both ***detached*** and ***on-body*** formats. This versatile approach allows users to receive demonstrations and feedback in various ways, catering to individual learning preferences. As articulated by an interviewee: *"... see what you are doing wrong from a **third point of view**, like a 3D recreation..."* and another one: *"... and I would like to see the coach's movement from **his view** as well..."*.
- C6: Visual cues that not only demonstrate the ideal execution of the target stroke, including paddle and full-body movement but also provide a simultaneous representation of the user's execution.
- C7: Visual cues not only to locate areas where the user made mistakes but also to offer clear guidance on how to refine their posture.

3.2.4 R4: Interface for visualization interaction. Incorporating insights from our interviews, particularly **I4**, which underscores the diverse perspectives individuals have during training, our fourth design requirement aims to address the challenges related to catering to varied visualization. This requirement underlines the need to create an intuitive interface that accommodates different viewing preferences, aligning with our user-centric approach. It comprises two pivotal modules:

- C8: Intuitive avatar manipulation, allowing users to interact with the visualized data, providing a sense of control over the analysis process. In the interviewee's own words: *"... So like, you can slow down time, **watch and play the stoke in slow motion** ..."*
- C9: Customization options that enable users to tailor the visualization to their specific needs, accommodating different skill levels and training objectives. Based on an interviewee's input: *"... And I would like to **change the viewpoint of the expert**..."*

4 AVATTAR SYSTEM WALK-THROUGH

We now elaborate on an example of using our system. Initially, an expert will use our authoring system to record various stroke movement videos and derive corresponding motion data. The expert could further trim and mark the keyframes for these videos, meanwhile, motion data are changed accordingly. The edited video and motion data will be saved in our AR system stroke database. When using our AR system, the system will first calibrate according to the user's height to adjust for the scale of the ***on-body*** and ***detached*** visualization cues. The user can then choose a desired stroke from the database to practice (as depicted in Figure 1 (b)). Subsequently, the system proceeds to visually depict the user's movements through a User Avatar (designated as UA). The UA accurately reflects the real-time posture, encompassing the entire body and the paddle's orientation (R1). Following this stage, the user selects the specific stroke they wish to learn, prompting the Expert's Avatar (referred to as EA) to appear correspondingly (as illustrated in Figure 1 (c)). Additionally, the user closely observes and emulates the movements of EA, thus acquiring proficiency in executing the stroke. Throughout the learning process, the system compares the user's movements and the ideal model represented by EA (R2). Any deviations or errors in the user's actions are highlighted on UA and the corresponding ***on-body*** guidance appears. This guidance is provided through visual representations, which include movement trajectories highlighting errors in body parts and paddle positioning. These are superimposed on the user's body based on the parameter of UA, aligning with the initially established reference points (R3, as shown in Figure 1 (d)). During training, users have the flexibility to manipulate the position, viewpoint, and scale of both UA and EA. Furthermore, they can adjust the playback speed of EA, allowing for a "slow-mo", detailed examination of the stroke. All of these interactions are seamlessly facilitated through our user-friendly interface (R4).

5 AVATTAR AUTHORING SYSTEM

We developed a simple interface similar to [1] as shown in Figure 4 for the expert to capture stroke movement and annotate key posture. The record mode (Figure 4 (a)) allows authors to record themselves performing the movement. The system captures the 3D body movement and paddle orientation data of the author. After starting the software in record mode, the expert is presented with a screen that has a "Start" button, a "Stop" button, as well as the current video stream from a webcam. To capture movement, the expert presses the "Start" button, performs the stroke, and then presses the "Stop" button. Note that we attached an IMU to the paddle to capture the orientation of the paddle. The edit mode (Figure 4 (b)) allows the expert to trim the recording to remove unwanted data with the "Start Frame" and "End Frame" buttons. The authors then specify the keyframes for the recorded movement with the "Keyframe" button. For our user study, keyframes are postures that reflect the key stages mentioned in Section 3, which helps the user understand the stroke in the big picture. The green rectangle marker on the video process bar reflects the start frame, the brown marker reflects the end frame, and the red triangle markers reflect the keyframes. The user could reset all markers with the "Reset" button.

6 AVATTAR LEARNING SYSTEM

The avaTTAR learning backend comprises several modules as shown in Figure 3 (b), these modules are: Motion Capture, Stroke Analysis, Cues Visualization, and AR Interface.

6.1 Motion Capture

Motion capture constitutes the foundational module of our system. The outcomes derived from our motion capture solution enable us to rebuild a 3D avatar for the player and furnish pose data for analysis. To track the player's movements, we developed an approach to gauge the 3D posture of both the player and the paddle (C1, C2). It comprises hardware for data acquisition and software for data processing.

6.1.1 Hardware Configuration. Our system relies on a webcam placed in front of the player for optimal data capture, providing an unobstructed view of movements and reactions (Figure 5). Additionally, we integrate an IMU on the player's paddle (similar to [68]), enabling real-time recording and analysis of paddle pose.

6.1.2 Software Setup. Using a real-time processing 3D pose estimation model, the webcam captures precise 3D joint locations, creating a detailed 3D avatar that mirrors the player's movements. Simultaneously, the IMU on the paddle provides pose data of the paddle.

6.2 Stroke Analysis

A fundamental requirement is to maintain a consistent posture for users, identical to that of the expert, throughout the stroke learning process. This facilitates a real-time comparison between learner data and expert data extracted from the ongoing frame. In this section, the two algorithms, one for frame-by-frame error comparison (C3) and the other for overall motion error comparison (C4) are presented.

6.2.1 Body. When studying body movements, we focus on joint angles, calculated from joint positions in the player's body. This data allows us to compare experts and users. We denote the user's joint sequence as $Q_1 \in \mathbf{R}^{N \times J \times 4}$, where N denotes the number of the sequence and J denotes the number of the joints, and 4 denotes the quaternions rotation. We consider all joints except the end joints, such as the head and toes.

To compare the expert and the user, we further align the user's joint angle series using the Dynamic Time Warping (DTW) algorithm [4, 51, 52] with the quaternions dissimilarity equation:

$$\text{quaternion\_dissimilarity}(q_1, q_2) = 1 - \langle q_1,\ q_2 \rangle$$

where q_1, q_2 represent two quaternions vectors. The details of the quaternion dissimilarity-based DTW are shown in Algorithm 1. Since the user may act slightly faster or slower than the EA, for real-time comparison, we used $N = M = 10$ to reduce the computation cost, otherwise, N and M refer to the length of the sequence. We introduce universal thresholds ξ_{joint} for all joints to identify incorrect movements. The user's movement is considered incorrect at joint k if:

$$D[N][M][k]/N < 1 - \xi_{\text{joint}}.$$

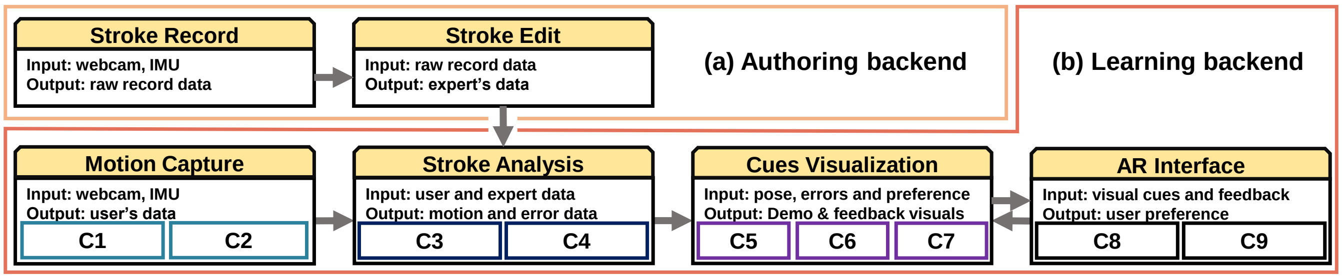

Figure 3: System's backend workflow. (a) The authoring backend: enables the expert to record and edit their movement posture data. (b) The learning backend: captures the user's motion and compares it with expert data, providing visual cues for stroke learning, and allows the user to adjust these visuals via the AR interface.

