Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence

Balasaravanan Thoravi Kumaravel[#+], Fraser Anderson[#], George Fitzmaurice[#], Björn Hartmann[+], Tovi Grossman[#‡]

[#]: Autodesk Research, Toronto, Canada; [+]: Computer Science Division, UC Berkeley, USA;
[‡]: University of Toronto, Toronto, Canada

bala@eecs.berkeley.edu, fraser.anderson@autodesk.com, george.fitzmaurice@autodesk.com, bjoern@eecs.berkeley.edu, tovi@dgp.toronto.edu

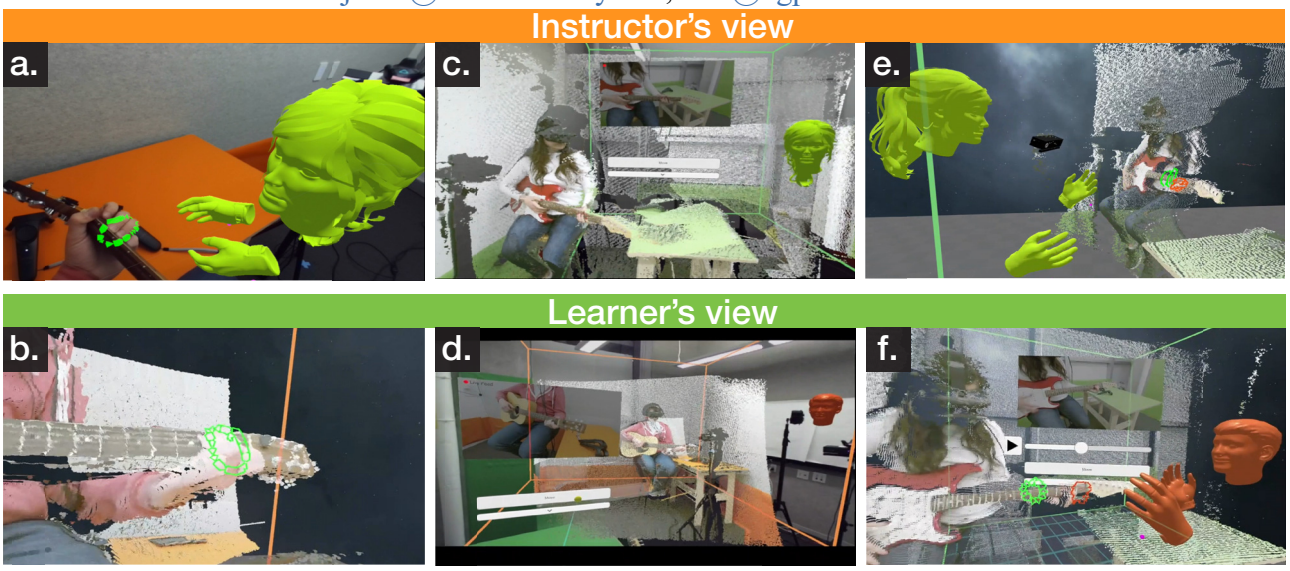

Figure 1: An instructor (orange) teaching a learner (green) how to play a chord on a guitar in mixed reality using Loki. The learner, who is in VR (d), observes the instructor who is in AR (a) demonstrating the chord. The learner uses spatial annotations to ask a question about the performance. Then, both enter AR and the learner begins to practice while the instructor provides occasional coaching (b, e). Lastly, the learner performance is recorded, and both instructor and learner review the recorded performance in VR and discuss the errors (c, f).

ABSTRACT

Remotely instructing and guiding users in physical tasks has offered promise across a wide variety of domains. While it has been the subject of many research projects, current approaches are often limited in the communication bandwidth (lacking context, spatial information) or interactivity (unidirectional, asynchronous) between the expert and the learner. Systems that use Mixed-Reality systems for this purpose have rigid configurations for the expert and the learner. We explore the design space of bi-directional mixed-reality telepresence systems for teaching physical tasks, and present Loki, a novel system which explores the various dimensions of this space. Loki leverages video, audio and spatial capture along with mixed-reality presentation methods to allow users to explore and annotate the local and remote environments, and record and review their own performance as well as their peer's. The system design of Loki also enables easy transitions between different configurations within the explored design space. We validate its utility through a varied set of scenarios and a qualitative user study.

Author Keywords

Mixed Reality; Physical Tasks; Learning; Remote Guidance;

CCS CONCEPTS

• Human-centered computing~Human computer interaction (HCI); • Human-centered computing~Mixed / augmented reality

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
UIST '19, October 20–23, 2019, New Orleans, LA, USA
© 2019 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-6816-2/19/10…$15.00
https://doi.org/10.1145/3332165.3347872

INTRODUCTION

The ability to remotely guide and instruct users in physical tasks has great value due to its ability to connect novices with

experts and improve the way people learn new skills and trades [12, 14]. The ability to re-skill and develop workers effectively is especially important as the nature of work changes and workforces become more dynamic [36]. Within the HCI community, researchers have leveraged novel interfaces such as AR, VR and other modalities to guide or teach physical skills and activities [1, 13, 43, 48, 51]. This research has driven commercial offerings which aim to direct users and provide guidance on job sites and during maintenance tasks [54, 55, 56]. However, these approaches have typically relied on asynchronous learning, tutorial generation, or presenting contextually relevant information such as guidance cues. Additionally, these approaches often rely on a single modality of capture and presentation data streams (e.g., 2D video, AR), in order to teach or guide the remote participant. While these methods can be effective, the spatial nature of physical tasks is often lost or reduced, as is the ability to interact with an instructor in a bi-directional manner. A rich interaction with an instructor can result in tailored guidance and can close the loop by supporting demonstrations by the instructor [7, 26, 41, 53].

In recent years, telepresence technology has advanced rapidly with commoditization of real-time spatial capture devices [57], more prevalent availability of VR and AR interfaces, and novel interactions for mixed-reality (MR) interfaces [27, 32]. These technologies have the potential to augment current training techniques and bridge the gap between instructor and learner by leveraging contextual cues, spatial information, allowing recording and playback of scenes, and enabling spatial annotations. However, it is not evident how to leverage these novel technologies in combination to exploit their unique value.

While prior work has introduced specific configurations of MR-based instruction, we present a broader design space exploring this domain and highlight the importance and utility of moving between the different configurations, based on the learning sciences literature. From this exploration, we develop and introduce Loki (Figure 1), a system for physical task training that supports operations across the different dimensions of the proposed design space. In this work, we will refer to them as '*transitions*'. Loki supports these transitions between the various modalities and data enabled by mixed-reality. These transitions are important to facilitate learning throughout the skill acquisition process, since the learner needs can change even within the course of a single session of learning a physical skill.

Loki is comprised of two symmetric spaces. Each space supports a single user and contains an immersive mixed-reality display utilizing pass-through AR to enable transitions between virtual and augmented reality. The physical environment of each user is spatially captured and streamed in real-time to the remote user alongside video, audio and annotation data. Both users are able to navigate between their local and remote environments in real-time, and also interact synchronously with live as well as recorded data. This flexibility allows for novel workflows that bring the instructor and learner closer together, which in turn allows for richer collaboration and improved training opportunities. We discuss and illustrate the value of Loki's mode transitions and corresponding system features through scenarios performed using the working Loki system. The scenarios illustrate that these additional affordances can allow users to learn a variety of physical tasks in a flexible manner. We also then discuss a qualitative user evaluation of Loki in which users learnt a 3D foam carving task remotely.

The primary contributions of this paper are:

- A design space that explores real-time bi-directional mixed-reality based remote training of physical tasks.
- A set of interaction techniques that allow users to navigate and utilize the breadth of information and presentation modalities within this space as well as enable effective learning workflows.
- The development of a real-time bi-directional depth-capture based mixed-reality telepresence system.
- An initial qualitative evaluation of the utility of mode transitions and the Loki system itself.

BACKGROUND: TEACHING PHYSICAL TASKS

The question of how to appropriately teach physical procedural skills has received significant attention in the learning sciences as well as in specialized domains where such skills are essential to job performance, for example in surgery [38, 39] and athletics [20, 29]. Several learning theories are particularly applicable when designing systems to support skill acquisition.

Fitts and Posner's three-stage model of motor skill acquisition [9] describes a process that begins with a cognitive stage (where movements are actively observed, reasoned and talked about); an associative phase, where some aspects of movement are controlled consciously and some aspects are automated; and the autonomous phase, where movements become fluid, accurate, and largely automatic. Focus shifts from acquiring gross, general movements to finer details through this process.

