

Innovation in Shared Virtual Spaces for Mental Health Therapy

Jake Backer

Advisor: Professor Soussan Djamasbi

Worcester Polytechnic Institute

May 23, 2021

This report represents the work of one or more WPI undergraduate students submitted to the faculty as evidence of completion of a degree requirement. WPI routinely publishes these reports on the web without editorial or peer review.

Contents

1	Abstract	3
2	Introduction	4
3	Background	5
3.1	Mental Health	5
3.2	Telehealth	5
3.3	Augmented Reality (AR)	6
3.4	AR Therapy	7
3.5	Avatars	7
4	Designing the Application	8
4.1	Infrastructure	8
4.1.1	Networking	9
4.1.2	Movement	11
4.1.3	Audio Communication	12
4.1.4	Immersive Experience	14
4.1.5	Telemetry	15
5	User Studies	17
5.1	Results	17
5.2	Discussion and Future User Studies	19
6	Contribution and Future Work	21
	References	24

1 Abstract

Rates of mental health disorders among adolescents and younger adults are on the rise with the lack of widespread access remaining a critical issue. It has been shown that teletherapy, defined as therapy delivered remotely with the use of a phone or computer system, may be a viable option to replace in-person therapy in situations where in-person therapy is not possible. Sponsored by the User Experience and Decision Making (UXDM) lab at WPI, this IQP is part of a larger project to address the need for mental health therapy in situations where patients do not have access to traditional in-person care. This project investigates the use of Augmented Reality technology to provide a similar experience to in-person care while remaining remote. Augmented Reality is an innovative technology that allows for virtual “holograms” to appear in the real world, blending the real and virtual worlds together. This IQP was focused on creating a minimum viable product (MVP) that could facilitate the basic level interaction between two or more users in a virtual space. By doing so, this IQP served as a basic step toward developing AR teletherapy prototypes for the larger project. A user study was also conducted at WPI (N=48) to solicit students’ opinions on AR technology as it relates to mental health therapy.

Keywords: mental health, therapy, anxiety, depression, augmented reality, HoloLens

2 Introduction

In today's digital economy technology plays a significant role in keeping people healthy. More and more people rely on digital devices to keep track of their physical activity, manage their eating habits, and track sleep. The role of technology in health and wellness became particularly prominent during the COVID-19 pandemic. During this time virtual visits became a regular activity. People used Zoom to meet with medical professionals for physical and mental health. The COVID-19 pandemic, perhaps more than ever, made us aware of the importance providing mental health remotely, when face-to-face visits are not possible. This IQP is part of a larger project that attempts to address the need for remote mental health services for those without access to traditional in-person care by providing an immersive therapy experience that may yield similar results to in-person therapy while providing the convenience of remote therapy. The focus of this IQP is to set up a minimum environment needed to facilitate interaction between a group of users. By doing so, this IQP explores the technical issues that need to be addressed for developing an AR application that enables users to interact in a shared virtual space, which could be used to provide mental health services.

The paper is structured in the following fashion. First, a brief discussion of mental health for younger adults, who tend to be tech savvy and benefit from such mental health services, is provided. Then, the paper provides a brief background about the effectiveness of teletherapy, discusses Augmented Reality technology, and discusses other research into using immersive technologies to deliver therapy remotely. Next, the design process for developing the environment for a basic application and its technical aspects are discussed. Then, the results of a preliminary survey study collecting opinions of students on AR technology for mental health therapy are discussed. Finally, the applications, limitations, and future work for the developed application are discussed.

3 Background

3.1 Mental Health

A major group of people who benefit from mental health services are younger adults. Between 2005 and 2017 rates of depression in adolescents and young adults in the United States rose from 8.7% to 13.2% (Twenge, Cooper, Joiner, Duffy, & Binau, 2019). Additionally, rates of anxiety in young adults rose from 7.9% to 14.6% between 2008 and 2018 (Goodwin, Weinberger, Kim, Wu, & Galea, 2020). Despite the increased rates of depression and anxiety, widespread lack of access to mental health services has remained. The need to access mental health services has become even more pronounced during the COVID-19 pandemic, as 83% of adolescents and young adults surveyed reported worsening of their conditions and 26% of those surveyed were unable to access services as in-person services were cancelled following COVID-19 precautions (Lee, 2020). Technology can help provide these services in remote locations and in situations where the patients are unable to seek in-person therapy. The typical method of delivering remote therapy is through video conferencing, but using other technologies such as Augmented-Reality (AR) may help provide more presence and improve the quality of care.

3.2 Telehealth

For the purposes of this paper telehealth is defined as realtime medical assessments or treatment that occurs when the patient is physically separated from the provider (VandenBos & Williams, 2000). A 2013 study performed by the Department of Psychosomatic Medicine and Psychotherapy at the University of Leipzig, Germany and the Department of Psychology at the University of Zurich, Switzerland showed that internet-based cognitive behavioral therapy for treating depression was just as effective or more effective than traditional in-person therapy (Wagner, Horn, & Maercker, 2014). Another study performed by Queensland Health, the University of New South Wales, and the University of California Davis also showed that both clients and case managers consider teletherapy acceptable with average scores ranging from average to slightly better than average (Griffiths, Blignault, & Yellowlees, 2006). Based on these studies,

it is reasonable to believe that telehealth is an area worth researching.

3.3 Augmented Reality (AR)

Augmented Reality (AR) is a new and innovative technology that allows for more immersive experiences that cross between the real and virtual world. AR headsets create “holograms” that the user can interact with in the real world. For these holograms to provide an immersive and life-like experience, the AR headset must take information from the environment and process it to produce better holograms. One such headset is the Microsoft HoloLens 2. The HoloLens has a large array of outward facing sensors including four tracking cameras and one Time-Of-Flight (ToF) depth sensing camera, as well as inward facing Infrared (IR) cameras for eye tracking and an Inertial Measurement Unit (IMU) (Cooley, Jaz, Miller, Jodben, & Paniagua, 2020). These sensors are used to make a virtual map of the environment which allows holograms to be placed more accurately. This virtual map is composed of a 3D triangle mesh of the environment which allows the software to occlude objects hidden behind the spatial mesh and for objects to be placed directly on the mesh, providing a more realistic experience (Hübner, Clintworth, Liu, Weinmann, & Wursthorn, 2020)¹. Figure 1 demonstrates the extremely high accuracy of the HoloLens. This highly accurate spatial mesh combined with IR cameras for eye tracking and an IMU for positional and rotational head tracking allows for an extremely immersive experience.

