Introduction

This rapid growing of reality technology has undeniably attracted a lot of limelight in recent years. Companies all over the world have been putting resources into VR/AR/MR development. Many existing applications in the market are for corporate uses. They are usually in the form of collaborations or simulations; the following are some well-known examples. *Dynamics 365 Remote Assist* [1] from Microsoft supplies remote aid to technicians on fixing machine problems, colleagues training and collaborations. *Microsoft Mesh* [2] supplies spatial space for collaborations. *Varjo Reality Cloud* [3] provides simulations of interactable 3D objects in a virtual environment, targeting product design and engineering use. According to the market research concerning mixed reality headsets conducted by the Grand View Research, the global market size is expected to have about 30% compound growth rate by 2030 [4]. The price of the headsets will eventually drop to a level that is affordable for everyone. We would like to take this opportunity to examine the feasibilities of mixed reality application. This will be done by building an app with immersive and interactive spatial experience.

Inspiration

Headsets are powerful tools with potential to revolutionize human lives once again like the smartphone did in the last decade. Out of curiosity, we wonder how this magical piece of hardware can transform human lives. What flashed in our mind is J.A.R.V.I.S. In the Iron Man movies, the main character Tony Stark has designed his artificial intelligence assistant with natural language processing capability- – “Just A Rather Very Intelligent System” to help him from work to battles.

As a computer science student, after learning technologies from artificial intelligence to the mixed reality, it is not hard to imagine everyone would like to bring this fantasy to reality. As a half-year project from October 2023 to April 2024, The J.A.R.V.I.S. is too complicated to be built in 6 months. Weighing all constraints, we want to build a beginner education assistant project.

Project Scope

In 6 months, we are going to deliver the following three interesting features.

Intelligent Interface

To enable easy voice input, a virtual helper is designed as a user interface to listen for voice input and provide feedback. As a preliminary idea, this helper will appear once the application transits from the starting logo page and begin to receive and answer to user's quest either in text or through voice. For demonstration purposes, this helper will be able to help supply VR training or AR solutions to limited topics. These topics will be further evaluated throughout the development process based on the availability of resources.

VR Training Classroom

There are things that needed to be trained through hands-on experience due to safety issues or limited availability of equipment. For example, experiments related to circuits assembling are hard to practice without access to school laboratories or supervision, despite its importance for understanding the subject matter. With the training space in the virtual reality environment, it is possible to provide on-demand practice opportunities for the learners to get familiar without access to the physical equipment and avoid danger.

AR Overlay Guidance

VR Training Classroom will never provide enough training as sometimes virtual experience could never replace genuine experience. For people who are practicing for the first time or want some interactive suggestion, the feature would show user step-by-step solutions at their

pace. Therefore, this application will create the overlay guidance on the real objects using AR technology.

Collaborations

It is always fun to learn with others. Collaborations allow others to join the room, to share information, to explain more about the relevant concepts, or more. This function will be enabled based on the preference of the user under the current proposal. This mechanism will be developed for both the VR Training Classroom and the AR Overlay Solutions.

Methodology

Targeting Platform

Among the platforms with mixed reality headsets available in the market, Meta has taken up to half of the market share by the end of the first quarter of 2023 [5]. Meta has also developed help pages for developers getting started with Meta Quest Development in Unity game engine [6]. Unity Software Inc, as a game engine company, has also developed plug-ins to accelerate the application developments for public [7]. Due to higher availability of tools and ease of development, we will build applications targeted for Meta Quest Pro in Unity game engine.

Intelligent Interface

Speech recognition and synthesis allow conversion between text and speech. Since voice and text input methods both have their advantages and disadvantages, allowing conversion between data from the two input methods will provide flexibility of users' input options. Therefore, our application would like to use speech recognition to convert user's instructions into text input for the chatbot, and speech synthesis to convert chatbot's response into voice. We will consider open-source libraries such as Mozilla DeepSpeech [8] and TensorFlowTTS [9] for implementing these features since open-source libraries tend to be more reliable, especially those from big companies like Mozilla and TensorFlow. We will also look for libraries that do not require paid subscription to online services.

Conversational AI chatbot provides human-like conversation which is usually powered by some large language model that is trained on dataset of text. There are many libraries for making AI chatbots, including Google Bard and Microsoft Bot Framework. To provide an immersive experience, a chatbot will be implemented to answer requests from users over

voice. First, a speech-to-text library is used to convert user's instructions into text. Then, the chatbot will generate a response to user's instructions. Finally, a text-to-speech library is used to convert responses into voice to provide a more human-like interaction.

VR Training Classroom

To create virtual classroom activities, a virtual classroom environment will be created using available online resources. If necessary, some assets or models will be made to tailor the use case. Functions normally shown as buttons in the UI will instead be invoked by interacting with in-app object models, for example, interacting with the door exit to return to the main page. Instructions for the activities will be delivered by chatbot through voice. The chatbot for instructing experiments will be customized to best suit each experiment. Necessary events and interactions will be implemented and tested to minimize motion sickness and mimic real-world like behaviors.

AR Overlay Guidance

To supply prompt solutions to the problem in front of the user, the application must first find the problem through analyzing the current scene that the user is looking at. Recognizing the objects and classifying them becomes the key. For demonstration purposes, some virtual copies of objects will be created from images of real objects. COLMAP [10], an open-source image-based 3D reconstruction pipeline software will be one choice for creating virtual objects. For this application, an AI model will be built and trained to detect and classify objects using relevant datasets. Then the model will be used to recognize and track the position of the tools and materials used in the real world. Overlayed virtual objects will be used as visual aids to provide instruction of the real object positions.