Collins et al.'s model of *cognitive apprenticeship* [7] highlights the changing role of the teacher in moving a learner from novice to expert performance. It starts with the teacher modeling a desired action while the learner observes, then offering feedback through coaching and scaffolding as learner performs it, and finally a phase for the learner for review and reflection.

Kolb's *experiential learning cycle* [26] also highlights the importance of action and reflection, and distinguishes four stages: having a *concrete experience*, reflecting on it, followed by abstract conceptualization (drawing conclusions), and active experimentation.

Taken together, these theories suggest that any system targeted at teaching physical tasks should be dynamic and fulfill several requirements: allow the learner to observe the teacher (for modeling) and vice versa (for coaching); allowing the teacher to provide effective feedback during a

task performance (live) or after a task (through a recording); enabling abstraction and conceptualization for both teacher and learner (e.g., through annotations and other ways of going beyond direct observation); and supporting a learner's reflection after a task performance, e.g., by jointly reviewing its recording. Loki satisfies these requirements by allowing the learner and instructor to select appropriate views and representations of each other's performance spaces.

RELATED WORK

Loki is grounded on past models of skill acquisition, and primarily builds on prior works in the areas of teaching and learning physical tasks, remote collaboration and immersive physical guidance systems.

Telepresence and Collaboration

There is a large corpus of work investigating remote collaboration and immersive telepresence [10]. Early work in the area focused on remote guidance primarily using 2D video call interfaces with integrated annotations [3, 16, 17]. There have also been prior work that identify the value of collaboration using a mixed-reality setup [47] and have proposed useful extensions such as spatial 3D annotations and tracked objects [33, 47, 50]. Further work by Ishii et. al identified that for a seamless remote collaborative experience, it is not sufficient to have only 2D annotations, but it is also important to have access to both physical as well as digital tools, awareness of gaze and gesture and a way to manage the digital and physical workspaces [18]. We build on these works and offer the ability to provide rich, 3D annotations in the fully captured spatial context on both sides of the telepresence experience.

Prior work in collaborative telepresence has shown the needs as well as benefits of access to multiple viewpoints of the remote collaborator [8, 11, 22, 30, 31, 35, 40, 44, 46]. There have been different techniques proposed to provide these viewpoints, such as providing controls for a mobile camera [8,40] but these approaches can add to the cognitive load of the user, requiring them to understand the spatial layout of each camera and actively manipulate them. Some systems automate viewpoint selection [30, 31, 37], however this may fail in the case of a real-time teaching systems as the learner may have a different learning goal that would require them to prefer one viewpoint over the other.

Lastly, real-time spatial capture has enabled novel interactions that are free from many technological constraints. Room-sized spatial data capture has enabled interactive, dynamic spaces [19, 21] that can capture and respond to users' intents and actions within them, or even modify the digital appearance of these spaces across time [28]. Recent advances to these underlying technologies have allowed for real-time 3D meshing and point cloud rendering enabling true, unencumbered room-scale telepresence [27, 32]. Loki builds on these technologies and leverages them to create a system which aims to improve training of physical tasks through unique combinations of, and transitions between, different technologies and data.

Physical Task Guidance Systems

There has been substantial prior work exploring the use of immersive environments and new technology for teaching physical tasks [1, 4, 5, 29, 48, 49, 51]. Most systems designed to support the teaching of physical tasks are automated systems, with the system providing automated feedback through heuristics or measures computed from comparisons to template actions [12, 48, 51]. These systems often focus on *experiential learning*, allowing the trainees to perform an action and then receive some feedback from their actions which they can reflect on. However, this feedback is often coarse, and may sometimes be inaccurate. This approach may not be appropriate for teaching physical tasks, which often requires the learner to first observe a demonstration by the instructor, perform it themselves alongside the instructor, and then get feedback [7, 15, 25, 49]. Additionally, providing learners with some control over the availability and modality of feedback has been shown to be beneficial for learning [20, 42].

Current AR/VR systems for teaching physical tasks primarily focus only on the psychomotor phase of learning [1, 14, 43]. However, this is just a part of learning physical tasks, another part being the elements of the environment and associated interactions with them. Currently, we are unaware of any systems that fully support *cognitive apprenticeship* [7]. To address this gap, Loki provides a bi-directional interface, where both the learner and instructor can transition between each other's environments to enable modeling, coaching and reflection in a meaningful manner.

DESIGNING REMOTE TEACHING OF PHYSICAL TASKS

The performance and instruction of physical tasks occurs in a particular space across a duration of time. In addition to these two fundamental dimensions, we explore how new AR, VR and spatial capture technology can be used to record, augment and facilitate the remote teaching of physical skills. We present these four dimensions of *perceived space*, *time*, the *display configuration* and *data modality* to form a design space (Figure 2) which can be used to categorize systems.

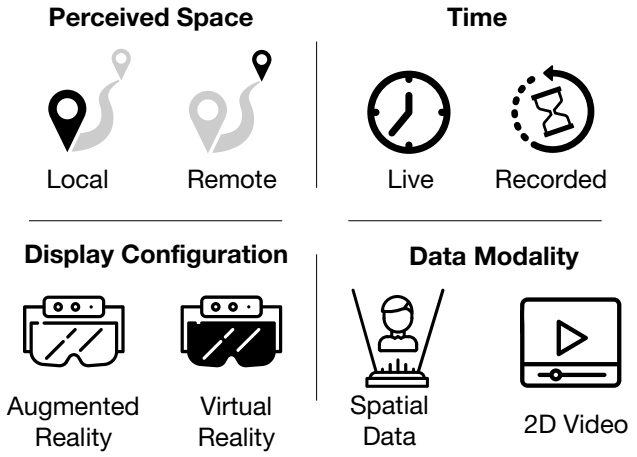

Figure 2: The dimensions of the design space for remote teaching of physical tasks.

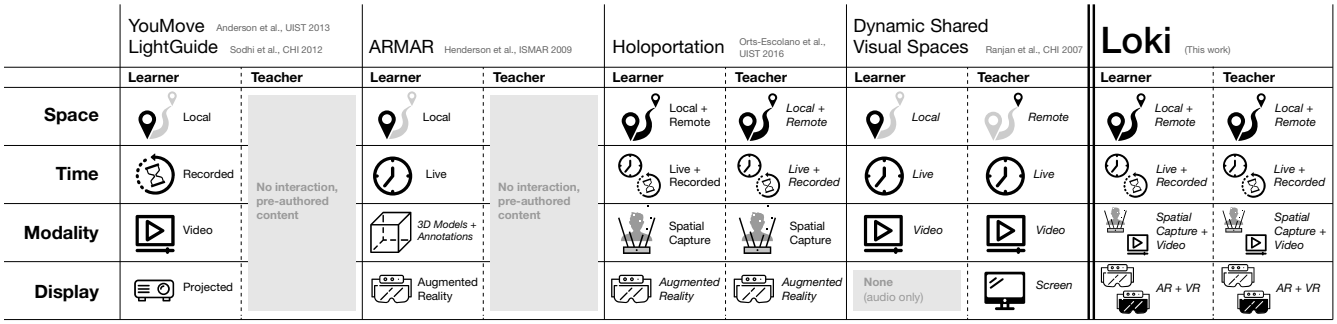

Figure 3: Prior work in the area as they fall within the design space outlined above.

There are other elements of this space that one could explore (e.g., haptics, hand-held AR, embedded sensors), however, this design space focuses primarily on mixed-reality displays which seem to have the largest potential for this domain, though this space could be expanded upon in future work.

Perceived Space

The *perceived space* dimension refers to which space(s), environments, and people the user can currently see and interact with. In the case of a bi-directional interface, each user has the potential to see and interact with their own *local* space, which is the environment that they are physically within and the objects within that space. The user would primarily interact in this space to execute the task or action in their own environment with their own objects or tools. The user may also see the other participant's *remote* space, which is the environment and objects of the other user. In this space, the user can observe, inspect and comment on the remote user's actions, body movements, and their interactions with tools or objects. Additionally, a user may see and interact with *both* their own local space as well as the remote space. With this configuration, a user can see and interact with the remote user as they perform a task within their own local environment, facilitating a 'work-along' scenario. A user may also choose to see *no environment*, and only render the audio and an avatar of the remote user. This configuration can provide a modality to have focused conversation about the task, free from other environmental distractions.

