Human-AI Teaming for Dynamic Interpersonal Skill Training

Xavian Alexander Ogletree

Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all

Part of the Computer Engineering Commons, and the Computer Sciences Commons

Repository Citation

https://corescholar.libraries.wright.edu/etd_all/2475

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.
HUMAN-AI TEAMING FOR DYNAMIC INTERPERSONAL SKILL TRAINING

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

by

XAVIAN ALEXANDER OGLETREE
B.S.C.S., Wright State University, 2020

2021
Wright State University
WRIGHT STATE UNIVERSITY
GRADUATE SCHOOL

April 30, 2021

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Xavian Alexander Ogletree ENTITLED Human-AI Teaming for Dynamic Interpersonal Skill Training BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

________________________________
Yong Pei, Ph.D.
Thesis Director

________________________________
Mateen M. Rizki, Ph.D.
Chair, Department of Computer Science and Engineering

Committee on Final Examination:

________________________________
Yong Pei, Ph.D.

________________________________
Mateen M. Rizki, Ph.D.

________________________________
Paul J. Hershberger, Ph.D.

________________________________
Barry Milligan, Ph.D.
Vice Provost for Academic Affairs
Dean of the Graduate School
ABSTRACT


In almost every field, there is a need for strong interpersonal skills. This is especially true in fields such as medicine, psychology, and education. For instance, healthcare providers need to show understanding and compassion for LGBTQ+ and BIPOC (Black, Indigenous, and People of Color), or individuals with unique developmental or mental health needs. Improving interpersonal skills often requires first-person experience with expert evaluation and guidance to achieve proficiency.

However, due to limited availability of assessment capabilities, professional standardized patients and instructional experts, students and professionals currently have inadequate opportunities for expert-guided training sessions. Therefore, this research aims to demonstrate leveraging technological advances in mobile computing (MC), automatic speech recognition (ASR), natural language processing (NLP) and augmented reality (AR) technology to address such limitations. Our result is a mobile application created in collaboration with a psychology expert to provide an innovative training/learning serious game.
# TABLE OF CONTENTS

1  INTRODUCTION .................................................................................................................. 1

1.1  OVERVIEW OF THESIS ................................................................................................. 2

2  REVIEW OF RESOURCES ............................................................................................... 3

2.1  WHAT IS AUGMENTED REALITY? ............................................................................... 3

2.2  BUILDING THE VIRTUAL SCENE ................................................................................. 3

2.3  AR TOOL KITS ................................................................................................................ 4

   2.3.1  Google’s ARCore API ............................................................................................ 4

   2.3.2  Vuforia API ........................................................................................................... 5

2.4  VOICE GENERATION .................................................................................................... 5

   2.4.1  Google Cloud TTS ............................................................................................... 5

   2.4.2  IBM Watson TTS .................................................................................................. 6

2.5  AUTOMATIC SPEECH RECOGNITION ........................................................................ 6

   2.5.1  IBM Watson Speech-to-Text ................................................................................ 7

   2.5.2  Google Speech-To-Text ......................................................................................... 7

2.6  ADVANTAGE OF CLOUD VOICE TECHNOLOGY ..................................................... 8

3  DESIGN AND DEVELOPMENT ....................................................................................... 9

3.1  INTEGRATION OF CLOUD AND AR ........................................................................ 9

3.2  EARLY WORK ON VIRTUAL INTERACTIVITY ........................................................ 10

3.3  DESIGNING INTERACTIONS ....................................................................................... 11

3.4  VR VERSION .................................................................................................................. 12

3.5  AUGMENTED REALITY VERSION ............................................................................. 15

4  FINAL RESULT .................................................................................................................. 17

5  CHALLENGES GOING FORWARD .............................................................................. 23

5.1  LIMITATIONS OF ASR ............................................................................................... 23

5.2  LIMITATIONS OF AR ................................................................................................ 23

