VR Dialogues: Navigating Conversations with Virtual Agents

Research Question: How do multiple virtual agents in a VR scene influence users' sense of social presence?

Summary: This project will involve creating a VR scene where the user can converse with 2 virtual agents. These virtual agents will have full-body avatars and will be able to respond to the user and each other's statements, providing a dynamic conversation for the user. This research aims to investigate the difference between a single-agent interaction and a multi-agent interaction in virtual reality. This research also seeks to see if multi-agent conversations can provide a better sense of social presence for users.

Methodology: The project will use the Unity engine and the Meta XR All-in-One SDK Unity package for the development of the VR scene. The avatars will come from the Ready Player Me website and will be animated using Blender. The project will also use the Meta Dictation Unity package for speech-to-text along with GPT 3 for text generation for agent responses. The project additionally will use GTTS for text-to-speech. The agents' performance will be evaluated using a 7-point Likert scale.

Resources Needed:

- VR Headset (provided by research advisor)
- OpenAI API key (provided by me)
- Text-to-Speech library (most likely GTTS, though may use Google Cloud Text-to-Speech which is $0.000016 per byte

- Powerful computer to run unity scene (provided by research advisor, additionally I have a good computer to run unity scene, worst case scenario, use Sidequest to load unity scene as .apk file on Quest headset)

Annotated Bibliography:

Y. Shibahara, K. Yamamoto and S. Nakagawa, "Effect of sympathetic relation and unsympathetic relation in multi-agent spoken dialogue system," 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA), Penang, Malaysia, 2016, pp. 1-6, doi: 10.1109/ICAICTA.2016.7803128.

This paper is about a multi-agent conversation and studying the effects of using sympathetic speech when engaging with users. Sympathetic speech is defined as speech from the agent that revolves around the user, for example, if the user were to talk about udon as the paper states, the agents would also focus on the topic of udon. The agents would discuss it with the user and with each other. The criteria for comparing sympathetic and unsympathetic speech systems were: natural conversation, interesting conversation, various opinions, familiarity with the agent, lively conversation, and like chatting, all on a 5-point Likert scale. The results indicated that sympathetic speech systems could provide significantly better lively conversations. A more interesting result was that unsympathetic speech systems performed better in providing various opinions, which may be due to sympathetic speech focusing on the user. The paper also compares multi-agent sympathetic speech systems with sing-agent sympathetic speech systems with the same criteria as before. The results here indicate that multi-agent sympathetic speech systems significantly perform better in providing lively conversation and in making the conversation feel like an actual chat. Given the existing literature on multi-agent systems and on

the effects of empathetic speech, the paper set out to find out how both concepts together could provide a livelier conversation. The main limitation of the system was in how their response generator could only work for a few general topics rather than any general topic the user chooses.

Manuel Guimarães, Rui Prada, Pedro A. Santos, João Dias, Arnav Jhala, and Samuel Mascarenhas. 2020. The Impact of Virtual Reality in the Social Presence of a Virtual Agent. In Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20). Association for Computing Machinery, New York, NY, USA, Article 23, 1–8. https://doi.org/10.1145/3383652.3423879

This paper explores the impact of virtual reality (VR) on social presence, which refers to the feeling of "being with another" in virtual environments. The authors designed a VR police interview simulation where participants interacted with a virtual suspect, comparing their experiences to a control group using a traditional computer setup. The study found that VR participants experienced a significantly higher level of social presence, particularly in the areas of co-presence (feeling like they were in the same space as the suspect) and perceived affective understanding (feeling like they understood the suspect's emotions). While the believability of the virtual suspect was high in both conditions, there was no significant difference between the two groups. The authors suggest that this research supports the idea that immersive environments can enhance social presence, which has implications for fields like social skills training. The main limitation of the study is the antagonistic nature of the police interview scenario.