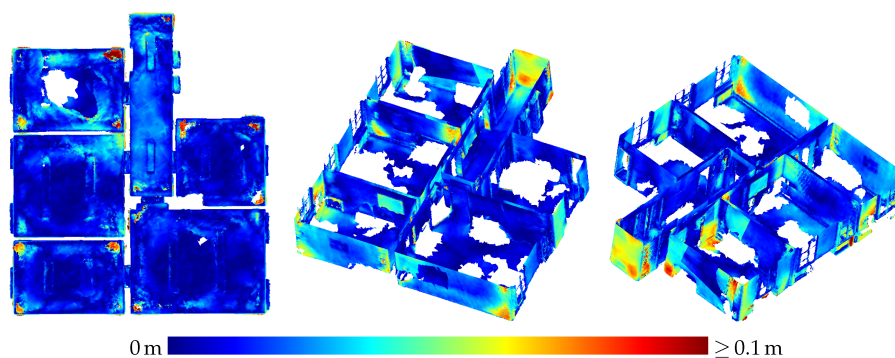


Figure 1: Accuracy of the HoloLens mesh compared to the ground truth. The mean error is 2.3cm (Hübner et al., 2020).

¹Though this article is for the HoloLens 1, these aspects have not changed for the HoloLens 2

3.4 AR Therapy

Immersive technologies such as Augmented Reality and Virtual Reality² are becoming more common and geared towards everyday use (Microsoft, 2020b). Despite consumer immersive devices being quite new, research into virtual reality therapy has been ongoing for the past 20 years (North & North, 2016). It has been shown that virtual reality therapy is extremely effective at treating specific phobias such as acrophobia and claustrophobia as well as social anxiety disorders (North & North, 2016). A key component to delivering effective therapy is the concept of presence (Price & Anderson, 2007). According to research conducted by the Massachusetts Institute of Technology, presence can be measured using 1) The extent of sensory information, 2) The control of relation of sensors to the environment, and 3) The ability to modify physical environment (Sheridan, 1992). Based off these criteria, presence can be increased by providing more sensory information, by making the sensory information more realistic to the environment, and by the ability to physically change the environment. Augmented reality technology provides significantly higher amounts of presence when compared to conventional teletherapy (North & North, 2016).

3.5 Avatars

For the purposes of this paper, avatars are defined as a digital representation of a user's presence in a virtual space (Davis, Murphy, Owens, Khazanchi, & Zigurs, 2009). Avatars allow two users to communicate in a shared virtual space while in two separate locations in the real world. According to a study by Davis et al. (2009), avatars affect a user's experience in a virtual space in three ways: **Representation**, **Presence**, and **Immersion**. **Representation** is defined as the appearance of avatars and their environment as well as how they interact with their environment, **presence** is defined as the sense of being in an environment, and **immersion** is defined as the sense of interacting with the environment (Davis et al., 2009). By providing an application where these senses are maximized, we can improve the quality of realism for a user. Despite this, even avatars which are not realistic still invoke feelings associated with social

²Virtual Reality is defined as an entirely virtual, immersive environment while Augmented Reality is a variation of VR that allows the user to see the both the real and virtual worlds (Azuma, 1997).

anxiety and could still be used to treat these conditions or in research (James, Lin, Steed, Swapp, & Slater, 2003).

4 Designing the Application

The application was built with intent to be used for delivering mental health services and therefore had to satisfy the following major goals:

- Real time avatar movement
- Low latency, synchronous audio
- Immersive experience
- Telemetry for the research team to build real-time features

To achieve these goals, the project team selected specific industry leading technology that would allow the team to perform the needed functions while retaining significant technical support. For the largest component, the target device, the team selected the HoloLens 2. The HoloLens 2 is one of the 3 leading augmented reality headsets along with the Magic Leap and the Vuzix Blade. The HoloLens was selected for its large range of sensors on the device and the extensive software support and libraries provided by Microsoft (Hübner et al., 2020).

4.1 Infrastructure

To develop the application, the Unity3D game engine was used. This is the standard method of developing applications for the HoloLens (Ferrone & Coulter, 2020). This also allows developers to easily adapt the software to work on other devices in the future if desired. Unity projects consist of Scenes which are composed of GameObjects. These GameObjects handle everything from the visual aspects to the scripting of the application. Each GameObject contains various Components which give the GameObjects functionality such as movement, a visual mesh, or a script. Along with Unity, the Mixed Reality Toolkit (MRTK) was used to develop specifically for AR. The MRTK is Microsoft's preferred method of developing for the HoloLens and has

extensive support and functionality compared to Unity's built in AR Foundation library. This library was built specifically for the HoloLens and therefore by using it the developer has access to its full potential.

4.1.1 Networking

The first goal was to achieve real time avatar movement. This goal is fundamental to the application. To achieve this goal, Mid-Level API (MLAPI) was chosen. MLAPI is Unity's new standard library for networking, succeeding High-Level API (HLAPI) (Technologies, 2021). This new library allows the developers to receive full support within the Unity community as well as providing useful high level and low level features. These features allow programmers to easily transmit Unity standard data, such as the positions of GameObjects, across the network. We can also develop our own low level network communications to transmit more specific data efficiently. These aspects combined give us great control over the networking of our application while providing a simple to use library.

All MLAPI projects consist of a NetworkManager and NetworkObjects. The NetworkManager is a required component of MLAPI that facilitates connections and sending data between clients. The NetworkManager can be heavily configured to meet the specific needs of any project. In addition to the NetworkManager, NetworkObjects enable individual GameObjects to have networked behavior. At its core, a NetworkObject component is simply a data marker for the NetworkManager to keep track of the object over the network correctly. It also enables the functionality of NetworkBehaviors and NetworkTransforms which allows the GameObject to have scriptable, networked behavior and for the Transform of the GameObject (its position, rotation, and scale) to be transmitted over the network, respectively.

To facilitate the connection between the users, a client-server model was used. This came as a result of analyzing the network layout between the two users. Assuming both users are on a typical residential network, we must assume a technology called Network Address Translation (NAT) is being used. This technique allows for multiple local devices to operate under a single public IP (Internet Protocol) address. When a client initiates a connection to an outside device on a specific port, such as port 80 which is the standard port for HTTP (the protocol used for

web pages), the router makes a temporary connection between a random outbound port and the internal address and port of the device. This allows the connection to successfully be established (Tsirtsis, 2000). Unfortunately, this also prevents inbound connections from occurring directly.

