Visual, Auditory, and Haptic Cue Navigation Techniques for Object Tracking in VR Gameplay

Mariusz Zbigniew Matyja* hkp680
Dennie Magomed Nuridov† fnp604
Junfei (Fiona) Zhang‡ stc505

https://gitlab.com/dikuvr2022/group-19/Project
commit hash: 3f17907c1ea343016bd7e332f9f295783eb26737
Group 19
The University of Copenhagen

ABSTRACT

The task of finding a hidden input is ubiquitous in 3D Virtual Reality (VR) games or applications, where a navigation technique is needed to guide the users to the target. However, integrating sensory cues to navigation is an arduous task for game developers due to the difficulties in balancing the sensory stimuli aroused to the user,as well as the unclear interactions between different types of sensory feedback when applied together. In this report, we investigate the impact of different types of sensory cue navigation techniques as well as their combinations to the users' in-game performance and their likability to the feedback approach. We present a study of visual, auditory, and haptic feedback navigation techniques in guiding the user to find a hidden target. We implement three interaction techniques and four different combinations of sensory cue navigation techniques to investigate the players' in-game performance and likeability. Our hypothesis is that multi-sensory feedback navigation is more efficient than simple visual cue navigation in terms of the users' performance. The results of this study is significant to VR game or application developers in providing more appropriate feedback cues when navigating the player.

Index Terms: Human-centered computing, Object Tracking, Interaction Techniques

1 INTRODUCTION AND MOTIVATION

Virtual reality (VR) is an interactive computer-generated experience taking place within a simulated environment. It immerses users by suspending their belief that what they are seeing is real. VR has the potential to transport people to worlds beyond their own imaginations, or recreate experiences from their past. The emergence of commercial VR headsets such as Oculus Quest explored the potential of this powerful technology and sparked its potential in entertainment and education, such as VitaSim [8]. As one such VR application, VR games engage the users with a stronger presence and realism [4]. This is due to the fact that VR games can heighten degree of flow and immersion and that players tend to have a positive reaction to such experiences [10]. Moreover, VR games not only serve as an essential configuration of contemporary entertainment technologies, but also as a way to learn new things. One such example can be learning a new language, where the highly engaging VR game can improve the culturally relevant physical interaction [2]. All those show that VR game has a practical potential, and is one of the most essential approaches to further popularize VR technology.

Most if not all VR-games where the user is supposed to follow a specific set of actions, a feedback object tracking technique (navigation technique) is necessary. Without that the user can not surely know if they are following the right path. Sensory cues including visual, auditory, and haptic are three essential dimensions for providing such feedbacks to the players [12]. Nevertheless, designing such interaction technique is a major challenge for game developers. To the best of our knowledge, the frontier studies about VR and VR games are focused on its technical potentials [10], but not the user's performance and preference. Therefore, it is worth studying the effect of different sensory feedback types to the users in-game performance and likeability to the navigation technique. In this study, we designed a game 'Finding Nemo' which integrates three types of sensory cues and four different combinations to investigate the user's performance and likeability to those variations. We compare our techniques with a simple pointer baseline.

Our results are essential for VR game or app developers and can serve as a guide to give more appropriate sensory feedback when navigating the user. One of the projects, that could potentially benefit from the results of this study is the Pain Distraction Game from Khora. Even though these two features are not in focus in this paper the game designers could use our results regarding feedback form preference and disturbance. It could be important for getting the patients as relaxed and pleased as possible by decreasing disturbance and choosing a more preferable feedback form. Another project for which our work can be relevant is Vitasim's Medication Administration Training Program. It is crucial to choose a feedback form such that the users can easily follow the cues and find the right room/shelf/pills without shifting their focus too much from the most important task - choosing the right medicine. It could be beneficial to choose a technique with low average user inactive time (the time that the user stays still) since inactivity could mean hesitance and/or confusion, since longer user inactive time could imply that the technique use too much of the users' attention.

The report is organised as follows: section one gives an introduction, aim, and background of this study, section two provides some literature review, section three explains the techniques of this study, section four describes the user tests, section five analyses the data, section six gives some concluding statements, and section seven details the contribution of each member of the group.