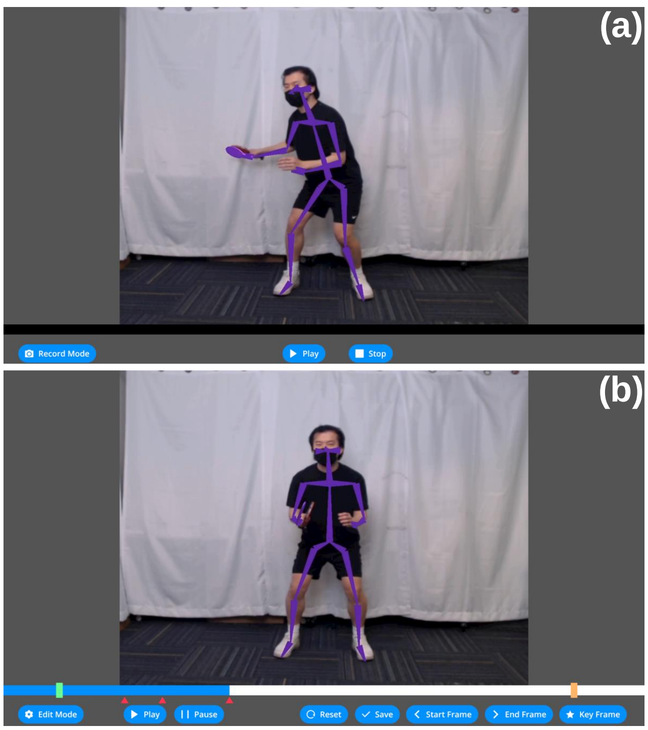

Figure 4: The authoring interface: (a) The record mode allows the user to record video and stroke. (b) The edit mode interface allows authors to trim video and specify keyframes.

6.2.2 Paddle. In contrast to the human body, we can directly use the quaternions received from the IMU to calculate the angle difference using DTW with a quaternion dissimilarity equation. Similarly, a threshold ξ_{paddle} is implemented to determine feedback intervals. The user's paddle movement is considered incorrect if:

$$D[N][M]/N < 1 - \xi_{\text{paddle}}.$$

6.3 Cues Visualization

As shown in Figure 6, we employ two distinct types (C5) of visual cues to enhance the user experience and provide demonstration

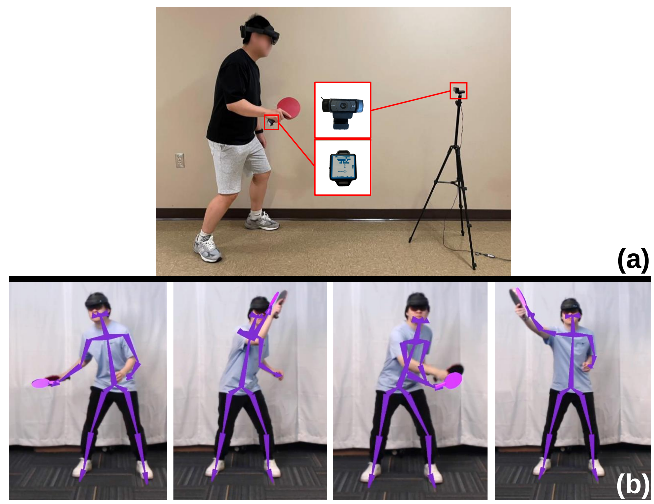

Figure 5: (a) Our motion capture module consists of one webcam placed in front of the user and one IMU attached to the table tennis paddle. The webcam provides the visual input for the system and the IMU provides the paddle orientation. (b) Examples of motion capture results with different body and paddle postures.

(C6) and feedback (C7) during execution: ***on-body*** cues and ***detached*** cues. Both cues are human skeleton-like avatars reconstructed based on the user's and experts' movements of both body and paddle that reflect the difference between them and guidance for the user.

6.3.1 Demonstration. In the demonstration phase of our system, our aim is to visually represent the stroke execution of both the expert and the user during training. The primary goal is not only to showcase the correct movement performed by the expert but also to create awareness for the user regarding their own motion execution.

- ***Detached* Cues**: For the desired movement, a ***detached*** expert avatar illustrates the ideal trajectory. Simultaneously, the user's avatar mirrors their real-time movements, providing a direct comparison.

Algorithm 1: Quaternion Dissimilarity-Based DTW for body

Input: Joint angle sequences: User's Q_1 with sequence length N and expert's Q_2 with sequence length M, number of joints J
Output: DTW distance matrix $D[N][M][J]$
$Q_1 \leftarrow \text{Kalman\_filter}(Q_1)$, $Q_2 \leftarrow \text{Kalman\_filter}(Q_2)$;
Initialize distance matrix D with dimensions $N \times M \times J$;
for $i = 0$ *to* N **do**
  for $j = 0$ *to* M **do**
    for $k = 0$ *to* J **do**
      $D[i][j][k] = \infty$;
for $i = 0$ *to* N **do**
  for $j = 0$ *to* M **do**
    $D[i][j][0] = 0$;
for $i = 1$ *to* N **do**
  for $j = 1$ *to* M **do**
    for $k = 1$ *to* J **do**
      Calculate quaternion dissimilarity $E = \text{quaternion\_dissimilarity}(Q_1[i][k], Q_2[j][k])$;
      $D[i][j][k] = E + \min(D[i-1][j][k], D[i][j-1][k], D[i-1][j-1][k])$;
Return $D[N][M][J]$;

Algorithm 2: Quaternion Dissimilarity-Based DTW for Paddle

Input: paddle quaternion sequences: User's Q_1 with sequence length N and expert's Q_2 with sequence length M
Output: DTW distance matrix $D[N][M]$
$Q_1 \leftarrow \text{Kalman\_filter}(Q_1)$, $Q_2 \leftarrow \text{Kalman\_filter}(Q_2)$;
Initialize distance matrix D with dimensions $N \times M$;
for $i = 0$ *to* N **do**
  for $j = 0$ *to* M **do**
    $D[i][j] = \infty$;
$D[0][0] = 0$ **for** $i = 1$ *to* N **do**
  for $j = 1$ *to* M **do**
    Calculate quaternion dissimilarity $E = \text{quaternion\_dissimilarity}(Q_1[i], Q_2[j])$;
    $D[i][j] = E + \min(D[i-1][j], D[i][j-1], D[i-1][j-1])$;
Return $D[N][M]$;

- ***On-body* Cues**: In addition, the user has the option to overlay an ***on-body*** expert avatar onto their physical body. Following the movements of this avatar can help reduce cognitive load, facilitating the transfer of the demonstration into the user's own physical actions.

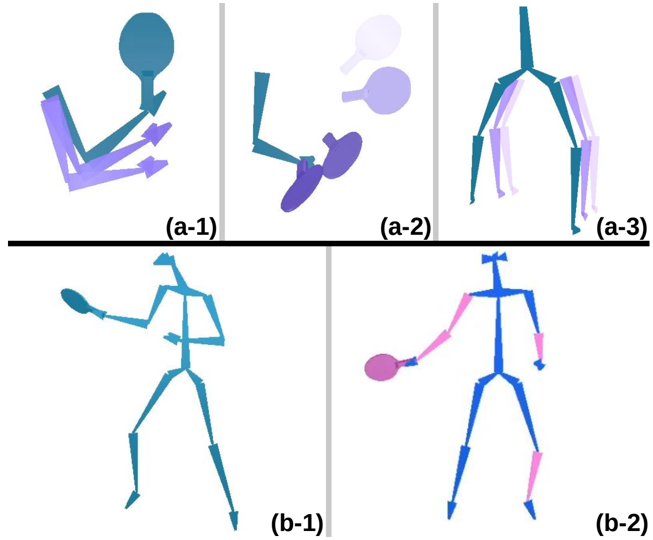

Figure 6: Showcase of our visualization. *On-body* cues: (a-1) movement trajectory; (a-2) paddle shadow shows paddle trajectory; (a-3) footwork trajectory. *Detached* cues: (b-1) expert avatar shows correct movement; (b-2) user avatar shows user's movement and highlights incorrect skeleton (in pink).