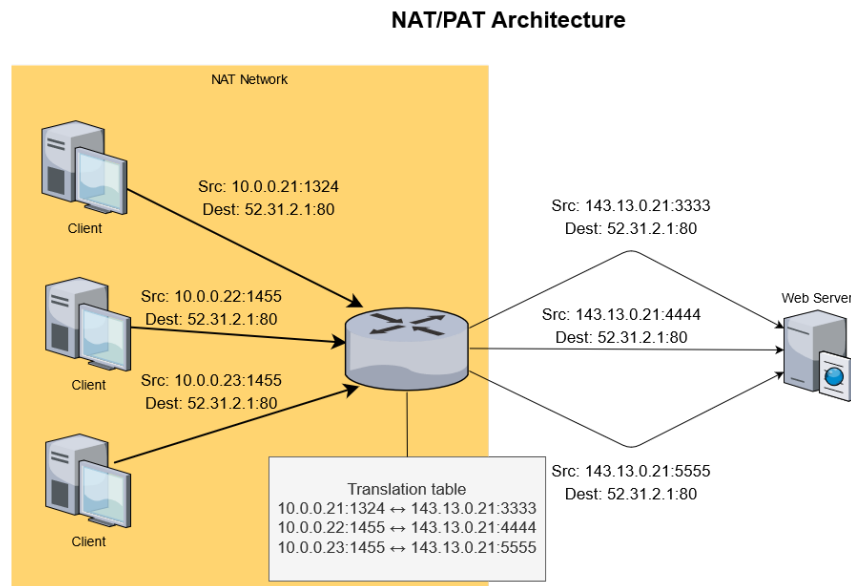


Figure 2: This diagram shows 3 client devices establishing a connection with a single web server. In order to distinguish between the 3 devices, the router creates a translation table from the internal IPs and ports to the external IPs and ports and assigns unique ports to each connection.

Initially, the plan was to use a peer-to-peer (P2P) model, but this plan changed after looking into the technical details specifically regarding NAT. Although MLAPI supports P2P connections, it would require setting up a relay server to bypass the restrictions of NAT. The initial plan of P2P was to cut out using a server entirely, but to use P2P a server must be used anyways. Another option we had was to use a NAT punching server to attempt to establish a P2P connection without a relay server, but this is not guaranteed to work. This technique works by using a server to facilitate a connection between two devices, but in networks with symmetric NAT or restricted cone NAT, this technique does not work (Kegel, Srisuresh, & Ford, 2008). If it succeeds, a direct connection can be established. If it does not, the client must fallback to using a relay server. Because of these restrictions, the client-server model became the easiest and most reliable to implement. This model also made it easy to collect telemetry in a centralized location.

4.1.2 Movement

The first step in real time avatar movement is initiating the connection between the users and the server. This was deceptively challenging as MLAPI and the MRTK interact in ways that cause the MRTK to not function properly. The main problem was that the MRTK camera object cannot be an MLAPI NetworkObject. To fix this, the MRTK camera object was kept completely separate from all NetworkObjects. This posed a new problem. We needed to take the player movement from the MRTK camera object and input it into the user's avatar, which is a NetworkObject. This allows the avatar to still follow the user while also allowing it to be a NetworkObject and have its position sent to the other user.

The final network model is composed of 3 main objects: The MLAPI NetworkManager, the PlayerController, and the PlayerMesh. With MLAPI, a single object prefab is marked as the "Default Player Object" in the NetworkController. This is the object that is automatically spawned when a client connects. In our case, this is the PlayerController. The PlayerController handles interactions between the MRTK camera object and the PlayerMesh, which is the avatar that both clients will be able to see. When the PlayerController is spawned, the client keeps track of the corresponding MRTK camera object while the server begins to instantiate the PlayerMesh. With MLAPI, instantiated NetworkObjects are not spawned across the network automatically (with the exception of the default player object). In our case, we need to send special data with the Spawn command. Specifically, this data is the client id of the player controller. This client id is then sent to all connected clients which check if it is the correct id. The correct client then requests ownership of the object as shown in Figure 3. Even though this may seem like a roundabout way to grant ownership of the object, it is necessary to avoid a race condition in which the instruction to grant ownership arrives before the object is fully spawned, causing an error. This ensures that the object is fully spawned before granting ownership. After ownership is granted, the client finds its PlayerMesh and begins to manipulate it.

Once connected, MLAPI sends the Unity Transform components over the network using the NetworkTransform components that are attached to the parent PlayerMesh object and the head. This replicates the position and rotation of the PlayerMesh over the network. Here, we found that it was important to not have a NetworkTransform component on the Body object as

it caused unintended behavior. Coupled with this, it was vital to reduce the minimum distance required to send the transform over the network. The lower this number is, the more fluid the motion will be, but it will also require sending data much more frequently.

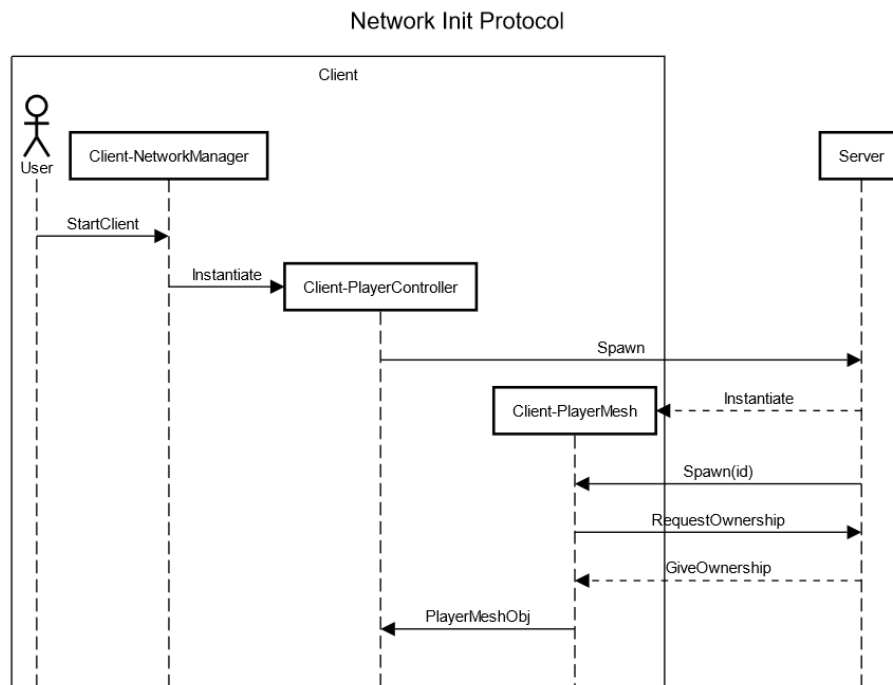


Figure 3: This is the sequence diagram for initiating a network connection. Some components are purposefully left out as they do not affect the flow.

