Spring 5-19-2017

Measuring Presence in a Police Use of Force Simulation

Dharmesh Rajendra Desai
drdesai@uno.edu

Measuring Presence in a Police Use of Force Simulation

A Thesis

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfilment of the
requirements for the degree of

Master of Science
in
Computer Science
Specialization in Artificial Intelligence

by

Dharmesh Desai

BS in Electronics & Telecommunication, Pune University, 2011

May 2017

Table of Contents

List of Figures

List of Tables

ABSTRACT

We have designed a simulation that can be used to train police officers. Digital simulations are more cost-effective than a human role play. Use of force decisions are complex and made quickly, so there is a need for better training and innovative methods. Using this simulation, we are measuring the degree of presence that a human experience in a virtual environment. More presence implies better training. Participants are divided into two groups in which one group performs the experiment using a screen, keyboard, and mouse, and another uses virtual reality controls. In this experiment, we use subjective measurements and physiological measurements. We offer a questionnaire to participants before and after play. We also record the participants change in heart rate, skin conductivity and skin temperature using Empatica device. By comparing the data collected from both groups, we prove that people experience more presence in the virtual environment.

KEYWORDS: planning, Classical Planning, Narrative Planning, virtual environment, virtual reality, presence.

INTRODUCTION

An interactive virtual environment is an effective medium for education, training, therapy, and entertainment. With the help of a virtual reality device, a participant can take the role of a character and collaboratively construct an interactive narrative.

Here, we created a police use of force training simulation, which has a high potential for border impact. In this simulation, a user acts as a police officer and deals with a suspect in an immersive virtual reality simulation. We are using a narrative planner for drama management. The narrative planner is a variant of classical planning. It searches the sequence of actions to achieve the desired goal in such a way that all actions are clearly motivated and goal-oriented for the agents who performs these actions. A narrative planner generates all the possible outcomes from the current state and which helps to achieve goal, then a drama manager decides which action to perform in response to the user's actions. The user's goal is to use minimal force to defuse a potentially violent situation. The Police Executive Research Forum (PERF) has noted a fact in two publications (Police Executive Research Forum, 2012) that some officers leave the academy with a bias towards force because some training exercises focus on the correct application of force rather than the decision of whether and how much force to use. The main purpose of our project is to improve training and provide innovative methods for police training exercise.

This project will have a direct impact on police training and the relationship between police and citizens. We implemented this simulation with a virtual reality device. We claim the virtual reality provides higher presence, which can be defined as an experience that the user feels of being inside the game environment and not being aware of the real world. Overall, this simulation creates a new foundation for subsequent intelligent training, tutoring, and entertainment systems.

RELATED WORK

In this section, we will cover the basic concepts of planning, narrative planning, drama management, virtual environments and different presence measurement techniques like subjective measurement, behavioural measurement, and physiological measurement.

Planning

Planning is the science of reasoning about a sequence of actions which achieve some goal. Three components represent a planning problem (Russell, S.J. and Norvig, P., 2016): state, action, and goal. A state is a logical sentence where everything is represented in true or false. An action is a step to perform a transition from one state to another. Actions are specified in terms of preconditions that must satisfy before it can be executed and effects which becomes true after an action occurs. The goal is the target which we want to achieve; it is a partially defined state. A valid plan is the sequence of actions that achieves the goal.

Classical Planning

In classical planning, it is assumed that the environment is fully observable, static, and deterministic (Russell, S.J. and Norvig, P., 2016). All the actions are deterministic, and the planner has complete knowledge and control over the world, but in the real world, the planner doesn't have complete knowledge of the world, neither complete control over it. Classical Planning at the essence discusses "What to do" and "in which order". Planning problems can be represented using a logic-like representation of states. In a domain dependent plan the facts presented will involve the domain about which the system is expected to reason (Ginsberg, Matthew; Geddis, Donald F., 1991) and it is designed to work efficiently in a single problem domain. However, a domain independent plan is generated by a planning technique which is applicable in many domains and provides general planning capabilities (Wilkins, David E., 1983). As Wilkins (Wilkins, David E., 1983) explains, domain-specific planners are designed to work efficiently in a single problem domain. Usually, a planner depends upon the structure of the domain, and because of that, the underlying ideas cannot be used in another domain. However, a domain-independent planner yields a planning technique that is applicable in many domains and provides general planning capabilities. There are widely renowned examples of domain independent planning such as SRI's STRIPS (Fikes, Richard E.; Hart, Peter E.; Nilsson, Nils J., 1972), Penberthy and Weld's UCPOP (Penberthy, J. Scott; Weld, Daniel S., 1992), Haffmann and Nebel's Fast Forward (Hoffmann, J; Nebel, Bernhard, 2001), etc.