6.3.2 Feedback. In the learning phase, the system offers guidance based on the user's execution, employing visualizations to highlight errors and generate instructive cues. The primary objective is to visually indicate areas that require refinement in specific aspects of movement. Similarly to the demonstration phase, the feedback visualization is presented in both ***detached*** and ***on-body*** formats.

- ***Detached* Cues**: The backend of the system analyzes the user's movement in comparison to the expert and identifies the error joints on the ***detached*** user avatar. These highlighted areas offer clues to where improvements can be made.
- ***On-body* Cues**: Concurrently, ***on-body*** cues visually represent the correct trajectory of joint movement, providing information on how the user can enhance their execution. This method helps users understand what improvements can be made.

6.4 AR Interface

We've designed an AR interface that offers customizable visualization options (C9), as outlined earlier. To begin practicing with avaTTAR, users simply click the "stroke selection" button located in the menu's first row, as seen in Figure 7(a). This action opens the stroke collection menu shown in Figure 7(b), where users can start their table tennis training by choosing a pre-recorded stroke from the system database. Once a specific stroke is selected, users can observe and mimic the movements of the virtual expert avatar. The avaTTAR system will provide both ***detached*** and ***on-body***visuals based on the user preference.

For both the expert and user avatars, users have the capability to adjust the avatar's scale through gestures (C8). During the training, as depicted in Figure 7(a), users can:

- modify the stroke play of the expert avatar; users can pause, resume, increase, or decrease the speed of the training.
- choose the ***detached*** elements provided by the system, such as enabling or disabling expert avatar and user avatar.
- customizes the ***on-body*** feedback provided by the system, such as enabling or disabling body and paddle feedback.

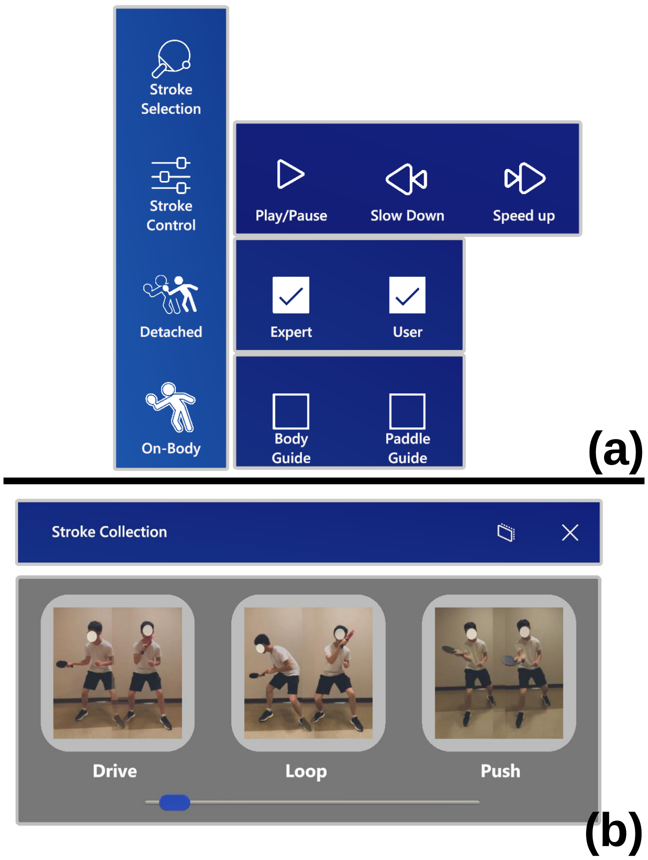

Figure 7: The AR interface of avaTTAR. The user could adjusting their visualization preferences (a) and initiate the training by selecting the desired stroke from the menu (b) and during the training session.

7 IMPLEMENTATION DETAILS

We implemented the avaTTAR system with a hardware setup consisting of a Microsoft Hololens 2 head-mounted display [33], a 1080p HD Logitech webcam, an IMU equipped with a high-precision 9-axis gyroscope and Bluetooth 5.0 connectivity, along with a local PC equipped with an NVIDIA 1080Ti graphics card. An expert in the formative interview process was invited to record the motion of three stroke movements and the key static poses of each of the strokes, which are used to compare with the user movement and generate the expert avatar movement and ***on-body*** cues. For the user's movement, the visual input captured by the webcam is processed on the local PC with the 3D pose estimation algorithm [49]. After receiving each frame from the camera, a request will be sent to the IMU (sample rate 250 Hz) to obtain the paddle pose. Motion analysis is then applied to compare the user's stroke with that of the expert. The entire motion capture, analysis process, and visual cue rendering (the process that includes the three blocks to the left in Figure 3) are completed in approximately 36 ms (around 28 FPS) for each frame of the ***detached*** user avatar and 17 ms (around 60 FPS) for the ***detached*** expert avatar and ***on-body*** cues which are faster because pose estimation is not needed. The resulting visualizations are then rendered and presented through the HMD, utilizing Holographic Remoting with a connection latency of around 55 ms, for an immersive user experience. The ***on-body*** cues are attached to the user based on their starting position, and the scales of the ***on-body*** cues are adapted based on the height of the user; meanwhile, ***detached*** cues are put 3 meters in front of the user, and the location and the scales of the ***detached*** cues are changeable during use, allowing flexibility in visual representation. The total latency of the system is around 91 ms, which is less than the visual reaction time of a table tennis player (around 260 ms [5, 28]). Both the authoring system and the learning system's user interface are constructed using Unity 3D, with some code components derived from [11]. We empirically set $\xi_{joint} = \xi_{paddle} = 0.1$ during system initialization.

8 USER STUDY

We conducted a two-session user study to evaluate user learning outcomes and usability of the system. 14 users (12 identified as male and 2 identified as female, 21 to 28 years old) were recruited. Of the 14 participants, 12 were familiar with VR applications on smartphones, tablets, or head-mounted devices, while the remaining two had prior exposure to both AR and VR technologies. 13 of the 14 participants did not have prior table tennis training experience, while one participant had received a moderate level of table tennis training before. The entire study lasted approximately 1.5 hours per participant, and each participant received compensation in the form of a $15 e-gift card. Before diving into the study, participants were asked to get acquainted with Hololens2's interaction modality through its built-in tutorial. There are two sessions in our study. After each session, users completed a 5-point Likert-type questionnaire (Strongly Disagree; Slightly Disagree; Neutral; Slightly Agree; Strongly Agree) about the training experience. At the end of the user study, each participant was interviewed and completed the standard System Usability Scale (SUS) questionnaire.