5.3  CHALLENGES OF AR UI DESIGN .......................................................................... 24

5.4  LIMITATIONS OF CHARACTER CREATION ........................................................... 25
5.5 THE UNCANNY VALLEY .......................................................... 25
6 CONCLUSION .................................................................................. 27
REFERENCES ...................................................................................... 28
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Flowchart of the AR/ASR system.</td>
<td>10</td>
</tr>
<tr>
<td>3.</td>
<td>Hand drawn storyboard drawn by the author.</td>
<td>12</td>
</tr>
<tr>
<td>4.</td>
<td>Erika presenting instruction to the user.</td>
<td>13</td>
</tr>
<tr>
<td>5.</td>
<td>The options the user can choose from.</td>
<td>14</td>
</tr>
<tr>
<td>6.</td>
<td>The user's character speaking to Alice.</td>
<td>14</td>
</tr>
<tr>
<td>7.</td>
<td>The QR code to launch the scene.</td>
<td>15</td>
</tr>
<tr>
<td>8.</td>
<td>Demonstration of Erika (left) and Alice (right)</td>
<td>16</td>
</tr>
<tr>
<td>9.</td>
<td>User aligning device with QR code.</td>
<td>17</td>
</tr>
<tr>
<td>10.</td>
<td>User engaging in conversation with Alice.</td>
<td>18</td>
</tr>
<tr>
<td>11.</td>
<td>An image showing Alice beginning repeated behavior.</td>
<td>19</td>
</tr>
<tr>
<td>12.</td>
<td>Erika providing information on needing to address the behavior.</td>
<td>20</td>
</tr>
<tr>
<td>13.</td>
<td>The user addressing the repeated behavior.</td>
<td>21</td>
</tr>
<tr>
<td>14.</td>
<td>Erika congratulating the player for appropriately addressing the behavior.</td>
<td>22</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENT

To start with, I want to thank Dr. Pei for giving me the opportunity to work on his team. Additionally, I would like to thank all the support he gave me throughout this project. Lastly, I would like to thank him for giving me numerous speaking opportunities and enabling to show my work to a multitude of people.

Additionally, I would like to thank Heather M. Rando for her help creating the story lines we used throughout our research. Her expert feedback on creating character interactions was core to building our application. Lastly, she provided key information in how students on the Autism Spectrum experience student life.

Next, I would like to thank Dixit Patel. His skill set in Unity development help solidify the design of this application.

I would also like to thank my parents for supporting me throughout my schooling. They provided continuous support as I navigated the maze of college. Through their support, I was able to commit myself to my final years of college.

Finally, I would like to thank my significant other Alex Grant. His support of me was invaluable in helping me find the drive to continue my degree. His kind and patient support of me throughout this project enabled me to push through countless difficulties.
1 INTRODUCTION

In almost every field, there is a need for strong interpersonal skills. This is especially true in fields such as medicine, psychology, and education. People in these careers need to be capable of working with a variety of individuals from a wide array of backgrounds. They need to be able to show understanding and compassion for LGBTQ+ and BIPOC (Black, Indigenous, and People of Color) experiences. Additionally, people in medicine, psychology, or education may need expert guided training to appropriately work with individuals that have unique mental health needs.

Becoming a medical expert requires multiple years of training. Thus, this creates a shortage of availability as demand increases. Additionally, such experts may need to utilize standardized patients. Standardized patients are specially trained people that emulate the symptoms and behavior of various medical issues [1]. These facts combined further limit the ability to provide sufficient training to individuals entering the field. Furthermore, events may arise where in-person training cannot be conducted. This is corroborated by previous work in the field of human-AI teaming [2].

The combination of these issues means students and professionals currently have inadequate opportunities for expert-guided training sessions. Therefore, this research aims to demonstrate leveraging technological advances in mobile computing (MC), automatic speech recognition (ASR), natural language processing (NLP) and augmented reality (AR) technology to address such limitations. Our result is a mobile application
created in collaboration with a psychology expert to provide an innovative
training/learning serious game.

