SAILS: A Synchronous Accessible Immersive Online Learning System for Young Learners

Yuran Sun, Zhuoying Zhang, Zhenxiao Luo, Alan William Dougherty, Man Ho Yip, Yi King Choi, Chuan Wu
The University of Hong Kong
yuransun@connect.hku.hk

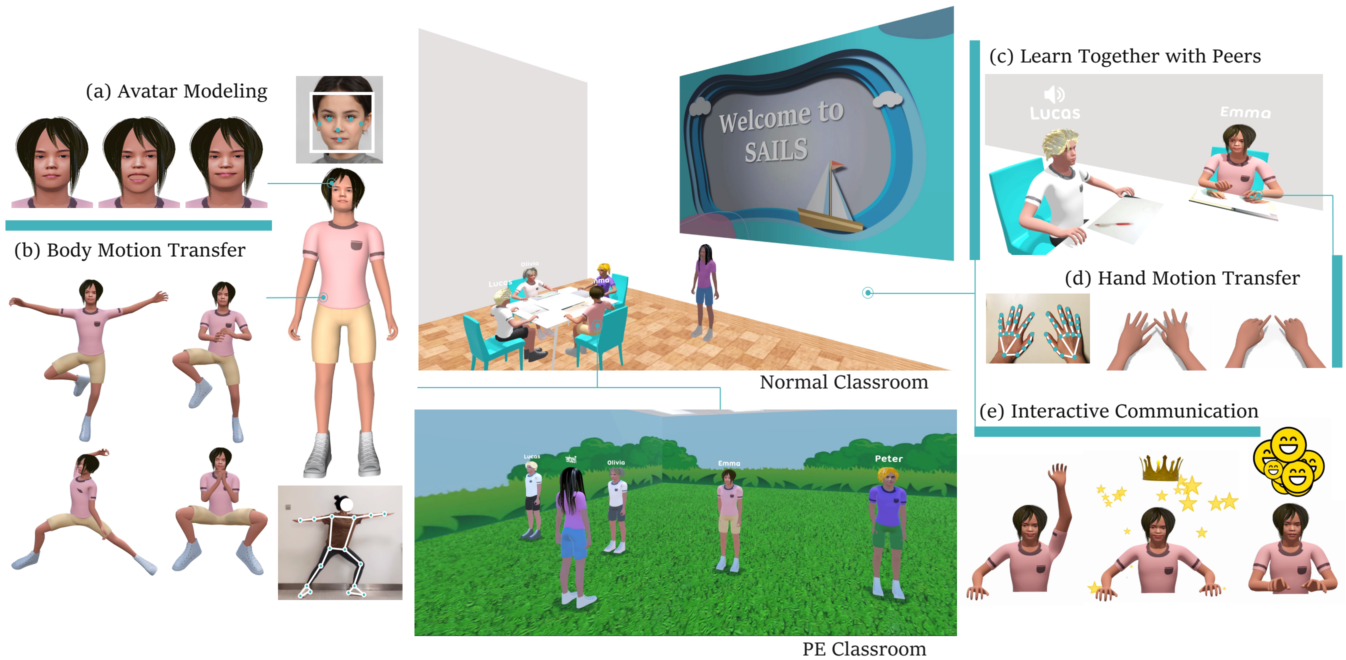

Figure 1: Core features of SAILS. The system supports two class types: a normal classroom and a PE classroom. Avatar modeling, based on a single input image, retains the user's facial features and serves as the fundamental representation of users within the application, as shown in (a). In the normal classroom, teachers can instruct using a virtual whiteboard while students sit around a table and learn together, observing each other's expressions, hand movements, and classwork. Various interaction methods are provided, such as chatting, raising hands, sending emojis, and offering rewards, as shown in (c), (d), and (e). In the PE classroom, students and teachers exercise together within a shared environment. Real-time body motion transfer allows everyone to view each other's movements, as shown in (b).

Abstract

Online distance learning emerged as a prominent means of education during the pandemic and is expected to continue as a long-lasting trend, enabling access to top education resources without the constraints of physical distance. However, conferencing software predominantly used for synchronous online teaching and learning is not suitable or sufficient for young students and teachers. The learning experiences of these students heavily rely on hands-on activities and interactions with peers and teachers that are challenging to replicate. Additionally, teachers face difficulties in effectively monitoring student progress. To address these challenges, we propose a Synchronous Accessible Immersive Online Learning System (SAILS). It offers an immersive and interactive platform for young learners by integrating real-life school activities with a virtual learning environment. Using the system, teachers can easily organize classes, assess students' work. We evaluated the effectiveness of SAILS through a user study involving 40 young learners from a kindergarten and several primary schools, along with their teachers. The results demonstrated a strong preference for our system among the participants, highlighting its ability to provide a more immersive and engaging online learning experience.

MMSys '25, March 31–April 4, 2025, Stellenbosch, South Africa

ACM ISBN 979-8-4007-1467-2/25/03
https://doi.org/10.1145/3712676.3714446

CCS Concepts

• Human-centered computing → User studies; **Interactive systems and tools**; **• Applied computing → Interactive learning environments**; **Distance learning**.

Keywords

Distance education, Classroom, Immersive virtual environments, Education, Learning effect, Full-body avatar, Character animation

ACM Reference Format:
Yuran Sun, Zhuoying Zhang, Zhenxiao Luo, Alan William Dougherty, Man Ho Yip, Yi King Choi, Chuan Wu. 2025. SAILS: A Synchronous Accessible Immersive Online Learning System for Young Learners. In *The 16th ACM Multimedia Systems Conference (MMSys '25), March 31–April 4, 2025, Stellenbosch, South Africa.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3712676.3714446

1 Introduction

During the pandemic, the way schools teach their students largely switched to online learning. Kindergarten or primary school children were provided with forms of online learning activities, such as pre-recorded videos or online classes. Many of their parents expressed concerns about learning efficiency and found their children lacked concentration and learning interests [39]. Even after the pandemic, synchronous distance education remains important, as it allows learners to participate in real-time classes from anywhere with an Internet connection, overcoming geographical barriers [13].

Existing solutions typically utilize conferencing tools such as Zoom [15] to facilitate synchronous online learning. However, these tools have limitations in engaging students and promoting interactions [7], especially for young learners. For instance, they do not effectively support hands-on and collaborative activities like drawing or physical education exercises [16]. In such scenarios, teachers are forced to check small windows one by one, making it difficult to monitor student participation intuitively and conveniently. While VR-based solutions offer better engagement, they require bulky headsets, which are not suitable for young children [21]. Therefore, it is essential to design a synchronous online learning system that is immersive, accessible, and specifically tailored to the needs of young students and their teachers.

We propose SAILS, a Synchronous Accessible Immersive Online Learning System. Our goal is to enable an interactive and immersive learning environment for young learners, fostering a sense of "togetherness" during classes while prioritizing practicality and affordability. The fundamental concept of our system is to replicate the physical classroom in a 3D virtual world and leverage both physical and virtual spaces to enhance the educational experience for young children in an online setting. The system allows students and teachers to join a virtual classroom using their home device (i.e., personal computers and webcams), where they are present alongside their classmates and teachers. By providing a structured environment that mimics traditional classroom settings, our system helps students concentrate more on their study tasks without frequent interruptions from teachers checking their work., while also facilitating the teacher's ability to conveniently monitor students' learning progress.