Time

The *time* dimension refers to "when" the data, that the users are interacting with, was captured. The data could be *live data*, in which case the users see a real-time view of their local or remote space. It could also be a *recorded data* stream of their actions which facilitates collaborative review and reflection on any action. In this case, users can navigate the recordings of the data streams and review the local, remote, or both environments. When viewing recorded data, the interactions between the two users are synchronous allowing them to communicate using voice, gesture, and gaze even though the data they are viewing is a recording of the past.

Display Configuration

The *display configuration* dimension refers to how the user can see and interact with the space, which can take many forms depending on the technology available. We explore *augmented reality* as a means to observe and interact with user's own space. This gives the user a direct view of the environment, the ability to interact with it naturally, as well as the ability to augment and annotate the local space. Additionally, augmented reality enables the user to situate the remote person in their own space and interact with them as if they are actually present there. However, when viewing the remote environment or reviewing a recorded data, having an AR view of their current space may be distracting. For these circumstances, Loki offers the ability to enter *virtual reality*, where only the data is rendered, thereby allowing the user to eliminate distractions from their current environment and focus on the data of interest for modeling and reflection. We focus on head-mounted displays, which enable switching between VR and AR. Other display configurations, such as projected or hand-held video-see-through AR, are also possible, but outside the scope of our current investigation.

Data Modality

The *data modality* dimension refers to the type of data collected and used to convey the information. Using spatial data capture, the user can see and interact with the local or remote space using a 3D reconstruction of that environment. This spatial information allows the user to obtain 3D information not readily available through other forms (e.g., video), to navigate to novel viewpoints to avoid issues with occlusion, and to add annotations in 3D space. Additionally, the user could see a *video* of the space. The video can provide a high resolution, easily understood mechanism to comprehend the environment and activities. In the local space, multiple videos can provide varied viewpoints to enable third-person views of the user's own actions and environment. There are many other technologies that can be leveraged to provide novel lenses to view and interact with the spaces (e.g., embedded sensors, infrared or other non-visual imaging, recording audio), however we limit our exploration in this paper to the ones described above.

Summary

This design space can be used to characterize key design decisions in prior work and elucidate important gaps (Figure 3). Notably, most prior work makes a single set of choices along these dimensions and enforces that the learner has one type of display while the instructor has another. For example, remote assistance systems [37] often focus exclusively on

real-time video information sent to the instructor, with only an audio channel for the learner. Some recent mixed reality projects offer access to both live and recorded spatial data [28, 32] but do not offer symmetric affordances for local and remote participants.

A key insight from our background review of the learning literature is that teaching physical tasks involves several distinct phases – observation of the teacher and the learner and vice versa; real-time feedback and reflection on past performance; and shifting foci from gross movements to detailed, subtle actions. Within a single training session, the ability to switch fluidly between these configurations could be very beneficial as the learning environment can be tailored to the optimal communication mechanism for that stage of learning. As a result, the interfaces for both the teacher and the learner need to be flexible enough to support these multiple different modes of interaction.

THE LOKI SYSTEM

Loki enables remote instruction of physical tasks using bi-directional, mixed-reality telepresence (Figure 4). The system comprises two spaces, each of which is equipped with multiple cameras that capture RGB and depth data from the respective spaces. These cameras are tracked, enabling them to be repositioned in the space to allow the remote participant to capture optimal viewing angles. This also enables focus + context interactions [2], providing higher resolution and/or overall context as needed. The user wears a mixed-reality display (HTC Vive + Zed Mini) allowing them to transition between AR and VR and adapt the interface to their needs, performing input via two 6-DOF Vive controllers.

Interface Components

To navigate between the data streams and presentation modalities, Loki supports a variety of interface primitives.

Synchronous Avatar – To facilitate collaboration, Loki offers a real-time rendering of the remote user's relative position from which they observe the local user's space. This is done by rendering a 3D model of a set of head and hands that correspond to the remote user's head and hand poses as tracked by the HTC Vive. The avatar allows the local user to have a sense of what the remote user is seeing. The finer hand movements and gestures need to be observed using the combination of video and a hologlyph, described below.

Synchronous Audio – A real-time audio connection allows the remote and local users to speak and communicate verbally. It also relays ambient audio of the room (e.g., the sounds of power tools, musical instruments, etc.). Currently, the audio is just a single channel and not spatially mapped.

Hologlyph – The spatial data is captured by depth cameras (Kinects) and rendered within a 3D widget (a hologlyph) that can be manipulated by the viewer of that data. In addition to the point cloud of the captured environment, the hologlyph also contains any annotations anchored to that space. A color-coded bounding box outlines the bounds of the hologlyph region. If multiple Kinects are capturing the scene, data from all of them is calibrated and consolidated within a single hologlyph. The hologlyph can be scaled, positioned and rotated within the space using the hand controllers.

Video – The 2D color video streams captured by the Kinects are presented in a floating window above the menu (Figure 4). The window is repositionable within the environment, allowing a user to position it in a convenient spot. The user can switch between the available camera feeds by raycasting at the video with the controller and pressing the trigger. Additionally, when teleporting around the hologlyph, if a user teleports to one of the snap-teleport spots (discussed later) associated with a camera, the camera feed will update with the relevant video feed. While viewing recorded data, the video widget has a scrubbing thumb to allow the user to navigate back and forth through time. Scrubbing this timeline updates the playback time for both the video as well as the data within the hologlyph.

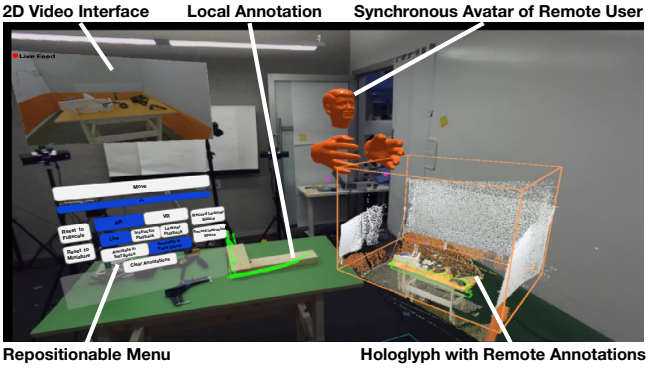

Figure 4: Loki overview, showing the view from the learner's perspective while they are in AR, viewing the remote location live, through the hologlyph and video. (Note: Menu UI text emphasized for figure clarity)

Menu – To control the various modes and features available within Loki, a simple 2D menu is available which is interacted with, via raycasting using the controllers (Figure 4). The menu is spatially linked to the video and can be repositioned or collapsed when not in use to reduce visual obstruction and complexity of the scene. It provides shortcuts to quickly reset viewpoints and scales, rendering the hologlyph in full-scale, or in a miniature view [45]. It also contains buttons to selectively enable recording and playback of the local and remote spaces as well as allows users to switch between AR and VR, depending on which mode might be more relevant to their current task.

Interactions

To navigate the spaces and interact with the content, Loki supports several interaction primitives.

Teleportation – To navigate the hologlyph, the user can transform it using the controllers, walk around in their own space to change their viewpoint, or they can use teleportation. Pressing the center of the trackpad activates a standard projectile-based teleportation ray that allows the user to navigate to any point within the hologlyph. The hologlyph would then enlarge to 1:1 scale placing the user in

the desired teleport location. Users can then adjust their orientation by manipulating the hologlyph with standard "grab, drag and scale" interactions in VR.

Loki additionally supports *snap-teleport* points, which are points where the Kinect cameras are present. These snap-teleports are visualized by green circles, and when the user teleports near these points, the teleport location snaps to that of the *snap-teleport* and the video in the video player switches to that of the corresponding Kinect in that location. This increases the spatial context when choosing an appropriate video feed in multi-camera settings.

Multi-Space Annotations – Loki features a novel bi-directional and context-specific 3D annotation system (Figure 5). Within Loki there are two types of annotations: Annotations by the user in the context of the local space are rendered as solid lines, and those in the context of hologlyph are rendered as outlined lines. For instance, annotations by the local user in the remote space, achieved by annotating within the hologlyph appear outlined to the local user, and in solid line to the remote user in their respective environment. These annotation types are available to both users, with the color (orange or green) denoting the author of the annotation. The type of annotation (local or remote) is determined by context, with annotations created within the space of the hologlyph defaulting to remote annotations and those outside the hologlyph defaulting to local annotations, however this behavior can be overridden using the menu.

These annotations can facilitate communication and feedback in a learning scenario. For instance, the learner can use them to indicate a particular region of interest in their workspace (e.g., where on a workbench they intend to place items), or they can help instructors give guidance or feedback on a learner's actions similar to telestration [6], but within a 3D space. Other forms of annotations, beyond free-hand curves, maybe interesting to explore in the future [34].