8.1 Procedure

8.1.1 Session 1: Movements Accuracy. In this session, we aimed to evaluate the improvement in movement accuracy when using avaTTAR compared to a baseline system. Session 1 revolved around two similar strokes, namely drive and loop, structured as a between-subjects and counterbalancing experiment. This means that each participant was exposed to only one of the systems to learn a specific stroke. This approach ensured that the training results were not biased by the order in which the systems were used. The two strokes were further decomposed into three key static poses of the stroke–preparation, back swing, and forward swing, as shown in Figure 8. The user was asked to learn the three static poses first and then the

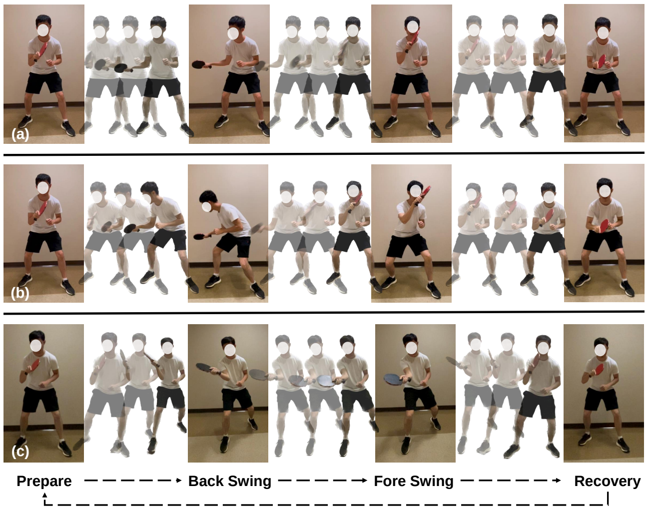

Figure 8: Forehand drive stroke (a), loop stroke (b), and push stroke (c) in the "prepare - back swing - fore swing-recovery" cycle (as mentioned in I1). We assess the entire movement and key postures, focusing on the prepare, back swing, and fore swing postures since the recovery posture matches the prepare posture.

whole stroke sequence. The total time spent with each system was the same. For the experiment:

- **Baseline System**: Participants observed an expert demonstration video. Subsequently, they attempted to adjust their posture and movement to mirror that of the expert. While doing this, they had access to a monitor displaying both their own video and the expert's, alongside detected skeletons. To facilitate learning, participants were allowed to adjust the camera angle and the monitor position by asking us. After a 5-minute practice session for each pose or sequence, their post-practice posture and movement were recorded. For each static posture, the demonstration time is 1 minute and the practice time is 3 minutes.
- **avaTTAR**: The participants began by observing the expert's demonstration video and the ***detached*** skeleton. This setup allowed them to walk around the skeleton, observing posture and movement from various angles. The learning phase followed, in which the participants initially watched the ***detached*** skeletons for 2 minutes, then the ***on-body*** skeletons for another 2 minutes. At the last minute, they were given the freedom to choose which skeleton to observe. After this learning phase, participants were asked to perform poses or sequence three times. We then recorded their skeletons for analysis. For each static posture, the demonstration time is 1 minute and the practice time is 1 minute for ***detached*** , 1 minute for ***on-body*** , and 1 minute for free observation.

8.1.2 Session 2: User Experience. Traditionally, learning a specific stroke in table tennis involves two primary steps: **Observation and Shadow Practice**: Trainees observe a coach executing the stroke and then practice it without a ball, mimicking the movements. **Multi-Ball Practice**: Trainees practice the stroke by returning multiple balls fed by a coach or a ball machine. These steps are repeated iteratively during the stroke training process.

In this session, our focus was on evaluating the user experience and overall usability of the avaTTAR system in the training cycle specifically for the first step, while utilizing a ball machine for multi-ball practice.

Our focus was on teaching the participants the "push" stroke with our platform. The experiment began with participants watching a video that briefly demonstrated the "push" stroke. After familiarizing themselves with the stroke through the video, they attempted to return balls launched from a ball machine. After an initial trial, participants engaged with the avaTTAR system. They observed and replicated the stroke using both ***detached*** and ***on-body*** cues, starting at a 50% play speed. The speed gradually increased to full speed, and the participants practiced over 50 repetitions, spanning approximately 4 minutes. Subsequently, an additional minute was allotted for participants to freely interact with and use the system according to their preferences.

After using our system, participants once again tried to return balls from the ball machine, then filled out the system usability questionnaire and were interviewed.

8.2 Results

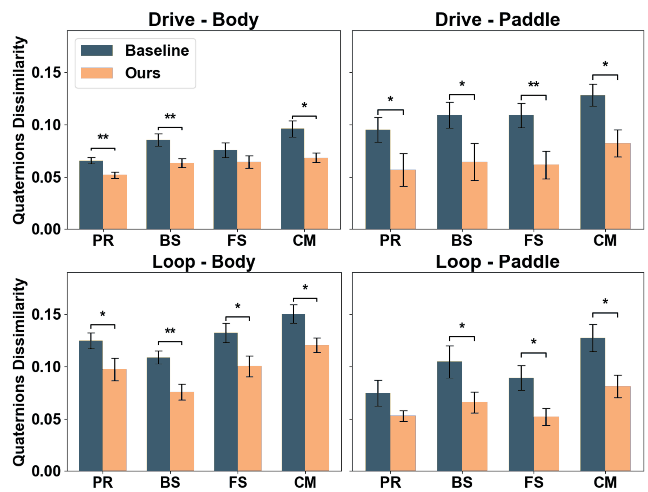

Figure 9: "prepare (PR) - back swing (BS) - fore swing (FS) - complete movement (CM)" error comparison, between the baseline approach and our method, "*" indicates p-value < 0.05, ** indicates p-value < 0.01.

8.2.1 Session 1 Results. The outcomes of this session are presented in Figure 9. Overall, user performance when performing static poses and movements was superior with avaTTAR compared to the baseline system. Specifically, for static poses, both body and paddle errors were reduced. This observation aligns with expectations, considering that a stroke sequence inherently incorporates multiple static poses. The average quaternion error for the baseline was 0.0987, while for avaTTAR , it was lower at 0.0754. After discovering that the collected data are not normally distributed by performing a *Shapiro-Wilk* test, a *Wilcoxon signed-rank* test revealed substantial differences in pose errors between the two systems. However, exceptions were noted in the drive stroke "FS" body error and the loop stroke "PR" paddle error. The results suggest that avaTTAR offers a clear advantage in facilitating users to accurately replicate static poses. The incorporation of embodied and detached avatars in avaTTAR appears to enhance users' ability to mimic poses with greater precision compared to the baseline system.

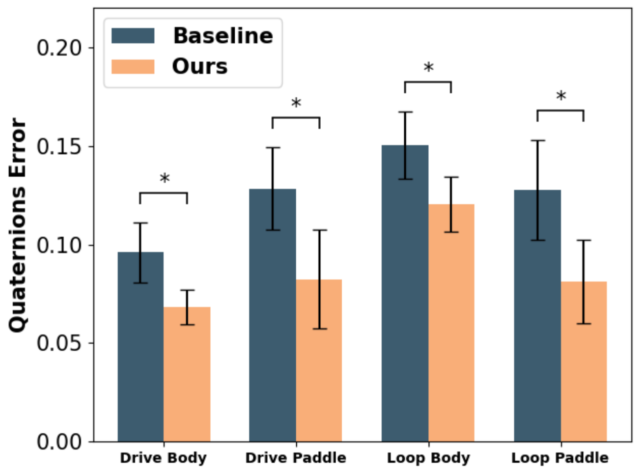

Figure 10: Overall the body and paddle movement error comparison (lower the better), between the baseline approach and our method, "*" indicates p-value < 0.05

When evaluating entire movements, the data in Figure 10 further supports avaTTAR 's efficacy. Users practicing with avaTTAR recorded a lower average error (drive stroke: body AVG = 0.0684, SD = 0.0249; paddle AVG = 0.0824, SD = 0.0709. loop stroke: body AVG = 0.1205, SD = 0.0391; paddle AVG = 0.0810, SD 0.0599)) in contrast to those who used the baseline system (drive stroke: body AVG = 0.0960, SD = 0.0427; paddle AVG = 0.1283, SD = 0.0587. loop stroke: body AVG = 0.1504, SD = 0.0479; paddle AVG = 0.1275, SD = 0.0712).

When comparing body and paddle errors for both strokes, we notice a more noticeable reduction in paddle error. This could be due to the ***on-body*** skeleton that helps users better understand the movement of the target paddle. This observation aligns with Figure 11, which shows that users utilizing avaTTAR generally have an enhanced self-awareness of their body (Q1), increased confidence in the accuracy of their motion (Q2), and find it easier to compare their movements with those of the expert (Q3).

8.2.2 Session 2 Results. All 14 users successfully completed the training for the "push" stroke using avaTTAR.

The features of the system were evaluated using Likert-type ratings collected at the end of this session, as shown in Figure 12. The questionnaire was divided into four parts, each part aiming to gather subjective user opinions on how avaTTAR fulfills the design requirements described in Section 3.2.