It uses the latest advances of human-AI teaming to showcase the potential of self-
guided training through a virtual expert and standard patient role-play session. Our
current work has the user interacting with a virtual standardized patient on the Autism
Spectrum. By building on past work with AR, ASR, NLP, mobile and cloud technologies,
we create an engaging conversational learning experience where the user practices their
direct communication skills with an individual on the Autism Spectrum. Both the virtual
expert and virtual patient are powered by AI and can react to the user’s choices, and the
guidance of the virtual expert is determined automatically based on the user’s action
through the usage of ASR and NLP.

1.1 OVERVIEW OF THESIS

This paper is broken into multiple chapters. First, Chapter 2 will present the resources
that went into building this application. Next, Chapter 3 will discuss the work that led to
the final application. Then, Chapter 4 will present the application we have built.
Afterwards, Chapter 5 will examine the unique challenges facing this work moving
forward. Finally, Chapter 6 will summarize all that has been discussed in this thesis and
the work to be done moving forward.
2 REVIEW OF RESOURCES

2.1 WHAT IS AUGMENTED REALITY?

Before beginning this section, we would like to take a moment to explain what augmented reality is for our use case. Augmented reality (AR) is the usage of virtual assets to enhance, or augment, the real world. In this paper, our augmentation is our 3D virtual expert, Dr. Erika Parker, and our virtual standardized patient, Alice. Our AR system then uses the camera of a mobile device to project these characters into the real world. These characters and their virtual scene are built using several tools as detailed in the following subsections.

2.2 BUILDING THE VIRTUAL SCENE

A multitude of tools were used for building both the virtual characters and scenes. One such tool was Reallusion’s Character Creator. Character Creator enables the rapid creation of realistic human characters. One can customize a character’s clothes, hair, and physical features through drag-and-drop menus, value sliders, and guided sculpting tools. These features allow for creating content without needing to understand 3D modeling, rigging, and texturing.

The next tool used was Reallusion’s iClone7. iClone7 is a powerful tool for animating characters created from Character Creator. Additionally, it provides tools for creating lip sync animations for dialogue. Additionally, iClone7 can automatically generate lip keyframes from audio. However, they still require manual touch ups before they are realistic.
Finally, Unity is used to tie everything together. It is the game engine that powers our application. It is used to script scenes, build environments, access the device microphone, and to render the AR scene. However, there are multiple choices for AR framework to use within Unity.

2.3 AR TOOL KITS

There exist a variety of augmented reality, or AR, tool kits. The main contenders in this space are ARCore, Vuforia, and ARKit. However, ARKit was eliminated from the very start of this project due to it being an iOS exclusive API. ARCore and Vuforia both support android and iOS allow for increased accessibility of training material. The following section examines both their strengths and weaknesses.

2.3.1 Google’s ARCore API

ARCore is Google’s first-class AR solution [3]. It is the main AR framework for android. However, ARCore requires devices to pass a certification process to insure it provides an optimal experience [4]. While this limits device support, it helps insure a better experience with augmented reality. Moreover, this list is continuously expanding through a combined effort of OEMs and Google. Furthermore, it is also capable of running on iOS devices, enabling the creation of cross-platform applications. Additionally, ARCore ships with a variety of AR features. It provides plane detection, image tracking, object occlusion, and lighting estimation. Lighting estimation is particularly useful in building immersive experiences [5]. With lighting estimation enabled, the 3D rending system will attempt to position and tint the virtual lighting so that it matches the real world [5]. Also, Google provides APIs for creating AR experiences in the Unity game engine [6]. All
these features are provided free of charge by google. Aside from the iOS exclusive AR
Kit, ARCore’s biggest competitor is PTC’s Vuforia API.

2.3.2 Vuforia API
Vuforia is a 3rd party AR solution provided by the company PTC, Inc. [7]. Their
framework provides similar functionality to ARCore. It even leverages ARCore
whenever it is available. However, when it is unavailable, Vuforia will use a custom AR
implementation. This allows for PTC to provide support for a wider selection of Android
6.0 and newer devices. It similarly ships support for iOS and includes support for AR
focused devices such as HoloLens [8]. However, such AR devices still have a long way
to go before reaching the consumer market in terms of price. Additionally, Vuforia
requires a subscription at a minimum of 42 $/month to use their APIs [9]. Since, both
Vuforia and ARCore have similar features, the difference in prices was the deciding
factor in choosing to develop with ARCore.