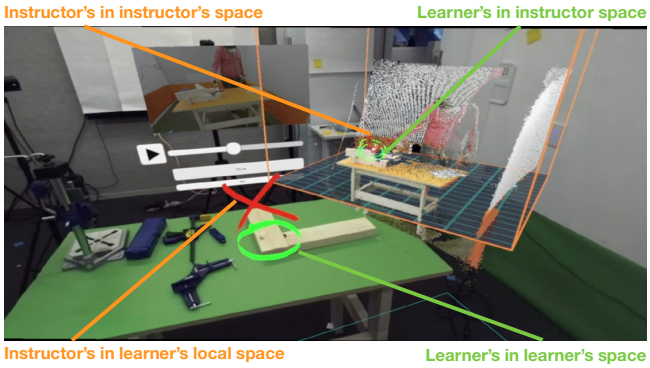

Figure 5: Annotations in the local (solid lines) and remote (outlined) spaces, as made by the instructor (orange) and the learner (green).

Sample Configurations

While the flexibility of Loki offers many different configurations that might be useful depending on the scenario, we believe that there are a few common configurations that will provide substantial value and enable more effective learning workflows.

Observation – this configuration is intended to support the instructor *modeling* the desired behavior and might be most useful within the *cognitive* phase of learning. Within this configuration, the instructor is in AR with a view of their environment, and the learner is in VR focusing on the hologlyph and the video (Figure 6). The instructor would perform the task they are intending to teach, potentially annotating the points of interest. The learner could navigate between videos and around the hologlyph to obtain novel viewpoints and can annotate the instructor's environment as they ask questions. As the instructor can see the avatar representing the learner's viewpoint, they can ensure that the learner is focusing on the right elements. The instructor can also use the avatar to understand the viewpoint that the learner would like to see, and maybe move a Kinect to that location to give the learner a clear video feed from there.

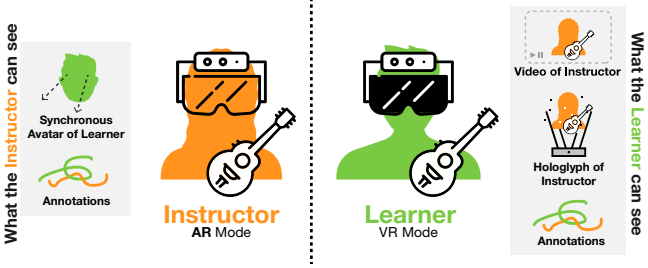

Figure 6: *Observation* enables learners to focus on the instructor who is modeling the desired behaviour.

Instructor Guidance – this configuration is intended to support *coaching and scaffolding*, while the learner has a *concrete experience* [26] and could be most useful in the *cognitive* and *associative* phases of learning. Technically, it is similar to '*Observation*', however the roles are reversed with the learner in AR and the instructor in VR (Figure 7). Within this mode, the instructor can scaffold the learner as they perform the task in their own environment, and provide proactive cues, guidance or feedback on the performance using annotations, voice and gesture.

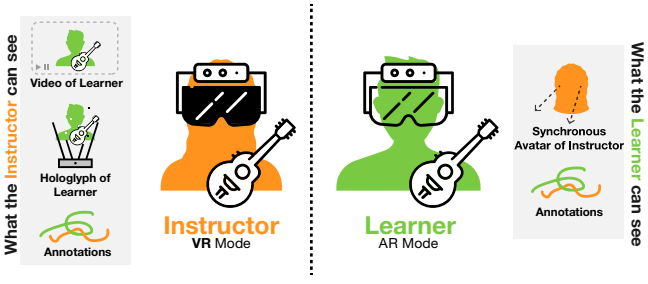

Figure 7: *Instructor guidance* enables the instructor to provide coaching and scaffolding, and for the learner to have a *concrete experience* [26].

Work Along – this configuration is intended to support *active experimentation*, and occasional *feedback through coaching*, and is likely most useful within the associative and automatic phases of learning. Within this configuration, both instructor

and learner are in AR, with the hologlyph and the video positioned so they do not interfere with the primary task (Figure 8). Depending on the spatial layout of the physical rooms and the type of details required, the hologlyph may be a small world-in-miniature sitting on a workbench, or it could be a full 1:1 scale rendering. This configuration allows for constant, low-touch collaboration while performing independent work. The instructor can occasionally check on the learner's progress and interrupt them if necessary, to provide guidance, or the learner can interrupt the instructor if they have a question or need assistance.

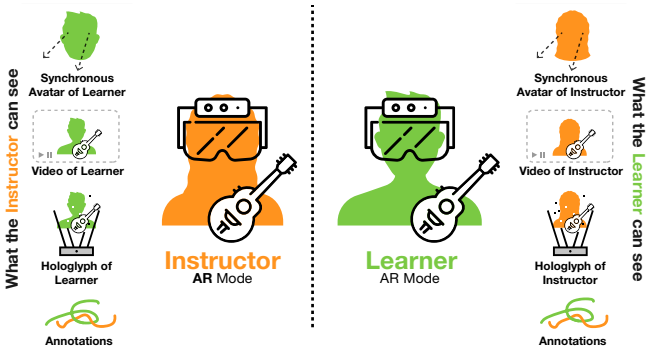

Figure 8: Work along enables feedback through coaching, and allows the learner to actively experiment with the task.

Collaborative review – this configuration is intended to support an *opportunity for reflection* and *abstract conceptualization* and is likely useful throughout all stages of learning. Within this configuration, both the instructor and learner are in VR, viewing a common hologlyph and video stream of the recorded data (Figure 9). This may be a recording of either the instructor or the learner. In this view, both users can see each other's avatar, speak with each other and collaboratively annotate the hologlyph. The instructor could annotate key areas of interest (e.g., errors the learner made, or parts of their own actions) and the learner can indicate locations where they have questions or where more details are needed for understanding the task.

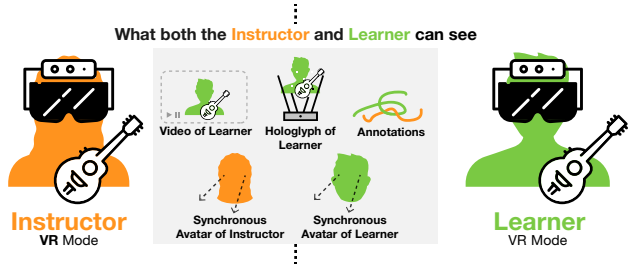

Figure 9: Collaborative review provides an opportunity for reflection, and allows for abstract conceptualization through a shared VR experience.

IMPLEMENTATION

Loki consists of two symmetric hardware systems that each leverage Kinect depth cameras for spatial capture, and an HTC Vive [58] and ZED Mini [59] for the mixed reality displays (Figure 10). The PCs that capture the Kinect data are laptops running Windows with intel i5 chip. The PCs that

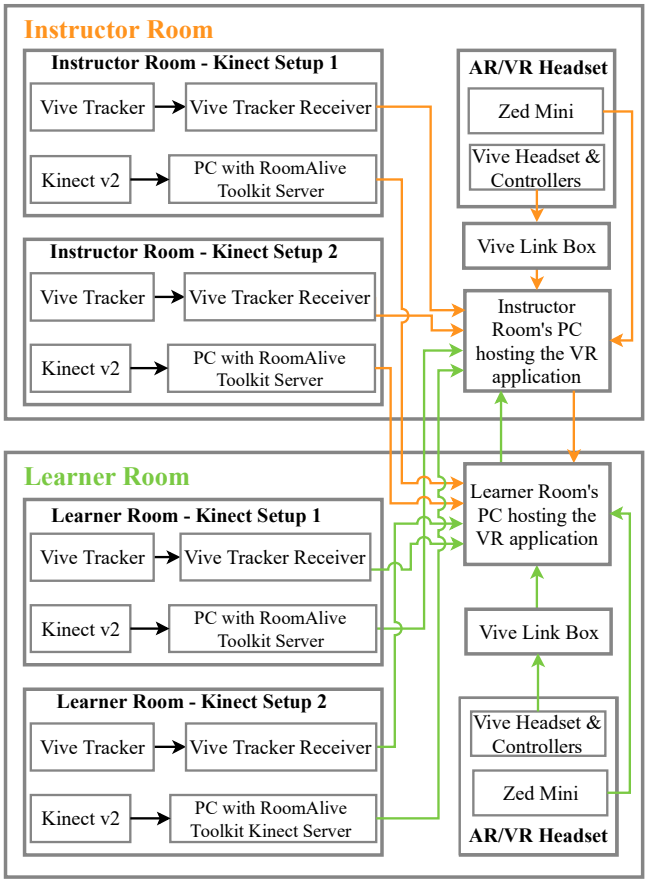

Figure 10: Overview of the hardware configuration of Loki across the two spaces.

render the MR content are gaming PCs with intel i7 chip, nVidia GTX1080 graphics card and Unity 2018.3.