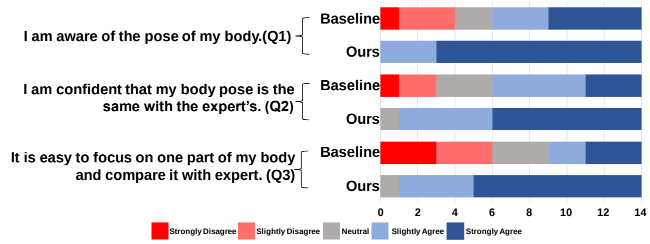

Figure 11: The results of the qualitative comparison between the baseline approach and our AR method.

Overall, users found our system to be highly user-friendly and navigable, with an average rating of 4.57 (SD = 0.51) for their ability to navigate through the system and manipulate avatars (Q12). Additionally, users reported that the hardware setup, including the webcam and IMU, did not significantly interfere with the training process, receiving an average rating of 4.79 (SD = 0.43) (Q11). A user commented, *"It is intuitive to select the strokes and start using the system to learn the movement of the stroke."* Regarding the motion reconstruction features of our system, users expressed satisfaction with the representation of the expert avatar's stroke motion (Q1: AVG = 4.64, SD = 0.50) and their own avatar representing their body (Q2: AVG = 4.50, SD = 0.52). One user remarked, *"The user avatar could follow my movement with low latency."* For stroke comparison, users' feedback was positive, indicating that the system effectively enhanced their understanding of their stroke movement performance (Q3: AVG = 4.36, SD = 0.63). According to one user, *"Both the **detached** avatars and the **on-body** feedback are useful, showing me what should do and how to correct."* Users also found the comparisons between their body and paddle movements with those of experts to be accurate and informative (Q4: AVG = 3.93, SD = 0.62) (Q5: AVG = 4.21, SD = 0.70). As mentioned by a user, *"I did some random movement that is not part of a stroke, the system immediately pointed out my error, that's fun."* Additionally, the system was reported to provide effective feedback when users made errors in their movements (Q6: AVG = 4.00, SD = 1.18), and the ***detached*** cues provided clear references for user movements (Q7: AVG = 4.29, SD = 0.73). One user pointed out, *"My avatar really highlighted where I was doing wrong during the stroke."* Users felt that the ***on-body*** cues were helpful in guiding their actions (Q8: AVG = 4.57, SD = 0.51), and real-time feedback enhanced their learning experience (Q9: AVG = 4.43, SD = 0.65). One user observed, *"The **on-body** virtual paddle helped me understand the stroke well."* Users also found visual cues valuable in understanding and replicating expert movements (Q10: AVG = 4.64, SD = 0.50). The standard SUS survey result is 80.89 with a standard deviation of 13.47, indicating the high usability of our system.

9 DISCUSSION AND FUTURE WORK

9.1 Cues for Table Tennis Training

The scope of our system is limited to the stroke training portion of table tennis, where accurate reproduction of stroke movement by the learner without guidance is of great importance. The result

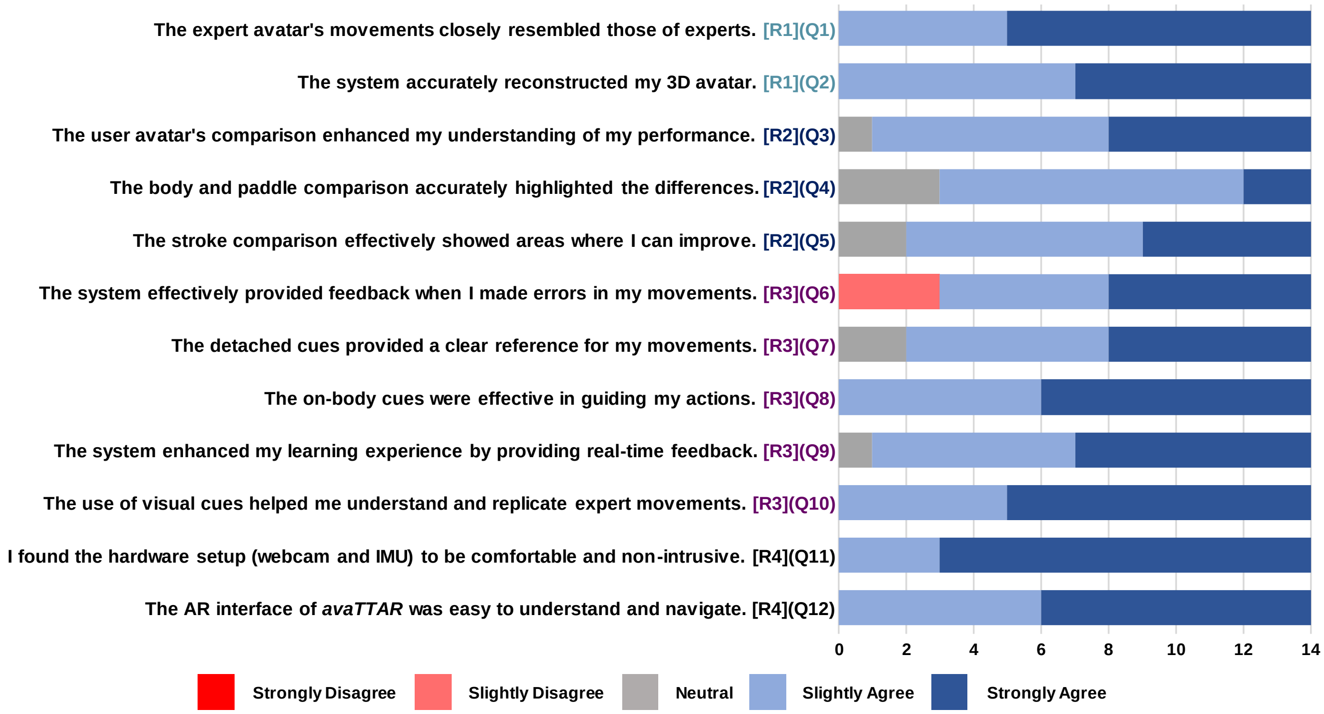

Figure 12: The results of the qualitative feedback on the system usability.

of our study (Section 8.2.1) indicates that the learners can accurately reproduce the stroke motion by using the system. Traditional practice can lead to reinforcement of incorrect techniques when practicing alone without accurate feedback, whereas avaTTAR , with its demonstrations and guidance, ensures that learners practice movement posture correctly during training.

Different levels of players can achieve different goals during stroke training. For beginner players, stroke training could help them develop and understand stroke techniques. For intermediate players, stroke training could help them reinforce their skills and adapt variants [65] of the same stroke to return incoming balls under different conditions such as different spins. For example, for the forehand drive stroke, the paddle angle is different with different kinds of spins. Another form of training, drills training, is designed to simulate match conditions and develop a combination of strokes (e.g. one forehand drive and one backhand drive). Our system also allows the user to record and train using these kinds of stroke variants and drills.

As training often presents consistent and stable strokes for practice, there are potential gaps between the static way of table tennis training and the dynamic nature of real-world gameplay situations. Trainees may face challenges when trying to transfer the skills acquired during practice to actual gameplay scenarios. This is intrinsic to various training methods [54, 71] that exclude practical training with real opponents, including our system. As one user noted, *"... the actual strokes are not always the same in real plays..."*.

Future work could involve adapting static target movements to different situations, such as changing ball trajectories and ball placements. Researchers could consider using generative models, such as GANs (Generative Adversarial Networks) [19, 20, 57] and diffusion models [25, 60], to enable such adaptations by synthesizing variant movements from static movements for the user to mimic. This dynamic training content could potentially adapt users to the situation in actual games.