2.4 VOICE GENERATION
In addition to a variety of AR tool kits, there are different cloud-based text-to-speech
(TTS) systems. Two such TTS systems examined for this project are Google Cloud TTS
and IBM Watson TTS.

2.4.1 Google Cloud TTS
Google Cloud TTS provides high quality voice generation. This is useful for providing
each virtual character a unique voice without needing a dedicated voice actor.
Additionally, Google’s service provides a wide selection of voices across multiple
languages and accents [10]. Furthermore, it provides access to WaveNet based voices.
WaveNet is a deep machine learning approach to generating voices. According to their own research, people rate WaveNet generated voices as more natural than other processes [11]. Additionally, the option to train WaveNet to generate custom voices is available. This would allow for quickly creating content when a desired voice actor is unavailable. These features are all provided at a competitive pricing.

Google allows for 1 million characters per month free [12]. After the limit of the free pricing has been reached, voice generation is billed at 16 $/million characters afterwards. These points make Google Cloud TTS a strong choice when compared against IBM Watson TTS.

2.4.2 IBM Watson TTS

IBM Watson TTS is the official text-to-speech system of IBM. It offers a much smaller selection of voice and languages compared to Google Cloud TTS. Additionally, it uses a voice generation technique called LPCNet. The researcher’s test results show it had slightly lower ratings than WaveNet when compared to similar voice generation techniques [13]. Finally, IBM offers a much more limited and costly character limits.

At the time of writing, IBM offers 10,000 characters per month free. Once the free limit is reach, IBM charges at a rate of 20 $/million characters, billed at a rate of 0.02 $/thousand character. The difference in billing rate and voice quality was why Google Text-to-Speech was ultimately chosen [14].

2.5 AUTOMATIC SPEECH RECOGNITION

Automatic speech recognition (ASR) provides us a way to build conversational interactions. Instead of a user pressing a button on a screen, they would instead be able to
speak to their device. Then, the ASR system will convert this into a transcript that can be processed locally. Additionally, the system needs to be capable of doing this in real-time to enable faster feedback. This section examines the two best candidate choices for this application: IBM’s Watson Speech-to-Text and Google’s Speech-to-Text.

2.5.1 IBM Watson Speech-to-Text

IBM provides a cloud-based ASR API through its Watson Speech-to-Text system [15]. With IBM having an official Unity plugin for accessing their API’s, it was initially a strong candidate [16]. However, in early development, we found it made consistent errors in recognition for some phonemes. It had trouble recognizing phoneme pair /sw/. Such a pair is present in words such as “s way”, “swing”, and “swim”. Additionally, it had challenges recognizing /ŋ/ in words such as “r unning”, “jumping”, “skipping”. Such consistent errors recognizing these phonemes where not present in Google’s solution. This ultimately led to choosing Google’s API. However, it is not without some drawbacks.

2.5.2 Google Speech-To-Text

Provided through Google Cloud, their Speech-to-Text solution posed some daunting challenges. The first and biggest hurdle to overcome was the lack of official Unity support. However, since it had a C# API, using it in Unity was possible because C# is Unity’s scripting language [17]. Although, implementing that support proved to be a long and arduous process of tediously finding compatible .DLL files. Additionally, extra code was needed to provide a compatibility layer to avoid fatal errors in the system. Fortunately, Google is working towards providing better support between Unity and its cloud technologies.
2.6 ADVANTAGE OF CLOUD VOICE TECHNOLOGY

By utilizing cloud technologies, we can create conversational interaction without further burdening the CPU of the device. Additionally, cloud technology allows us to have access to high-quality voice models that would be inaccessible otherwise. It could even enable procedural generation of audio. The system could procedurally generated sections of dialogue as text and send it to the chosen cloud-service. The cloud-service then sends back generated audio that can be integrated with the animated characters to produce the synchronized audiovisual effects. The result would be the virtual characters appearing to speak the generated audio. This will allow for emergent combinations of handcrafted material greater than the sum of their parts. Moreover, it would further enhance the immersive power of AR/VR technology, helping teach people critical and compassionate interpersonal skills.
3 DESIGN AND DEVELOPMENT

The following section will layout the related research and exploration leading to the final prototype. It covers the integration of Cloud and AR components, with the early development leading to an AR training application, and the later collaborative efforts to provide virtual expert guided training.