Our innovative approach seamlessly integrates physical school activities with virtual learning experiences. When students physically works on their learning activities at their desk, their work is automatically shared with others in the virtual classroom. Meanwhile, the teacher can present teaching materials on a virtual whiteboard and effortlessly monitor each student's work by "walking" around, all in the virtual space. Additionally, our system includes a virtual physical education (PE) classroom, allowing students to exercise together in the same virtual environment with their teacher leading the session from the front, rather than being confined to separate small windows as in traditional video conferencing systems.

The main contributions of our system are summarized as follows:

▹ We design a Synchronous Accessible Immersive Online Learning System, named SAILS, for young learners. It runs on a single home computer equipped with two RGB cameras. The system mimics traditional classroom settings, allowing multiple students to sit around tables and share the individual work that they are working on the physical table, creating a virtual "learning-together" experience in 3D space. We also implement a PE classroom, where students can exercise together virtually.

▹ We have developed realistic 3D head modeling and implemented real-time facial expression and body/hand movement transfer in SAILS. Users are represented by their virtual avatars within the system. These avatars can reflect users' appearance and be dynamically animated according to users' actual movements. Our system also integrates various multimedia components, such as voice chat, live streaming videos, and clicking-based interactivity features to enhance the virtual teaching and learning experience. By involving young learners and teachers in the iterative design process, we ensure that the system is tailored to their needs and preferences, enhancing usability and user satisfaction.

▹ We conducted a user study involving 40 kindergarten and primary school students aged from 5 to 12, along with their teachers, to evaluate the effectiveness of SAILS. More than 90% of the participants reported that they enjoyed the experience and were excited to use our system. The results indicated that SAILS was highly regarded by participants, providing a more immersive and enjoyable learning experience than video conferencing solutions. Participants also reported increased in-class interest and engagement, as well as a sense of "togetherness" while using SAILS.

2 Related Work

Synchronous distance education applications have gained significant interest in recent years, offering a virtual platform for real-time interaction between students and educators. However, many existing applications are not specifically designed for young children.

2.1 Online Conferencing Applications (2D)

Virtual conferencing applications, such as Zoom [15], Microsoft Teams [22], and Google Meet [14], have been widely adopted for synchronous distance education [8]. These tools allow participants to communicate in real-time through video and audio, share screens, and collaborate on virtual whiteboards. When having classes, students can ask questions through chat or virtual hand-raising. To facilitate group activities, breakout rooms are available where students can collaborate in smaller groups.

However, these applications, primarily designed for group meetings, may not provide the level of engagement and interaction necessary for effective online learning, especially for young students. Teaching with such tools typically centers around the teacher delivering lectures and sharing their screen, which often results in passive listening by students and a lack of a cohesive classroom experience. Students may choose not to turn on their webcams, further diminishing interaction and making it challenging for teachers to monitor participation.

Furthermore, assessing hands-on activities, such as drawing, becomes problematic. Teachers must rely on students holding their work up to the camera, which is time-consuming and often results in unclear visuals. This process can disrupt students' workflow and hinder effective assessment.

While virtual conferencing tools offer convenient and accessible solutions for online classes, they have limitations in creating immersive and engaging learning experiences and in allowing easy monitoring by teachers.

2.2 Immersive Applications (3D)

3D applications create a sense of depth and realism, making them more engaging than 2D applications. They can be used on virtual reality devices or computers.

2.2.1 VR-based. Virtual Reality (VR) has emerged as a technology that enables immersive and interactive experiences. It can serve as an advanced platform for synchronous online learning, offering a 3D environment where students and teachers can interact in real time, similar to a physical classroom [10, 25, 36, 38]. Some applications utilize VR headsets to enhance the immersive experience and offer educational features such as online classes [36, 38], lab simulations [25], and virtual field trips [25].

However, VR also comes with challenges such as the need for specialized hardware and potential discomfort or nausea for some users. Prolonged use of VR headsets, especially in applications that require significant attention and movement, can cause nausea or dizziness [37]. Moreover, manufacturers typically recommend that VR headsets be used by older children, generally 13 years and up, and under adult supervision to minimize risks [21]. They also emphasize that children under 10 cannot use VR headsets.

2.2.2 Computer-based. Computer-based applications allow users to navigate in a 3D virtual environment [12, 17, 23, 35]. Users can create rooms and invite others to join, enabling interaction in a shared virtual space. For online teaching, a teacher can create a virtual classroom that resembles a real one, sharing their screen and camera feed. This setup provides an intuitive overview of the class structure, allowing the teacher to see all virtual avatars in the classroom.

However, such applications are not specifically designed for a class of younger children. They offer a high degree of freedom, requiring each individual to control their own avatar. This can be distracting to the students. The teacher would have limited control over the class without the cooperation of the students.

Furthermore, these applications often use pixel-based [17] or cartoon-like [35] avatar representations, with some even lacking the lower body [12, 23]. Such designs lack crucial non-verbal information when users make facial expressions or perform movements, leading to less immersive and intuitive interactions. The absence of realistic avatar modeling, including detailed face modeling and accurate hand and body movement transfer, along with the inability to share real-time hands-on activities, makes these applications less suitable for young learners.

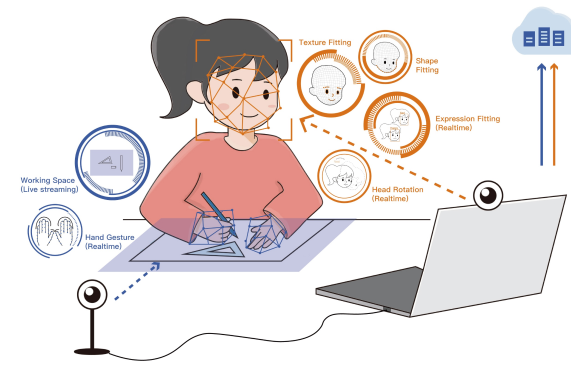

Figure 2: Device setup and client side illustration. Users need to equip their personal computers with two cameras: one front-facing camera to capture the user's face and another camera to capture the working area on the user's desk. This setup enables real-time facial expression, hand gesture, hands-on activities transfer to the virtual environment. All real-time data is processed by the client application and sent to the cloud server for distribution.

3 SAILS

3.1 System Overview

Our synchronous accessible immersive online learning system, SAILS, targets young learners and their educators, aiming to provide a sense of "togetherness" and lively engagement in classroom activities. The system is easy to set up on home devices and intuitive to use. We have designed two types of virtual classrooms in the system to cover most class types: a normal classroom and a PE classroom.

In the normal classroom, students are seated together around tables and have classes with their classmates. They can perform study tasks on their physical desks while sharing their work automatically in the virtual class. Their facial expressions and hand gestures are transferred in real-time to the virtual environment as well, allowing teachers to easily monitor students' learning progress and attentiveness. Teachers can present teaching materials or demonstrate hands-on activities on a virtual whiteboard.

