HoloCook: A Real-Time Remote Mixed Reality Cooking Tutoring System

Liuchuan Yu(✉), Bo Han, Songqing Chen, and Lap-Fai Yu

George Mason University, Fairfax, VA 22030, USA
{lyu20,bohan,sqchen,craigyu}@gmu.edu

Abstract. Advancements in extended reality (XR) technology have spurred research into XR-based training and collaboration. On the other hand, mixed reality (MR) fuses the real and the virtual world in real time and provides interaction, which brings the possibility of completing real-world tasks collaboratively through MR headsets. We present HoloCook, a novel real-time remote cooking tutoring system utilizing HoloLens 2. HoloCook is a lightweight system that doesn't require any additional devices. HoloCook can not only synchronize the coach's action with the trainee in real time but also provide the trainee with animations and 3D annotations to aid in tutoring process. HoloCook supports tutoring two recipes: pancakes and cocktails. Our user evaluation with one coach and four trainees establishes HoloCook as a feasible and usable remote cooking tutoring system in mixed-reality environments.

Keywords: Mixed reality · Cooking tutoring · Real-time remote system

1 Introduction

Virtual reality (VR) and augmented reality (AR) have been demonstrated effectiveness comparable to conventional training methods [14]. Creating real-world training setups can be costly and time-intensive [15], while VR training significantly reduces costs and enhances performance [16]. Additionally, AR holds promising potential for future training applications [18]. Beyond training, VR and AR find applications in education [17], assistive technology [3], manufacturing [34], and collaboration [36]. However, VR environments are entirely virtual, posing challenges for authoring, while AR primarily overlays virtual objects onto real ones, limiting interactivity.

Mixed reality (MR) combines aspects of both VR and AR, integrating digital and physical objects to interact and coexist in the same space [26]. With the progression of MR headsets like Microsoft HoloLens, Meta Quest Series (Quest Pro/Quest 3), and the newly launched Apple Vision Pro, MR is gaining considerable research interest, often surpassing VR and AR in various aspects. Because of its seamless blending of the digital and physical realms, MR provides a more immersive and interactive experience.

J. Y. C. Chen et al. (Eds.): HCII 2024, LNCS 15377, pp. 244–264, 2025.
https://doi.org/10.1007/978-3-031-76812-5_17

Fig. 1. An illustration of HoloCook. Two individuals, one as a coach and the other as a trainee, each equipped with HoloLens 2, are situated in separate kitchens. As the system starts, the coach's actions are live-streamed to the trainee's end. The trainee observes and replicates the actions to complete tasks. (a) and (b) depict the coach's and the trainee's perspectives, with side camera views indicated by the top-left orange boxes. In this cocktail-making instance, the coach pours lime juice (0.75 oz) into a measuring cup. Simultaneously, the trainee sees an overlay of the lime juice, the measuring cup, and 3D texts indicating the amount. The trainee follows suit accordingly.

The initial release of the Microsoft HoloLens occurred in 2016, followed by its successor, HoloLens 2, in 2019. Researchers have extensively employed HoloLens for investigations across various domains [30]. HoloLens 2 serves as an apt platform for mixed-reality research due to its Windows-like operating system and provision of multimodal interactions, including hand-tracking, gaze input, and voice recognition. While some research works utilize HoloLens 2, certain projects necessitate additional devices or sensors [28, 42], are limited to single-user scenarios [29, 45], or entail complex setup procedures [31, 41].

To explore multiplayer mixed-reality systems with simplified setup, we introduce HoloCook, a real-time remote mixed-reality cooking tutoring system utilizing HoloLens 2. HoloCook is a lightweight peer-to-peer tutoring system comprising two clients: one running on a standard PC and the other on HoloLens 2, without the need for additional devices or sensors. Additionally, HoloCook features a symmetric design, where both coach and trainee utilize identical clients, ensuring ease of extension and maintenance. To enhance the tutoring experience, we incorporate animations and 3D annotations. Key contributions of HoloCook include:

1. Proposing a novel lightweight real-time remote cooking tutoring system, symmetrically designed and devoid of additional devices or sensors;
2. Implementing two recipes, pancake and cocktail, based on our system; and
3. Conducting user evaluation experiments involving one coach and four trainees for pancake-making and cocktail-making tutoring.

HoloCook code is available at https://github.com/luffy-yu/HoloCook.

2 Related Work

2.1 XR Training

VR and AR serve as effective training mechanisms comparable to conventional methods [14]. Establishing real-world training environments can be costly and time-intensive [15], whereas VR training significantly reduces costs and enhances performance [16]. VR finds applications across various fields, including first responder training [11], medical training [12], military training [39], sports [22], recycling behaviors [6], construction safety [20], mining [24,25], and evacuation [19,21]. Additionally, AR demonstrates positive potential for future [18], such as medical training [38], industry training [37], corporate training [23], and vocational training [4].

Unlike prior XR training approaches, HoloCook focuses on real-time tutoring, enabling learners to complete real-world tasks in authentic environments, diverging from training solely within synthetic environments.

2.2 XR Education and Tutoring

XR demonstrates positive impacts in education, accommodating various learning styles and aiding in teaching methods [35]. Several studies highlight XR's utility in secondary education: Ray et al. [33] developed a VR application for teaching micro-controllers and Arduino boards using Google Cardboard headsets, showcasing VR's affordability in education. Dieker et al. [5] described TLE TeachLivE, a VR application from the University of Central Florida. Villanueva et al. [43] introduced Meta-AR-App, a collaborative AR platform leveraging cloud computing. Villanueva et al. [44] presented ColabAR, enhancing AR laboratories with physical proxies.

While previous works relied on additional infrastructure [43], sensors [44], or complex setups [5,33], HoloCook stands out as a lightweight system requiring only a PC and HoloLens 2 headset on one side. No extra devices or sensors are necessary, and both coach and trainee can share the same PC, provided it's networked and accessible to both HoloLens 2 headsets.