Narrative Planning

A narrative (Riedl, Mark; R, Michael Young, 2014) is a predetermined, temporally ordered set of actions or events. An interactive narrative (Riedl, Mark; R, Michael Young, 2014) is a form of digital entertainment in which users create or influence a dramatic storyline through actions, either by assuming the role of a character in a fictional world or by issuing commands to an autonomous, virtual non-player character. Narrative planning is a variant of classical planning which searches for a sequence of actions to achieve the author's goal such that all actions are clearly motivated and goal-oriented for agents who take them. In 2012, Haslum (Haslum, Patrik, 2012) and Riedl and Young (Riedl, Mark; R, Michael Young, 2014) explained that Narrative Planning is a type of planning with additional conditions. It places additional constraints on a planner's solution: some system level goal called the author's goal must be achieved, but agents must only act in service in their individual goals, possibly cooperating and competing with one another in the process. The main difference between narrative planning and classical planning is the notion of intentionality. Agents behave intentionally in such a way that the agent has some motivation behind the actions.

Drama Management

An Interactive Drama (Roberts, David L.; Isbell, Charles L., 2008) is one where a player is an active participant in how the story unfolds. Here the user makes the decisions for one agent. A players' exercise is to explore different parts of the environment and to engage other players and non-player characters by taking specific actions which make the non-player characters react to the behaviour of the player. This helps in making the exercise more engaging and player-driven. Thus, the author of the exercise designs specific situations that can be expected to happen during play. A Drama Manager (Roberts, David L.; Isbell, Charles L., 2008) is a coordinator. It tracks the narrative progress by directing roles and responses of objects for achieving specific narrative or training goals. The user makes the decision for one agent and the drama manager makes decisions for all other agents. Currently, computer games of skill are widely popular, and there is an increasing desire for immersive experience that is more akin to stories. These experiences are complex and deliver agency (Wardrip-Fruin; Mateas, Noah; Dow, Michael; Sali, Steven; Serdar, 2009) to the player to influence the way in which the experience unfolds. Drama management approaches are based on a set of plot points, a set of drama manager actions that can be taken in the game world, a model of player responses to DM actions, and a model of the authors intent.

Presence in Virtual Environment / Virtual Reality System

The exercise can be made more interactive and player-driven using a virtual reality (VR). Presence (Witmer, Bob J; Singer, Michael J, 1998) is an experience where the user feels that he is at a certain place even if he is somewhere else. Witmer and Singer (Witmer, Bob J; Singer, Michael J, 1998) explained presence as a normal awareness phenomenon which requires directed attention of the user and is based on the interaction between sensory stimulation and environmental factors. It encourages involvement, enables immersion, and involves internal tendencies (Witmer, Bob J; Singer, Michael J, 1998). Slater et al. (Slater, Mel; Linakis, Vasilis; Usoh, Martin; Kooper, Rob; Street, Gower, 1996) state that presence is a state of consciousness, the (psychological) sense of being in the virtual environment, and corresponding modes of behaviour.

Alexander at el. (Alexander, L Amy; Brunye Tad; Sidman, Jason; Weil, Shawn, 2005), demonstrates that presence increases engagement with training content. Heightened engagement should increase student time on training tasks, and time on task is one of the strongest predictors of acquisition and retention of knowledge and skill. In addition, there is some suggestion from the literature that presence is valuable in training because it increases motivation and provides a more engaging experience (Lombard, M. and Ditton, T., 1997). So, from all the above information, we assume that better presence indicates better training.