In a PE classroom, the teacher demonstrates some physical exercises while students follow along in front of their respective cameras; the body movements of students and the teacher are transferred to their virtual avatars in the virtual classroom, creating an "exercising-together" scene in the system, as shown in Figure 1.

To use SAILS, users need to equip their personal computers with two cameras, as shown in Figure 2. The computer should have basic

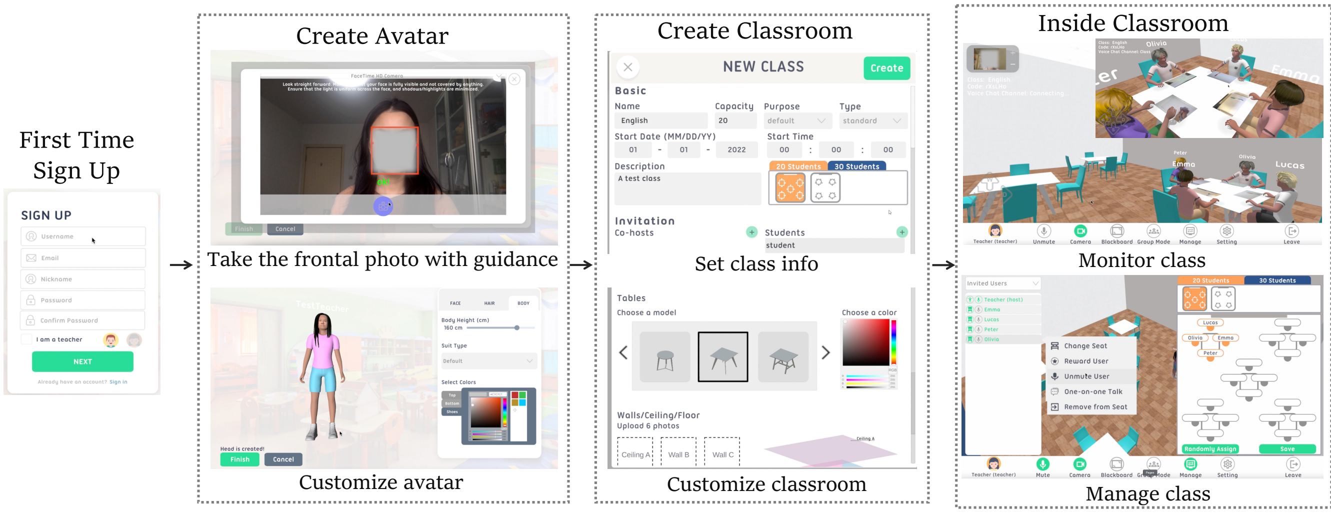

Figure 3: Application flow for a first-time teacher user in SAILS. Upon sign-up, users create and customize their avatars with guided instructions. Teachers can set up and personalize classrooms, and invite students to join. In the classroom, teachers can instruct, monitor, and interact with students.

graphic rendering capabilities. One camera should be front-facing to capture the user's face when they are sitting in front of the desk. This camera can be either the built-in camera or a webcam attached. The other camera should capture the working area on the user's desk. It can be an additional webcam connected to the computer. For the PE class, only one camera is required.

The application workflow for first-time users is illustrated in Figure 3. We use a teacher account as an example, as it includes more management functions. Upon sign-up, users follow the guidance to create and customize their avatars. Teachers can set up new classrooms by filling in basic class information and customizing the classroom to resemble a physical one. Students invited by teachers can then access the classroom. In the classroom, teachers can conduct lessons using a virtual whiteboard, manage the class in the management portal, and walk around to monitor and interact with students. Students can see their peers working together around a table, chat with them, raise hands to ask questions, send emojis to express their feelings, and receive rewards from teachers.

In the following subsections, we present the core components of SAILS: immersive modeling, interactive communication, and the "togetherness" design.

3.2 Immersive Modeling

Students would have a higher degree of classroom engagement if they were allowed a sense of presence with their teachers and classmates [1], possibly in a familiar classroom environment, and a visual realism with virtual avatars representing their teachers and peers. In addition, non-verbal cues such as body movements and facial expressions are vital for humans to interact and communicate with each other [2]. In a traditional classroom, teachers heavily rely on these cues for comprehending students' responses, especially when managing multiple students simultaneously.

To create an immersive virtual learning environment, we utilize 3D modeling to replicate the real classroom in a virtual space. We aim to ensure that the users' virtual representations (i.e., avatars), particularly their heads, are rendered as realistically as possible to promote engagement. Furthermore, we incorporate behavioral hints with real-time facial expressions and movement transfer techniques to enable a more genuine classroom learning experience.

3.2.1 Classroom Modeling. The SAILS platform provides a virtual 3D classroom environment that can be personalized to meet diverse teaching needs and create a sense of familiarity to students. To achieve this, our system offer a virtual reconstruction feature that allows teachers to recreate the physical classroom. This process involves collecting images of the walls, floor, and ceiling of the actual classroom, and applying them as textures in the virtual classroom. Teachers can also choose from various 3D table and chair model and customize colors based on their preferences. Moreover, they can select from different table/chair layouts, each with varying group division settings.

3.2.2 Virtual Avatar Modeling. Our technology allows users to create realistic and personalized avatars through advanced head modeling, hairstyle classification, and appearance customization.

Head modeling is a core component of the virtual avatar, as it enables users to recognize each other within the virtual environment. We adopt the FLAME [18, 41] morphable model structure, which generates a complete head model in one step from a single frontal face image. The process involves performing landmark detection to identify positions of facial features, then extracting the skin texture and constructing a 3D mesh that approximates the face in the 2D image. For facial texture generation, we segment the face area to create a facial mask and apply post-processing to obtain a natural skin color representation.

As the generated head model does not include hair, we implement a hairstyle classification function utilizing a Vision Transformer (ViT) [9] trained on a custom hairstyle dataset. It automatically suggests the closest matching hairstyle, represented by a realistic 3D hair model, based on the input image. This process reuses the single headshot photograph captured during the head modeling phase. Young learners can choose a hairstyle from our selection based on their preferences for a more enjoyable and diverse range of avatar appearances.

Beyond hairstyle, a user can adjust their hair color, set the height of their avatar body, select clothing styles, and customize clothes colors accordingly. This level of avatar customization not only enhances interaction within the system but also suits individual user preferences, establishing their individuality.

3.2.3 Facial Expression Transfer. The transfer of user expressions and movements to the virtual classroom heavily relies on camera capturing and image processing techniques. To detect facial, hand, and body landmarks and orientations from RGB video frames, we have chosen MediaPipe [20] and its provided models. The face landmark model, BlazeFace [4], provides real-time facial information that can be transferred to the virtual environment. To facilitate the representation of user expressions, we have developed new facial blend shapes that encompass a range of expressions, such as mouth opening, smiling, and blinking. By utilizing facial landmark data and calculating the vector difference between these blend shapes and the base model, we are able to dynamically display users' various expressions in real-time, enabling a more immersive virtual classroom experience.

3.2.4 Body Movement Transfer. Reproducing the user's body movements on the virtual avatar is an essential feature of our system, aimed at supporting students and teachers in PE courses. We utilize body pose landmarks obtained through Mediapipe's BlazePose [3] from the user's captured videos, which provides position information for each joint in image coordinates. We apply basic operations such as smoothing and normalization to the body joint landmark data as initial preparation for animation.