2.3 XR Collaboration

With network connectivity, XR finds applications in collaborative human scenarios. Various frameworks have been proposed to facilitate collaboration in virtual environments. For instance, Elvezio et al. [7] introduced a method supporting remote collaboration in AR and VR through virtual replicas, enabling remote users to manipulate virtual replicas of physical objects locally. Teo et al. [40] implemented a similar framework for MR remote collaboration, combining 360-degree video and 3D reconstruction. Nebeling et al. [27] presented XRDirector, a collaborative immersive authoring system allowing designers to express interactions through AR and VR devices, effectively manipulating virtual objects within physical space. Piumsomboon et al. [32] introduced CoVAR, a remote collaborative system blending VR and AR seamlessly, employing natural communication cues to foster novel collaboration. Liu et al. [21] proposed a networked training drill system supporting remote collaboration in virtual environments and locomotion in large shared virtual environments using treadmills.

Previous works often exhibit asymmetric designs [7], require additional devices [21,40], or involve different hardware and software on both ends [27,32]. In contrast, HoloCook is symmetrically designed, with both coach and trainee utilizing the same hardware and software. Each end can assume either role, selected before tutoring begins. This symmetry enhances HoloCook's extendability and maintainability.

2.4 MR Applications on HoloLens

Microsoft HoloLens debuted in 2016, followed by the release of HoloLens 2 in 2019, serving as a pivotal tool for research across diverse domains. Orts-Escolano et al. [28] developed Holoportation, showcasing real-time 3D reconstructions of spaces using new depth cameras. Piumsomboon et al. [31] introduced Mini-Me, an adaptive avatar enhancing mixed-reality collaboration between local AR users and remote VR users. Loki [41] facilitated remote instruction for physical tasks through bi-directional mixed-reality telepresence, with one side in AR and the other in VR. ARTEMIS [9] enabled immersive surgical telementoring, employing AR for novices and VR for experts. Zhao et al. [46] developed a visual and audio wayfinding guidance app on HoloLens for individuals with low vision. Yu et al. [45] introduced HoloAAC, an augmentative and alternative communication app on HoloLens 2 that aids individuals with expressive language difficulties in grocery shopping. Farouk et al. [8] presented an app on HoloLens 2 enhancing collaboration on visualized data through Kinect-captured remote user movements. Ihara et al. [13] introduced HoloBots, a mixed-reality collaboration platform improving holographic telepresence via synchronized mobile robots, integrating Kinect and tabletop robots.

These applications often involve disparate hardware and technologies [9,31,41] or necessitate additional devices and sensors [8,13,28]. Some lack multiplayer support [45,46]. In contrast, HoloCook is a networked real-time tutoring system with identical hardware and software on both ends, supporting peer-to-peer coaching in real time. Furthermore, HoloCook's design facilitates easy extension for one-to-many coaching applications.

3 Overview

The overall workflow of HoloCook is depicted in Fig. 2. HoloCook comprises two clients at each end, one running on HoloLens 2 and the other on PC. It's noteworthy that the PC clients can operate on the same PC, provided both HoloLens clients can access it. Initially, both the coach and the trainee wear HoloLens 2. Subsequently, the PC client can initiate, followed by the coach and

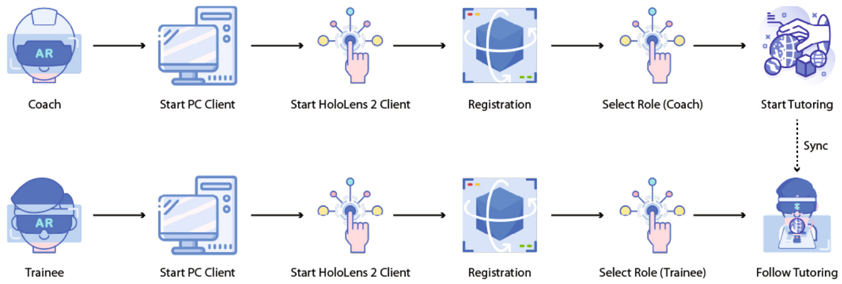

Fig. 2. Overall system workflow. Please refer to the main text for explanation.

the trainee starting their respective HoloLens clients. Registration is necessary on both ends before tutoring commences. The coach can initiate tutoring after selecting the coach role, while the trainee can view and follow tutoring upon selecting the trainee role. An example of this tutoring process is illustrated in Fig. 1.

4 Technical Approach

4.1 Network Data Flow

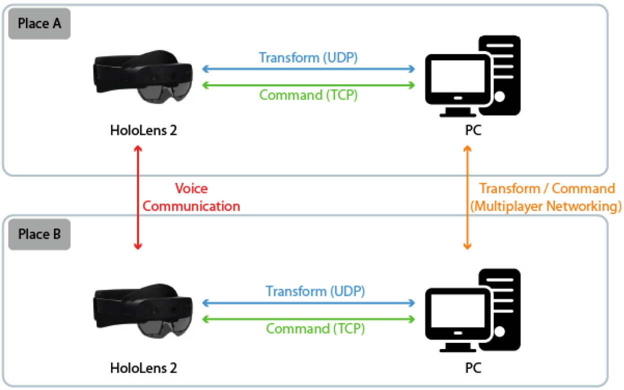

Fig. 3. Network data flow. Please refer to the main text for explanation.

In order to facilitate communication between two HoloLens 2 headsets situated in separate physical locations (kitchens in our scenario), we employ the PC as an intermediary. Figure 3 illustrates the network data flow and communication mechanisms employed. In our setup, both place A and place B are equipped with identical hardware, comprising a HoloLens 2 and a PC. Each endpoint, comprising the HoloLens and the PC, can engage in bidirectional communication. We employ two distinct network protocols for data transmission: TCP (Transmission Control Protocol) and UDP (User Datagram Protocol). TCP is utilized for sending command data due to its reliability. Here, "command" refers to actions such as controlling the HoloLens and the PC, enabling/disabling virtual objects, altering transform synchronization, displaying animations, etc. UDP, on the other hand, is used for synchronizing transform data of objects at a specific frequency. Given that the volume of data is relatively larger compared to command data and the loss of some packets is tolerable, UDP proves to be more suitable than TCP.