Our system focuses only on the stroke training aspect of table tennis. In other aspects of table tennis training, more cues are needed in addition to body posture. Players could be interested in precise feedback on the contact points when the paddle hits the ball. A user with said: *"... I want to know where I hit the ball if it was within the 'sweet spot'..."*. Such cues provide detailed feedback for fine-tuning strokes. Additionally, timing cues are valuable to users, as mentioned by another user who expressed the need for *"... an incoming virtual ball to help learn the timing to hit the ball..."*. These cues play a vital role in refining the timing and coordination required for advanced gameplay, improving the training experience for skilled players, which could lead to better performance or faster skill acquisition.

To fulfill these requirements, we envision future work to consider outcome investigation, such as statistical analysis of ball-paddle contact area and hit timing. Furthermore, as an extension of previous work in the field of AR table [35, 48], visualizing the trajectory and placement of the ball could be integrated into the HMD-based AR to enhance the training experience. Additionally, multimodal cues could be considered to integrate to avaTTAR to improve user experience, such as haptic [40, 71] and acoustic [58] feedback.

9.2 *On-body* Visuals for Other Sports and Areas

The motivation of the ***on-body*** visual is to provide a closer look at the stroke from the coach's viewpoint (first-person view of the coach) with the ***on-body*** cues. This setup offers a similar hands-on instruction experience that some coaches have when correcting trainees, as one of the users *"... the first-person view guidance, just like a coach, hand-to-hand correcting my posture..."*. The equivocation caused by the viewpoint was mentioned in [44], where they provide a part-based visualization with viewpoint suggestions. When relying on ***detached*** cues, users must first mentally translate movements to their own viewpoint, which could lead to ambiguity. The ***on-body*** cues, on the other hand, illustrate the precise trajectories of both the body and the paddle to follow. A user said *"... I found the **detached** cues ambiguous sometimes, but **on-body** ones are straightforward ..."*.

The ***on-body*** visuals that overlay on the physical body may also be applied to other sports beyond table tennis. Especially those with basic skills that demand precise control of body movements. For example, sports involve a combination of fundamental skills that are complex movements or fine motor skills: racket sports that have multiple elemental strokes similar to table tennis; basketball which involves basis movements such as dribbling, shooting, passing, and more. On the other hand, these visuals could be applied to train the rhythm in sports such as figure skating, where timing and rhythm are essential to execute jumps and spins with precision and control.

Similarly, we also notice that the ***on-body*** cues for the paddle improve the spatial understanding of the object for the learner, as mentioned by a user, "... I found the ***on-body*** cues for the paddle alone are very useful to understand the orientation and path of it ..." Training scenarios that involve the use of tools or hand-object interaction [36] in various domains could also consider using ***on-body*** instructions in future studies, such as welding [34], carpentry [41], or surgery [66], could also benefit from a similar MR training approach, improving skill acquisition, safety, and precision in professions where effective tool usage is critical.

9.3 Movement Reconstruction Limitation

The motion reconstruction component of avaTTAR provides promising results overall, according to the user study. However, there were cases where the user's physical movements mismatched their virtual representation within the system. A user who wore clothes in a white color that is close to the color of the background mentioned that *"... my avatar's leg is jittering when my leg is not..."*. Also, since the pose estimation network operates at a frequency of around 30 frames per second (FPS), there might be latency issues with rapid movement. As expressed by another user, *"...my avatar cannot follow me when I move my hand fast..."*. These examples illustrate the potential challenges posed by pose estimation errors. While in the training phase, the latency and the low FPS might be ignored, they more or less interfere with the user experience.

Limited by the graphic card, avaTTAR adopted a lightweight neural network with acceptable performance. To obtain better reconstruction accuracy, future studies should consider using advanced algorithms [17, 18, 42, 76]. However, these methods may require more computational resources and result in a low FPS. Future solutions may involve the use of cloud services to process data to solve resource challenges.

In table tennis, the grip remains unchanged during stroke execution, therefore, the hand posture is also fixed. The position and orientation of the paddle can also deduce the position and orientation of the wrist. Therefore, we did not incorporate the detection of the hand pose. However, this might limit the understanding of the wrist and hand movement. Future studies could apply whole-body estimation [78] to detect and provide the visualization of both body and hand.

9.4 Field of View

Intrinsic to the ***on-body*** and ***detached*** visualization method, only when the natural viewing area aligns with the movement that the user attempts to learn, visualization can be applied efficiently. Otherwise, the user might not be able to view and practice the correct motion at the same time. This viewing area problem has been investigated in [37], where the researchers fixed the visualization to various positions relative to the user.

Besides the software limitation on the field of view, the current system may suffer from a limited field of view with HoloLens which may compromise the overall training experience. In particular, a comment arose with the ***on-body*** cues, *"... I can only see parts of the **on-body** visual cues, such as my hand and paddle..."*. Future work could explore the use of novel AR/MR headsets, such as the Meta Quest Pro [32] and Apple Vision Pro [31]. These headsets offer a wider field of view, potentially improving the immersive quality of the system and addressing this limitation. Additionally, the weight and comfort of the HMD would also be one of the possible reasons that hinder the experience of the user, which can also be investigated in the future.

9.5 Visual Presentation

Based on the design rationale, we adopted the skeletal visualization by following YouMove [1] which also includes learning body movement. According to [75], based on the level of indirection there are three types of motion guidance visualization which are explicit (e.g. YouMove [1]), implicit (e.g. SleeveAR [61]), and abstract (e.g. LightGuide [59]). Future work could consider studying alternative visualization methods based on these three categories. For example, for players who already understand stroke movement, explicit skeletal motion guidance could be redundant. Using only the key joint movement path or racket movement, implicit (or abstract) visualization could also be applied. Additionally, for explicit visualization, it is unknown whether the skeletal method or the avatar method [29] is more effective, researchers could consider exploring user preferences for body movement visualization.

Visual attention guidance techniques [14] could also be beneficial when visualizing the stroke. Although our system did not employ visual attention guidance methods, studies [3] have demonstrated that directing the user's attention to critical areas, such as errors or difficult segments of the stroke, can help reduce cognitive load.

10 CONCLUSION

In this work, we presented avaTTAR, an AR system that provides ***on-body*** and ***detached*** visual cues for the training of table tennis strokes. This dual visualization approach not only strengthens the user's understanding of the correct techniques but also offers immediate feedback for refinement. Our contributions include a design rationale extracted from interviews with experienced players. Based on the design rationale, we derived the design requirements for our system and then decomposed them into detailed components to implement. Our system integrated a camera and IMU setup to capture 3D body and paddle movements, and an AR interface that enables users to practice with personalized visual cues. Furthermore, our user study first highlights the potential of avaTTAR to improve movement posture accuracy, then the usability of the system to practice stroke movements. We envision that future research can further investigate other types of visual cues and apply the cues we propose to other areas.

ACKNOWLEDGMENTS

We thank the reviewers for their invaluable feedback. This work is partially supported by the NSF under the Future of Work at the Human Technology Frontier (FW-HTF) 1839971. We also acknowledge the Feddersen Distinguished Professorship Funds. Additionally, we would like to thank the coach and players from the Purdue Table Tennis Club and the Indy Pong Table Tennis Center for their invaluable support.

REFERENCES

[1] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: Enhancing Movement Training with an Augmented Reality Mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. Association for Computing Machinery, New York, NY, USA, 311–320. https://doi.org/10.1145/2501988.2502045

[2] Farwa Babar, M Farhan Tabassum, Sumera Sattar, Saadia Hassan, Rabia Karim, et al. 2021. Analysis of table tennis skills: an assessment of shadow practice in learning forehand and backhand drive. *PalArch's Journal of Archaeology of Egypt/Egyptology* 18, 08 (2021), 4488–4502.

[3] Luis Bautista, Fernanda Maradei, and Gabriel Pedraza. 2023. Strategies to reduce visual attention changes while learning and training in extended reality environments. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 17, 1 (2023), 17–43.

[4] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd international conference on knowledge discovery and data mining*. 359–370.