3.1 INTEGRATION OF CLOUD AND AR

The AR and cloud technologies are combined through the Unity game engine. The flowchart of this system is shown in Figure 1. Unity’s main application loop continuously listens for the user’s speech. Once speech is detected, it streams the audio in real time for the automatic speech recognition. A background process waits for Google to start returning the detected speech. As the transcribed text is received, it fires events to any listening component. In this case it is a component we refer to as the story. The story registers function callbacks, which we call passages, within the story engine that can be fired based on the received text. Whenever the story receives a transcript, it analyzes it to determine what passage to tell the story engine to execute. When the story engine executes a passage, the passage plays its associated animations and dialogue. The character’s dialogue is generated using Google’s cloud-based TTS system. While currently pre-recorded, future development could generate this speech on the fly for procedural AI reactions.
3.2 EARLY WORK ON VIRTUAL INTERACTIVITY

An early project showing off this combined AR-ASR system was a story book application developed by this paper’s author. The AR-ASR system took the form of a virtual popup book. The book featured an interactive fox as the main actor and a tracker that showed the user’s progress through the story. The tracker provided live feedback to the user in the form of green, orange, and red text highlighting. The green represents correctly read text, the orange represents text that has been skipped over, and red represents text that has been read incorrectly. Additionally, the fox’s animation would only progress as the user reads the story. The goal of the AR capability was to provide an engaging motivation to continue reading. An image of this system can be seen in Figure 2. While this application was the catalyst for our development, it is missing key features such as the virtual characters giving direct feedback to the user.
3.3 DESIGNING INTERACTIONS

Before starting development on our system, we had built a story board of events. The story board served to roughly visualize each possible route the player could take. We worked with an expert in education, Heather M. Rando, to construct the script used in the training interaction. Our script included the following key points:

1. The possible choices the player can make.
2. The various way our virtual standardized patient would respond to the user’s action.
3. The virtual expert’s feedback on each choice.

We used this outline to develop an initial VR prototype to demonstrate the storyline we had created. The story board can be seen in Figure 3.
Additionally, this is the stage where we constructed the narrative we used throughout our development. The role-play sessions were built to emulate the face-to-face setting of Wright State University’s RASE program. In this setting, our standardized patient, Alice, takes the role of a college student on the Autism Spectrum. Meanwhile, the user takes the role of one of the student advisors. Our application focuses on guiding the player through getting to know Alice while addressing behavior she might exhibit. This role-play session emulates what might take place on the first RASE advising session.

3.4 VR VERSION

To test our storyline, we created a virtual reality version of our application. It placed the users into a fully 3D environment. While the characters would speak to the user, the user
was limited to a selection of buttons to press. When a button was pressed, the character representing the player would speak a pre-recorded line of dialogue. This is presented in Figure 4, Figure 5 and Figure 6.

Figure 4 shows the player receiving instructions from our virtual expert Dr. Erika Parker. Erika provides the user information about the current role-play session created with the help of experts in the field. Figure 5 shows the options the player can pick to address the behavior shown by Alice. Figure 6 shows the caption of the dialogue spoken by the player’s character. This dialogue however is not spoken by the player.

![Figure 4 - Erika presenting instruction to the user.](image-url)
However, this created an asymmetric interaction. While the character would speak to each other, the player would be left out of participating in the conversation. Their part of the dialogue is spoken for them. Therefore, our next and final iteration sought to resolve that.
3.5 AUGMENTED REALITY VERSION

The augmented reality (AR) version of our application used ARCore’s image recognition system to provide a reliable anchor point. Using a QR code printed on paper, we can launch the virtual scene into the real world. This QR code can be seen in Figure 7. This image is stored inside ARCore’s image database. When it is recognized, the ARCore framework gives Unity a unique 3D transform that represents the estimated distance, rotation, and scale of the anchor. Additionally, this would allow for multiple anchor points each displaying a unique object in future work.