4.1.3 Audio Communication

Communication is an essential part of building an application for use in teletherapy. To ensure sufficient communication between the patient and the therapist, networked audio for the two users to hear each other speak in the virtual environment was implemented. To achieve this goal, the Mixed Reality WebRTC library was used. WebRTC is a protocol developed by Google that enables simple real-time communication (RTC) between two clients (Sredojev, Samardzija, & Posarac, 2015). To enable this, clients first discover and connect to each other using a Signaling server. This server allows two clients to initiate the connection between them and exchange information regarding what data will be transmitted. Despite being a major component of WebRTC, Signaling is not defined by WebRTC to maximize potential as implementations are not forced to conform to a potentially limiting standard (Jennings, Boström, & Bruaroey, 2021). After this initial connection, the clients begin to exchange data. Mixed Reality WebRTC

is an API that enables WebRTC communication on the HoloLens. According to the Mixed Reality WebRTC documentation, each client has four Unity Components that are used to manage the audio. The PeerConnection object facilitates the networking between clients. The Signaler component manages signaling with a Web RTC Signaler server. The MicrophoneSource component simply captures the audio which the PeerConnection object sends to the remote client. The AudioReceiver component takes audio received by the PeerConnection object and sends it to a Unity AudioSource to be played.

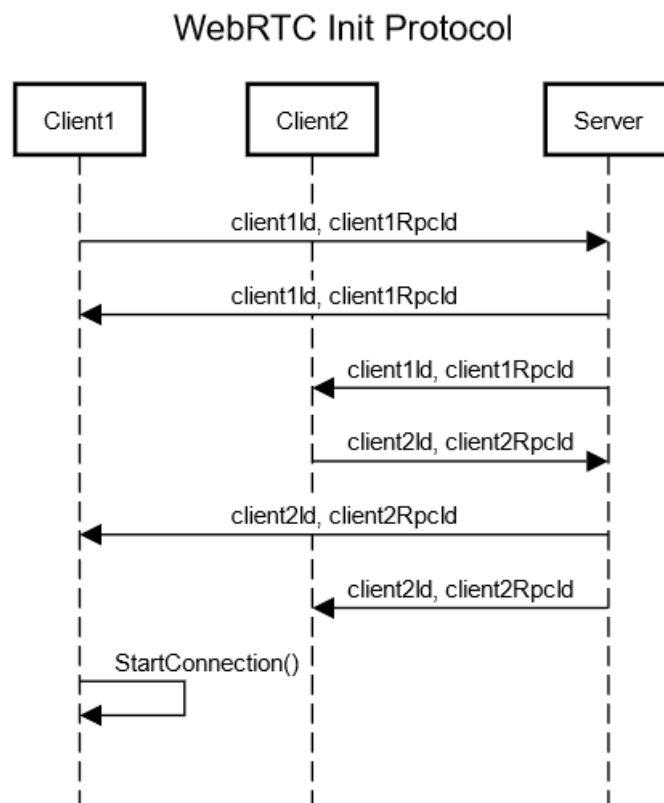


Figure 4: This is the sequence diagram for initiating the WebRTC connection.

To facilitate peer discovery, the existing network connection is used between the clients to exchange WebRTC client IDs. After the connection is initiated, the new client invokes a ServerRPC which sends the randomly assigned GUID of the client to the server. After the server receives this, the server then sends the GUID along with the original client ID to all the clients. The other client will then send its ID back to the initial client in a similar way as demonstrated by Figure 3.

Outside of the Unity and server code, a basic signaling server using node-dss was also setup. Node-dss runs on NodeJS, a server-side Javascript framework. Node-dss is an extremely simple

Signaling server that does not offer many useful features, but it worked well for the development of our application. In addition to the signaling server, we also set up an ICE server. ICE, which stands for Interactive Connectivity Establishment, is a protocol for NAT traversal, allowing two clients to communicate directly which is normally not possible when NAT is being used (Rosenberg, 2010). ICE combines the STUN (Session Traversal Utilities for NAT) and TURN (Traversal Using Relay NAT) protocols to create and maintain this connection. ICE first attempts to facilitate a direct connection between the devices, but if that does not work it falls back to using STUN. STUN is a protocol that allows devices to determine their outwards facing IP and port through NAT as well as other identifying information (Matthews, Mahy, et al., 2020). STUN alone does not solve the problem of breaking through NAT though. STUN does not work on networks with Symmetric NAT, Full Cone NAT, Restricted Cone NAT, and Port Restricted Cone NAT (Rosenberg, Mahy, Huitema, & Weinberger, 2003). To alleviate this problem, ICE uses TURN. TURN is a set of extensions for STUN that allow for relay functionality through a STUN server (Matthews, Johnston, Rosenberg, & Reddy. K, 2020). This relay functionality enables the clients to talk directly to each other by relaying their data first to the server, which it can create a connection to and send data, after which the server relays that data to the other clients. This implementation of a relay can also support relaying data to multiple clients over a single address, improving speed and simplicity (Matthews, Johnston, et al., 2020). Combining STUN and TURN together, ICE allows for reliable data transfer between two clients.

4.1.4 Immersive Experience

Providing an immersive experience is essential for using AR to deliver therapy remotely. For the purposes of developing a Minimum Viable Product (MVP), we aimed to implement placing PlayerMeshes on the floor to remove floating meshes and provide a more immersive experience. This could be implemented using the Spatial Mapping and Scene Understanding APIs. The Spatial Mapping API is part of the base MRTK. According to the Microsoft documentation, this API generates a detailed spatial mesh of the real-world environment allowing applications to change and provide a more immersive experience. The common usages of Spatial Mapping are placement of virtual objects in the real environment, occlusion of virtual objects behind real

objects, realistic physics of virtual objects interacting with real objects, and the navigation of virtual objects through a real environment (Zeller & Coulter, 2018). This project was particularly focused on the placement of virtual objects.