Virtual Reality (VR) is the system which sets the virtual environment. In 2003, Insko explained that Virtual Reality systems enable the user to feel as if the user is present in a computer-generated environment (Insko, Breat E., 2003). As per Insko's (Insko, Breat E., 2003) theory, there are three different ways to determine the user's presence in the virtual environment. These methods are Subjective, Behavioural, and Physiological.

Subjective Measurement

This method depends on the self-assessment of the user. Witmer and Singer (Witmer, Bob J; Singer, Michael J, 1998) created a Presence Questionnaire (PQ) to measure user presence in a VE. In addition, Witmer and Singer developed the Immersive Tendencies Questionnaire (ITQ) to measure differences in the tendencies of individuals who experience presence. The combination of these two questionnaires helps us to measure the presence of the user in a virtual environment (Witmer, Bob J; Singer, Michael J, 1998).

Behavioral Measurement

This is an objective approach to calculating the user's presence. The more a user is involved in a VE, the better he will respond. For example, suppose an object in a virtual environment is about to tackle the user, then user tries to save himself from that object by taking a step backwards. Insko (Insko, Breat E., 2003) explains that in this experiment we need to examine postural response as a possible way to measure presence. The more the user feels as if he was in that environment, the more the postural adjustment he would make. These observations are helpful in two ways (Freeman, Jonathan; Avons, S E; Meddis, Ray; Pearson, Don E; IJsselsteijn, Wijnand, 2000). The player is normally not aware of his postural response, and the postural measures have the capacity to produce differential levels of responses. Postural measures do not generate a binary result; Hence, it is easier to relate them to a graded subjective presence. We need to observe the degree in which the participant swayed back and forth.

Physiological Measurement

There are various physiological responses that can be measured.

Change in Heartrate (Insko, Breat E., 2003)*:* Many things can affect a person's heart rate, such as stress, fear, exertion, emotions etc. Heartrate increases when a person is under stress and heartrate decrees when one is relaxed. In general, heartrate increases in positive emotions and decreases in negative emotions. We can observe an increase or decrease in heartbeat per minute using an electrocardiogram (ECG).

Change in Skin Conductance (Insko, Breat E., 2003): It is the measure of the change in conductivity of person's skin. As stress increases, sweat on palm increases, and as sweat increases the conductivity of skin increases. Skin conductance also increases with the presentation of an unexpected stimulus, which can be measured on the fingertips of the user.

Change in Skin Temperature (Insko, Breat E., 2003): In this experiment Insko (Insko, Breat E., 2003) measure the skin temperature at the extremities of the body. As stress increase the temperature on the fingertips decreases. Skin temperature can be measured by placing a thermistor at the end of one of the figure and holding it in place with a thin porous tape.

As Insko (Insko, Breat E., 2003) mentioned these three parameters can also be used to measure physiological changes in the body when experiencing virtual environment situations. Meehan at el. (Meehan, Michael; Insko, Brent; Whitton, Mary; Brooks Jr, Frederick P, 2002) conclude that physiological reaction is reliable, valid, sensitive and objective presence measure.

THE SIMULATION

A narrative planner produces a story which guarantees to achieves the author's goal while ensuring that all character actions appear believable to the audience. Here the participant acts as a police officer and his goal is to use the minimum amount of force to catch the suspect. The user can perform several actions in the simulation like walk, talk, draw a gun, fire a gun etc., which helps the user to achieve the desired goal.

Training Narratives for Best Practices in Use of Force

Simulations and role-playing exercises are frequently used to train professionals (Hays, Robert T; Singer, Michael J, 1989) including nurses, military personnel, firefighters, and police. A police officer is expected to use the minimum level of force which is necessary to catch a suspect while ensuring his own safety and safety of civilians. This project provides a safe, immersive, repeatable virtual environment in which officers interact with the virtual agent to understand the consequences of their use of force decisions.