Regarding synchronization between the local and remote ends, we employ an off-the-shelf multiplayer networking framework. This framework automatically synchronizes objects' transforms (position and rotation). For command information, we utilize RPC (Remote Procedure Call) between two PCs. To enhance the tutoring experience, we enable verbal communication between the coach and the trainee.

One potential advantage of using PCs as bridges is the visualization of transform synchronization through the PC. Additionally, various post-processing tasks such as video/image processing and trajectory mapping can be applied.

Consider the scenario where the coach is situated in place A, and the trainee is in place B. During tutoring sessions, the data flow unfolds as follows: transforms are sent from the HoloLens 2 in place A to the PC in place A; these transforms are synchronized from the PC in place A to the PC in place B; finally, the transforms are sent from the PC in place B to the HoloLens 2 in place B. Command information follows a similar schema.

4.2 Clients

(a) Main UI (b) Recipe Menu (c)Workspace (d) Screenshot of PC Client

Fig. 4. Screenshots of HoloCook. (a), (b), and (c) refer to the HoloLens 2 client. (d) shows the PC client. Please refer to the main text for their explanation.

As aforementioned, at each end, HoloCook comprises two clients, of which one runs on HoloLens 2 (HoloLens 2 client) and the other runs on Windows (PC client). Figure 4 shows some screenshots. Figure 4(a), Fig. 4(b), and Fig. 4(c) belong to the HoloLens 2 client. Figure 4(d) shows the screenshot of the PC client.

HoloLens 2 Client. Figure 4(a) illustrates the screenshot of the main menu, which is comprised of three primary components arranged from top to bottom:

(a) **Title:** The title dynamically changes throughout interactions. Initially, it displays as "HoloCook." The format of the title follows the pattern *HoloCook-role-direction*, where *role* denotes either "Coach" or "Trainee," and *direction* indicates the synchronization direction: $HL2->PC$ (default, from HoloLens 2 to PC) or $PC->HL2$ (from PC to HoloLens 2).
(b) **Buttons:** This panel hosts nine buttons:
  (1) $HL2->PC$: Adjusts synchronization direction to from HoloLens 2 to PC.
  (2) $PC->HL2$: Adjusts synchronization direction inversely.
  (3) *Coach*: Sets this side as the coach and initiates the tutoring process.
  (4) *On*: Enables transformation synchronization.
  (5) *Off*: Disables transformation synchronization.
  (6) *Trainee*: Sets this side as the trainee and joins the tutoring process.
  (7) *Lock*: Locks the workspace (Fig. 4(c)), preventing user manipulation.
  (8) *Unlock*: Unlocks the workspace, allowing user manipulation.
  (9) *Quit*: Exits the application.
(c) **Status Bar:** A line of text displays the IP address and port of the PC client.

Figure 4(b) illustrates the recipe menu. Currently, HoloCook supports tutoring in making pancakes and cocktails. However, it is designed to be flexible and can be easily extended to accommodate other recipes.

Figure 4(c) showcases the workspace, which essentially constitutes a plane with its origin positioned at the bottom left corner. We devised this workspace for two primary purposes: 1) It facilitates the conversion of coordinate systems between the coach and trainee ends. 2) It aids in detecting the drop-off action. Upon launching the HoloLens client, a coordinate system is initialized based on its own spatial context. Consequently, there are intrinsic variations in coordinate systems between the coach and trainee ends. Hence, it is imperative to incorporate such a conversion to ensure that tutoring appears correctly. Furthermore, although hand-tracking is well-supported on HoloLens 2, detecting the drop-off action poses a challenge. The workspace functions as a collider, enabling the detection of drop-offs. When objects, including hands, intersect with the plane, it triggers the drop-off action, releasing the object. In the bottom-left corner, we have incorporated an "L" symbol with the words **Top** and **Right**, denoting the origin. This signifies that the user should orient the plane correctly, with **Right** indicating the right direction and **Top** indicating the upward direction. The workspace allows for translation, rotation, and scaling. Additionally, it can be rendered invisible during tutoring sessions.

PC Client. The PC client interface is depicted in Fig. 4(d). During the tutoring process, there is no direct interaction with the PC client; instead, it functions solely as a conduit for transferring transform and command data between the local and remote ends. At the top left corner, four lines of text are displayed: 1) The first line exhibits the application title, HoloCook. 2) The second line displays the Role, with three possible values: Unknown, Coach, and Trainee.

In the current screenshot, it indicates Unknown because the HoloLens 2 device hasn't connected to it yet. 3) The third line reveals the IP address and port of the PC, facilitating connectivity for HoloLens 2. 4)The fourth line presents the IP address and port of the HoloLens 2 device. It's noteworthy that even though the HoloLens 2 hasn't established a connection with the PC at this moment, it is essential to have prior knowledge of both the PC and HoloLens 2 addresses.

4.3 3D Models

Fig. 5. Screenshot of the real objects (kitchen wares, tools, and ingredients) and their virtual counterparts. The first row and the second row show the real and virtual objects, respectively. From left to right and from top to bottom: pan, whisk, turner, knife, spoon, oil, banana, egg, egg, cutting board, plate, bowl, ice, measuring cup (small cup), large cup, lime juice, beer, and rum.

In order to enhance the tutoring experience, particularly for the trainee, it is imperative to utilize realistic 3D models (Fig. 5). While some models can be readily acquired from online sources, certain objects may not be readily available (e.g., oil, lime juice, beer, and rum). To address this issue, we employ an iPad application called *Reality Composer*[1] to first scan these objects and then refine the scanned models using Blender. For the models obtained online, we manually adjust their scales to ensure they match real-world objects.

[1] https://apps.apple.com/au/app/reality-composer/id1462358802.

4.4 Pose Estimation

Pose estimation poses a significant challenge in tutoring scenarios, as poor estimation can severely degrade the user experience. Achieving real-time tutoring necessitates a pose estimation algorithm that operates swiftly and can effectively handle various lighting conditions and occlusions. Despite conducting initial research, we found no off-the-shelf solution that met our expectations. Consequently, we pursued an alternative approach: hand tracking.