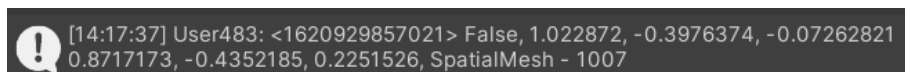
On top of the Spatial Mapping API, the Scene Understanding API provides a simplified, but more useful approach to environment recognition. The high-level representation of the environment provided by the API allows for a more efficient way to place objects in a scene, similar to the lower-level Spatial Mapping API. According to the Microsoft documentation, the Scene Understanding API simplifies the spatial meshes generated by the Spatial Mapping API into Quads, which are flat, rectangular surfaces. This allows for flat surfaces such as walls or floors to be represented in a simplified way. The API can also infer areas that were not scanned and add them to a Quad, creating a water-tight model of the scene allowing for the efficient placement of objects (SzymonS & Coulter, 2019).

4.1.5 Telemetry

Telemetry is required to provide real-time personalization and/or to conduct useful, quantitative research into the effectiveness of AR therapy. The sensors on the HoloLens allow for large amounts of data to be collected, but this project focused only on head and eye tracking data. Eye tracking has been shown to be an important component to studying affective disorders such as anxiety and depression as they provide continuous measurements on visual attention (Armstrong & Olatunji, 2012). By integrating the capturing of this telemetry, the application will provide valuable information for developing real-time responses to users as well as measuring user reactions to provided responses. Such real-time interactions are likely to enable therapists to provide better services.

The eye tracking data we collect is primarily composed of the 3D direction of eye gaze in virtual space. In addition to the gaze direction, we also collect a timestamp of the data, if the current gaze data is valid ("Valid" gaze data is when the data returned references the eye gaze of the user. When the HoloLens is unable to determine the gaze direction, it defaults to the head position. According to the Microsoft documentation, potential reasons include: "The system failed to calibrate the user, The user skipped the calibration, The user is calibrated, but decided

to not give permission to your app to use their eye tracking data, The user has unique eyeglasses or some eye condition that the system doesn't yet support, External factors inhibiting reliable eye tracking such as smudges on the HoloLens visor or eyeglasses, intense direct sunlight, and occlusions because of hair in front of the eyes.” (Sostel, Keveleigh, & Coulter, 2019)), the virtual, 3D position of the user's gaze, and the virtual object the user is looking at. This could be used to produce a 3D heat-map of gaze locations. This is not completely accurate as no pupil dilation data is available to help assess the depth that a user is looking at. We can also determine what virtual object the user is looking at³. This uses the same technique to find the gaze location, but instead of recording the hit position we record the GameObject that was hit. Combining this with the spatial awareness mesh generated by the MRTK, we can also determine if the user is looking at a specific part of the spatial mesh. Using the part of the spatial mesh hit we can determine if the user is looking at a wall, floor, or a platform (such as a table). This could also be further expanded to guess other real-world objects.



```
[14:17:37] User483: <1620929857021> False, 1.022872, -0.3976374, -0.07262821 0.8717173, -0.4352185, 0.2251526, SpatialMesh - 1007
```

Figure 5: This figure shows an example of data output using the DebugDataCollector. The data is formatted as: UserRandomNum: <timestamp>isGazeValid, gazePosX, gazePosY, gazePosZ, gazeDirX, gazeDirY, gazeDirZ, hitObj. The position are in virtual Unity units. With MRTK, 1 unit is 1 meter in physical space.

To collect this data in a centralized location, MLAPI's NetworkedVariable system was used. This system allows for variables to automatically be sent over the network with little overhead (Technologies, 2021). A simple API was implemented on the server side to gather data received by the NetworkedVariables in a generic way to provide ease of implementing new methods of recording the data. Two simple implementations were created: The DebugDataCollector and the CSVDataCollector. The DebugDataCollector simply prints the collected data to the console to make testing more simple. The CSVDataCollector accumulates the data and writes it to a comma-separated value (CSV) file. This simple format allows the raw data to be analyzed easily in custom programs or to be imported into a spreadsheet program.

³This is done by casting a ray in the direction of the gaze until it hits an object. The first object the ray hits is recorded as the object the user is looking at, but due to pupil dilation, the user may be focused on another object in that direction. The HoloLens 2 cannot sense pupil dilation.

5 User Studies

To gain a general understating about younger users' reactions to AR therapy, a preliminary survey was conducted at WPI (N=48). The survey questions solicited WPI Students' opinions on telehealth and in-person therapy as well as the students' opinions on AR technology as it relates to mental health therapy. This survey was open for a week to collect a sufficient number of responses for analysis.

5.1 Results

As shown by the results in Table 1, majority of the respondents in the study (who were recruited from WPI student population) had some form of therapy, about 67% had in-person therapy and 50% used teletherapy.

Table 1: Experience with Therapy

Which methods of therapy have you used?	Percent of respondents
In-Person	66.67%
Teletherapy	50.00%
Group Therapy	22.92%
Other	0.21%

As expected from college students in a technical university, the results showed that respondents were quite tech savvy. For example, the results in Table 2 shows that majority of respondents were interested in using advanced technologies such as virtual and augmented reality. However, they did not show preference for artificial intelligence (AI) personal assistants, which some people described in the comment section as unrealistic.

Table 2: Technology Preference

Should you have the opportunity to use the following technologies, which would you be interested in using?	Percent of respondents
Virtual Reality (VR)	79.17%
Augmented Reality (AR)	64.58%
AI Personal Assistants	39.58%

To interpret the rest of the results, as in a prior research (Djamasbi, Mortazavi, & Shojaeizadeh, 2015) the Likert scales were divided in three equal sections representing low, aver-

age, and high rating ranges. For example, on a 5-point scale rating, a score above 3.67 was considered in the high range, between 2.33 and 3.66 is in the average range, and below 2.33 is in the low range. Similarly, on a 7-point scale rating, a score above 5 was considered in the high range, between 3 and 5 is in the average range, and below 3 is in the low range.

As shown in Table 3, the rating for meeting in-person with a therapist was in the high range (4.19 on 5-point scale). The rating for meeting with a therapist online or in a shared virtual space were in the average range (3.04 and 2.40 respectively).