Rendering Mixed Reality

Loki is able to transition between AR and VR using the HTC Vive headset and the ZED Mini. The ZED Mini is a stereo pass-through camera (Having FOV - 85º(H) and 54º(V)) designed for AR applications, capable of depth mapping and lighting estimation. These features allow it to process real-time object occlusions between the virtual and real worlds. To maintain a consistent coordinate system between VR and AR, the ZED Mini's native inside-out tracking is disabled and the Vive's tracking is used instead. As the user switches from AR to VR, the camera feeds from the ZED Mini are disabled, and replaced with the VR camera's render.

As rendering large point clouds in AR can be taxing on the computer, Loki uses custom shaders as well as reduces the update frequency of the point cloud from its native 30fps to 10fps. This allows the AR experience to remain high quality and responsive, while still giving the user enough context about the remote environment. When switching to VR, the framerate is increased providing a better experience when the user's attention is likely focused on the hologlyph.

Communication

The two PCs communicate with each other through a custom TCP/IP Unity plugin. The plugin serializes, sends and

receives, and deserializes the custom data frame that Loki uses to synchronize the experience across both PCs at approximately 66Hz. This data frame includes the users' hands and head positions, their current modes and controller states, the tracked positions of the Kinects, and other lightweight metadata needed to synchronize the systems. The Kinect data (RGB-D data) is transmitted to the remote PC using RoomAlive Toolkit's KinectV2 Server [21] program through a router. The RoomAlive toolkit handles the data capture, compression and decompression of the Kinect data. Audio between the two rooms was transmitted through an IP telecom system.

Aligning Point Clouds

The HTC Vive Headset, controllers and trackers are tracked by referencing HTC's IR emitters mounted in each room. The trackers are mounted to the Kinect cameras and track their positions, which are then used to dynamically auto-calibrate the multiple Kinect feeds at runtime. While the original RoomAlive Toolkit renders a mesh of the scene, the mesh tends to distort the finer details in the scene. This distortion is problematic for teaching some physical tasks where these finer details could play an important role. Therefore, we instead render the raw-point clouds using the Kinect's RGB-D data frames assembled using the toolkit and rendered using a custom shader. We then use a custom auto-calibration script that uses the position data from the trackers mounted on the Kinects to assemble and calibrate the individual point-cloud of each of the Kinect. While there is some offset between the point cloud captured by each Kinect, for many tasks this error does not play a large role, and for tasks where precision matters Loki can be run with a single Kinect to eliminate this offset.

Telepresence and Spatial Synchronization

Spatial consistency is essential to maintain coherent telepresence between the two environments. The user's avatar, as well as annotations made within one hologlyph must be accurately mapped to the augmented reality environment of the other user. To perform this mapping, Loki first computes the relative transform of the desired object to one of the Kinects that render these point clouds, in the coordinate system of the hologlyph. Once that relative transform is computed, Loki renders a virtual copy of the object of interest in the real environment of the remote partner by computing its position with respect to the position of the corresponding real Kinect in the remote room in its own Vive coordinate system. We know the position of the real Kinect in the remote room through the Vive tracker attached to it. Once we have the basic pose, we can then compute and render the scale of the rendered object as per the use case scenario. To transfer an annotation from real space to a corresponding hologlyph, we use the same pipeline of transform but in the reverse direction. It is again important to note that, since we use the dynamically Vive-tracked position of Kinects as references between the two coordinate systems, we could change the position of Kinects and this pipeline ensures that the spatial transforms would still function, enabling a robust telepresence experience.

Shared Playback Space

As the Kinect data and video is bandwidth intensive, it is only stored on one of the PCs. During shared playback, the data needs to be synchronized across both PCs. When a user initiates playback of the remote user's recording, the stream of data is serialized in a binary file on a shared network drive. Once it is copied successfully and the file stream is closed, a 'ready for playback' flag is updated to synchronize both programs. The length of time to synchronize the binary file data varies depending on the length of the recording but is always under 10s for the durations tested (around 30-60 seconds of recorded content). The file is then opened in a read-only mode and copied into a local buffer, while also deserializing and processing the stream to an appropriate data structure to support playback operations such as quickly seeking to an instance.

The synchronized coordination of network read-only streams for playback ensures that both users operate with the same set of file streams when they are in a playback. Following the initial synchronization, a shared immersive telepresence experience is facilitated through the sharing of playback metadata like the time of playback and the video player state as well as the spatially synchronized rendering of virtual avatars and annotations in the coordinate space of the respective playback rooms.

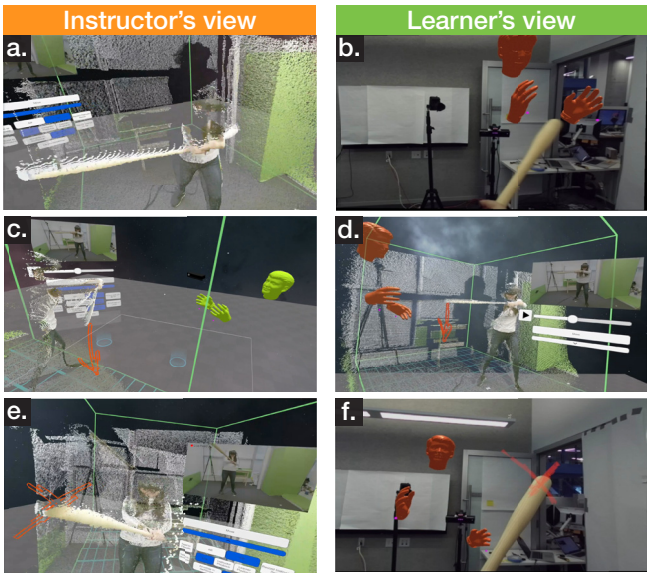

Figure 11: Overview of instructor coaching a learner through learning to swing a baseball bat. The instructor observes the learner's initial swings through VR (a, b), then records their performance for them to reflect on (c, d). After coaching, the instructor guides the swing through a target placed in the learner's AR space (e, f).

SCENARIOS

To validate Loki and the utility of being able to transition between various modes and data within a mixed reality, bi-directional, synchronous instructional experience we implemented and assessed a number of instructional scenarios that spanned Loki's functionality. These scenarios were carried out and tested with Loki by the authors. The

screenshots from those tests are included as figures in the respective scenario subsections.

Teaching Guitar

To instruct a learner on how to play a certain chord (Figure 1), an instructor positions one Kinect near the neck of the guitar, so the learner can view a high resolution video and depth map of the fretboard. The other Kinect is placed in such a manner so as to capture context and body pose in which the guitar is being held and used. Next, they enter an *observation* configuration where the instructor demonstrates the proper fingering for the chord that they want the learner to hold and strums the strings as they play the chord (Figure 1a, d). The learner, in VR, carefully watches the video and point cloud and annotates to ask a question. Next, both users enter AR in a work-along configuration, each with a live fullscale point cloud in front of them (Figure 1b, e). The instructor can watch the learner perform and offers feedback. For instance, when the instructor hears a muffled note, they quickly inspect the point cloud in real time and verbally coach the learner on which finger needs to be moved. Later, the learner is still playing incorrectly, so the instructor and learner enter a collaborative review of the learner's performance, where the instructor scrubs to a particular point in time where the finger looks like it's touching the string and they highlight the error for the learner (Figure 1c, f).