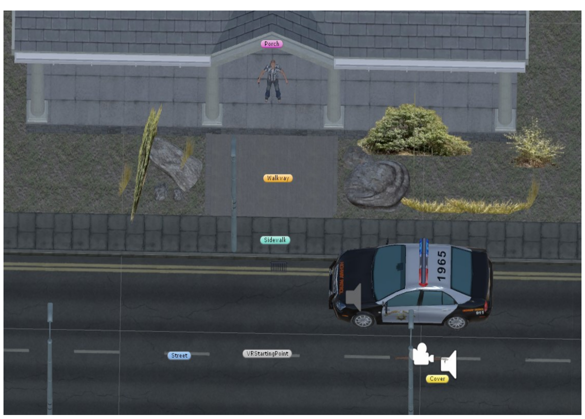

Figure 1: Screenshot of simulation from the top view and first person perspective.

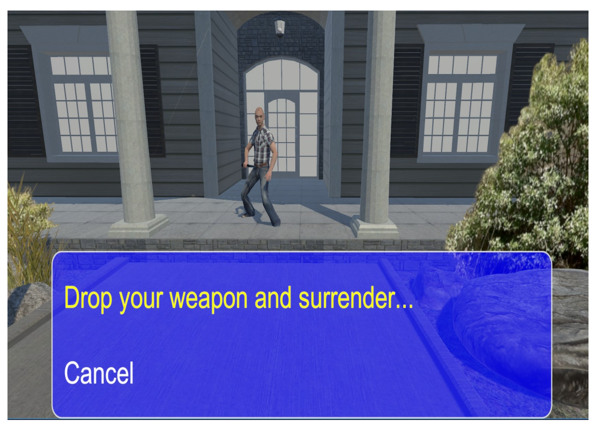

Figure 2: The way we can talk in simulation

Here we built a virtual reality training simulation as shown in figure 1. We used the drama manager and narrative planner to control all the non-player character (NPC) actions. This training simulation takes about 1 minute and designed to teach the officer a critical lesson identified by PERF (Police Executive Research Forum, 2012). In the scenario, we are trying to teach officers to keep distance and cover between themselves and the suspect who is armed with a knife. The officer can de-escalate the situation using the simple policy *distance + cover = time*. This simulation provides 9 different actions which a user can perform such as walk, draw, point, shoot, talk etc. This simulation includes 5 endings that range from the officer dying to the suspect surrendering peacefully. We perform this experiment on a small scale because highly interactive narratives are difficult to manage and each choice presented to the player contributes to the combinatorial explosion of possible stories that can be made. To overcome this drawback we could use automatic story generation techniques like narrative planning which can produce and manage much larger spaces than a human author.

Modular Architecture Of Intelligent Police Training Simulation

Here the goal of the experiment is that the user should understand the policy that distance + cover = time. In the simulation, if the officer does not understand this policy and will try to approach suspect immediately, then the suspect will attack the officer and the officer does not have time to evade, which forces the officer to use force or suffer harm. Here the officer learns the negative consequences of approaching an armed suspect. Then the officer replays the simulation, keeps distance, sees a peaceful ending, and learns the positive consequences of this policy. In this way, the officer learns the domain and understands the preconditions and effects of their actions. As per Mestre's (Mestre, Daniel R, 2011) theory, presence is the sensation of being there (part of Virtual Environment VE). In addition, presence has often been found as a sign of "ecological validity" of virtual reality (VR) devices, also as a sign of potential positive transfer of skills or knowledge learned in VE to the real world. As Herbelin et. al. (Herbelin, Bruno; Vexo, Frederic; Thalmann, Daniel, November 2002) discussed, VR finds effective application in many different fields where simulation training is preferable to real training, for example, aviation, surgery etc. Also, VR is used for entering mental areas. A therapist could help people understand, accept, and control the sensation to fight their phobias (cognitive therapy). Hence it looks like presence is an important factor in the exposure success. Social Anxiety Disorder (SAD) is a concrete application of this idea. In future experiments, we will test learning in some other ways like behavioural measurements.