Table 3: Meeting Preference

Rate your agreement with the following statements	Average Scores*
If I needed therapy, I would like to meet with my counselor in-person.	4.19
If I needed therapy, I would like to meet with my counselor online.	3.04
If teletherapy was the only option, I would like to meet with my counselor represented as an avatar in a shared virtual space.	2.40

*Ratings were captured on 1 to 5 scale where 1 is strongly disagree and 5 is strongly agree.

The results shown in Table 4 show that the ratings for perceptions that are important for therapy were rated in the high range for teletherapy. As displayed in Table 4, the scores for feeling comfort, trust, connected, safe, and “being myself” were all in the high range (>3.67) for both in-person and teletherapy. The ratings for convenience and accessibility were in the high range for teletherapy but in the average range for the in-person therapy.

Table 4: Comfort, Trust, Convenience, and Access

Rate your agreement with the following statements	In-Person	Teletherapy
I am comfortable with <in-person/tele>therapy.	4.17	4.25
I am trusting of <in-person/tele>therapy.	4.31	4.38
I feel connected during <in-person/tele>therapy.	4.21	3.67
I feel safe during <in-person/tele>therapy.	4.41	4.54
I feel like I could “be myself” during <in-person/tele>therapy.	3.86	4.04
I feel <in-person/tele>therapy is convenient.	3.38	4.75
I feel <in-person/tele>therapy is accessible.	3.35	4.5

*Ratings were captured on 1 to 5 scale where 1 is strongly disagree and 5 is strongly agree.

The paired t-test comparing the items in Table 4 for only those who had received both in-person and teletherapy (n=21) show significant differences in perceptions of connectedness, convenience, and accessibility between the two methods of therapy. The results, which are

summarized in Table 5, show that the ratings for feeling connected were significantly higher for in-person therapy while the ratings for convenience and accessibility were significantly higher for teletherapy.

Table 5: Results of paired t-test for those who had both in-person and teletherapy

Rate your agreement with the following statements	In-Person	Teletherapy	p-values
I am comfortable with <in-person/tele>therapy.	4.38	4.19	0.21
I am trusting of <in-person/tele>therapy.	4.57	4.33	0.06
I feel connected during <in-person/tele>therapy.	4.52	3.57	0.00
I feel safe during <in-person/tele>therapy.	4.57	4.48	0.54
I feel like I could “be myself” during <in-person/tele>therapy.	4.24	4.14	0.43
I feel <in-person/tele>therapy is convenient.	3.71	4.76	0.00
I feel <in-person/tele>therapy is accessible.	3.52	4.52	0.00

*Ratings were captured on 1 to 5 scale where 1 is strongly disagree and 5 is strongly agree.

Table 6 displays responses to a set of questions that were asked to gain insight about what features would be more important to developing an engaging experience for AR therapy. Not surprisingly, “Ease of connection” and “Low latency” were ranked as the most desirable features. The rest of the features received average rankings (values in 4 to 5 range).

Table 6: Feature Preference

How would you rank the importance of the following features for an AR therapy application?	Average Scores*
Ease of connection between patient and therapist	1.81
Low latency-little lag between clients	2.91
Spatial audio	4.23
Virtual emoticons, showing emotions virtually	4.26
Group therapy, connect multiple patients at the same time	4.74
Virtual Objects	4.83
Realistic Avatars	5.21

*Ratings were captured on 1 to 7 scale where 1 is the most desirable and 7 is the least desirable feature.

5.2 Discussion and Future User Studies

The above results together provide interesting insight. For example, while the results displayed in Table 1 suggest that teletherapy (3.04 and 2.40 in the average range) is rated less favorably than in-person therapy (4.19 in the high range), responses in Table 3 reveal high rat-

ings (above 3.6 on a 5-point scale) for teletherapy not only for its convenience and accessibility but also for comfort, safety, and trust.

As indicated by the results in Table 6, ease of connection and low latency were regarded as the most important aspect for an AR therapy project. Given the importance of the sense of presence in user engagement for AR applications (Sheridan, 1992), it is not surprising that participants rated having an easily attainable natural real-time interaction as the most important aspect of the AR therapy. The next important feature, according to participants' responses, was the spatial audio and ability to see the emotions of the person a user is interacting with in the virtual space. Again, as discussed earlier, this response supports previous research that asserts sensory information enriches the experience of presence in AR (Sheridan, 1992). In particular, the ability to see emotions is something that benefits in-person and tele-conferencing therapy. Thus, enabling avatars to reveal emotion (e.g., through facial expressions and/or body language) provides sensory information that is likely to improve the sense of presence (Davis et al., 2009). While Table 4 shows high rating for many important aspects of teletherapy, the results in Table 1 indicate that people may feel lesser emotional connection during teletherapy when compared with in-person therapy. Similarly, the results displayed in Table 5, showing those who experienced both in-person and teletherapy, provided significantly lower ratings for feeling connected ($p=0.00$) and almost significantly lower ratings ($p=0.06$) for trust in teletherapy. These results suggest that AR may serve as an excellent tool to address these shortcomings in teletherapy. By enhancing the sense of presence, AR therapy is likely to improve the emotional connection between the client and therapist that seems to be lacking in current teletherapy methods.

The results showed that ability to show emotions in the AR environment was rated in the average range (Table 6). Even respondents who selected "realistic avatars" as the least important or close to the least important demonstrated interest in realistic avatars to show emotion clearly. It has been shown that seeing therapists' emotions and reactions causes an important impact on cognitive behavioral therapy (Westra, Aviram, Connors, Kertes, & Ahmed, 2012). These results back this claim. It is also clear from the results that people are not interested in talking to an AI driven therapist. Despite AI not being the focus of the study, participants expressed (unsolicited) concern about how an Artificial Intelligence based therapy program would feel

unrealistic.

Other features such as having realistic avatars, virtual objects, and group therapy were not rated as important as ease of connection, low latency, and emotions. Future user studies (including generative interviews, scenario and basic user-testing of a working product) are needed to explore these aspects more thoroughly. Studies show that AR can be particularly effective in providing certain types of therapy. Future studies can investigate what kind of mental health services delivered with AR would be desirable by students.