To animate the virtual avatar, we utilize Unity's Animation Rigging package [31], which uses Multi-Aim Constraints [32]. These constraints enable procedural motion by rotating the parent joint to face the target position specified by the child joint, at each parent-child joint level. However, the absence of ground-truth bone length measurements for users introduces variability in the scaling relationship between the user and their virtual avatar, particularly when the user is moving towards or away from the camera. Directly using the detected joint landmark position as the aim target can result in incorrect bone rotations. To address this issue, we introduce a vectorization process that generates a direction vector from the parent joint's landmark position to the child joint's landmark position. We set the aim target of the parent joint to its current position plus the direction vector. This process effectively maps rotation information from the detection results to the virtual avatar using position data only, without prior knowledge of bone length.

In conjunction with the vectorization process for joint mapping, we have integrated the Unity physics engine [33] to resolve differences between landmark position coordinates and the Unity game scene coordinates. Specifically, we apply gravitational forces to the avatars within the scene, utilizing the unity rigid body and colliders to simulate physical interactions between the avatar and the ground surface. We also incorporate velocity-based foot state detection to determine the vertical position of the virtual avatar. This integration ensures a more realistic representation of the virtual avatar's movements.

3.2.5 Hand Movement Transfer. In classes where students engage in hands-on activities, capturing and reconstructing the movement of their hands becomes crucial as it allows teachers to easily monitor students' activities and provide assistance if necessary.

The Mediapipe Hands model [40] is used to produce hand landmarks and predict the relative position between the hand and the camera. Considering that each person's palm size and finger length differ, and the virtual character's hands are predefined during the 3D modeling process, we implement a vectorization process, similar to the method used on pose detection data to ensure accurate alignment between the captured hand data and the virtual character's hands. This involves generating a direction vector based on the hand joint data and establishing an aiming target using Animation Rigging's Multi-Aim Constraint [32]. This guarantees precise local movements of the virtual character's hands.

In the real world, the joints of the fingers are limited by certain angles, typically ranging from 0 to 90 degrees. We incorporate this physical constraint into the virtual hands by calculating the angles between each joint vector derived from the hand landmarks. By imposing restrictions on the angles of the hand joints, we prevent excessive bending of the finger joints and ensure that the virtual hand movements are more natural and realistic.

Furthermore, we apply constraints on the virtual hand positions. The virtual hands closely follow the movements of real hands in the captured video shown on the virtual table. We also smooth the data to eliminate jittering.

3.3 Interactive Communication

We implement various communication means using images, videos, audio, and others for excellent interactivity in SAILS.

3.3.1 Virtual Whiteboard. Teachers can upload slides, deliver presentations, and perform live demonstrations on the virtual whiteboard from their local devices. The virtual whiteboard is displayed at the front, mirroring the physical classroom setup.

3.3.2 Desk Video. Each student's physical desk workspace is projected onto the virtual desk. Teachers can always have an unobstructed view of the students' work areas, regardless of the camera angle. Utilizing OpenCV [5], we have implemented customization of workspace calibration and selection, allowing users to specify specific areas they wish to be featured in the live video. This respects their privacy, as all processing related to this camera runs only on their local devices. Only the cropped image/video is uploaded to the cloud and shared with teachers.

3.3.3 Voice Chat. Beyond visual elements, we also implement voice chat functionalities using Vivox [34]. This feature allows teachers to communicate with the entire class, selected groups of students, or individual students using voice channels. By providing voice communication, we promote effective audio information exchange and facilitate diverse modes of voice communication.

3.3.4 Others. We further design a variety of clicking-based means to enhance teacher-student interactions and simulate essential class features such as raising hands and giving rewards. Students can raise hands and express their sentiments in emojis by clicking dedicated buttons in the virtual classroom scene. Teachers can reward outstanding students in a similar way.

3.4 "Togetherness" Design

Video conferencing software provides essential tools for fulfilling basic teaching and learning requirements. Nevertheless, the absence of "togetherness" between teachers and students during online classes is a significant challenge. For instance, teachers often find themselves navigating through numerous small windows, monitoring individual students one by one. This fragmented view not only consumes valuable class time but also hampers meaningful interactions. The majority of classroom interactions tend to become one-way, leading to a shift towards passive learning.

Our system aims to infuse synchronous online teaching and learning with a sense of togetherness. Our design presented above has enabled the immersive and interactive experience in SAILS. We create a virtual environment that closely resembles physical classroom settings. This replication fosters a shared sense of belonging among students, allowing them to feel connected and engaged in online classes. Meanwhile, teachers can walk around, monitor student attentiveness and closely track learning progress, in a virtual classroom.

3.4.1 Gather Together. Students' engagement level increases when they experience a deep sense of "togetherness". Imagine a classroom where spatial orientation mirrors the physical setting — a space where teachers and classmates occupy familiar positions, akin to the arrangement in a physical room. This shared spatial context cultivates a feeling of unity, creating an environment where students feel connected and familiar.

By employing realistic classroom modeling techniques, teachers can design the virtual classroom to closely resemble real-life settings. The class is divided into groups, with each group sitting around a table within the same virtual space. Throughout the class, students are assigned specific seats and cannot move freely. Only the teacher has the authority to assign students to different seats and walk around. Although students cannot move around in the virtual classroom, they can visually explore their surroundings by looking around. They can observe the facial expressions and body gestures of their groupmates as in a physical classroom, mirroring the dynamics in a physical classroom.

3.4.2 Learn Together. SAILS utilizes both physical and virtual spaces to enhance online education for young learners. The students and teachers exist in a shared virtual space within the system, and aspects of their physical space are projected into this virtual space. This allows students to actively participate in class activities in a physical manner, while having an intuitive way to interact with their teacher and classmates through the virtual space.

During classes, teachers can utilize the virtual whiteboard to display and illustrate lecture materials. They also have the option to enable a full-screen mode, temporarily disabling the 3D scene and directing students' attention solely to the virtual whiteboard. This ensures students focus on the lecture material being presented.

In addition to one-way teaching, in-class activities are beneficial for students to understand the knowledge deeply and allow excellent opportunities for teachers to offer feedback and support to students based on their individual needs. For young learners, these activities often involve hands-on activities such as writing, drawing, or making handcrafts. To accommodate these activities, SAILS allows students to physically work on their desks, with the second camera capturing their activities and displaying them in real-time on the virtual table. Students view the 3D scene from a first-person perspective without seeing their own avatars.

With real-time hand movement transfer, behavioral cues are available inside the system. Teachers can monitor multiple students' engagement by observing their virtual hand movements simultaneously. Teachers can also thoroughly check each student's work by "walking" to their virtual avatar and viewing the live video displayed on the table. This enables teachers to assess student progress and provide personalized instructions and tailored education to individual students. Additionally, the system enables one-on-one chat between a teacher and a selected student to further enhance communication and address specific needs.