[5] Mahesh K Bhabhor, Kalpesh Vidja, Priti Bhanderi, Shital Dodhia, Rajesh Kathrotia, and Varsha Joshi. 2013. Short Communication A comparative study of visual reaction time in table tennis players and healthy controls. *Indian J Physiol Pharmacol* 57, 4 (2013), 439–442.

[6] Guido Brunnett, Stephan Rusdorf, and Mario Lorenz. 2006. V-Pong: an immersive table tennis simulation. *IEEE Computer Graphics and Applications* 26, 4 (2006), 10–13.

[7] 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.

[8] Hua-Tsung Chen, Yu-Zhen He, and Chun-Chieh Hsu. 2018. Computer-assisted yoga training system. *Multimedia Tools and Applications* 77 (2018), 23969–23991.

[9] Zhutian Chen, Shuainan Ye, Xiangtong Chu, Haijun Xia, Hui Zhang, Huamin Qu, and Yingcai Wu. 2021. Augmenting sports videos with viscommentator. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 824–834.

[10] 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 33rd Annual ACM Symposium on User Interface Software and Technology*. 196–208.

[11] Digital Standard Co.. 2023. ThreeDPoseUnityBarracuda. https://github.com/digital-standard/ThreeDPoseUnityBarracuda.

[12] Dazhen Deng, Jiang Wu, Jiachen Wang, Yihong Wu, Xiao Xie, Zheng Zhou, Hui Zhang, Xiaolong Zhang, and Yingcai Wu. 2021. Eventanchor: Reducing human interactions in event annotation of racket sports videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[13] Samson Dubina. 2015. *100 Days of Table Tennis*. CreateSpace.

[14] Magy Seif El-Nasr, Athanasios Vasilakos, Chinmay Rao, and Joseph Zupko. 2009. Dynamic intelligent lighting for directing visual attention in interactive 3-d scenes. *IEEE Transactions on Computational Intelligence and AI in Games* 1, 2 (2009), 145–153.

[15] 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9919–9928.

[16] Mark Andrew Flores, Dave Bercades, and Fernando Florendo. 2010. Effectiveness of shadow practice in learning the standard table tennis backhand drive. *Editorial Board* 105 (2010).

[17] Lin Geng Foo, Tianjiao Li, Hossein Rahmani, Qiuhong Ke, and Jun Liu. 2023. Unified pose sequence modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13019–13030.

[18] Jia Gong, Lin Geng Foo, Zhipeng Fan, Qiuhong Ke, Hossein Rahmani, and Jun Liu. 2023. Diffpose: Toward more reliable 3d pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13041–13051.

[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.

[21] Yaodong Gu, Changxiao Yu, Shirui Shao, and Julien S Baker. 2019. Effects of table tennis multi-ball training on dynamic posture control. *PeerJ* 6 (2019), e6262.

[22] Hong Guo, ShanChen Zou, YiLin Xu, Han Yang, Jian Wang, HongXin Zhang, and Wei Chen. 2022. DanceVis: toward better understanding of online cheer and dance training. *Journal of Visualization* 25, 1 (2022), 159–174.

[23] Ping-Hsuan Han, Kuan-Wen Chen, Chen-Hsin Hsieh, Yu-Jie Huang, and Yi-Ping Hung. 2016. Ar-arm: Augmented visualization for guiding arm movement in the first-person perspective. In *Proceedings of the 7th Augmented Human International Conference 2016*. 1–4.

[24] Ping-Hsuan Han, Yang-Sheng Chen, Yilun Zhong, Han-Lei Wang, and Yi-Ping Hung. 2017. My Tai-Chi coaches: an augmented-learning tool for practicing Tai-Chi Chuan. In *Proceedings of the 8th Augmented Human International Conference*. 1–4.

[25] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.

[26] Thuong N Hoang, Martin Reinoso, Frank Vetere, and Egemen Tanin. 2016. Onebody: remote posture guidance system using first person view in virtual environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. 1–10.

[27] Larry Hodges. 2014. *Table Tennis Tips*. Createspace Independent Publishing Platform.

[28] Thorben Hülsdünker, Martin Ostermann, and Andreas Mierau. 2019. The speed of neural visual motion perception and processing determines the visuomotor reaction time of young elite table tennis athletes. *Frontiers in behavioral neuroscience* 13 (2019), 165.

[29] Atsuki Ikeda, Dong-Hyun Hwang, Hideki Koike, Gerd Bruder, Shunsuke Yoshimoto, and Sue Cobb. 2018. AR based Self-sports Learning System using Decayed Dynamic TimeWarping Algorithm.. In *ICAT-EGVE*. 171–174.

[30] Atsuki Ikeda, Yuka Tanaka, Dong-Hyun Hwang, Homare Kon, and Hideki Koike. 2019. Golf training system using sonification and virtual shadow. In *ACM SIGGRAPH 2019 Emerging Technologies*. 1–2.

[31] Apple Inc. 2023. Apple Vision Pro. https://www.apple.com/apple-vision-pro

[32] Meta Inc. 2023. Meta Quest Pro. https://www.meta.com/quest/quest-pro/

[33] Microsoft Inc. 2023. Microsoft HoloLens | Mixed Reality Technology for Business. https://www.microsoft.com/en-us/hololens

[34] Ananya Ipsita, Levi Erickson, Yangzi Dong, Joey Huang, Alexa K Bushinski, Sraven Saradhi, Ana M Villanueva, Kylie A Peppler, Thomas S Redick, and Karthik Ramani. 2022. Towards modeling of virtual reality welding simulators to promote accessible and scalable training. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–21.

[35] Hiroshi Ishii, Craig Wisneski, Julian Orbanes, Ben Chun, and Joe Paradiso. 1999. PingPongPlus: Design of an Athletic-Tangible Interface for Computer-Supported Cooperative Play. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 394–401. https://doi.org/10.1145/302979.303115

[36] Rahul Jain, Jingyu Shi, Runlin Duan, Zhengzhe Zhu, Xun Qian, and Karthik Ramani. 2023. Ubi-TOUCH: Ubiquitous Tangible Object Utilization through Consistent Hand-object interaction in Augmented Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[37] Hye-Young Jo, Laurenz Seidel, Michel Pahud, Mike Sinclair, and Andrea Bianchi. 2023. Flowar: How different augmented reality visualizations of online fitness videos support flow for at-home yoga exercises. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[38] 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 2016 CHI conference on human factors in computing systems*. 758–769.

[39] Eren Karatas, Kissinger Sunday, Sude Erva Apak, Yiwei Li, Junwei Sun, Anil Ufuk Batmaz, and Mayra Donaji Barrera Machuca. 2023. \" I consider VR Table Tennis to be my secret weapon!\": An Analysis of the VR Table Tennis Players' Experiences Outside the Lab. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*. 1–12.

[40] Benjamin Knoerlein, Gábor Székely, and Matthias Harders. 2007. Visuo-Haptic Collaborative Augmented Reality Ping-Pong. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology* (Salzburg, Austria) *(ACE '07)*. Association for Computing Machinery, New York, NY, USA, 91–94. https://doi.org/10.1145/1255047.1255065

[41] I-Jui Lee. 2020. Using augmented reality to train students to visualize three-dimensional drawings of mortise–tenon joints in furniture carpentry. *Interactive Learning Environments* 28, 7 (2020), 930–944.

[42] Wenhao Li, Hong Liu, Hao Tang, Pichao Wang, and Luc Van Gool. 2022. Mhformer: Multi-hypothesis transformer for 3d human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13147–13156.

[43] 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 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 461, 13 pages. https://doi.org/10.1145/3411764.3445649

[44] 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), 1–14.

[45] Wu Liu, Qian Bao, Yu Sun, and Tao Mei. 2022. Recent advances of monocular 2d and 3d human pose estimation: A deep learning perspective. *Comput. Surveys* 55, 4 (2022), 1–41.