4.5 Hand Tracking

Microsoft HoloLens 2 leverages hand tracking to facilitate various interactions, including object manipulation (translating, rotating, and scaling), button clicks, and air-tap gestures. However, determining the pose of objects held by the hand poses a challenge, given that the hand pose is generally known. To address this challenge, we propose the following approach:

Upon object registration, the object intended for manipulation attaches to the hand when the hand collides with it. Consequently, the object tracks hand movements, including translation and rotation. To prevent unintentional pickups resulting from collisions, we employ two strategies: 1) Defining handedness. Certain objects are designated as right-handed, meaning they can only be picked up by the right hand. For instance, a banana may be defined as right-handed, allowing only the right hand to grab it. 2) Determining effective collisions. Due to the limited workspace and close proximity of objects, full-hand collisions may lead to inadvertent pickups. Therefore, we consider a hand-object collision effective for grabbing only if the *ThumbTip* joint collides with the object. Here, we refer to the *ThumbTip* joint as defined in the default hand joint representation (Fig. 6).

As mentioned earlier, we utilize the workspace (depicted in Fig. 4(c)) to aid in detecting the drop-off action. When a virtual object is already attached to the hand, it becomes detached once the object collides with the workspace. Initially, we perform manual registration to ensure that the virtual object models align accurately with their real-world counterparts.

5 Experiments and Results

5.1 Development Environment

We developed HoloCook using a Windows 11 PC equipped with a Nvidia GTX 3070 GPU, running Unity 2020.3.20, Microsoft Visual Studio 2019, and JetBrains Rider 2023.2.

5.2 Implementation

Networking. For communication between HoloLens 2 and PC, we utilize an off-the-shelf Unity Asset called Netly[2]. To facilitate communication between PCs, we employ Photon Fusion. Additionally, we integrate Photon Voice, an off-the-shelf implementation, for voice communication purposes.

[2] https://assetstore.unity.com/packages/tools/network/netly-tcp-udp-225473.

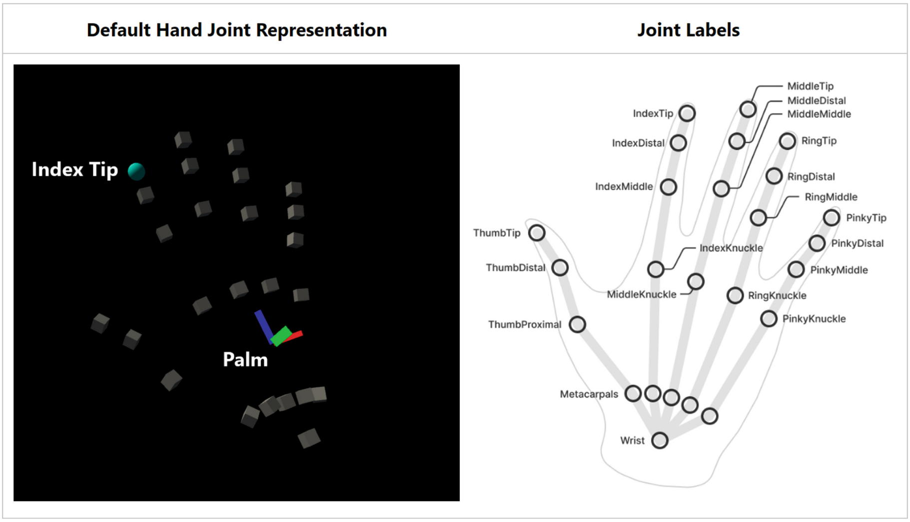

Fig. 6. Default hand joint representation (source: MRTK 2 official website).

Fig. 7. Banana slices-falling animation. This dynamic animation is triggered as the coach maneuvers banana slices atop the bowl using the knife (Fig. 12(c)).

Animation. Transformation synchronization via the network effectively handles many aspects of the tutoring process. However, certain behaviors, such as cracking eggs and dropping off ice, pose challenges for synchronization. To enhance the tutoring experience, particularly for the trainee, we have developed several animations including dropping banana slices (Fig. 7), cracking eggs (Fig. 8), pouring cooked pancakes (Fig. 9), and dropping ices (Fig. 10).

To ensure that these animations function as intended, we have implemented trigger areas positioned strategically above their expected destinations. For instance, for the banana and eggs animations, the trigger area is located above the bowl; for the pancake animation, it is positioned above the plate; and for the ice animation, it is situated above the large cup. These trigger areas facilitate the seamless execution of animations in response to specific actions or events during the tutoring process.

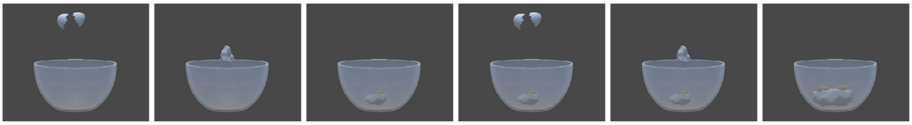

Fig. 8. Eggs-cracking animation. This animated sequence is activated as the coach cracks the two eggs. The first three frames depict the cracking of the left egg above the bowl (Fig. 12(d)). Subsequently, the following three frames illustrate the cracking of the right egg above the bowl (Fig. 12(e)).

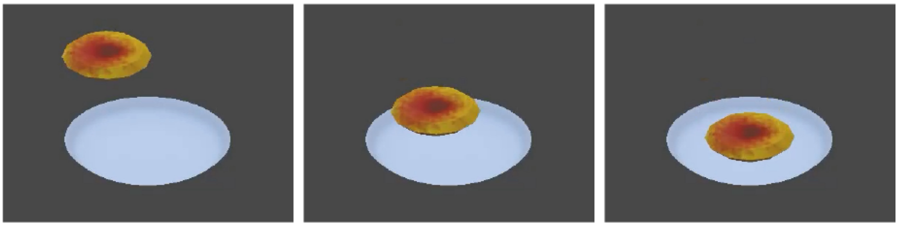

Fig. 9. Pancake-pouring animation. This animation will be activated as the coach holds the pan and prepares to pour the pancake onto the plate (to Fig. 12(g)).