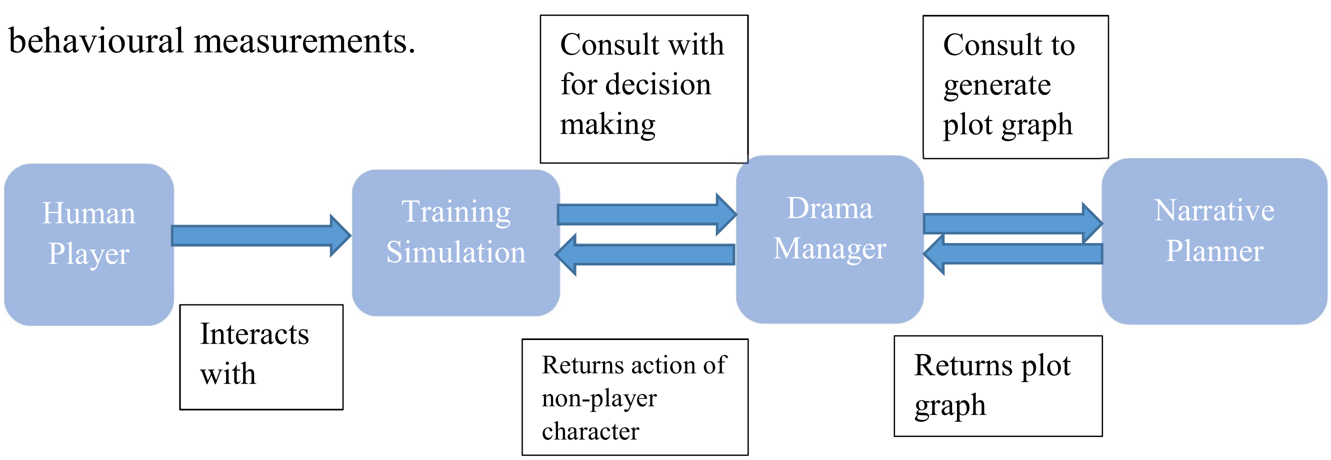

Figure 3: Modular Architecture of Intelligent Police Training Simulation

Figure 3 shows the architecture of the police training simulation. In this simulation, the user first starts with a tutorial so that the user is familiar with the VR controllers to play the simulation and keyboard controls to play using the keyboard. In the next step, the officer interacts with the simulation, which directly consults with the drama manager to decide how non-player characters will act in response to the user's action. The drama manager gets all the possible actions from the narrative planner, then the drama manager chooses the best action for a non-player character. A plot graph defines the space of legal story progression and ultimately determines the possible events at any given point in time. Our simulation uses a plot graph which has 2408 nodes and 9458 edges. A plot graph defines the space of legal story progression and ultimately determines possible events at any given point in time. A plot graph (Li, Boyang; Lee-Urban, Stephen; Johnston, George; Riedl, Mark O, 2013) is a commonly used representation in story generation systems. The ending is scored based on the final state of the game world. The state where the officer is dead has the lowest score and the state where all the characters are alive and unharmed and the suspect surrendered with minimum use of force has the highest score. The officer who correctly understands all the policies will be able to get the highest score.

EVALUATION

Figure 4: HTC Viva and Empatica E4 device

We developed the police use of force training simulation for two different platforms, a virtual reality which uses the HTC Vive VR device and Empatica E4 device as shown in Figure 4, to collect physiological data like heart rate, skin temperature, skin conductivity etc., and a computer (Windows OS) by using the Unity game engine.

As this experiment is very time consuming, so here in this experiment, a total of 22 people from the University of New Orleans, LA participated. To perform the experiment with just 22 participants took approximately 23 hours. So, because of shortage of time we just used a data set of 22 participant. Half of the participants played the screen keyboard version first, and the other half played the VR version first to control for an ordering effect. At the beginning of the simulation, we assign a unique id to each participant then we show them a short video introduce the participant about the controls and some precautionary aspects which the participant needs to be aware of while playing the simulation. After the video was finished, we strapped the Empatica wrist watch on their hand, which provides the physiological data. Then we introduce the participant to the controllers of the first simulation (screen keyboard or virtual reality). After they play, we give them a questionnaire which is developed by Witmer and Singer (Witmer, Bob J; Singer, Michael J, 1998) called the Presence Questionnaire (PQ) which measures the degree to which the individual experienced presence in the virtual environment. As this questionnaire is dependent on the virtual environment experience, we give this questionnaire to the user after the user played the first simulation. This questionnaire relies exclusively on self-report information. PQ uses a seven-point scale format which is based on the semantic differential principle (Dyer, Robert F., 1976). The PQ asks users to select an appropriate scale in the accordance with the question content and descriptive labels. An example item from PQ is shown as follows.