6 Contribution and Future Work

This IQP focused on developing a minimum viable product (MVP) to showcase the base capabilities of a shared virtual space in Augmented Reality that are essential in providing AR therapy. This MVP had to satisfy four major design goals: provide 1) Real time avatar movement, 2) Low latency, synchronous audio, 3) Immersive experience, and 4) Telemetry for the research. The application developed in this IQP satisfied these design goals by providing 1) the ability for multiple users to connect to a central server and view each other's avatars in the same virtual space with real time avatar movement, 2) implementing networked audio to allow the connected users to hear each other, 3) providing a framework to improve immersive experience through Scene Understanding and a simple to change avatar system, 4) providing the capability to collect eye tracking telemetry data for research. Studies show that eye movement data can reveal a great deal of information about a user's cognitive state unobtrusively (Shojaeizadeh, Djamasbi, Paffenroth, & Trapp, 2019; Trapp, Liu, & Djamasbi, 2019; Fehrenbacher & Djamasbi, 2017). Hence, including eye tracking telemetry in AR therapy is likely to provide invaluable insight for designing effective AR therapy systems. In addition to the product produced, the technical challenges that were solved can be used in future work to assist in the development process and ease problem solving.

In the area of networking, there are some improvements that could be made to provide an easier-to-use application that covers a larger variety of scenarios. An important feature that was not developed is to create a simpler and more robust way for users to connect to a server.

Currently, the server's address is hard coded in the application which prevents clients from connecting to arbitrary servers. Similar to this, each server can only handle a single shared space at once. These issues combined allow for only one pair of people to connect at once. Another improvement would be to properly allow more than two clients connect at once. This is currently limited by WebRTC as it is intrinsically peer-to-peer which only allows for two clients to directly communicate with each other. There are, however, ways to enable group peer-to-peer scenarios (Turner & Coulter, 2019). There is also currently an issue with the MixedReality-WebRTC library that requires the application to be built for the ARM architecture to be used (Microsoft, 2020a). This prevented us from deploying the application with audio support to the physical HoloLens as other libraries we used depended on being built for ARM64, though deployment did succeed when Holographic Remoting was used. Currently, the MixedReality-WebRTC library uses the WebRTC UWP SDK project, which is built on an outdated version of Google's libwebrtc implementation. This version does not have ARM64 support while newer versions do. The WebRTC UWP SDK project uses a fork of the libwebrtc implementation and therefore is not up to date. The developers at Microsoft are working to resolve this issue. Once this bug is fixed, full audio support can be implemented into the final deployment.

There is significant room for growth throughout this application to add more functionality and provide a more immersive experience to the users. Using Scene Understanding to place objects on the floor is very important to provide an immersive experience. Due to time constraints, this IQP did not implement this feature, but the created framework provides an easier way to enable this feature in future projects. A key area of growth is providing more realistic avatars. Currently, the basic avatars provided simply represent a "generic user" in the space and they provide very little realism. Microsoft Mesh is a framework that is actively being developed with the goal of providing a more realistic sense of presence in shared virtual spaces with the use of semi-realistic avatars or "holoportation", which allows a lifelike projection of a person to be shared with others using a custom 3D camera setup or an Azure Kinect (Thacker, 2021). After the release of the Mesh SDK, holoportation could be integrated into this application to provide more realistic avatars. Providing a lifelike projection of the therapist and patient could resolve some concerns revealed by the survey study regarding the inability to see real emotions

with avatars. Implementing shared virtual objects may also improve the immersive experience by allowing users to interact with the same virtual object. This could be as simple as decorations for the shared virtual space or as complex as fully animated objects. These features may all be used to provide a more useful therapy experience. Finally, displaying the physical boundaries of the users in the shared space could improve immersion by discouraging users from entering a space in which another user has an object.

The minimum viable product developed by this IQP has produced resources to allow future projects to progress more smoothly and provided preliminary feedback from younger users which sets a path for future research in this area.

References

- Armstrong, T., & Olatunji, B. O. (2012, Dec). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review, 32*(8), 704–723. doi: 10.1016/j.cpr.2012.09.004
- Azuma, R. T. (1997, Aug). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355–385. doi: 10.1162/pres.1997.6.4.355
- Cooley, S., Jaz, J., Miller, E., Jodben, V., & Paniagua, S. (2020, Oct). *Hololens 2 hardware*. Microsoft. Retrieved from <https://docs.microsoft.com/en-us/hololens/hololens2-hardware>
- Davis, A., Murphy, J. D., Owens, D., Khazanchi, D., & Zigurs, I. (2009). Avatars, people, and virtual worlds: Foundations for research in metaverses. *Journal of the Association for Information Systems, 10*(2), 90.
- Djamasbi, S., Mortazavi, S., & Shojaeizadeh, M. (2015). Baby boomers and gaze enabled gaming. In J. Zhou & G. Salvendy (Eds.), *Human aspects of it for the aged population. design for everyday life* (Vol. 9194, p. 479–487). Springer International Publishing. Retrieved from http://link.springer.com/10.1007/978-3-319-20913-5_44 doi: 10.1007/978-3-319-20913-5_44
- Fehrenbacher, D. D., & Djamasbi, S. (2017, May). Information systems and task demand: An exploratory pupillometry study of computerized decision making. *Decision Support Systems, 97*, 1–11. doi: 10.1016/j.dss.2017.02.007
- Ferrone, H., & Coulter, D. (2020, Dec). *Unity development for hololens - mixed reality*. Microsoft. Retrieved from <https://docs.microsoft.com/en-us/windows/mixed-reality/develop/unity/unity-development-overview>
- Goodwin, R. D., Weinberger, A. H., Kim, J. H., Wu, M., & Galea, S. (2020, Nov). Trends in anxiety among adults in the united states, 2008–2018: Rapid increases among young adults. *Journal of Psychiatric Research, 130*, 441–446. doi: 10.1016/j.jpsychires.2020.08.014
- Griffiths, L., Blignault, I., & Yellowlees, P. (2006, Apr). Telemedicine as a means of delivering cognitive-behavioural therapy to rural and remote mental health clients. *Journal of*