2 RELATED WORK

2.1 Feedback techniques for navigation

A lot of virtual reality applications involve locomotion in a virtual environment. For the sake of immersion, freedom to explore the environment is often given. Freedom of such kind can come with certain consequences, which in some cases involve confusion in finding the right path towards the goal of the simulation. To deal with this consequence, without sacrificing the given freedom and thereby potential immersion, a series of navigation techniques are

*e-mail: hkp680@alumni.ku.dk
†e-mail:fnp604@alumni.ku.dk
‡e-mail:stc505@alumni.ku.dk

developed or being developed. One thing these techniques have in common is the usage of sensory stimulus's to navigate the user. This sensory stimulus's usually involve either positive or negative feedback depending on the users actions.

Many studies show the advantages and disadvantage of certain senses being invoked. Vision being the dominant sense, makes using sight the common choice in most application. This would suggest that vision is always the better choice, but other studies [1] would beg to differ. They concluded that out of the tree tested senses (vision, audio and tactile), the response time of visual stimuli was the lowest, while the auditory stimuli was 5% shorter and tactile stimuli was 34% shorter. This means that even though sight is the dominant sense, it is not the best if fast response time is a requirement in the simulation.

Knowing this we can go through different already used navigation techniques and have a better understanding of the reasons behind the usage of those techniques. The most common technique that most already know of is the dynamic arrow. An arrow that point towards the direction you need to go, follows you and changes angle depending on your location. Its simplicity makes it not only less confusing, but also easy to implement, because of this it is a highly used technique which shows high efficiency in medium to long distance travel. One example of the arrow being used in VR is the game "The Lab" which uses a blue arrow to direct players to their next destination. Techniques using audio are usually not done alone, but in combination vision the game "Audioshield" is a good example since uses audio cues to indicate, when it is the right time to block incoming projectiles.

2.2 Visual Navigation

From pointing arrow and lit up paths to objects shining when selected, visual feedback is one of the most commonly used feedback technique. A previous study [11] goes into dept on ways to improve selection techniques by changing the visual feedback. By analysing the Visualization Patterns, Visualization Size and Visualization Versatility, they found the advantages and disadvantages of different techniques. One of their conclusions, was that by making the visual indication around a potentially selected object more apparent, an increase in efficiency when selecting the desired object happens.

2.3 Auditory Navigation

A technique that uses audio to navigate towards the location of a object is spatial audio navigation [6], which guides the user through audio cues that pinpoint the direction of the desired object or location. The initial reasoning behind creating this technology lies in the principle behind the biological construction of the human body. The human body has two ears, the way we detect the location of a sound source, is by considering the difference in the time it takes the sound to reach each ear. The calculation done to pinpoint the location using this time difference happens in the subconsciousness and momentarily. The human eye, on the other hand, is not equipped with this kind of functionality, so we can only locate the direction of a visual object using the angle from which the object is viewed. This type of localizing is slower then with sound and therefore the ears are believed to be better at detecting the location of a sound source than the eye is, at detecting the location of a visual object.

This type of technology even though promising, is not fully integrated into VR yet. One of the mayor reason for that, is the requirement for the sound to be able to come from all places surrounding the 3d space around the ear. The commercial design of the headset does not support that requirement yet, so its practical functionality is still in question.

2.4 Haptic Navigation

Haptic or tactile feedback in guiding the user to perform certain tasks is well developed in previous research studies. Kaul et al. [5] presented HapticHead in their study, which is a device using many vibrotactile actuators distributed and attached around the head to provide guidance cues. They showed that HapticHead feedback is faster and more precise than spatial audio in guiding the user to find objects. They also showed the practical potential of the device through successively guided blindfolded people to find the items in households. Another study by Marquardt et al. [7] developed a forearm-and-glove tactile interface to guide hand planning and coordination in selection and manipulation tasks. They showed that tactile patterns can trigger the user's hand pose as well as motion changes. Tactile instructions are also used in correcting wrong posture in physical activities [9].

2.5 Multi-sensory Navigation

Gao et al. [3] combined visual and auditory spatial cues in their study of redirected walking. They showed that human perceives the final location of the target according to the reliability of each cue if there is an in-congruence. For instance, the users perceive auditory cue as a more reliable way to infer the location of the target object when there is a decrease in vision. Zhang et al. [12] argued in their study that multiple sensory modalities can influence people to act sustainably. They found that sensory cues in games can impact the perception and immersion of the users. They also found that sensory stimuli reduces the cost of acquiring information and shortens the decision-making process.