Fig. 10. Ice-dropping animation. This animation plays as the coach prepares to release the ice from his hand above the large cup (Fig. 13(c)).

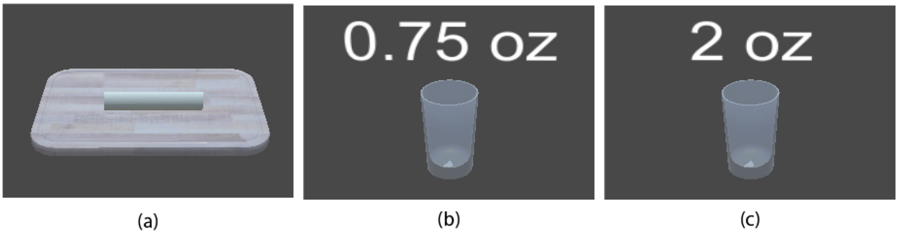

Fig. 11. Annotations. (a) occurs when the torn banana is placed on the board. (b) occurs when the lime juice is grabbed. (c) occurs when the rum is grabbed.

Annotations. In addition to animations, we have incorporated several annotations to enrich the tutoring experience. Tearing a banana is challenging to simulate and visualize, so we have introduced a virtual representation of a torn banana to illustrate the action when the coach tears the banana and places it on the board. This visualization is demonstrated in Fig. 11(a).

Furthermore, to alleviate the memorization burden for the trainee during tutoring sessions, we have integrated 3D texts onto the measuring cup. These texts are displayed when the coach retrieves the lime juice and the rum, respectively. Figure 11(b) and Fig. 11(c) showcase these annotations. They serve to provide helpful guidance and instruction, enhancing the trainee's learning experience.

5.3 Cooking Recipes

HoloCook currently supports tutoring two recipes: pancakes and cocktails.

Table 1. The procedure for making a pancake[4]

Step	Action
1	Grabbing the banana and tearing it (Fig. 12(a))
2	Placing the torn banana on the board
3	Using a knife to cut the torn banana into pieces (Fig. 12(b))
4	Transferring banana slices into the bowl by placing them on the side of the knife (Fig. 12(c))
5	Mashing the banana slices with a spoon
6	Grabbing the left egg and cracking it above the bowl (Fig. 12(d))
7	Grabbing the right egg and cracking it above the bowl (Fig. 12(e))
8	Whisking with a whisk (Fig. 12(f))
9	Grabbing the oil bottle and pouring it into the pan
10	Grabbing the bowl and pouring its contents into the pan
11	Cooking with a turner
12	Pouring the cooked pancake onto the plate (Fig. 12(g))

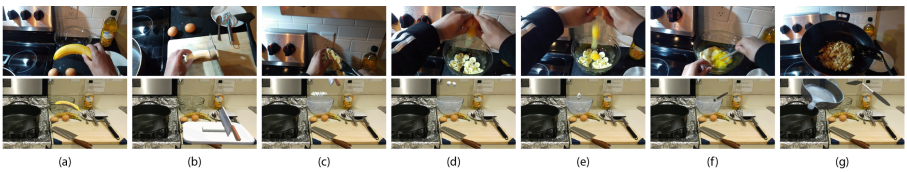

Fig. 12. An illustration of pancake-making tutoring. The first row depicts the coach's view. The second row represents the trainee's view. (a) shows the coach grabbing the banana; (b) illustrates the coach cutting the banana using a knife; (c) displays the movement of banana slices into the bowl; (d) shows the cracking of the left egg; (e) captures the cracking of the right egg; (f) demonstrates the coach whisking; and (g) depicts the coach pouring the pancake onto the plate.

[4] Adapted from: https://www.instagram.com/p/Bv4AWJ6nedf.

Pancake. Following the preparation depicted in Fig. 14, the step-by-step process of making pancakes is outlined in Table 1. Certain steps illustrating the perspectives of both the coach and the trainee are depicted in Fig. 12.

Cocktail. Table 2 shows the procedures to make the cocktail. Some steps are visualized in Fig. 13.

Table 2. The procedure for making a cocktail[6]

Step	Action
1	Pouring $0.75oz$ lime juice into the measuring cup (Fig. 13(a))
2	Pouring lime juice from the measuring cup into the large cup (Fig. 13(b))
3	Grabbing some ice from the bowl and placing it into the large cup (Fig. 13(c))
4	Pouring some beer (Fig. 13(d))
5	Pouring $2oz$ rum into the measuring cup (Fig. 13(e))
6	Pouring rum from the measuring cup into the large cup

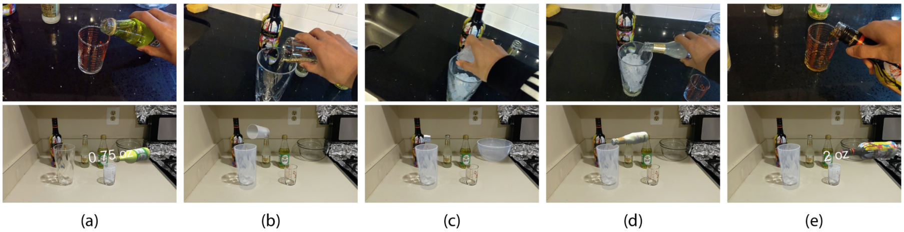

Fig. 13. An illustration of cocktail-making tutoring. The first row represents the coach's view. The second row depicts what the trainee will observe. (a) depicts the coach measuring $0.75oz$ lime juice; (b) shows the coach pouring lime juice into the large cup; (c) illustrates the coach grabbing ice; (d) displays the coach pouring beer; and (e) captures the coach measuring $2oz$ rum.

6 User Evaluation

We conducted case studies to validate the usability of HoloCook and gather feedback. We used two kitchens as depicted in Fig. 14 for our experiments. The coach's kitchen was situated in an apartment, while the trainee's kitchen was located in a townhouse. During tutoring sessions, both the coach and the trainee wore the HoloLens 2. Additionally, each side was equipped with a common PC capable of running the PC client.

[6] Source: https://www.youtube.com/watch?v=c6GV_vRlIIA&t=462.