![Figure 7 - The QR code to launch the scene.](image)

Once the anchor point is established, the characters can be projected into the scene. ARCore utilizes the phone’s gyroscopes and accelerometers to determine the...
position of the anchor. Then any object that is attached to that anchor is subsequently transformed. This allows for the user to freely look around without needing to have the anchor in view. Additionally, if object start to drift, the anchor image can be used to quickly align them. Demonstration of this can be seen in Figure 8 taken from an early stage of AR testing. It shows the two actors of the scene positioned in different parts of the same room.

*Figure 8 - Demonstration of Erika (left) and Alice (right)*
The final result of this process is a short role-play session with our virtual college student on the Autism Spectrum. However, before the interaction can start, we first must locate the anchor image. This process is as simple as pointing the device camera at the printed QR code. This code is then automatically recognized by ARCore. This step is shown in Figure 9.

*Figure 9 - User aligning device with QR code.*
Then, Alice will appear sitting in her chair. The character waits for the user to
great her with “good morning” or just “morning”. She will response similarly with “good
morning”. The player can then proceed to ask Alice various questions about herself. This
is to emulate the process of getting to know the individual on the Autism Spectrum. This
interaction can be seen in Figure 10. Additionally, Figure 10 shows the ASR system
automatically captioning the user’s dialogue.

Figure 10 - User engaging in conversation with Alice.
However, after this exchange Alice will begin exhibiting one of the many types of repeated behavior those on the Autism Spectrum show. The behavior we focused on in our development is body swaying. Alice will proceed to rock from side to side in the middle of the conversation. However, the user is free to continue the conversation ignoring this behavior. This can be seen in Figure 11.

![Figure 11 - An image showing Alice beginning repeated behavior.](image)

If the player chooses to ignore addressing Alice’s behavior, the expert – Erika will appear to inform and guide the player. She appears to the left of Alice and delivers
information about why the user needs to address the behavior. This can be seen in Figure 12.

![Image](what kind of movies do you like to watch)

*Figure 12 - Erika providing information on needing to address the behavior.*

After that, the player can then address the behavior using the appropriately detailed response. In this case, working with someone on the Autism Spectrum requires using direct speech. A succinct explanation of direct speech is a manner of talking that removes ambiguities and implied request. Therefore to successfully preform this task, the user needs to directly state what they want to happen and why. In this case, the user uses
the expression “I noticed that you are swaying. Can you please stop so that we can continue our conversation?” The NLP system will detect that as direct communication based on keywords in the sentence. Then, Alice will kindly apologize and proceed to stop swaying. However, as you can see, the ASR system had trouble with detecting “notice”, instead recognizing it as “know that”. This is shown in Figure 13.
Finally, Erika will congratulate the player on successfully navigating the interaction. This will then conclude the session between the user, Erika, and Alice. Figure 14 shows this interaction.

*Figure 14 - Erika congratulating the player for appropriately addressing the behavior.*
5 CHALLENGES GOING FORWARD

Developing an application with such a variety of technology comes with a multitude of challenges for others looking to contribute. Some issues come from limitations of technology, and other issues come from limitations of inclusivity. This section will look at the challenges in using ASR technology, AR technology, and character creation technologies.

5.1 LIMITATIONS OF ASR

Automatic speech recognition has a multitude of limitations. The first limitation is that accents heavily effect the system. This limitation was noted by Saurab in his work and continues to be a challenge 3 years later [19]. Additionally, it was previously noted that some syllable structures can cause confusion between words. Furthermore, reduction of syllables can cause issues. Human speech, when naturally spoken, is not perfectly enunciated. When syllables start to blur together, get drop, or cannot be detected, the ASR will struggle. Additionally, background noise, especially noise from a background speaker, will cause issues with recognition. However, IBM has shown promising work in this field with their speaker diarization technology [20]. Additionally, Google has a working preview of such a feature [21]. With the growing demand for voice-controlled systems, ASR technology will only continue to improve with due time. However, ASR is not the only technology that is still growing.