3 Technique

We have designed three basic navigation techniques with respect to three feedback dimensions as well as four ways of combining them, making a total of seven techniques. We compare all the seven techniques relative to a simple pointer navigation which serves as the baseline. We designed a game 'Finding Nemo' which integrates all the navigation techniques, and the targeting object is Nemo (a cute fish). The navigation techniques are as follows (in the same order as the game levels):

3.1 Simple Pointer Navigation

The pointer is implemented with a knife prefab (Figure 1), and is set as a child to the player to follow the transformations and rotations of the player. The player sees the knife in the upper center of the screen. The hilt of the knife is fixed, whereas the nose of the knife will serve as a pointer to the direction of Nemo.

3.2 Visual Cue Navigation

The visual cue is implemented with respect to brightness change (Figure 2). A linear function of the intensity $I(d) = (1.5 - \frac{d}{d_i}) \times i$ is used, where d is the current distance between the player and Nemo, d_i is the initial distance of between the player and Nemo, and i is the intensity interval constant, which is set to 2 in this experiment. As a result, the scene gets brighter as the user approaches Nemo.

3.3 Haptic Cue Navigation

The left controller vibrates intermittently. The vibration gets stronger and more frequent as user approaches Nemo. The strength and frequency follows a predefined quadratic function with respect to the distance between the user and Nemo.

3.4 Auditory Cue Navigation

The user will hear a poking sound effect intermittently. The sound increases frequency as the user approaches Nemo, following a predefined quadratic function with respect to the distance.

Figure 1: Simple Pointer Navigation

(a) User is far away

(b) User is close

Figure 2: Visual Cue Navigation

3.5 Multi-sensory Navigation

We also implement various combinations of the previous three techniques to study their interactions. They are combined in four ways:
Visual-haptic Navigation
Visual-auditory Navigation
Haptic-auditory Navigation
Visual-haptic-auditory Navigation.

3.6 Game Design

The aim of the game design is to study the players' in-game performance and likeability towards different sensory feedback. We used low-poly style and the scene is modelling Nyhavn. The platform of the scene is modelled to have a crossed box shape (Figure 3) so that the users have more options when choosing their direction from the feedback. The starting position of the user of each level is always the lower left corner. The game starts with a set up interface to select the preset order. There are two orders available, but in the experiment only one order is used: pointer, visual, haptic, auditory, visual-haptic, visual-auditory, haptic-auditory, visual-auditory-haptic.

The game consists of eight levels, integrating the previously mentioned feedback navigation techniques. For each level, there is a

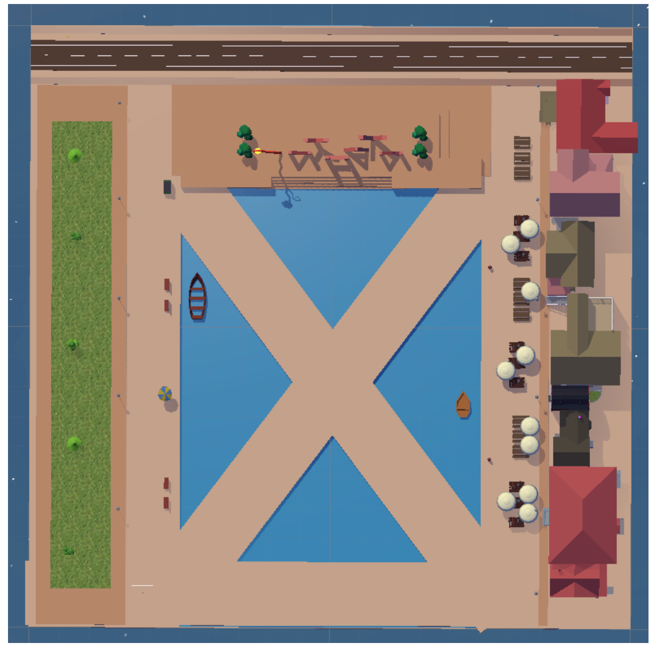

Figure 3: Game Scene

staging phase where the user mush press button 'A' to start the time logging when they are ready. When the user gets close to Nemo, the curent game level is completed and the user is sent back to the starting position, entering the staging phase of the next level. The game is very self-explanatory as we created tutorial canvas to show the instructions at each step. After eight levels are completed, the game shows the time log of both active time (the time taken in each level for the user to find Nemo) and the inactive time (the time in each level that the user stays still).