Table 3. Participants' demographics. Note that P5 served as the coach and the others served as the trainees in our user evaluation.

Participant	Years of Cooking Experience	Years of Making Pancake	Years of Making Cocktail	VR/AR Experience	Way of Learning Cooking
P1	5	0	0	2	Youtube Video
P2	15	10	5	3	Apps
P3	10	0	0	0	Youtube Video
P4	1	0.5	0	0	Friends
P5	5	2	0	1	Youtube Video

Fig. 14. Kitchen layouts for the user evaluation. The first row and the second row show the coach side and the trainee side, respectively. The first column refers to making a pancake, and the second column refers to making a cocktail.

6.1 Participants

We recruited five participants, each with specific roles: one participant (P5) served as the coach, while the remaining four participants (P1, P2, P3, P4) acted as trainees. The participants' ages ranged from 23 to 32, with a mean age of 28 and a standard deviation of 3.03. Their demographic details are presented in Table 3.

6.2 Experiment Objects

The majority of the ingredients were purchased from local stores or markets, including bananas, eggs, oil, lime juice, beer, and rum. Additionally, we utilized various kitchen tools that were readily available in the kitchen such as pans, cutting boards, and knives.

6.3 Questionnaire

Each participant was asked to fill out a questionnaire. The questionnaire was composed of four sections: demographics, NASA Task Load Index (TLX) [10], System Usability Scale (SUS) [2], and general feedback.

6.4 Procedure

At the outset of each case study, all items were arranged neatly as depicted in Fig. 14. Subsequently, we provided instructions to the coach on registration and coaching procedures. For the trainees, we advised them to register objects first and then observed and followed the coach's actions. Both the coach and trainee were encouraged to communicate during the tasks. The first task involved making pancakes, followed by preparing cocktails. At the conclusion of each case study, we organized all items in preparation for the next session.

Trainees were instructed to finish a questionnaire after both tasks, while the coach was asked to finish the same questionnaire after coaching all four trainees.

6.5 Result Analysis

We collected all questionnaire responses and analyzed them from three aspects: NASA TLX, SUS, and general feedback.

NASA TLX. The original ratings are depicted in Fig. 15. Observation reveals that a majority of ratings fall at or below 5. The top three highest ratings pertain to *Performance Dissatisfaction*, provided by P1, P3, and P4. This is largely due to their minimal experience in making pancakes and cocktails, except for P4 who possesses 0.5 years of pancake-making experience, as shown in Table 3.

Mental Demand. Trainees (P1, P3, and P4) rated mental demand at 2, while P2 rated it 5. The average rating across all trainees stands at 2.8, indicating a better than neutral (4) perception. The coach's rating is 5, implying a need for increased attention to ensure effective tutoring.

Physical Demand. Ratings for physical demand include two 1 s (from P2 and P4), one 2 (from P1), and one 3 (from P3), yielding an average rating of 1.8. This suggests a level of physical demand nearing very low (1). The coach's rating of 5 indicates a higher demand perceived by them compared to the trainees.

Temporal Demand. All ratings for temporal demand are at or below 4 (Neutral). The average rating across all trainees is 2.0, implying that participants do not feel rushed during the experience.

Performance Dissatisfaction. Ratings for performance dissatisfaction are uniformly at or above 4 (Neutral). Trainees' average rating is 5.5, the highest among all aspects. The coach also rated it 5. Notably, P2, with significant prior experience, provided the lowest rating. This suggests that participants with prior experience tend to be more satisfied.

Effort. All ratings for effort are at or below 4 (Neutral). The average rating across all trainees is 3.0, indicating a perception better than Neutral (4).

Frustration. Trainees' ratings for frustration are uniformly at or below 2, with the coach rating it 5. The trainees' average rating is 1.3, significantly better than Neutral (4).

Overall, trainees consistently provided lower ratings compared to the coach. Trainees rated five out of six aspects better than neutral: mental demand (2.8), physical demand (1.8), temporal demand (2.0), effort (3.0), and frustration (1.3). They rated performance dissatisfaction higher (5.5) likely due to their lack of experience in pancake and cocktail making.

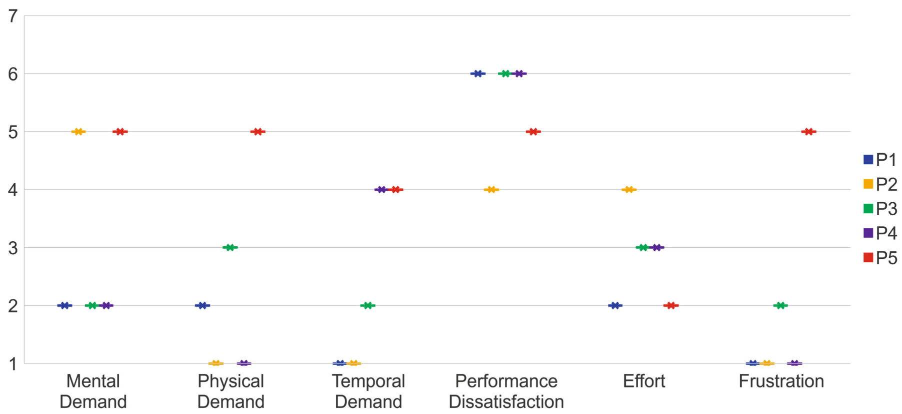

Fig. 15. NASA TLX ratings given by the five participants. 1 means very low, 4 means neutral, and 7 means very high. Refer to the main text for an explanation.

Table 4. Participants' answers to the ten SUS questions.

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
P1	4	2	4	2	4	2	3	3	4	2
P2	3	1	5	2	2	3	5	3	5	1
P3	3	3	2	2	4	1	5	1	5	1
P4	4	1	4	3	4	2	3	1	4	3
P5	4	3	3	5	4	2	3	2	4	5

The coach assigned four 5-ratings for mental demand, physical demand, performance dissatisfaction, and frustration, a 1-rating for temporal demand, and a

2-rating for effort. This indicates the coach's need to ensure synchronization with the trainees' actions, emphasizing the importance of object registration offset in coaching experiences.