- Telemedicine and Telecare*, 12(3), 136–140. doi: 10.1258/135763306776738567
- Hübner, P., Clintworth, K., Liu, Q., Weinmann, M., & Wursthorn, S. (2020). Evaluation of hololens tracking and depth sensing for indoor mapping applications. *Sensors*, 20(4). Retrieved from <https://www.mdpi.com/1424-8220/20/4/1021> doi: 10.3390/s20041021
- James, L. K., Lin, C.-Y., Steed, A., Swapp, D., & Slater, M. (2003, Jun). Social anxiety in virtual environments: Results of a pilot study. *CyberPsychology & Behavior*, 6(3), 237–243. doi: 10.1089/109493103322011515
- Jennings, C., Boström, H., & Bruaroey, J.-I. (2021, Jan). *Webrtc 1.0: Real-time communication between browsers*. W3C. Retrieved from <https://www.w3.org/TR/webrtc/>
- Kegel, D., Srisuresh, P., & Ford, B. (2008, Mar). *State of peer-to-peer (p2p) communication across network address translators (nats)*. IETF. Retrieved from <https://tools.ietf.org/html/rfc5128>
- Lee, J. (2020, Jun). Mental health effects of school closures during covid-19. *The Lancet Child & Adolescent Health*, 4(6), 421. doi: 10.1016/S2352-4642(20)30109-7
- Matthews, P., Johnston, A., Rosenberg, J., & Reddy, K. T. (2020, Feb). *Traversal using relays around nat (turn): Relay extensions to session traversal utilities for nat (stun)*. IETF. Retrieved from <https://tools.ietf.org/html/rfc8656>
- Matthews, P., Mahy, R., Wing, D., Petit-Huguenin, M., Rosenberg, J., & Salgueiro, G. (2020, Feb). *Session traversal utilities for nat (stun)*. IETF. Retrieved from <https://tools.ietf.org/html/rfc8489>
- Microsoft. (2020a, Jun). *Arm64 support progress · issue #414 · microsoft/mixedreality-webrtc*. GitHub. Retrieved from <https://github.com/microsoft/MixedReality-WebRTC/issues/414>
- Microsoft. (2020b). *Hololens 2: Find specs and features - microsoft hololens 2*. Author. Retrieved from <https://www.microsoft.com/en-us/p/hololens-2/91pnzzznzwcq>
- North, M. M., & North, S. M. (2016). Chapter 6 - virtual reality therapy. In J. K. Luiselli & A. J. Fischer (Eds.), *Computer-assisted and web-based innovations in psychology, special education, and health* (p. 141-156). San Diego: Academic

- Press. Retrieved from <https://www.sciencedirect.com/science/article/pii/B9780128020753000061> doi: <https://doi.org/10.1016/B978-0-12-802075-3.00006-1>
- Price, M., & Anderson, P. (2007, Jan). The role of presence in virtual reality exposure therapy. *Journal of Anxiety Disorders*, 21(5), 742–751. doi: 10.1016/j.janxdis.2006.11.002
- Rosenberg, J. (2010, Apr). *Interactive connectivity establishment (ice): A protocol for network address translator (nat) traversal for offer/answer protocols*. IETF. Retrieved from <https://tools.ietf.org/html/rfc5245>
- Rosenberg, J., Mahy, R., Huitema, C., & Weinberger, J. (2003, Mar). *Stun - simple traversal of user datagram protocol (udp) through network address translators (nats)*. IETF. Retrieved from <https://tools.ietf.org/html/rfc3489>
- Sheridan, T. B. (1992, Jan). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1(1), 120–126. doi: 10.1162/pres.1992.1.1.120
- Shojaeizadeh, M., Djamasbi, S., Paffenroth, R. C., & Trapp, A. C. (2019, Jan). Detecting task demand via an eye tracking machine learning system. *Decision Support Systems*, 116, 91–101. doi: 10.1016/j.dss.2018.10.012
- sostel, keveleigh, & Coulter, D. (2019, Oct). *Eye tracking - mixed reality*. Microsoft. Retrieved from <https://docs.microsoft.com/en-us/windows/mixed-reality/design/eye-tracking>
- Sredojev, B., Samardzija, D., & Posarac, D. (2015, May). Webrtc technology overview and signaling solution design and implementation. In *2015 38th international convention on information and communication technology, electronics and microelectronics (mipro)* (p. 1006-1009). doi: 10.1109/MIPRO.2015.7160422
- SzymonS, & Coulter, D. (2019, Jul). *Scene understanding - mixed reality*. Microsoft. Retrieved from <https://docs.microsoft.com/en-us/windows/mixed-reality/design/scene-understanding>
- Technologies, U. (2021). *Unity multiplayer networking: Unity multiplayer networking*. Unity Technologies. Retrieved from <https://docs-multiplayer.unity3d.com/>
- Thacker, N. (2021, Mar). *Microsoft mesh - a technical overview*. Microsoft. Retrieved from <https://techcommunity.microsoft.com/t5/mixed-reality-blog/>

microsoft-mesh-a-technical-overview/ba-p/2176004

- Trapp, A. C., Liu, W., & Djamasbi, S. (2019, Jul). Identifying fixations in gaze data via inner density and optimization. *INFORMS Journal on Computing*, *31*(3), 459–476. doi: 10.1287/ijoc.2018.0859
- Tsirsis, G. (2000, Feb). *Network address translation - protocol translation (nat-pt)*. IETF. Retrieved from <https://tools.ietf.org/html/rfc2766>
- Turner, A., & Coulter, D. (2019, Feb). *Shared experiences in mixed reality - mixed reality*. Microsoft. Retrieved from <https://docs.microsoft.com/en-us/windows/mixed-reality/develop/platform-capabilities-and-apis/shared-experiences-in-mixed-reality>
- Twenge, J. M., Cooper, A. B., Joiner, T. E., Duffy, M. E., & Binau, S. G. (2019, Apr). Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. *Journal of Abnormal Psychology*, *128*(3), 185–199. doi: 10.1037/abn0000410
- VandenBos, G. R., & Williams, S. (2000). The internet versus the telephone: What is telehealth anyway? *Professional Psychology: Research and Practice*, *31*(5), 490–492. doi: 10.1037/0735-7028.31.5.490
- Wagner, B., Horn, A. B., & Maercker, A. (2014, Jan). Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial. *Journal of Affective Disorders*, *152–154*, 113–121. doi: 10.1016/j.jad.2013.06.032
- Westra, H. A., Aviram, A., Connors, L., Kertes, A., & Ahmed, M. (2012). Therapist emotional reactions and client resistance in cognitive behavioral therapy. *Psychotherapy*, *49*(2), 163–172. doi: 10.1037/a0023200
- Zeller, M., & Coulter, D. (2018, Mar). *Spatial mapping - mixed reality*. Microsoft. Retrieved from <https://docs.microsoft.com/en-us/windows/mixed-reality/design/spatial-mapping>