[46] Tom Lodziak. 2020. *Spin: Tips and Tactics to Win at Table Tennis.*

[47] Guillaume Martinent, Eirik Ansnes, et al. 2020. A Literature review on coach-athlete relationship in table tennis. (2020).

[48] Thomas Mayer. 2019. http://thomas-mayer.de/portfolio/table-tennis-trainer

[49] Dushyant Mehta, Srinath Sridhar, Oleksandr Sotnychenko, Helge Rhodin, Mohammad Shafiei, Hans-Peter Seidel, Weipeng Xu, Dan Casas, and Christian Theobalt. 2017. Vnect: Real-time 3d human pose estimation with a single rgb camera. *Acm transactions on graphics (tog)* 36, 4 (2017), 1–14.

[50] Stefan Carlo Michalski, Ancret Szpak, Dimitrios Saredakis, Tyler James Ross, Mark Billinghurst, and Tobias Loetscher. 2019. Getting your game on: Using virtual reality to improve real table tennis skills. *PloS one* 14, 9 (2019), e0222351.

[51] Meinard Müller. 2007. Dynamic time warping. *Information retrieval for music and motion* (2007), 69–84.

[52] Meinard Müller. 2007. *Information retrieval for music and motion*. Vol. 2. Springer.

[53] Takayuki Nozawa, Erwin Wu, and Hideki Koike. 2019. Vr ski coach: Indoor ski training system visualizing difference from leading skier. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1341–1342.

[54] Hawkar Oagaz, Breawn Schoun, and Min-Hyung Choi. 2021. Performance improvement and skill transfer in table tennis through training in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2021), 4332–4343.

[55] Markus Raab, Rich SW Masters, and Jonathan P Maxwell. 2005. Improving the 'how'and 'what'decisions of elite table tennis players. *Human movement science* 24, 3 (2005), 326–344.

[56] Nikolaos Sarafianos, Bogdan Boteanu, Bogdan Ionescu, and Ioannis A Kakadiaris. 2016. 3d human pose estimation: A review of the literature and analysis of covariates. *Computer Vision and Image Understanding* 152 (2016), 1–20.

[57] Jingyu Shi, Rahul Jain, Hyungjun Doh, Ryo Suzuki, and Karthik Ramani. 2023. An HCI-Centric Survey and Taxonomy of Human-Generative-AI Interactions. *arXiv preprint arXiv:2310.07127* (2023).

[58] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review* 20 (2013), 21–53.

[59] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 179–188.

[60] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).

[61] Maurício Sousa, João Vieira, Daniel Medeiros, Artur Arsenio, and Joaquim Jorge. 2016. SleeveAR: Augmented reality for rehabilitation using realtime feedback. In *Proceedings of the 21st international conference on intelligent user interfaces*. 175–185.

[62] Manuel Stein, Thorsten Breitkreutz, Johannes Haussler, Daniel Seebacher, Christoph Niederberger, Tobias Schreck, Michael Grossniklaus, Daniel Keim, and Halldor Janetzko. 2018. Revealing the invisible: Visual analytics and explanatory storytelling for advanced team sport analysis. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*. IEEE, 1–9.

[63] Manuel Stein, Halldor Janetzko, Andreas Lamprecht, Thorsten Breitkreutz, Philipp Zimmermann, Bastian Goldlücke, Tobias Schreck, Gennady Andrienko, Michael Grossniklaus, and Daniel A Keim. 2017. Bring it to the pitch: Combining video and movement data to enhance team sport analysis. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 13–22.

[64] Kosuke Takahashi, Dan Mikami, Mariko Isogawa, Yoshinori Kusachi, and Naoki Saijo. 2019. Vr-based batter training system with motion sensing and performance visualization. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1353–1354.

[65] Artyom Utochkin, Vasilii Zhdanov, and Ivan Zhdanov. 2018. *Modern table tennis: strokes, trainings, strategies*. Litres.

[66] Petr Vávra, Jan Roman, Pavel Zonča, Peter Ihnát, Martin Němec, Jayant Kumar, Nagy Habib, Ahmed El-Gendi, et al. 2017. Recent development of augmented reality in surgery: a review. *Journal of healthcare engineering* 2017 (2017).

[67] Eleven VR. 2023. Eleven VR. https://elevenvr.com/

[68] Jiachen Wang, Ji Ma, Kangping Hu, Zheng Zhou, Hui Zhang, Xiao Xie, and Yingcai Wu. 2022. Tac-Trainer: A Visual Analytics System for IoT-based Racket Sports Training. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 951–961.

[69] 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 27th ACM international conference on multimedia*. 374–382.

[70] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. 2021. Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding* 210 (2021), 103225.

[71] 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.

[72] Techeng Wu and Piren Su. 2010. How to coach world-class athletes of table tennis. *International Journal of Table Tennis Sciences* 6 (2010), 195–199.

[73] 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.

[74] Xingyao Yu, Katrin Angerbauer, Peter Mohr, Denis Kalkofen, and Michael Sedlmair. 2020. Perspective matters: Design implications for motion guidance in mixed reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 577–587.

[75] Xingyao Yu, Benjamin Lee, and Michael Sedlmair. 2024. Design Space of Visual Feedforward And Corrective Feedback in XR-Based Motion Guidance Systems. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.

[76] Ce Zheng, Sijie Zhu, Matias Mendieta, Taojiannan Yang, Chen Chen, and Zhengming Ding. 2021. 3d human pose estimation with spatial and temporal transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 11656–11665.

[77] Qiushi Zhou, Andrew Irlitti, Difeng Yu, Jorge Goncalves, and Eduardo Velloso. 2022. Movement guidance using a mixed reality mirror. In *Designing Interactive Systems Conference*. 821–834.

[78] Yue Zhu, Nermin Samet, and David Picard. 2023. H3wb: Human3. 6m 3d wholebody dataset and benchmark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 20166–20177.

A FORMATIVE INTERVIEW

Questions asked in the interview:

(1) Do you have any specific stroke training experience in table tennis? If yes, please briefly describe the type of stroke training you have undergone and any notable outcomes.
(2) What other types of training or practice have you engaged in to improve your table tennis skills (e.g., footwork, serve and receive practice, tactics and strategy training, physical conditioning, ball control)?
(3) In your opinion, what are the critical components of an effective stroke training program?

Table 1: Demographics of interviewees, including age, table tennis experience in years, training experience with a coach in years, and the insights they mentioned. Only P6 and P10 have AR / VR experience.

ID	Age	TT (Years)	Training (Years)	Insights
P1	69	54	2	I2, I3
P2	29	19	1	I2, I3
P3	65	25	5	I2, I3
P4	66	20	9	I1, I2, I3
P5	47	35	10	I1, I2, I3
P6	27	20	12	I1, I2, I3
P7	26	20	4	I1, I2, I3, I4
P8	27	8	0	I2, I4
P9	23	6	0	I2, I3, I4
P10	22	1	0	I2, I3
P11	24	12	9	I1, I2, I3, I4

(4) What specific aspects of table tennis stroke training do you find most challenging?
(5) What are your preferred methods or resources for learning new table tennis techniques? (Videos, books, in-person coaching, etc.)
(6) Have you ever had experience coaching or mentoring others in table tennis or received formal training or coaching from a table tennis coach?
(7) Briefly describe your coaching (from) others' experience (e.g., duration, level of players, specific areas of focus).
(8) (For coach) Based on your coaching experience, what are some common challenges you have encountered when coaching table tennis strokes? How do you currently address or overcome these challenges?
(9) (For trainee) How do you believe training with a coach has impacted your table tennis stroke technique and overall performance? Please provide specific examples or instances where you felt the coaching had a significant influence.
(10) What kind of feedback or guidance do you find most helpful when practicing table tennis strokes?
(11) Have you ever used any technology or applications to aid in your table tennis stroke training? If yes, please provide details.
(12) Have you ever used virtual reality (VR) or augmented reality (AR) technologies for sports training? Have you heard about VR or AR for sports training? If yes, please describe your experience and its impact on your training.