Peer-to-peer interactions play a crucial role in student development as well, which promote learning from peers, collaborative problem-solving, and the development of social skills [30]. Though not being able to move freely, students can look around and observe their peers' activities and facial expressions, creating an environment similar to a physical classroom. This setup helps keep the class organized. Additionally, the design of group-mode voice chat further enhances group exchange. Students seated together at a virtual table can engage in group discussions and collaborative conversations without disrupting other groups.

3.4.3 Exercise Together. SAILS extends the traditional classroom by offering virtual PE classes. The teacher can assign a standing point for each student. Students can do workouts in front of their respective device's camera. Their movements are captured in real time and transferred to their virtual avatars. All virtual avatars exercise in the same virtual classroom, forming an "exercise-together" scene. Similar to normal classes, students can only stay in place, while watching the live demonstration from the teacher or seeing all the classmates working out together.

3.5 Implementation and Capability

3.5.1 Client Application. The above sections mainly describe the client application of our SAILS system. Developed in Unity, our client application is designed to operate on both Windows (64-bit, Windows 10 or above) and Mac computers (MacOS 12 or above) with two cameras (360p or above). The home devices running our application do not have to be equipped with high computational power but should possess basic graphics rendering capabilities.

3.5.2 Server Implementation. The primary functions of our system's server backend include managing classrooms and users, distributing data to clients for individual 3D scene rendering, and executing high computational tasks (e.g., avatar head modeling).

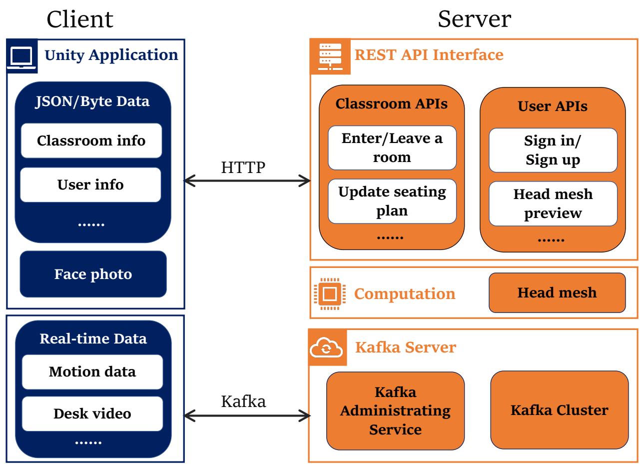

Figure 4: Client-server architecture of SAILS. The client Unity application sends JSON/byte format classroom and user information, as well as face photos, to the server via HTTP. The server processes the data (e.g., generating head meshes) and sends back the appropriate responses via REST APIs. Real-time data, such as motion data and desk video, are sent to the server via Kafka for distribution, ensuring efficient and prompt data exchange.

We employ Apache Kafka [11] as the backend framework, enabling efficient and effective handling of multimedia data sharing with client applications in real-time. Our system adaptively allocates processing and rendering tasks according to the processing capacity of client devices and the available network bandwidth between the client and the server. Computationally demanding offline tasks are delegated to the server. For instance, during the head modeling phase, the client application captures a frontal facial photo, which is then sent to the server for further processing and modeling. The server returns the result to the client once it has finished computation. To ensure smooth interactions, we strategically shift most of the real-time rendering and processing tasks from the server to the client. The client application handles real-time data detection and processing, including facial expression, and hand and body movement. The processed data is continuously transmitted to the server backend for distribution, ensuring synchronized experiences across users. A similar approach is adopted for desk video live streaming.

3.5.3 Client-Server Architecture. The server and client are connected through a combination of HTTP and Kafka protocols, as shown in Figure 4. We designed different communication protocols based on the data type.

For data that does not require continuous real-time synchronization, such as classroom and user information in JSON/byte format, we use the HTTP protocol for communication between the client and server REST API interface. When the client (Unity application) needs to send or request information (such as classroom info, user info), it makes HTTP requests to the server. The server processes these requests and sends back the appropriate responses. Specifically, the face photo is sent to the server via HTTP, where high computational head mesh computation is performed, and the generated head mesh result is returned to the client.

For data requiring real-time synchronization, such as motion data and desk video, we use the Kafka protocol. The client application processes motion data and desk video from webcam inputs and sends them to the server via Kafka. Data are queued and distributed by the server, which handles clustering and services for data transmission, maintaining seamless communication. Kafka serves as a message broker to manage real-time data streams between the client and server, ensuring efficient and timely data exchange.

3.5.4 Design Iteration and User Feedback. During development, we conducted iterative participatory design processes to effectively address users' needs and facilitate the adoption of technology in the classroom. To create user-friendly designs that meet the online teaching and learning requirements of young learners, we carried out periodic pilot trials with three young student teachers, four teenage students, and two young students under 10. By gathering their insights, preferences, and requirements on usability and in-class functions, we were able to refine and improve the design. We applied and enhanced existing technologies to address the specific challenges of this scenario, ensuring that the system is tailored to the needs and preferences of its users.

For instance, to ensure accurate body movement transfer for underage users with different skeletal proportions compared to adults, we developed a vectorization process to map their poses to 3D avatars accurately. Considering various user environments, such as suboptimal lighting conditions, we implemented a skin correction process for face modeling to obtain a more robust and natural skin texture.

In addition to addressing their specific needs, we made adjustments based on their feedback to ensure ease of use and interactivity. For example, we ensured that moving the avatar is intuitive and incorporated suggestions from students to refine interactive methods. These iterative cycles of gathering user feedback and making improvements were repeated several times until we finalized the design.

4 Experiments

To comprehensively evaluate the performance of SAILS, we conducted a user study involving 40 kindergarten and primary school students as well as their teachers. The study consisted of both subjective and objective evaluations. For the subjective user study, participants were asked to share their experiences of taking classes on the system, focusing on aspects such as system design, usability, and their feelings. For the objective evaluation, we established a server statistics monitoring stack to capture and analyze system data. Metrics such as latency, throughput and CPU utilization were measured to assess system efficiency and stability.

4.1 Subjective User Study

4.1.1 Participants. The user test involved 40 students ($M = 8.4, SD = 1.985$) and 6 teachers ($M = 34.3, SD = 13.426$). Participants included 10 kindergarten students ($M = 5.4, SD = 0.516$), 10 lower primary students ($M = 8.6, SD = 0.699$), and 20 upper primary students ($M = 9.8, SD = 0.951$). Half of the taught general subjects, while the

Table 1: Questionnaire. (x) indicates the question is excluded from the kindergarten questionnaire set.

Criteria	Questions
User Experience	
Enjoyment	I enjoy using the app.
Excitement	It is exciting to use the app.
Recommendation	I would like to recommend the app to a friend.
System Usability	
Simplicity	The app is easy to use. (x)
Frequency	I would like to use the app frequently.
System Design	
UI	The app UI is delightful.
Smoothness	The app runs without bugs or errors. (x)
Trustworthiness	I trust the app with my personal information. (x)
Special Features	
Attractiveness	I find the app to be attractive.
Functionality	The app meets my teaching/learning requirements.
Uniqueness	The app offers features that other online teaching tools don't. (e.g., Zoom) (x)

other half taught PE classes. All participants had prior experience with Zoom-like online tools.