The buttons are designed as follows: 'B' to select the order of preset, 'X' to confirm the selected order, 'Y' to show/hide the canvas (instructions), 'A' is overloaded: it starts the time log in the staging phase of each level, and is the button for jumping when the level starts, and 'Left Index Trigger' is to skip this level (which we told the user not to press).

4 Experiment

There are four measurements that we take in the user tests:

- Likeability: the user's preference to the feedback technique, from 1 to 5 with 1 being the least preferred.
- Disturbance: how much does the user think is annoyed or disturbed by the feedback technique, from 1 to 5 with 1 being not disturbed.
- Active time (s): the time it takes for the user to find Nemo in each level.
- Inactive time (s): the time that the user is not moving in each level.

The user test consists of seven participants. The procedure is as follows:

1. The user enters the room
2. We explain the rules of the game and the types of feedbacks that the user will receive

3. We ask for the user's nickname and VR proficiency

4. The user put on the headset and starts the game

5. After each level, we ask the user for a preference score and a disturbance score of the feedback technique of this level, both ranked from 1 to 5 with 1 being the least and 5 being the most.

6. The game ends, we copy down the time logged.

5 EXPERIMENT RESULTS

5.1 Obtained Data

In order to make improve readability of the experiment results we have created a simple mapping of level types to numbers. Now each level has a number and the mapping is as follows:

- Level 1: Simple Visual (Pointer) feedback
- Level 2: Visual (Brightness) feedback
- Level 3: Haptic feedback
- Level 4: Auditory feedback
- Level 5: Haptic and Visual(Brightness) feedback
- Level 6: Auditory and Visual(Brightness) feedback
- Level 7: Auditory and Haptic feedback
- Level 8: Auditory, Haptic and Visual(Brightness) feedback

We have performed the experiment with help of seven participants . The data was collected directly by us or under our supervision in the way we wanted it, therefore it was not necessary to do much of data cleansing and feature engineering of the obtain data. However, two of our seven participants experienced unexpected problems during the experiment. One of them has fallen behind the game map but recovered, i.e. came back on the map, by using the jump functionality. Unfortunately, this has strongly influenced the participant's time results for that game level. The second participant could not get one of the controllers to work while in the middle of the experiment. As in the previous case this has also affected the person's time score. We have chosen to not include these two scores in the data analysis part since even a few strongly faulty scores can be crucial for a small data-set. Instead, because of the generally low amount of data points we have chosen to exchange the faulty results with the average score for these two game levels. This way we could still include the other results of these two participants in our data pool.

5.2 Analysis of The Results

When analyzing the obtain data we were limited by the low amount of data-points. We have decided to not consider variance nor standard deviation in our analysis since in such a small data set few artifacts can strongly influence variation. Instead we have focused on analyzing the best, the worst, average and median time of level completion and time of the participants' inactivity. Because level 1 is the baseline feedback method, in this paper we chose to analyze how much the other techniques improve/worsen the results observed in level 1. Figures 4 and 9 represents the results of the analysis. Figures 5, 6, 7 and 8 show more intuitive and comprehensible visualization of some of the results.

One can easily observe a trend in Figures from 5 to 8. Feedback forms 3,4,5,7 and 8 seem to be better than the baseline in level 1. Especially methods used in levels 4 and 5 i.e. auditory feedback and haptic + visual(Brightness) feedback, seem to be much better. On the other hand feedback techniques 2 and 6, which are visual(Brightness) feedback and auditory + visual(Brightness) feedback, tend to have much worse results than the baseline method.