SUS. Table 4 presents participants' responses to the ten SUS questions. Following the original methodology by Brooke et al. [2], the SUS scores for the five participants are 70 ("OK"), 75 ("Good"), 77.5 ("Good"), 72.5 ("Good"), and 52.5 ("OK") with corresponding approximate adjectives as per Bangor et al. [1]. According to conventional standards, a score exceeding 68 indicates above-average usability, while a score below 68 indicates below-average usability. As the results show, all trainees rated HoloCook's usability above-average, whereas the coach's rating is below-average. This pattern mirrors NASA TLX ratings. Given that the coach typically requires more attention than the trainees during tutoring sessions, it is reasonable to infer that the trainees would have a more positive experience than the coach. Overall, all participants perceived HoloCook to be **OK** or better, with three of them rating it as **Good**.

General Feedback. P1 appreciated the overall organization but expressed dissatisfaction with the limited field of view. P2 admired the real-time tutoring approach and recommended providing textual instructions alongside. Both P3 and P4 enjoyed the visual guidance provided. P3 eagerly expressed, **"I can't wait to use it to learn some new cuisines."** Meanwhile, P4 exclaimed, **"This experience is fantastic and new to me."** P5 appreciated the straightforward communication of complex actions but found hand tracking to be unreliable sometimes. P5 proposed enhancing user experience by incorporating textual or visual prompts for feature activation or instructions and optimizing the user interface to let the coach conveniently reset tutoring and register objects.

7 Limitations and Future Work

Accurate object registration forms the foundation of HoloCook's functionality. Presently, manual registration suffices; however, enhancing precision through computer vision techniques, such as semi-automatic or automatic approaches, represents a potential solution. For instance, initial rough registration by the user followed by refinement using computer vision techniques can significantly improve accuracy, leveraging the locatable feature of HoloLens 2 camera.

HoloCook operates on identical relative layouts for objects on both the coach and trainee sides, enabling real-time operation streaming but limiting deployability. Extending HoloCook to comprehend the semantic significance of coach operations and intentions presents an opportunity for enhancement. Additionally, implementing recording and relaying features can further augment the system's utility. For example, coaches can record operations for some recipes, which trainees can then replay and learn from independently.

To address complex actions involving deformable objects, HoloCook incorporates animations to convey meaning, albeit not as realistically as synchronized transformations. Future research avenues may explore segmenting objects of interest from camera streams and transferring relevant clips to the trainee side, leveraging HoloCook's PC client for efficient video/image processing.

8 Conclusion

We introduced HoloCook, a novel lightweight real-time remote cooking tutoring system harnessing HoloLens 2. HoloCook exhibits several distinctive features. First, it adopts a symmetric design, eliminating the need for additional devices or sensors. Both the coach and the trainee use the same client interface, simplifying setup and operation. Second, HoloCook offers comprehensive instruction for two complete recipes: pancake and cocktail. These recipes are meticulously crafted to be taught seamlessly via the HoloCook system. Third, we incorporated animations and annotations into HoloCook to enhance tutoring, providing visual aids for clearer understanding and guidance. We conducted a user evaluation session to assess HoloCook's usability and gather valuable user feedback, which helps refine HoloCook and inspires future research.

Acknowledgments. We are grateful to the participants for their feedback on HoloCook. We thank Huizhen Zhou for his generously providing the kitchen. This project was supported by NSF grants (award numbers: 2235049 and 2128867) and the GMU IDIA P3 Faculty Fellowship.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bangor, A., Kortum, P., Miller, J.: Determining what individual sus scores mean: adding an adjective rating scale. J. Usability Stud. **4**(3), 114–123 (2009)
2. Brooke, J., et al.: Sus-a quick and dirty usability scale. Usability Evaluation Ind. **189**(194), 4–7 (1996)
3. Bryant, L., Brunner, M., Hemsley, B.: A review of virtual reality technologies in the field of communication disability: Implications for practice and research. Disabil. Rehabil. Assist. Technol. **15**(4), 365–372 (2020)
4. Chiang, F.K., Shang, X., Qiao, L.: Augmented reality in vocational training: a systematic review of research and applications. Comput. Hum. Behav. **129**, 107125 (2022)
5. Dieker, L.A., Hynes, M.C., Hughes, C.E., Hardin, S., Becht, K.: Tle teachlive^{TM}: using technology to provide quality professional development in rural schools. Rural Special Educ. Quart. **34**(3), 11–16 (2015)
6. Do, T.D., Dylan, S.Y., Katz, A., McMahan, R.P.: Virtual reality training for proper recycling behaviors. In: ICAT-EGVE (Posters and Demos), pp. 31–32 (2020)