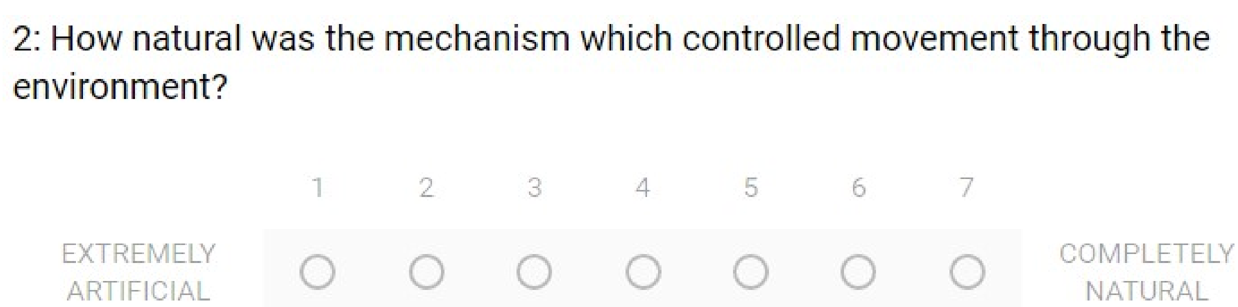

Figure 5: An example item from PQ

The complete questionnaire is given in appendix 1.

After completion of the PQ, the participant plays the second version of the simulation. Then we give another questionnaire which we created to measure the difference in the participant's presence in both versions of simulation.

In addition to subjective measures of presence, we also use objective physiological measurements. We monitored the change in heart rate, skin conductivity and skin temperature of the participant while playing the simulation. Because some other factors may influence a subjective measurement technique, for example, the participant is tense before they start playing the simulation, so the participant's feedback may be affected as per their mood, heart rate will give more accurate readings.

We created a checklist of steps which we followed for each participant. We explained the important steps above. However, the complete checklist is provided in the appendix 3.

Experimental Design

Experiment 1

We performed this experiment within subjects, meaning we compare the user's different experiences in the virtual reality and screen keyboard versions. So, for this part, we consider the data that asks the participants to compare the two experiences.

We used a modified version of Presence Questionnaire (Witmer, Bob J; Singer, Michael J, 1998) that asks subjects to compare the two experiences.

An example question is given below:

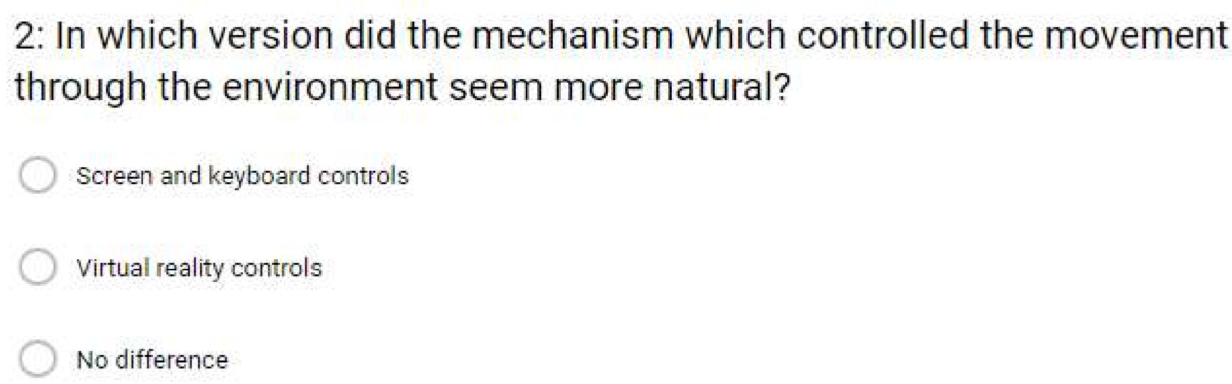

Figure 6: An example question for comparative study

We also used physiological data here. For each participant, we compare the difference between average heart rate, skin conductivity, and skin temperature for both versions of the simulation.