Total time improvment w.r.t. lvl 1				
lvl number	Best Time	Worst TIme	Median	Mean
1	0.0%	0.0%	0.0%	0.0%
2	-40.4%	-53.6%	-164.7%	-82.0%
3	27.6%	27.8%	-27.6%	5.5%
4	1.8%	35.5%	-6.6%	14.4%
5	34.6%	25.8%	4.4%	20.7%
6	-4.9%	-48.1%	-89.8%	-43.3%
7	3.8%	30.1%	-36.9%	3.0%
8	20.0%	35.9%	-49.1%	5.6%

(a) Total time improvement

Inactive time improvement w.r.t. lvl 1				
lvl number	Best Time	Worst TIme	Median	Mean
1	0.0%	0.0%	0.0%	0.0%
2	47.1%	-64.8%	-134.9%	-66.8%
3	46.1%	42.6%	-11.1%	30.2%
4	58.2%	76.8%	23.0%	59.4%
5	79.1%	62.1%	50.0%	56.7%
6	9.8%	0.5%	-63.9%	-36.1%
7	24.2%	82.2%	12.1%	60.9%
8	49.3%	75.5%	0.8%	46.5%

(b) Inactive time improvement

Figure 4: Represents results of the data analysis.

For all of the features presented in Figures 4 and Figures from 5 to 8 one can observe, that the mean values are higher than medians. Since the earlier mentioned pattern is visible for both the median and for the mean values we chose to look only at the median values. The reason to why the mean values are higher than the median values could be, that one of the participants a highly skilled gamer, or is exceptionally good in reacting to various types of feedback.

6 DISCUSSION

By looking at Figure 4 one can observe, that feedback form 5, i.e. the haptic and visual feedback, obtained only positive score. The method exceeded the baseline level for all of the 4 features. It is an interesting observation since feedback method 2 consisting only of visual clues is the worst of all 8 feedback forms. It has obtain only one positive result improving the best(i.e. the least) inactive time value. Level 6 having both visual and auditory feedback performed as the second worst. The last feedback technique involving brightness change is on the third worse place w.r.t total time improvement with the score equal -49%. Level 8 has however scored positive notes for every other feature. In this paper we focus most on the total time it takes to complete a level. To sum up, there are four levels using visual feedback: 2,5,6 and 8. Three of them are the three worst feedback forms if looking at the total level completion time, yet the fourth scores the highest notes.

The wining level 2 consists also of haptic feedback, thus one could think it is this technique, that gives it so high scores and not the visual part of it. If it were to believe level 3, i.e. only haptic feedback, should score somewhat as high scores as level 5, but it does not seem to be true. Level 3 scores a negative note in total completion time equal -27.6%. This points to the following conclusion: The visual feedback alone is one of the worst feedback forms. The haptic feedback is also worse than the baseline when used solo. However a mix of these two techniques works particularly good. This conclusion is directly supporting our hypothesis stated in the `Abstract` section.

The preference and disturbance scores showed in Figure 9 does not imply, that technique 5 is the most preferable type of feedback. Also the method has scored a rather high note in the disturbance table. Unfortunately due to the low amount of participants, and a number of experiment improvements we have omitted we can not be sure of our conclusions. Because of the time constrains we were not able to implement any type of level permutation. The results would be more reliable if each participant would go through the different level forms in different order, as well as the fish should be hidden in different places for each player. The lack of it could have significantly influenced our data leading to false conclusions. Therefore we can neither accept nor disprove our hypothesis.[1]

[1] Usually it would also require to conduct statistical analysis on the results of the experiment. We have been however instructed to omit this step in this paper.

Figure 5: Results of level finish total time analysis

Figure 6: Results of inactive time analysis

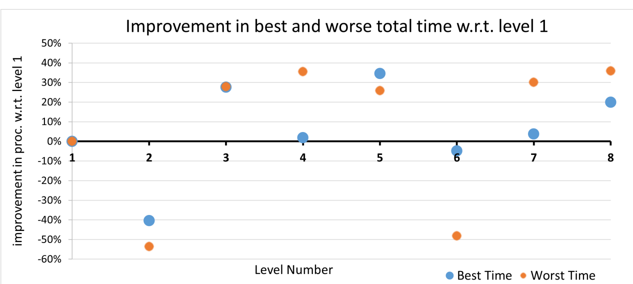

Figure 7: Results of level finish worst/best time analysis, where worst means the longest total time.

Figure 8: Results of inactive worst/best time analysis, where worst means the longest time of inactivity.

lvl number	Average Preference	Average Disturbance
1	3.43	1.43
2	2.86	1.86
3	3.00	2.14
4	3.57	1.43
5	3.00	2.29
6	3.29	1.86
7	4.14	1.71
8	2.57	3.00

Figure 9: Shows the average preference and disturbance ratings.