Collaborations

To implement multiplayer functionalities, the Photon Unity Networking (PUN 2) package [11] will be used in the application. PUN 2 is a free and easy-to-use networking solution to Unity Multiplayer which enables up to maximum 20 players in one game. To enable real-time communication, Photon Voice 2 [12], a voice chat SDK with integration with PUN 2, will be used.

Project Schedule

Starting from 1st October 2023 to 31st March 2024, it is estimated to have 28 weeks for the software development. As a two-member group, it is estimated that each person may spend around 12 hours per week on this project. This project kicks off from the implementation of suitable chatbot in the virtual environment, and the development completes at AR overlay solution. The middle phrase involves the creation of single player and multiplayer virtual reality experiments. Through the development, the application will be tested at every milestone and progress will constantly be assessed for rescheduling. The project development will adopt an agile approach as MR is a fast-changing technology. The use of API, libraries, models, dataset, and all related project materials may be subject to change.

Timeline / date	Learning hours/person	Milestones
September – October 2023	48	Learn & test chatbot
		Customize chatbot
		Apply text-to-voice
		Apply voice-to-text
		Create VR Chatbot
October – November 2023	48	Design experiments
		Prepare 3D models
		Build 3D UI
		Create events & interactions
		Create Lab Room (Single Player)
		Integrate Netcode

November – December 2023	60	Build voice-chat function
		Update chatbot Integration
		Create Lab Room (Multiplayer)
December 2023 – January 2024	48	Build image classification model
		Train model with datasets
		Test and modify for object recognition
		Object Recognition Model
January - February 2024	48	Test 3D reconstruction libraries
		Create 3D virtual objects
		Overlay objects on top of pre-defined targets
		Overlay Virtual Objects to Real-world Objects
February - March 2024	60	Object recognition and tracking
		Create events & interactions
		AR Overlay Guidance

Risks and Challenges

Mixed reality development requires a wide range of techniques including heavy software integration and user experience design. To create a smooth and user-friendly experience, this application will be tested after every new feature is added. However, unlikely other matured technologies like web, app, and AI, mixed reality is still very immature. A lot of features are subject to limitation of hardware and software.

Hardware limitations

The most advanced headset available currently in the market is Meta Quest Pro. Although we have done some preliminary testing of the headset, we are still unsure whether the headset's hardware such as camera quality and computational power are good enough for our proposal. In case of inferior performance of the hardware, we will downsize our 3D models quality for optimization.

Software limitations

We would like to integrate generative AI language models like ChatGPT and Google Bard into the application if it is possible. These models are well-funded and equipped with powerful algorithms and databases. Therefore, we would try to integrate them into our project. However, since most of the generative AI language models are currently restricted in Hong Kong, this project will bet on other API for chatbot function if the abovementioned language models cannot be accessed in Hong Kong. The rapid change of APIs and Plug-ins are also concerning issues we might switch to other APIs and Plug-ins if necessary.

Challenges

The creation of 3D models is infamously requiring considerable amount of time. We will explore available AI methods or software to generate virtual models in hope of saving development time.

Summary

To summarize, this project is to build an educational MR assistant with Unity engine. Different APIs and plug-ins will be employed to implement other necessary features. This project is an explorative project. We do not intend to produce anything marketable in this stage as many technologies are still immature. However, this exploration will surely bring enormous insights both for us and hopefully society.

References

[1] D. Pinch *et al.* "Overview of Dynamics 365 Remote Assist." Microsoft Learn. https://learn.microsoft.com/en-ca/dynamics365/mixed-reality/remote-assist/ra-overview (accessed Sep. 26, 2023).

[2] *Microsoft Mesh.* (2021), Microsoft. [Online]. Available: https://www.microsoft.com/en-us/mesh

[3] *Varjo Reality Cloud.* (2022), Varjo. [Online]. Available: https://varjo.com/products/realitycloud/

[4] Grand View Research. "Mixed Reality Headset Market Size, Share & Trends Analysis Report By Component, By Operating System, By Storage, By Charging Technology, By End-user, By Industry, By Resolution, By Distribution Channel, By Region, And Segment Forecasts, 2023 – 2030." Grand View Research. https://www.grandviewresearch.com/industry-analysis/mixed-reality-headset-market-report (accessed Sep. 26, 2023).

[5] ETTelecom. "Global XR headset shipments fall 33% in Q1 2023: Counterpoint." ETTelecom.com. https://telecom.economictimes.indiatimes.com/news/devices/global-xr-headset-shipments-fall-33-in-q1-2023-counterpoint/101520906 (accessed Sep. 27, 2023).

[6] Meta. "Get Started with Meta Quest Development in Unity." Meta Quest Developer Center. https://developer.oculus.com/documentation/unity/unity-gs-overview/ (accessed Sep. 29, 2023).

[7] Unity. "About the Oculus XR Plugin." Unity User Manual. https://docs.unity3d.com/Packages/com.unity.xr.oculus@4.0/manual/index.html (accessed Sep. 29, 2023).

[8] *DeepSpeech.* (2020), Mozilla. [Online]. Available: https://github.com/mozilla/DeepSpeech

[9] *TensorFlowTTS.* (2021). [Online]. Available: https://github.com/TensorSpeech/TensorFlowTTS

[10] *COLMAP.* (2023). [Online]. Available: https://github.com/colmap/colmap

[11] *Photon Unity Networking.* (version 2). [Online]. Available: https://www.photonengine.com/pun

[12] *Photon Voice.* (version 2). [Online]. Available: https://www.photonengine.com/voice