4.1.2 Procedures. We provided on-site installation and comprehensive instructions to all participants on how to use the SAILS. Common personal computers, such as MacBook Air, were used. Each participant received a brief introduction to navigate and operate within the system. Additionally, teachers were given specific instructions on how to conduct and manage classes efficiently within the application.

All participants were required to create their own virtual avatars with in-app guidance. They were instructed to take a frontal facial photo and personalize their avatars by adjusting hair color, clothing style, and clothing color. Teachers were also responsible for creating a virtual classroom specifically for testing purposes.

The student participants were divided into batches of ten. They engaged in the same normal and PE classes led by their respective teachers within SAILS. In normal classes, teachers used slides for teaching and included hands-on activities such as writing and drawing related to the learning materials. Teachers were also encouraged to move around within the virtual environment to monitor students' progress during in-class exercises. In PE classes, students actively participated in a series of exercises guided by their PE teachers.

The testing session with each participant group (10 students and 2 teachers) lasted approximately 30 minutes. After experiencing the application, each participant was asked to complete a questionnaire to gather their feedback and insights. For kindergarten participants, instead of having them write their feedback, we conducted short interviews to ask them to rate their experience.

4.1.3 Measurements. We designed a questionnaire to assess the user experience of SAILS, which mostly consisted of rating-based questions. Users were asked to provide ratings on a 5-point Likert scale [19], ranging from 1 (strongly disagree) to 5 (strongly agree). The questions were selected from two established questionnaires, the System Usability Scale (SUS) [6] and the Standardized User Experience Percentile Rank Questionnaire (SUPR-Qm) [28].

We tailored the questionnaire to different participant groups. Kindergarten participants were given a subset of the questionnaire (7 in total) related to their user experience, as they were expected to use the application under parental supervision. Other participants, including teachers, were given a broader set of 11 questions, which included questions about app design in addition to user experience. The questions were organized into four categories, with each category containing two to three questions, as listed in Table 1.

Alongside rating-based questions, we included an optional open-ended question inviting participants to share what they liked most about SAILS. This qualitative feedback provided valuable context and complemented the quantitative ratings.

4.1.4 Results from Students. We utilized statistical inferences to evaluate the rating responses from students and gain insights. Our analysis focused on two key aspects for each question: the mean rating, which represents the general group opinion, and the frequency distribution, which reveals the degree of variability in the groups' opinions.

In analyzing the mean value of 5-point Likert scale questions, the scale interval of each agreement level is defined at 0.8 [29]. If the mean rating falls into the interval $(4.2, 5]$, it indicates that the group "strongly agrees" with the statement. If the mean rating is within the interval of $(3.4, 4.2]$, it signifies the group generally "agrees" with the statement and so on. The results are as follows:

- User Experience: All student participants reported a positive experience using the app, finding it enjoyable ($M = 4.55, SD = 0.783$) and exciting($M = 4.7, SD = 0.648$), and would likely recommend the app for distance learning ($M = 4.25, SD = 1.149$).
- App Usability: The participants agreed that the application was easy to use ($M = 3.87, SD = 1.279$) and that they would use it frequently($M = 4.25, SD = 1.171$).
- App Design: The student group strongly agreed that the user interface was well-designed and delightful ($M = 4.4, SD = 0.810$). The group also generally agreed that the application ran smoothly ($M = 3.7, SD = 1.291$) and could protect privacy ($M = 3.97, SD = 1.272$).
- Special Features: Participants strongly agreed that the new features were attractive ($M = 4.4, SD = 0.900$) and unique compared with other online conference tools ($M = 4.25, SD = 1.080$). The group also generally agreed that SAILS can satisfy their learning requirements ($M = 4.07, SD = 1.388$).

Overall, the mean rating of every question was above 3.7 ("agree" level), indicating that participants had a positive perception of our system on average. Our system was highly rated by the students, particularly in terms of user experience. Students highly rated our system in terms of user experience, with all criteria under this category receiving a rating above 4.2 ("strongly agree" level). This suggests that our system can provide a positive experience. The system design, usability, and special features also received positive feedback from the participants.

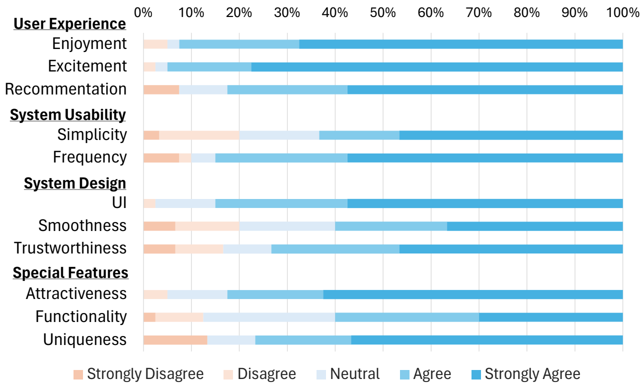

Figure 5: Frequency distribution of students' responses.

To analyze the frequency distribution of individual questions, we counted the number of responses for each rating and represented it in a stacked bar chart, as shown in Figure 5. The results revealed that all distributions leaned towards the positive side. We further conducted an analysis of interrater agreement among all student participants by calculating the Finn coefficient (two-way) as recommended in [24]. The computed value was 0.708, indicating a substantial level of agreement among the participants.

Specifically, less than 20% of respondents expressed disagreement with our system features across all evaluation criteria, indicating that our system was generally accepted by the majority. Notably, more than 90% of participants expressed enjoyment and excitement when using the system. Over 80% of participants were drawn to the special features of our application and expressed a liking for our UI design. The majority of participants expressed a desire to frequently use SAILS as an online learning tool and were willing to recommend it to others. They also highly approved of the unique design. However, we identified areas for improvement. Around one-third of the students agreed that our app had bugs and encountered difficulties in operations. This feedback has motivated us to enhance the stability and user-friendliness of the system.

As for the open question, several young participants expressed their enthusiasm for the customized avatar creation feature. Some also enjoyed using emojis in the normal class and found the PE class fun and interactive. Some mentioned that they could easily know what other classmates were working on, and the virtual classroom environment made them feel like staying in the physical classroom. Their feedback proved that SAILS could raise their interest and improve their engagement in classroom activities. Below are some representative comments from the students, with their participant number and age in parentheses:

- *"I'm excited about creating my avatar."* (Student1, 5)

- *"I enjoy using emojis and participating in the PE class. It's fun and interactive."* (Student2, 6)

- *"I love the pink hair and clothes on my avatar."* (Student3, 5)

- *""In normal class, I can see my friends at the same table and the teacher walking around. It's exciting."* (Student10, 5)

- *"I like the avatar design. I can customize it based on my preferences."* (Student11, 10)

- *"It is convenient and interesting."* (Student29, 12)

- *"I love PE classes the most. They are fun."* (Student35, 8)

4.1.5 Comparison among Different Student Age Groups. We observed that some questions had a high standard deviation, suggesting that some student participants had varying opinions. We then calculated the mean values and standard deviation for the kindergarten, lower primary, and upper primary groups. We found that the kindergarten and lower primary groups had similar results.