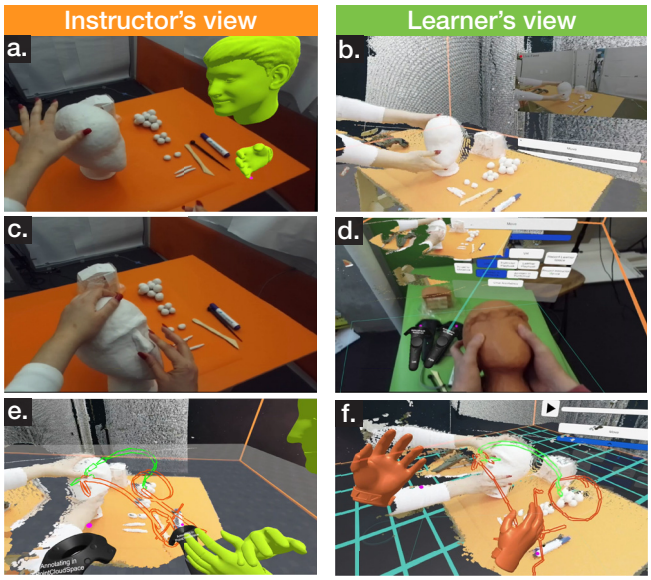

Figure 12: Instructor teaching a novice, the sculpting techniques for modeling a head. The instructor models the behaviour for the learner, who is viewing the demonstration in VR (a, b). Both users work alongside each other, with light coaching and feedback (c, d). After the learner misses a critical step, the instructor enters a collaborative review to highlight the essential elements (e, f).

Coaching Baseball

In coaching a person's baseball swinging action, the learner and instructor enter the *instructor guidance* configuration where the instructor observes the learner and comments on their performance (Figure 11a, b). After a few recorded demonstrations from the learner, the instructor and learner enter a collaborative review to comment on the learner's swing action and indicates through annotations that the action needs to be lowered (Figure 11c, d). Both sides then switch to live data and the instructor annotates a target for the learner to aim for, and offers real-time corrections to overcome their repeated error (Figure 11e, f).

Sculpting

In mentoring a learner on clay sculpting, the instructor and learner enter into an *observation* configuration where the learner gets an overview of the task and the instructor begins by forming the initial shape (Figure 12a , b). Following the introduction, they then switch to the *work-along* configuration, and the learner places the instructor's point cloud and video off to the side as they both work on their own (Figure 12c, d). Since the learner was focused on their own sculpture, they miss a critical step from the instructor. Rather than re-perform the step and spoil the instructor's sculpture, both instructor and learner enter a collaborative review of the instructor's performance in VR where the instructor reviews the steps they took and annotates them to highlight important actions and tools used (Figure 12e, f).

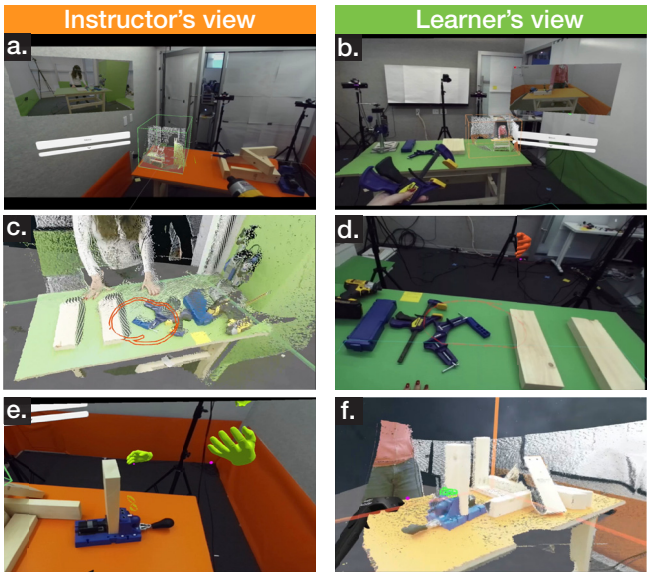

Figure 13: Overview of two peers working independently, providing on-demand mentorship. One peer encounters an issue joining two pieces of wood, and asks his peer for help, who becomes the instructor. The instructor visits the peer in VR to examine their environment and understand the problem (a, b). The instructor then enters a demonstration mode, where they switch to AR and demonstrate possible solutions for joining the wood (c, d). The instructor then provides guidance directly in the learner's space to coach them in how to use the tools they suggested (e, f).

Workshop Learning through Peers

Two members of a woodworking community use Loki while they are working as a means of convenient communication. They have varied skillsets and often share tips with each other. Primarily working in the *work-along* configuration, both users focus on their task and place the remote peer in a miniature scale to the side of their workbench, occasionally

observing their remote peer (Figure 13a, b). When one user encounters an issue, such as uncertainty in how two wooden pieces should be joined, they ask their remote peer for assistance. The remote peer (now the instructor) then enters instructor guidance to observe the environment and context of the user through video and spatial capture. They see that there are several options for the joint, such as metal brackets, pocket holes, or more complex joinery (Figure 13c, d). Both then transition to *observation* where the expert demonstrates various types of points and coaches the user in how to use a pocket hole jig to drill holes in their boards (Figure 13e, f).

EXPLORATORY USER STUDY

We evaluated Loki and the utility of mode transitions by an evaluation where participants learned foam carving to create a 3D foam pyramid shown in Figure 14.

Procedure

The study required participants to learn a hot wire 3D foam carving task over a 30-minute session from a remotely located instructor (an author of this paper), using Loki. Prior to the session, all users were given a 5-minute safety training on the usage of the foam cutter, as well as a 15-minute training session on the Loki system itself. We recruited 8 participants (2 male, 6 female, age range 22-34 years) from within our institution. Participants were compensated with a 50CAD gift card.

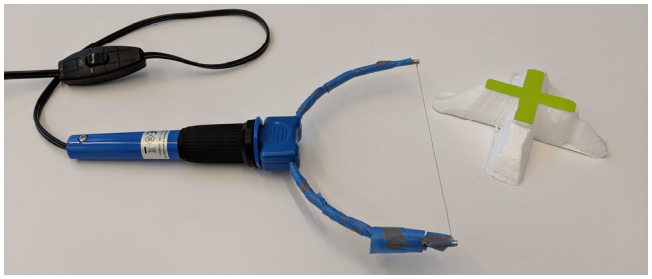

Figure 14: The end goal of the task - A 3D foam pyramid.

Measurement

After the session, the users completed a questionnaire regarding their ability to understand the different elements of the instruction as well as the utility of the different features and modes of Loki. The ratings were based on a 5-point Likert scale. Following the questionnaire, we then conducted semi-structured interviews to better understand the value of the different features of Loki to the participants.

Results and Discussion

In the post-task questionnaire, all users agreed or strongly agreed that the entire system was useful in understanding what was being taught. Using this same metric, 7 of 8 users found value in collaborative review, 6 in point clouds, 4 in videos, 5 in instructors' annotations and 5 in the ability to switch between AR-VR modes. There were variations in how participants preferred to use our system and it varied depending on their personal learning styles and comfort. Users reported that, they would use the system differently depending on the task at hand. Most users first observed in VR, then transitioned to AR to perform the task; P3: *"if I was ... actually building myself and like I definitely need to be in AR mode. But I think VR, it was nice if I was simply watching and didn't want the table and other things to be in the way."*

Participants varied in the way they positioned, scaled and used the hologlyphs during the different stages of learning. Some liked to keep it small and kept it on the side as a reference material, while others preferred it in 1:1 scale directly opposite or beside them. They reasoned about the tradeoffs between the point clouds and the video; P2: *"the point cloud was good because if I miss something in real time, I could just turn around and see a slightly different perspective…..and if you're in a video, you don't want to switch between perspectives, toggle between several videos just to find the right one."*; as well as how those tradeoffs affect the usage of other features like annotations and collaborative playback review; P5: "*point cloud has benefit, you get more 3D perception… you can annotate it in context of the 3D scene*."

Participants appreciated Loki's ability to combine benefits of different features such as annotations, videos, 3D models, collaborative review and playback while also allowing for easy transition across them. Most of them also felt that the system helped them engage better with their partners in the one-on-one learning setting of the study and made the learning process enjoyable. P1 stated that the engagement helps the learning process in an indirect manner; "*It feels more like you're connected to them…I think that it makes the learning process more enjoyable, which would probably help me learn.*" Another participant talked about how this engagement and telepresence gave rise to social dynamics of movements in local versus remote spaces; P7: "*If you invite someone to your house, you feel more comfortable because it's your space and, but if you visit your friend's house, you feel less comfortable because that's another's house…[similarly] when [instructor] visits my space, I feel very comfortable. But when I visit [instructor]'s reconstructed space...I feel like I wanted to keep a social distance and to move in a certain distance that does not make him uncomfortable…Even though I know both are virtual spaces, but I feel different.*"

The study found that participants successfully used Loki in nuanced ways that exercised the different modes to communicate with an instructor within a single learning session. At the end of the study, participants came up with interesting use case scenarios for Loki such as learning activities like cooking, swinging a bat, arts and crafts, origami folds, musical instruments such as flute, and discussing sitting postures with physicians. For these different use cases participants described how the different features of Loki, can be used to accomplish the wide variety of learning outcomes present in these tasks.