Hypothesis for Experiment 1

The null hypothesis is the default position which says that there is no relation between two measured phenomena. For this simulation, the null hypothesis is that there is no difference between presence in the virtual reality and screen keyboard versions of simulation, and because of that participants felt the same in both versions. Hence, we should get the probability as a 50% for VR and 50% for others. An alternate hypothesis is that significant preference should be given to the virtual reality simulation. User's feel more presence in virtual reality than a computer simulation. In contrast, for negatively worded question significant preference should be given to screen keyboard version. We consider results for which $p <= 0.1$ marginally significant, and results for which $p < 0.05$ significant. In the questionnaire, some questions are positively worded and some are negatively worded. For example, a positively worded question is, "In which version did your interactions with the environment seem more natural?" However, the negatively worded question is "Which version's visual display quality distracted you more from performing assigned tasks or required activities?"

Experiment 2

We performed this experiment between subjects, meaning we compare the user's experience with another user in the different version of simulation. That means we only consider the first half of the data collected for each participant. We will consider the experience of people who played the screen and keyboard version and compare those to the experiences of those who played virtual reality version. We used the original Presence Questionnaire here. This questionnaire measures presence based on control factors, sensory factors, distraction factors and realism factors. We cannot use the physiological data in this experiment, because each person's baseline heart rate, skin conductivity, and skin temperature are different, so it doesn't make sense to compare one person to another.

Hypothesis for Experiment 2

The null hypothesis for experiment 2 is that we get the same response from participants who played the screen and keyboard version and the virtual reality version. An alternate hypothesis is that the participants who played the virtual reality version should feel more presence.

RESULTS

Experiment 1

In this experiment, each participant plays both versions of the simulation.

The binomial test is a statistical test of the significance of deviations from a theoretically expected distribution of observations into two group. Using this we calculate the p-value to prove our hypothesis.

We calculated p-value using binomial test in R programming language.

Steps to calculate p-value:

$$P(X) = \frac{n!}{(n-X)!\,X!} \cdot (p)^X \cdot (q)^{n-X}$$

n = number of trials

X = number of successor

Step 1: first calculate n! /(n-X)! X!

Step 2: find p (probability of successor) and q (probability of failure), here calculate percentage of it.

Step 3: find p^X

Step 4: find q^(n-X)

Step 5: multiply step 1,3,4.

This is the set of questions where VR should be preferred.

Table 1: Questionnaire table which is positive questions for our hypothesis

Questions	p-value	Support?
In which version did your interactions with the environment seem more natural?	0.0006	Success
In which version did the mechanism which controlled the movement through the environment seem more natural?	0.0006	Success
Which environment seemed more consistent with your real-world experiences?	0.0004	Success
In which version, did you adjust to the environment more quickly?	0.4159	Fail
In which version were all your senses more engaged?	0.0005	Success
In which version, did you feel more involved?	0.0002	Success
In which version were you better able to learn new techniques that enabled you to improve your performance?	0.5840	Fail
In which version were you more likely to have lost track of time?	0.0002	Success
In which environment was it easier to survey or search the environment using vision?	0.0021	Success
Which environment's visual aspects involved you more?	0.0006	Success
In which version were you better able to examine objects?	0.0006	Success
In which version was it easier to examine the objects from multiple viewpoints?	0.0262	Success
In which version was your sense of objects moving through space more compelling?	0.0002	Success
In which environment was your sense of moving around more compelling?	0.0005	Success
Which environment allowed you to control events more?	0.7383	Fail
Which environment was more responsive to actions that you initiated?	0.7383	Fail