7. Elvezio, C., Sukan, M., Oda, O., Feiner, S., Tversky, B.: Remote collaboration in ar and vr using virtual replicas. In: ACM SIGGRAPH 2017 VR Village, pp. 1–2 (2017)
8. Farouk, P., Faransawy, N., Sharaf, N.: Using hololens for remote collaboration in extended data visualization. In: 2022 26th International Conference Information Visualisation (IV), pp. 209–214. IEEE (2022)
9. Gasques, D., et al.: Artemis: a collaborative mixed-reality system for immersive surgical telementoring. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2021)
10. Hart, S.G.: Nasa task load index (tlx) (1986)
11. Heirman, J., et al.: Exploring the possibilities of extended reality in the world of firefighting. In: 2020 IEEE international conference on artificial intelligence and virtual reality (AIVR), pp. 266–273. IEEE (2020)
12. Hurd, O., Kurniawan, S., Teodorescu, M.: Virtual reality video game paired with physical monocular blurring as accessible therapy for amblyopia. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE (2019)
13. Ihara, K., Faridan, M., Ichikawa, A., Kawaguchi, I., Suzuki, R.: Holobots: Augmenting holographic telepresence with mobile robots for tangible remote collaboration in mixed reality. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1–12 (2023)
14. Kaplan, A.D., Cruit, J., Endsley, M., Beers, S.M., Sawyer, B.D., Hancock, P.: The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis. Hum. Factors **63**(4), 706–726 (2021)
15. Karabiyik, U., Mousas, C., Sirota, D., Iwai, T., Akdere, M.: A virtual reality framework for training incident first responders and digital forensic investigators. In: International Symposium on Visual Computing, pp. 469–480. Springer (2019)
16. Koutitas, G., Smith, S., Lawrence, G.: Performance evaluation of ar/vr training technologies for ems first responders. Virtual Reality, pp. 1–12 (2020)
17. Law, E.L.C., Heintz, M.: Augmented reality applications for k-12 education: a systematic review from the usability and user experience perspective. Int. J. Child-Comput. Inter. **30**, 100321 (2021)
18. Lee, K.: Augmented reality in education and training. TechTrends **56**(2), 13–21 (2012)
19. Li, C., Liang, W., Quigley, C., Zhao, Y., Yu, L.F.: Earthquake safety training through virtual drills. IEEE Trans. Visual Comput. Graphics **23**(4), 1275–1284 (2017)
20. Li, W., Huang, H., Solomon, T., Esmaeili, B., Yu, L.F.: Synthesizing personalized construction safety training scenarios for vr training. IEEE Trans. Visual Comput. Graphics **28**(5), 1993–2002 (2022)
21. Liu, H., Choi, M., Yu, L., Koilias, A., Yu, L.F., Mousas, C.: Synthesizing shared space virtual reality fire evacuation training drills. In: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 459–464. IEEE (2022)
22. Liu, H., Wang, Z., Mousas, C., Kao, D.: Virtual reality racket sports: virtual drills for exercise and training. In: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 566–576. IEEE (2020)
23. Martins, B.R., Jorge, J.A., Zorzal, E.R.: Towards augmented reality for corporate training. Interactive Learning Environments, pp. 1–19 (2021)
24. McMahan, R., Bowman, D., Schafrik, S., Karmis, M.: Virtual environment training for preshift inspections of haul trucks to improve mining safety. In: International Future Mining Conference & Exhibition, pp. 520–528 (2008)

25. McMahan, R.P., Schafrik, S., Bowman, D.A., Karmis, M., Brune, J.: Virtual environments for surface mining powered haulage training. In: Proceedings of SME Symposium Celebrating, vol. 100, pp. 520–528 (2010)
26. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. **77**(12), 1321–1329 (1994)
27. Nebeling, M., et al.: Xrdirector: a role-based collaborative immersive authoring system. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2020)
28. Orts-Escolano, S., et al.: Holoportation: virtual 3d teleportation in real-time. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 741–754 (2016)
29. Park, B.J., Hunt, S.J., Nadolski, G.J., Gade, T.P.: Augmented reality improves procedural efficiency and reduces radiation dose for ct-guided lesion targeting: a phantom study using hololens 2. Sci. Rep. **10**(1), 18620 (2020)
30. Park, S., Bokijonov, S., Choi, Y.: Review of microsoft hololens applications over the past five years. Appl. Sci. **11**(16), 7259 (2021)
31. Piumsomboon, T., et al.: Mini-me: an adaptive avatar for mixed reality remote collaboration. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2018)
32. Piumsomboon, T., Lee, Y., Lee, G., Billinghurst, M.: Covar: a collaborative virtual and augmented reality system for remote collaboration. In: SIGGRAPH Asia 2017 Emerging Technologies, pp. 1–2 (2017)
33. Ray, A.B., Deb, S.: Smartphone based virtual reality systems in classroom teaching-a study on the effects of learning outcome. In: 2016 IEEE Eighth International Conference on Technology for Education (T4E), pp. 68–71. IEEE (2016)
34. Sahu, C.K., Young, C., Rai, R.: Artificial intelligence (ai) in augmented reality (ar)-assisted manufacturing applications: a review. Int. J. Prod. Res. **59**(16), 4903–4959 (2021)
35. Sala, N.: Virtual reality, augmented reality, and mixed reality in education: a brief overview. Current and prospective applications of virtual reality in higher education, pp. 48–73 (2021)
36. Schäfer, A., Reis, G., Stricker, D.: A survey on synchronous augmented, virtual and mixed reality remote collaboration systems. ACM Comput. Surv. (CSUR) (2021)
37. Sorko, S.R., Brunnhofer, M.: Potentials of augmented reality in training. Procedia Manufacturing **31**, 85–90 (2019)
38. Tang, K.S., Cheng, D.L., Mi, E., Greenberg, P.B.: Augmented reality in medical education: a systematic review. Canadian Med. Educ. J. **11**(1), e81 (2020)
39. Taupiac, J.D., Rodriguez, N., Strauss, O., Rabier, M.: Ad-hoc study on soldiers calibration procedure in virtual reality. In: 2019 IEEE Conference on Virtual Reality and 3d User Interfaces (VR), pp. 190–199. IEEE (2019)
40. Teo, T., Lawrence, L., Lee, G.A., Billinghurst, M., Adcock, M.: Mixed reality remote collaboration combining 360 video and 3d reconstruction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2019)
41. Thoravi Kumaravel, B., Anderson, F., Fitzmaurice, G., Hartmann, B., Grossman, T.: Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, pp. 161–174 (2019)
42. Vidal-Balea, A., et al.: Analysis, design and practical validation of an augmented reality teaching system based on microsoft hololens 2 and edge computing. Eng. Proc. **2**(1), 52 (2020)

43. Villanueva, A., Zhu, Z., Liu, Z., Peppler, K., Redick, T., Ramani, K.: Meta-ar-app: an authoring platform for collaborative augmented reality in stem classrooms. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2020)
44. Villanueva, A., Zhu, Z., Liu, Z., Wang, F., Chidambaram, S., Ramani, K.: Colabar: a toolkit for remote collaboration in tangible augmented reality laboratories. Proc. ACM Hum.-Comput. Interact. **6**(CSCW1), April 2022
45. Yu, L., et al.: Holoaac: a mixed reality aac application for people with expressive language difficulties. In: International Conference on Human-Computer Interaction. Springer (2024)
46. Zhao, Y., Kupferstein, E., Rojnirun, H., Findlater, L., Azenkot, S.: The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2020)