LIMITATIONS AND FUTURE WORK

Through our explorations, we have found that there is utility and value in a system that is able to capture and relay spatial data. As different scenarios were examined and developed, a

number of limitations were uncovered and some interesting avenues for future work emerged.

There are some current technological limitations with the system as implemented that we anticipate will be resolved in the near future. The mixed reality hardware itself is somewhat limiting: the headset can be cumbersome, with cables occasionally interfering with the primary task, latency of AR headset and the reduced field of view were restricting to some users. Additionally, the controller interface itself occupies the users' hands and interferes with their performance of the task. In the future, we anticipate that headsets will become wireless and less intrusive, and voice and gesture interactions will become more robust and reliable allowing the system to be used in a hands-free manner. The avatars are currently passive with no gaze and finger movements. This is primarily due to commercially available VR hardware. Gaze tracking is not yet common in VR headsets. While systems (Kinect, Leap) can track body/hand pose, these approaches usually fail when users interact with physical objects, as they do in most physical tasks Loki addresses. Once tracking is reliable, it would be a valuable extension to the avatars. Additionally, in our implementation, users occasionally experience interference between the Vive headset and the Kinect which caused a temporary loss of tracking. Hardware that uses light in different bands could alleviate this issue.

There is a rich space to explore with annotations within this context. Currently, the utility of live annotations is somewhat limited, and they are primarily useful for static objects, as moving objects become misaligned with their static annotations. One area of interest would be examining annotations that snap to content and stay attached even as that content moves through space and time. Additionally, authoring of temporal annotations seem like a rich space to explore. Adapting some of the techniques proposed in prior dynamic illustration work [23, 24, 34] could allow for very rich annotations, or even annotations that the expert and learner could interact with (e.g., creating a virtual baseball that moves along a trajectory and varies its speed).

Currently Loki is bi-directional and only supports connecting two remote spaces. There are several use cases [32] where a one-to-many or many-to-many connection may be useful, such as a distributed peer learning scenario where a number of people are connected in a spatially aware group chat, or a scenario where one instructor is teaching a distributed cohort of learners. While there is some apparent value in these scenarios, managing these spaces and providing intuitive and effective ways of interacting with and managing these spaces remains an open research question.

Lastly, Loki explored the use of spatial data and 2D video as a means to capture and relay the people, objects and environmental context between remote users. While this is a rich set of data, there are many other channels that may be useful, especially when conveying skills that may contain a lot of embedded or tacit knowledge. Sensors to detect force or torque profiles, actuators to enable haptic experiences, or novel methods of abstracting or presenting the captured data may prove to be valuable in capturing and relaying skill-related information.

CONCLUSION

In this work, we have introduced a broader design space for exploring the domain of MR-based live instruction. We then presented Loki, a system that supports this flexible exploration for remote teaching of physical skills. By supporting a range of modalities and various mechanisms for data capture and rendering, Loki provides a rich communication medium that leverages spatial data, video, annotations and playback that helps connect people as they teach and learn real-world tasks. We showed the value of these different features by describing a variety of scenarios we carried out, from teaching guitar to aiding in sculpting and peer learning. We then described a qualitative user evaluation which showed that users were able to use Loki and found the different features and modes of Loki valuable. While some limitations to this technology exist, there is a range of interesting research questions that have emerged from this exploration.

ACKNOWLEDGEMENTS

We thank Roya Shams-Zadeh-Amiri for being a super-user of Loki and helping us with testing and carrying out the different Loki Scenarios used in the paper. We thank Justin Matejka for assistance with figures and for feedback.

REFERENCES

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

[2] Patrick Baudisch, Nathaniel Good, and Paul Stewart. 2001. Focus Plus Context Screens: Combining Display Technology with Visualization Techniques. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology* (UIST '01), 31–40. https://doi.org/10.1145/502348.502354

[3] M. Bauer, T. Heiber, G. Kortuem, and Z. Segall. 1998. A collaborative wearable system with remote sensing. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, 10–17. https://doi.org/10.1109/ISWC.1998.729524

[4] S. M. B. I. Botden and J. J. Jakimowicz. 2009. What is going on in augmented reality simulation in laparoscopic surgery? *Surgical endoscopy* 23, 8: 1693–1700.

https://dx.doi.org/10.1007%2Fs00464-008-0144-1

[5] Sanne M B I Botden, Sonja N Buzink, Marlies P Schijven, and Jack J Jakimowicz. 2008. ProMIS augmented reality training of laparoscopic procedures face validity. *Simulation in healthcare: journal of the Society for*

Simulation in Healthcare 3, 2: 97–102. https://doi.org/10.1097/SIH.0b013e3181659e91

[6] Andrius Budrionis, Knut Magne Augestad, Hiten RH Patel, and Johan Gustav Bellika. 2013. An evaluation framework for defining the contributions of telestration in surgical telementoring. *Interactive journal of medical research* 2, 2. https://doi.org/10.2196/ijmr.2611

[7] Allan Collins, John Seely Brown, and Susan E. Newman. 1988. Cognitive Apprenticeship: Teaching the Craft of Reading, Writing and Mathematics. *Thinking: The Journal of Philosophy for Children*. https://doi.org/10.5840/thinking19888129

[8] Fabian Lorenzo Dayrit, Yuta Nakashima, Tomokazu Sato, and Naokazu Yokoya. 2014. Free-viewpoint AR human-motion reenactment based on a single RGB-D video stream. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, 1–6. https://doi.org/10.1109/ICME.2014.6890243

[9] P.M. Fitts and M.I. Posner. 1967. *Human performance*. Brooks/Cole, Oxford, England.

[10] Henry Fuchs, Andrei State, and Jean Charles Bazin. 2014. Immersive 3D Telepresence. *IEEE Computer* 47, 7: 46–52. https://doi.org/10.1109/MC.2014.185

[11] Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of Head-mounted and Scene-oriented Video Systems on Remote Collaboration on Physical Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03), 513–520. https://doi.org/10.1145/642611.642701

[12] T. P. Grantcharov, V. B. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch Jensen. 2004. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British Journal of Surgery* 91, 2: 146–150. https://doi.org/10.1002/bjs.4407

[13] Steven J. Henderson and Steven K. Feiner. 2002. *Augmented Reality for Maintenance and Repair (ARMAR)*. DTIC Document.

[14] Steven J. Henderson and Steven K. Feiner. 2011. Augmented reality in the psychomotor phase of a procedural task. In *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality* (ISMAR '11), 191–200. https://doi.org/10.1109/ISMAR.2011.6092386

[15] Julian Hough, Iwan de Kok, David Schlangen, and Stefan Kopp. 2015. Timing and Grounding in Motor Skill Coaching Interaction: Consequences for the Information State. *Proceedings of the 19th SemDial Workshop on the Semantics and Pragmatics of Dialogue (goDIAL),* 86–94, Gothenburg, August 2015.

[16] Hiroshi Ishii. 1990. TeamWorkStation: Towards a Seamless Shared Workspace. In *Proceedings of the 1990 ACM Conference on Computer-supported Cooperative Work* (CSCW '90), 13–26. https://doi.org/10.1145/99332.99337

[17] Hiroshi Ishii and Minoru Kobayashi. 1992. ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '92), 525–532. https://doi.org/10.1145/142750.142977

[18] Hiroshi Ishii, Minoru Kobayashi, and Kazuho Arita. 1994. Interactive design of seamless collaboration media. *Communications of the ACM*. https://doi.org/10.1145/179606.179687

[19] Shahram Izadi, Richard A. Newcombe, David Kim, Otmar Hilliges, David Molyneaux, Steve Hodges, Pushmeet Kohli, Jamie Shotton, Andrew J. Davison, and Andrew Fitzgibbon. 2011. KinectFusion: Real-time Dynamic 3D Surface Reconstruction and Interaction. In *ACM SIGGRAPH 2011 Talks* (SIGGRAPH '11), 23:1–23:1. https://doi.org/10.1145/2037826.2037857

[20] C M Janelle, D A Barba, S G Frehlich, L K Tennant, and J H Cauraugh. 1997. Maximizing performance feedback effectiveness through videotape replay and a self-controlled learning environment. *Research quarterly for exercise and sport* 68, 4: 269–279. https://doi.org/10.1080/02701367.1997.10608008

[21] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. 2014. RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-camera Units. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (UIST '14), 637–644. https://doi.org/10.1145/2642918.2647383

[22] Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference on - AH '14*, 1–8. https://doi.org/10.1145/2582051.2582097

[23] Rubaiat Habib Kazi, Fanny Chevalier, Tovi Grossman, and George W. Fitzmaurice. 2014. Kitty: sketching dynamic and interactive illustrations. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 395–405. https://doi.org/10.1145/2642918.2647375

[24] Rubaiat Habib Kazi, Fanny Chevalier, Tovi Grossman, Shengdong Zhao, and George Fitzmaurice. 2014. Draco: Bringing Life to Illustrations with Kinetic Textures. In *Proceedings of the 32Nd Annual ACM Conference on*

Human Factors in Computing Systems (CHI '14), 351–360. https://doi.org/10.1145/2556288.2556987

[25] Iwan de Kok, Julian Hough, Felix Hülsmann, Mario Botsch, David Schlangen, and Stefan Kopp. 2015. A Multimodal System for Real-Time Action Instruction in Motor Skill Learning. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (ICMI '15), 355–362. https://doi.org/10.1145/2818346.2820746

[26] David A. Kolb. 2014. *Experiential Learning: Experience as the Source of Learning and Development*. FT Press.