For comparison, we categorized participants into two distinct age groups: kindergarten and the lower primary section as the younger group ($M = 7.0, SD = 1.747$) and the upper primary section as the older group ($M = 9.8, SD = 0.951$). The aim was to identify any significant differences between their responses to rating-based questions with large variances.

We conducted a t-test and discovered that the aspect related to system functionality (i.e., "*The app's capabilities meet my learning requirements*") had statistically significant differences ($p < 0.05$) between the two groups. The average rating of the younger group was 4.05 ($SD = 1.099$), and the older group rated the same aspect with an average of 3.45 ($SD = 0.999$).

The results suggested that older students perceived our app as less aligned with their learning requirements compared to their younger counterparts. We hypothesize that this discrepancy arises due to differences in classroom activities: Upper primary students are less engaged in hands-on activities such as drawing or crafting; young learners benefit more from our system, especially when engaged in hands-on activities.

4.1.6 Feedback from Teachers. Teacher participants praised SAILS for providing several advantages that enhanced the overall teaching experience. Below we present a summary of their feedback from both normal classes and PE classes.

Teacher participants in normal classes appreciated the ability to create and customize virtual avatars based on individual appearances and preferences. This feature attracted the interest and curiosity of children, making the classroom environment more engaging. Additionally, the classroom environment settings met fundamental teaching needs, creating a learning space that felt like a real classroom. Students could concentrate on their hands-on activities, and teachers could easily check progress without interrupting students.

- *"SAILS allows users to create virtual avatars based on individual appearances and customize according to preferences. This feature is interesting and attracts children's interest and curiosity. I can see that students are engaged and excited. The classroom environment settings are also very complete, meeting basic teaching needs."* (Teacher1)

- *"Students can attend classes in a virtual environment that closely resembles a physical classroom. Compared to Zoom, using SAILS feels more like participating in an actual classroom setting."* (Teacher2)

- *"It is beneficial to see several students' work displayed on the same virtual table, making it easy to quickly check their progress."* (Teacher3)

Teachers leading PE classes said that the system provided an innovative way to deliver PE classes by capturing movements and transferring them to avatars in real-time. This feature allowed students to engage in physical activities while enjoying the virtual environment. They observed that students were more engaged in exercise when using SAILS compared to Zoom-like software.

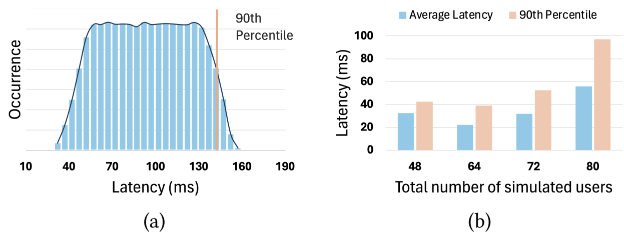

Figure 6: Server result. (a) Server latency distribution in the user study. (b) Average server latency in the stress test.

- *"The idea of capturing movements and applying them to avatars is very appealing."* (Teacher4)

- *"The user interface is interesting. sers can customize avatars, including hair, clothes, and more."* (Teacher5)

- *"Conducting classes in SAILS is more engaging than using Zoom. I noticed that students are more involved and attentive."* (Teacher6)

Due to page limits, rating-based questionnaire results from teachers are included in the supplementary material.

4.1.7 Observations. SAILS is primarily designed for online synchronous learning, where students participate from separate locations. However, during the user study, participants were in the same room and connected to the same Wi-Fi network. This sometimes led to slower rendering and occasional echoes when they simultaneously enabled voice chat. This setting may not have fully showcased the features of SAILS. Nevertheless, we highly value the insightful feedback provided by the participants, as it serves as a driving force for us to further enhance the robustness and usability of the system.

4.2 Objective Evaluation

4.2.1 Procedure. We extracted latency data from server logs and used JMX Exporter [26] and Prometheus [27] to gather throughput data and monitor CPU usage during user study. To evaluate the scalability of the server for real-time applications, we conducted a stress test. Three cloud servers were utilized to simulate several tens of clients in four virtual classrooms. Each server ran multiple instances of Python clients to simulate real-time data transmission. The server utilized was a virtual machine equipped with a 4-core CPU (64-bit Intel Core Broadwell IBRS) and 16 GB of memory, operating on Ubuntu 20.04.6 LTS. It had a maximum disk speed of 150 Mbps and a bandwidth capacity of 10 Gbps.

4.2.2 Measurements. We measured the system's performance using three evaluation metrics: latency, data throughput, and CPU usage. These metrics helped us determine the system's responsiveness in handling high-volume data transmission and complex processing. Latency was measured as the time difference between the server's message distribution timestamp and the client's acknowledgment receipt. The data throughput was the total amount of data output from the server.

4.2.3 Results. We analyzed the distribution of latency obtained during the user study, as shown in Figure 6 (a). 90% of the recorded latency values were below 135ms, indicating SAILS efficiently handled the user load during the study. Addtionally, the average data throughput was 78.96 Mbps, within the network capacity and the maximum disk speed. However, the CPU usage exceeded 250% during the avatar creation process, because of the significant computational overhead associated with head modeling, which also contributed to latency.

The latency results of stress tests under different settings are presented in Figure 6 (b). During the test, the network bandwidth was maintained at 80 Mbps. In our setting, the size of the data transmitted was slightly reduced with an increasing number of simulated users. The 90th percentile latency of SAILS remained below 100ms with 80 users. However, the average latency increased significantly to 1418.3ms with 96 users. SAILS was capable of supporting around 80 users for real-time streaming, with CPU usage remaining below 55%. Detailed experiments are available in the supplementary material.

4.2.4 Observations. Due to the user study setup, users may experience a less stable network connection, resulting in higher latency. Additionally, the high-computational head modeling process overloaded the CPU when multiple users were creating avatars simultaneously. This led to system hangs, where users experienced substantial delays in head modeling. Offloading the head modeling process to a GPU backbone would alleviate this system bottleneck and further enhance the user experience.

4.3 Future Work

4.3.1 User Study. The current evaluation was limited to a single session; future work will include longitudinal studies to assess the system's long-term capabilities and evolving user behaviors. To improve feedback quality, we will integrate structured interviews with post-study questionnaires, as younger participants may express their perspectives more effectively verbally. We will also refine question phrasing to reduce ambiguities. Quantitative metrics (e.g., immersion, engagement, learnability) will be introduced to supplementing subjective feedback.

4.3.2 App Design. To improve accessibility and reliability, we will enhance the app design for diverse age groups, optimize stability under network congestion, and ensure cross-device compatibility.

5 Conclusion

In this paper, we propose SAILS, a Synchronous Accessible Immersive Online Learning System. We incorporate immersive modeling, interactive communication, and "togetherness" design to provide an engaging virtual learning experience in a realistic setting. SAILS can support hands-on activities and in-class interactions for students while enabling easy monitoring for teachers. We discovered a positive preference for SAILS through a user study involving young learners of various ages and their instructors, showcasing the system's advantages in enhancing remote classroom engagement and effectiveness for young learners.

Acknowledgments

This work was supported by Hong Kong Innovation and Technology Commission's Innovation and Technology Fund (Award No. ITS/332/21FP) and Hong Kong RGC grants CRS_PolyU50123 and T43-51323-N (TRS).