5.2 LIMITATIONS OF AR

AR technology is still relatively young and rapidly developing. The workflows around AR can be precarious and slow. Additionally, it relies heavily on the quality of the
device. If the device gyroscopic sensors are low quality, then the AR scene may be more susceptible to drifting when anchors go out of focus. Additionally, difference in camera resolution will produce difference in recognition quality. Furthermore, lighting drastically effects the image recognition process. If the image is to under or overly lit, then the AR tool kits can struggle to find key features in the image. Finally, AR is CPU intensive. It requires a multitude of calculations to interpret how to align objects. Therefore, a slow CPU would result in poor performance of an AR application. However, AR presents challenges that exist outside of hardware requirements. One such challenge is the creation of user interfaces (UIs).

5.3 CHALLENGES OF AR UI DESIGN

Designing UIs is universally challenging. There are few ways to do it right, and there are countless ways to do it wrong. Building a UI for a mobile application can be challenging do to screen size limitations. The addition of AR presents multiple new dimensions to this challenge. Augmented reality is inherently a visual experience, and a good UI must enhance this nature.

Luckily, Google themselves have provided some helpful tips. They note to avoid full-screen popups [22]. One way to incorporate this principle is to integrate virtual signage such as a whiteboard that displace extra information. This signage should be position at a comfortable distance from the viewer. They also note to create visual cues to help the users know where to look [22]. A phone held vertically provides a very limited horizontal field of view. So, integrating such visual information will become critical. Multiple iterations of repositioning Erika had to be done to assure that the user noticed when she appears.
Finally, as the needs of our application grow, the challenges of maintaining the AR UI will also expand. It will be important to regularly evaluate the user experience with the UI to maintain an enjoyable and immersive session. While ASR and AR are limited by technology, this line of research is also challenged by a unique factor: the limitation of character creating software.

5.4 LIMITATIONS OF CHARACTER CREATION

Since this work focuses on inclusivity training, limitations in the ability to create diverse characters will be a hurdle. While Reallusion’s Character Creator greatly simplifies the creation process, it also limits it. We found in our work many of the hair and facial features catered toward individuals of European descent. Additionally, Reallusion’s software does not ship with any natural African hairstyles. It required external collaboration with the authors parents, both of whom are black, to finalize the look of Erika. This issue limits the ability to accurately depicted BIPOC individuals. Such limitations will need to be carefully worked around to create respectful and accurate characters. However, one final challenge related to character creation is the uncanny valley.

5.5 THE UNCANNY VALLEY

The uncanny valley is a familiar term. It is the phenomenon describing the feeling of unease from things that are very realistic but fall short in key points [22]. Since our application works with human figures, this effect will be present throughout this project. Some people that have been shown this project have noted Erika and Alice can be creepy. Therefore, we feel it is important to note that building naturalistic animations for gesture and, more importantly, facial expressions will be core to future development. Bad lip
syncing, jerky movement, and head motion/facial expression during speaking are all factors in actors appearing “creepy”. Additionally, lighting and the gaze of avatars should be fine-tuned throughout testing.


6 CONCLUSION

Our application demonstrates the combined power of augmented reality and automatic speech recognition when paired with expert guided feedback. It allows for a user to have the experience of face-to-face training while only needing access to a mobile device. AR allows for creating immersive scenes that appear to take place in the real world. ASR and NLP allow for AI technology to engage with the player in a conversational manner. This allows for a naturalistic approach to virtual training. Instead of having the user listen to audio and press buttons, they can directly engage with the content.

While such technology is still emerging, it can be successfully used to develop valuable interpersonal communication skills. As the fields of medicine, psychology, and education continue to expand, such application will be vital in keeping up with demand for training. They can provide supplemental on-demand training with material created by experts. Further work would include more expansive interaction with the player, dynamic generation of AI responses, a more robust NLP system, and more characters each with a unique and diverse background.
REFERENCES