[27] Jan Kolkmeier, Emiel Harmsen, Sander Giesselink, Dennis Reidsma, Mariët Theune, and Dirk Heylen. 2018. With a little help from a holographic friend: the OpenIMPRESS mixed reality telepresence toolkit for remote collaboration systems. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology - VRST '18*, 1–11. https://doi.org/10.1145/3281505.3281542

[28] David Lindlbauer and Andy D. Wilson. 2018. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 129:1–129:13. https://doi.org/10.1145/3173574.3173703

[29] Helen C. Miles, Serban R. Pop, Simon J. Watt, Gavin P. Lawrence, and Nigel W. John. 2012. A review of virtual environments for training in ball sports. *Computers & Graphics* 36, 6: 714–726. https://doi.org/10.1016/j.cag.2012.04.007

[30] Kazuya Nakae and Koji Tsukada. 2018. Support System to Review Manufacturing Workshop Through Multiple Videos. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion* (IUI '18 Companion), 4:1–4:2. https://doi.org/10.1145/3180308.3180312

[31] Mamoun Nawahdah and Tomoo Inoue. 2012. Motion Adaptive Orientation Adjustment of a Virtual Teacher to Support Physical Task Learning. *Information and Media Technologies* 7, 1: 506–515. https://doi.org/10.11185/imt.7.506

[32] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (UIST '16), 741–754. https://doi.org/10.1145/2984511.2984517

[33] Doug Palmer, Matt Adcock, Jocelyn Smith, Matthew Hutchins, Chris Gunn, Duncan Stevenson, and Ken Taylor. 2007. Annotating with Light for Remote Guidance. In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces* (OZCHI '07), 103–110. https://doi.org/10.1145/1324892.1324911

[34] Ken Perlin, Zhenyi He, and Karl Rosenberg. 2018. Chalktalk : A Visualization and Communication Language -- As a Tool in the Domain of Computer Science Education. *arXiv:1809.07166 [cs]*

[35] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–13. https://doi.org/10.1145/3173574.3173620

[36] Kausik Rajgopal and Steve Westly. 2018. How Tech Companies Can Help Upskill the U.S. Workforce. *Harvard Business Review*. Retrieved April 5, 2019 from https://hbr.org/2018/02/how-tech-companies-can-help-upskill-the-u-s-workforce

[37] Abhishek Ranjan, Jeremy P. Birnholtz, and Ravin Balakrishnan. 2007. Dynamic Shared Visual Spaces: Experimenting with Automatic Camera Control in a Remote Repair Task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), 1177–1186. https://doi.org/10.1145/1240624.1240802

[38] Richard K. Reznick. 1993. Teaching and testing technical skills. *The American Journal of Surgery* 165, 3: 358–361. https://doi.org/10.1016/S0002-9610(05)80843-8

[39] Hazim Sadideen and Roger Kneebone. 2012. Practical skills teaching in contemporary surgical education: how can educational theory be applied to promote effective learning? *The American Journal of Surgery* 204, 3: 396–401. https://doi.org/10.1016/j.amjsurg.2011.12.020

[40] Mhd Yamen Saraiji, Charith Lasantha Fernando, Kouta Minamizawa, and Susumu Tachi. 2015. Development of Mutual Telexistence System Using Virtual Projection of Operator's Egocentric Body Images. In *Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments* (ICAT - EGVE '15), 125–132. https://doi.org/10.2312/egve.20151319

[41] R. Sigrist, G. Rauter, R. Riener, and P. Wolf. 2012. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review*: 1–33.

https://doi.org/10.3758/s13423-012-0333-8

[42] Roland Sigrist, Jürg Schellenberg, Georg Rauter, Simon Broggi, Robert Riener, and Peter Wolf. 2011. Visual and Auditory Augmented Concurrent Feedback in a Complex Motor Task. *Presence: Teleoperators and Virtual Environments* 20, 1: 15–32. https://doi.org/10.1162/pres_a_00032

[43] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 179-188. https://doi.org/10.1145/2207676.2207702

[44] Maximilian Speicher, Jingchen Cao, Ao Yu, Haihua Zhang, and Michael Nebeling. 2018. 360Anywhere: Mobile Ad-hoc Collaboration in Any Environment using 360 Video and Augmented Reality. *Proceedings of the ACM on Human-Computer Interaction* 2, EICS: 1–20. https://doi.org/10.1145/3229091

[45] Richard Stoakley, Matthew J. Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 265–272. http://dl.acm.org/citation.cfm?id=223938

[46] Hongling Sun, Yue Liu, Zhenliang Zhang, Xiaoxu Liu, and Yongtian Wang. 2018. Employing Different Viewpoints for Remote Guidance in a Collaborative Augmented Environment. In *Proceedings of the Sixth International Symposium of Chinese CHI*, 64–70. https://doi.org/10.1145/3202667.3202676

[47] Zsolt Szalavári, Dieter Schmalstieg, Anton L. Fuhrmann, and Michael Gervautz. 1998. “Studierstube” - An Environment for Collaboration in Augmented Reality. *Virtual Reality* 3, 1: 37–48. https://doi.org/10.1007/BF01409796

[48] Richard Tang, Xing-Dong Yang, Scott Bateman, Joaquim Jorge, and Anthony Tang. 2015. Physio@ Home: Exploring visual guidance and feedback techniques for physiotherapy exercises. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 4123–4132. https://doi.org/10.1145/2702123.2702401

[49] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Björn Hartmann. 2019. TutoriVR: A Video-Based Tutorial System for Design Applications in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). https://doi.org/10.1145/3290605.3300514

[50] Michael Tsang, George W. Fitzmaurice, Gordon Kurtenbach, Azam Khan, and Bill Buxton. 2002. Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. In *Proceedings of the 15th annual ACM symposium on User interface software and technology* (UIST ’02), 111–120. https://doi.org/10.1145/571985.572001

[51] Eduardo Velloso, Bulling, Andreas, and Gellerson, Hans. MotionMA: Motion Modelling and Analysis by Demonstration. *Proceedings of CHI 2013*. https://doi.org/10.1145/2470654.2466171

[52] Joan N. Vickers. 1990. *Instructional Design for Teaching Physical Activities: A Knowledge Structures Approach*. Human Kinetics Publishers, Inc.

[53] Gabriele Wulf, Charles Shea, and Rebecca Lewthwaite. 2010. Motor skill learning and performance: a review of influential factors. *Medical Education* 44, 1: 75–84. https://doi.org/10.1111/j.1365-2923.2009.03421.x

[54] Industrial Augmented Reality | Vuforia | PTC. Retrieved April 4, 2019 from https://www.ptc.com/en/products/augmented-reality

[55] Manufacturing | Microsoft Industry. Retrieved April 4, 2019 from https://www.microsoft.com/he-il/enterprise/manufacturing

[56] Fologram. Retrieved April 4, 2019 from https://fologram.com/

[57] Kinect - Windows app development. Retrieved April 4, 2019 from https://developer.microsoft.com/en-us/windows/kinect

[58] VIVE™ | VIVE Virtual Reality System. Retrieved April 4, 2019 from https://www.vive.com/us/product/vive-virtual-reality-system/

[59] ZED Mini Stereo Camera - Stereolabs. Retrieved April 4, 2019 from https://www.stereolabs.com/zed-mini/