References

[1] Leonard A. Annetta. 2006. Creating Presence and Community in a Synchronous Virtual Learning Environment Using Avatars. *International Journal of Instructional Technology and Distance Learning* 3 (2006), 27–41. https://api.semanticscholar.org/CorpusID:208126489

[2] German Barquero, Johnny Núñez, Sergio Escalera, Zhen Xu, Wei-Wei Tu, Isabelle Guyon, and Cristina Palmero. 2022. Didn't see that coming: a survey on non-verbal social human behavior forecasting. In *Understanding Social Behavior in Dyadic and Small Group Interactions (Proceedings of Machine Learning Research, Vol. 173)*, Cristina Palmero, Julio C. S. Jacques Junior, Albert Clapés, Isabelle Guyon, Wei-Wei Tu, Thomas B. Moeslund, and Sergio Escalera (Eds.). PMLR, 139–178. https://proceedings.mlr.press/v173/barquero22b.html

[3] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. 2020. Blazepose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204* (2020).

[4] Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov, Karthik Raveendran, and Matthias Grundmann. 2019. Blazeface: Sub-millisecond neural face detection on mobile gpus. *arXiv preprint arXiv:1907.05047* (2019).

[5] Gary Bradski. 2000. The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 25, 11 (2000), 120–123.

[6] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Evaluation In Industry* 189 (11 1995).

[7] Zhilong Chen, Hancheng Cao, Yuting Deng, Xuan Gao, Jinghua Piao, Fengli Xu, Yu Zhang, and Yong Li. 2021. Learning from Home: A Mixed-Methods Analysis of Live Streaming Based Remote Education Experience in Chinese Colleges during the COVID-19 Pandemic. 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 348, 16 pages. https://doi.org/10.1145/3411764.3445428

[8] Ceren Çubukçu and Cemal Aktürk. 2020. The rise of distance education during Covid-19 Pandemic and the related data threats: A study about Zoom. *Iğdır Üniversitesi Sosyal Bilimler Dergisi* EK SAYI (2020) (2020), 127–144.

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*. https://openreview.net/forum?id=YicbFdNTTy

[10] Haihan Duan, Jiaye Li, Sizheng Fan, Zhonghao Lin, Xiao Wu, and Wei Cai. 2021. Metaverse for Social Good: A University Campus Prototype. In *Proceedings of the 29th ACM International Conference on Multimedia* (Virtual Event, China) *(MM '21)*. Association for Computing Machinery, New York, NY, USA, 153–161. https://doi.org/10.1145/3474085.3479238

[11] Apache Software Foundation. 2023. Apache Kafka. https://kafka.apache.org.

[12] FRAME. 2024. FRAME : Virtual Learning in a browser based campus. https://learn.framevr.io/education-campus.

[13] Magdalena Garlinska, Magdalena Osial, Klaudia Proniewska, and Agnieszka Pregowska. 2023. The Influence of Emerging Technologies on Distance Education. *Electronics* 12, 7 (2023). https://doi.org/10.3390/electronics12071550

[14] Google. 2024. Google Meet. https://meet.google.com/.

[15] Zoom Video Communications Inc. 2024. Zoom. https://zoom.us/.

[16] Jinyoung Kim. 2020. Learning and teaching online during Covid-19: Experiences of student teachers in an early childhood education practicum. *International Journal of Early Childhood* 52, 2 (2020), 145–158.

[17] FUTURE WORK LAND. 2024. Virtual office and class software: Remote team and students management. https://futurework.land/classroom.

[18] Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. 2017. Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)* 36, 6 (2017), 194:1–194:17. https://doi.org/10.1145/3130800.3130813

[19] Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of Psychology* (1932).

[20] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).

[21] Meta. 2024. Meta Quest Safety Information for Parents and Pre-teens. https://www.meta.com/quest/parent-info/.

[22] Microsoft. 2024. Microsoft Teams. https://www.microsoft.com/en-us/microsoft-teams/group-chat-software.

[23] Mozilla. 2024. Hubs - Private, virtual 3D worlds in your browse. https://hubs.mozilla.com/.

[24] Thomas A. O'Neill. 2017. An Overview of Interrater Agreement on Likert Scales for Researchers and Practitioners. *Frontiers in psychology* 8 (2017), 777.

[25] Optima Academy Online. 2024. Optima Academy Online - Online School Reimagined. https://www.optimaacademy.online.

[26] Prometheus. 2023. JMX Exporter. https://github.com/prometheus/jmx_exporter.

[27] Prometheus. 2023. Prometheus. https://prometheus.io.

[28] Jeff Sauro and Pareezad Zarolia. 2017. SUPR-Qm: a questionnaire to measure the mobile app user experience. *Journal of Usability Studies* 13, 1 (nov 2017), 17–37.

[29] Erol Sözen and Ufuk Guven. 2019. The Effect of Online Assessments on Students' Attitudes Towards Undergraduate-Level Geography Courses. *International Education Studies* 12 (09 2019), 1. https://doi.org/10.5539/ies.v12n10p1

[30] Harriet R Tenenbaum, Naomi E Winstone, Patrick J Leman, and Rachel E Avery. 2020. How effective is peer interaction in facilitating learning? A meta-analysis. *Journal of Educational Psychology* 112, 7 (2020), 1303.

[31] Unity. 2024. Animation Rigging. https://docs.unity3d.com/Packages/com.unity.animation.rigging@1.0/manual/index.html.

[32] Unity. 2024. Multi-Aim Constraint. https://docs.unity3d.com/Packages/com.unity.animation.rigging@1.0/manual/constraints/MultiAimConstraint.html.

[33] Unity. 2024. Physics Engine. https://docs.unity3d.com/Manual/PhysicsSection.html.

[34] Unity. 2024. Vivox voice chat. https://unity.com/cn/products/vivox-voice-chat.

[35] ViewSonic. 2024. UNIVERSE by ViewSonic. https://universe.viewsonic.io/.

[36] Virbela. 2024. Virbela: A Virtual World for Work, Education and Events. https://www.virbela.com.

[37] Yuyang Wang, Jean-Rémy Chardonnet, and Frédéric Merienne. 2021. Development of a speed protector to optimize user experience in 3D virtual environments. *International Journal of Human-Computer Studies* 147 (2021), 102578.

[38] Yuyang Wang, Lik-Hang Lee, Tristan Braud, and Pan Hui. 2022. Re-shaping Post-COVID-19 Teaching and Learning: A Blueprint of Virtual-Physical Blended Classrooms in the Metaverse Era. In *2022 IEEE 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW)*. IEEE, 241–247. https://doi.org/10.1109/ICDCSW56584.2022.00053

[39] Eva Lau Yi-hung. 2022. Survey: Nearly 70% of Parents Find Their Children Have Difficulty Learning at Home. https://www.eduhk.hk/en/press-releases/survey-nearly-70-of-parents-find-their-children-have-difficulty-learning-at-home.

[40] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214* (2020).

[41] Wojciech Zielonka, Timo Bolkart, and Justus Thies. 2022. Towards Metrical Reconstruction of Human Faces. In *European Conference on Computer Vision*. https://api.semanticscholar.org/CorpusID:248177832