Interacting with Video Call Settings Through Hand Gestures

1st Jackie Zhou
100750922
Ontario Tech University
jackie.zhou1@ontariotechu.net

2nd Alvin Ng
100754129
Ontario Tech University
alvin.ng@ontariotechu.net

3rd Maddin Thom
100751351
Ontario Tech University
maddin.thom@ontariotechu.net

***Abstract*—Usage of video call software has risen due to the pandemic. As the pandemic continued, many in-person activities were moved to online platforms. Some examples include lectures being moved to video calls which created a rise in usage of video call software. Some users may have found the radical shift from in person to online difficult to adjust to or that the video conferencing software was hard to use. Our prototype will allow users to control video calls with more ease by allowing them to use simple and common hand gestures that are typically used in-person.**

***Index Terms*—gesture recognition, video call, computer vision, hands**

I. Introduction

Video conferencing platforms, such as Google Meets and Zoom, have experienced significant increases in daily users due to the pandemic [1]. Even with the pandemic dying down, experts believe that the use of videoconferencing will continue long after the pandemic ends, with Gartner -a management consulting company- predicts that only 75 percent of business meetings will take virtually by 2024 [2]. As such our goal is to make controlling video calls a more immersive and natural experience. The use of hand gestures can benefit the experience of the user in any environment and would allow users to act more like how you would outside of online calls, because it is increasing more engaging to interact fully with the video call capabilities rather than simply moving around a computer mouse. People will often use their hands while communicating in-person with someone, whether it be to convey direct non-verbal communication or to accompany verbal interaction with further indicators [3]. Therefore, when someone is in a video call, it can be beneficial to be able to convey what they would like to do using gestures. The objective is to determine the interest of users in having this capability, how can it be best implemented and build an effective way to interact with video calls, the settings and options, through the use of hand gestures.

II. Methods

The first thing we did was a literature review, we came up with a list of potential keywords and topics that would help us in our project, such as how gesture recognition works and the usage of keyboard shortcuts compared to UI elements to control video calls. We proceeded by searching for papers in trusted databases such as the school library that relate to hand-tracking and interacting with video calls. For design thinking we first came up with interview questions that would be relevant to what we want to do for our project. Then we each interviewed a person to gather data for our initial design ideas. We shared this data with each other in our internal team discord server so that it can be compiled and examined as a whole.

The questions for the interview focused initially on the ease of use of well-known video calling software like Zoom. The participants described how they felt about the current controls, what was good or bad about the controls. From there, the questions dive into the participants' interest in potentially using their hands to control aspects of the video call. To get ideas of the expectations or common gestures, participants were asked if they could think of example gestures they would associate with an action. For example, raising their hand could associate with getting attention or virtually raising your hand to ask to speak or answer a question. Finally, they were asked for any final concerns or comments, to allow the participants a chance to give their opinion outside of focused questions, and provide insight into their general thoughts on the idea. A Canban/Trello board of the tasks we did can be seen in Appendix A.

III. Results

During the literature review, while we found several papers that explained how gesture recognition in video conferencing could work through methods such as AI or machine learning [4]. However we found next to nothing on the actual usage statistics of gesture recognition and whether or not people actually used the feature on a regular basis.

After collecting the data from our interviews, the results show that people thought that current designs of video call controls were already simple and easy to use but that there was still room for improvement. Improvements of video calls can be done using hand gestures and hand tracking technologies. These designs should be easier to use compared to current video call functions by using common hand gestures. Our results show that there is a mixed interest for our prototype depending on how well the hand features are implemented. Further study may be needed as we found that many users associate different hand gestures with different functions.

Our research showed us we needed to be careful in not expecting or needing too much human-technology interaction with hand gestures as it could cause issues for users over time [5]. Continuous strenuous movements and actions will cause users to become tired, and potentially experience aches and pains in their arms and hands. This is not the desired impact and therefore, we will have to figure out how to avoid this issue.

Conclusion

From what we saw in the literature review and design thinking interviews, interest in hand gesture controls seemed to be present but depended on how well implemented the software is. During our literature review we did see a gap where there simply weren't any statistics on how common gesture recognition would be used over UI elements or keyboard shortcuts.

To address this gap, one of the ideas we had for our prototype would be to try and implement a system that would gather data on when a gesture would be used and compare it to when an UI element or keyboard shortcut would be used. This would allow us to determine whether or not people are actually using our system and whether or not they think it's better over the other existing controls. It would also allow us to see what gestures are popular and what gestures aren't, so that we could potentially tweak or get rid of unused gestures.

In terms of next steps, we are looking to use what we have learned from our literature review and design thinking interviews to build our first prototype. As we progress towards this prototype, we will begin testing out gestures suggested by the interviews for certain actions.

References

[1] Karl, K. A., Peluchette, J. V., Aghakhani, N. (2022). Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly. Small Group Research, 53(3), 343–365. https://doi.org/10.1177/10464964211015286

[2] Standaert, W., Muylle, S., Basu, A. (2021). How shall we meet? Understanding the importance of meeting mode capabilities for different meeting objectives. Information Management, 58(1), 103393–14. https://doi.org/10.1016/j.im.2020.103393

[3] G. Buckingham, "Hand tracking for immersive virtual reality: Opportunities and challenges," Frontiers in Virtual Reality, 24-September-2021.

[4] Obaid, F., Babadi, A., Yoosofan, A. (2020). Hand Gesture Recognition in Video Sequences Using Deep Convolutional and Recurrent Neural Networks. Applied Computer Systems (Online), 25(1), 57–61. https://doi.org/10.2478/acss-2020-0007

[5] M. J. Camilleri, D. Rempel, and D. L. Lee, "The design of hand gestures for Human–Computer Interaction: Lessons from sign language interpreters," ScienceDirect, 23-May-2014.

Appendix A
Process Flowchart