7 Conclusion

In this paper we have conducted a study of visual, auditory, and haptic feedback navigation techniques in guiding the user to find a hidden target. Our hypothesis is that multi-sensory feedback is more efficient in terms of the users' performance. Even though the obtained results support our hypothesis we advice to take under consideration the low amount of participants in this study, as well as the lack of permutations of the levels and the impact of the position of the target. Also the time constrains forced us to omit multiple techniques, which could potentially improve the quality of the obtained data. Therefore despite the satisfactory results the topic still needs some future work.

Acknowledgments

Here is a summary of the work done by each member of this group:

Mariusz:
Implemented: player object + scripts, movement script, UI objects and scripts, game logic class, baseline feedback, haptic feedback, auditory feedback. Performed data analysis, data visualization. Conduct experiment. Writing: Khora+Vitasim cases, Experiment Results, Discussion, Conclusion

Junfei:
Modelled the scene; implemented Visual Navigation Technique; 'Finding Nemo' game design and prefabs; updated the game logic: added staging phase of each level and updated UI, implemented time logging; designed user survey; conduct experiment; writing report: abstract, introduction, related work (2.3-2.5), technique, experiment; recorded, narrated and edited the demonstration video

Dennie:
Choosing of measurements. Game design in consideration to chosen measurements. Conducting experiment on test subjects. Writing report: Introduction, related work.

The authors wish to thank all participants, and all teachers from 5100-B2-2E22 Virtual Reality course.

References

[1] A. Chan and A. Ng. Finger response times to visual, auditory and tactile modality stimuli. *Lecture Notes in Engineering and Computer Science*, 2196:1449–1454, 03 2012.

[2] A. Cheng, L. Yang, and E. Andersen. Teaching language and culture with a virtual reality game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 541–549. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025857

[3] P. Gao, K. Matsumoto, T. Narumi, and M. Hirose. Visual-auditory redirection: Multimodal integration of incongruent visual and auditory cues for redirected walking. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 639–648, 2020. doi: 10.1109/ISMAR50242.2020.00092

[4] J. Hvass, O. Larsen, K. Vendelbo, N. Nilsson, R. Nordahl, and S. Serafin. Visual realism and presence in a virtual reality game. In *2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1–4, 2017. doi: 10.1109/3DTV.2017.8280421

[5] O. B. Kaul and M. Rohs. Haptichead: A spherical vibrotactile grid around the head for 3d guidance in virtual and augmented reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 3729–3740. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025684

[6] A. F. . F. L. L. Fialho, J. Oliveira. Soundspace vr: spatial navigation using sound in virtual reality. *Virtual Reality*, Nov. 2021. doi: 10.1007/s10055-021-00597-0

[7] A. Marquardt, J. Maiero, E. Kruijff, C. Trepkowski, A. Schwandt, A. Hinkenjann, J. Schöning, and W. Stuerzlinger. Tactile hand motion and pose guidance for 3d interaction. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, VRST '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3281505.3281526

[8] J. R. M. A. P. T. G. O. Rosenfeldt Nielsen M, Kristensen EQ. Clinical ultrasound education for medical students: Virtual reality versus e-learning, a randomized controlled pilot trial. *Ultrasound Q*, 37(3):292–296, Sept. 2021. doi: 10.1097/RUQ.0000000000000558

[9] E. Ruffaldi, A. Filippeschi, A. Frisoli, O. Sandoval, C. A. Avizzano, and M. Bergamasco. Vibrotactile perception assessment for a rowing training system. In *World Haptics 2009 - Third Joint EuroHaptics conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 350–355, 2009. doi: 10.1109/WHC.2009.4810849

[10] C. T. Tan, T. W. Leong, S. Shen, C. Dubravs, and C. Si. Exploring gameplay experiences on the oculus rift. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '15, p. 253–263. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2793107.2793117

[11] D. Yu, Q. Zhou, J. Newn, T. Dingler, E. Velloso, and J. Goncalves. Fully-occluded target selection in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3402–3413, 2020. doi: 10.1109/TVCG.2020.3023606

[12] Y. Zhang and Y. Song. The effects of sensory cues on immersive experiences for fostering technology-assisted sustainable behavior: A systematic review. *Behavioral Sciences*, 12(10), 2022. doi: 10.3390/bs12100361