Notes from User Testing

Facebook Reality Labs
Remote Collaboration in VR
2019.11.20

Description:
I saw a paper flyer posted in my neighborhood, ostensibly by Oculus, looking for user testers. The purpose of the study was to evaluate how people could use VR for remote collaboration. I was obviously interested and went to the link to apply and was screened with a few questions. Not sure if I was being screened on the content of my answers, my ability to give clear written feedback, or even whether I was human or a bot. Nonetheless I was accepted and scheduled a date to attend.

This happened when we were still “full steam ahead” with Lilypad so the topic of the study promised to be pretty informative.

I attended the study and was paired with two other participants

Setup and Hardware
We were each seated in separate partitions next to each other, but out of view. I was seated in a chair facing an array of cameras ~10 feet in front of me. One of them was a Kinect sensor but I didn’t recognize the other hardware. Behind me and out of view was a workstation running their Unity app with a modified Oculus Rift

There were five tests in VR, each about 10-15 minutes long, and before each test we had to calibrate the hardware with an A-pose, followed by a gaze detection calibration (I think) and lastly a facial expression calibration where I was asked to mimic five facial expressions.

Once I was in the VR environment, in each test I was seated at the same small round table with the other two participants sitting at my 9 and 3 o’clock positions. We had avatars that the research team made for us on the spot that were pretty so-so in terms of likeness but which had drab skin textures and pretty flat, basic features and clothing. I could see the other two participants but not my own avatar. One of the other participants said she felt like we looked like Sims characters.

The hardware setup was such that we had the headset, and hand tracking from the modified hardware on the HMD. At no point did I hold a controller. Also, the HMD had some gaze tracking but did not appear any better than our experiments with the same tech using the Vive with SMI hardware.

Also, the cameras in front of me were tracking my body movements fairly well, but I didn’t really notice the quality of the tracking until later in the study.

Initial Reaction:
I think the first thing I noticed was that the hand tracking was pretty good and I could

reproduce hand gestures pretty well, but with some noticeable latency. I would guess that the latency was in the neighborhood of half a second for hand gestures.

The avatars had some issues with their eyelids that were unpleasant, like blinking out of sync, looking often like the person's eyes were closed, or like the eyes might not be looking at exactly what the user was looking at. We tried briefly to look cross-eyed but the did not show the expected eye movement.

The audio was definitely not as comfortable as what I'm used to in Hifi. It was possible occasionally to hear the other users IRL start speaking well before I heard it in-world and there were times where people's statements were clipped or choppy. It was on par with other real-time audio I guess, but not as good as ours

The Tests:
In each of the five tests we were asked to complete a collaborative process in our small room. In all but one of the five, there was a component where we would be offered individual or group bonuses in the form of gift cards if we completed the test to our own advantage or worked together as a group to find consensus.

The five tests were:

1. We were told we had crashed on a plane in Alaska and were 85 miles from the nearest town. In three minutes we had to choose from several items which item we would want to aid us post-crash. If we all agreed on one item we would get bonus points.

2. We were tasked with planning a party and each represented different roles: heads of Finance, Social, and Security. We each had conflicting objectives which we had to meet based on a sheet they gave us in world. We had to compromise on each of four objectives to come to agreement, but were rewarded with bonus points based on how well we negotiated in our own favor w/r/t our objectives

3. We were told that we were three roommates choosing an apartment and were given a floorpan in-world on the table in front of us. We were again told to negotiate the most favorable outcome for ourselves based on a set of objective we each had about size and choice of room, and how much we would want to spend for our share.

4. We were told that we were an HR committee convened to resolve our sale's team's violation of company policy. We were instructed to come to an unanimous set of conclusions about how the company should punish the responsible parties. Our framework was to address and balance the needs of the team, the company, and the management

5. We were given several questions were we had to estimate together the answers to questions based on very little information (e.g.: how many dogs are there in the US, or How many people arrived at SFO in 2017)

Responses
Between each test we were given a battery of multiple choice questions with a Disagree Strongly/Disagree Moderately/Neutral/Agree Moderately/Agree Strongly scale. Many of the questions were around how easily and effectively we were able to communicate with this tool.

There were also questions about how well I could read the emotions, mood, and attitude of the other testers and how well I thought they could read mine.

At the end of the test, we were given a post-test questionnaire that asked what we liked and didn't like as well as whether we would use a tool like this for work. These are very similar to questions we've asked Lilypad testers, but I think because Facebook/Oculus gave users such a focused task AND because they didn't give users the option to actually use the technology afterwards, (nor subsequently see how/if they actually would use it on their own devices,) that their responses would look on paper somewhat more favorable than ours.

Conclusions
I think that the Facebook team is disadvantaged by how limited this testing is. It relies on very specialized hardware, a controlled workspace, focused and very short tasks, and no context. I think they will get very positive-looking data and may proceed to develop products that are not accounting for some of the very significant factors we discovered through trying to develop Lilypad.

Absent from their post-study questions:

- They did not ask about whether the users liked or would want to be represented as avatars at work
- They did not ask about whether they would be able to get whole teams on such a platform
- They did not ask about what kinds of computers we had at work
- They did not ask how much time of our day we might use a product like this
- They did not ask whether we preferred this experience to video chat services
- They did not ask what else we would need feature-wise to use this tool for work
- They did not ask whether we'd want to have a 2D desktop mode

I tried to cover as many of the takeaways and thoughts I had during this process but if anyone has questions, I'll be happy to go into further detail if you're interested.