In which environment was it easier to anticipate what would happen next in response to the actions that you performed?	0.7383	Fail
In which environment, did you experience less delay between your actions and expected outcomes?	0.8568	Fail
At the end of which version did you feel more proficient in moving and interacting with the environment?	0.0669	Marginal success
In which version did the auditory aspects of the environment involve you more?	0.0669	Marginal success
In which version were you better able to identify sounds?	0.2617	Fail
In which version were you better able to localise sounds?	0.2617	Fail
In which version was it easier to concentrate on the assigned task or required activity rather than the mechanism used to perform that task or activity?	0.1431	Fail

Below are the negative questions so in this case the screen keyboard version should preferred.

Table 2: Questionnaire table which is negative questions for our hypothesis

Questions	**p-value**	**Support?**
Which version's visual display quality distracted you more from performing assigned tasks or required activities?	0.9915	Fail
Which version's control devices interfered with the performance of assigned tasks or with other activities more?	0.9915	Fail
In which version was the information coming from your senses more inconsistent or disconnected?	0.7382	Fail
In which version were you more aware of events occurring in the real world around you?	0.7383	Fail

Our Measurement of Physiological Data

Table 3: Table to show physiological data

Participant No	**Controls**	**Average Heart Rate**	**Average Skin Conductivity**	**Average Skin Temperature**

1	screen and keyboard controls	93.13296	1.365878	29.69736334
1	virtual reality controls	87.66618	2.408592	29.23580101
2	screen and keyboard controls	74.46801	2.15846	30.65599678
2	virtual reality controls	99.6205	0.007473	30.12836228
3	screen and keyboard controls	69.1	0.168976	29.55513274
3	virtual reality controls	93.61898	0.143896	29.86265306
4	screen and keyboard controls	102.4244	0.507215	31.20089166
4	virtual reality controls	78.95063	1.040144	30.66124011
5	screen and keyboard controls	68.96916	11.9486	32.10810427
5	virtual reality controls	76.35813	12.50022	31.47917517
6	screen and keyboard controls	71.80843	3.095694	32.71871901
6	virtual reality controls	86.88489	5.52379	31.83867021
7	screen and keyboard controls	67.44764	4.392367	32.65645217
7	virtual reality controls	78.3685	7.075468	32.29128898
8	screen and keyboard controls	92.00448	1.522156	30.6822707
8	virtual reality controls	94.04915	1.813999	30.0840688
9	screen and keyboard controls	53.97053	2.082882	30.50658863
9	virtual reality controls	86.92129	1.833984	29.8316163
10	screen and keyboard controls	89.22948	0.206194	32.7043675

10	virtual reality controls	82.9113	0.378399	31.98153693
11	screen and keyboard controls	84.85676	0.218196	33.03330882
11	virtual reality controls	91.87771	0.175697	31.85036068
12	screen and keyboard controls	90.0364	4.060165	31.72409091
12	virtual reality controls	91.94308	2.100154	31.34811765
13	screen and keyboard controls	84.89549	0.130559	32.50706577
13	virtual reality controls	90.35211	0.26589	31.65132502
14	screen and keyboard controls	78.12844	0.548435	30.68305483
14	virtual reality controls	84.93288	0.774215	31.07085011
15	screen and keyboard controls	85.49168	10.59248	30.36062331
15	virtual reality controls	78.3774	13.94834	30.43367751
16	screen and keyboard controls	93.64552	0.596234	32.68051121
16	virtual reality controls	88.35546	0.255295	32.06626923
17	screen and keyboard controls	108.1138	0.70888	33.10072368
17	virtual reality controls	96.9003	0.531216	33.59030395
18	screen and keyboard controls	68.0432	0.173165	27.81529557
18	virtual reality controls	87.44283	0.214145	28.05603279
19	screen and keyboard controls	69.95897	0.276633	31.6670418
19	virtual reality controls	97.30571	0.251751	31.29666997

20	screen and keyboard controls	72.46048	2.237906	32.15057711
20	virtual reality controls	74.91057	2.612564	30.98967403
21	screen and keyboard controls	76.73655	3.012486	32.82266055
21	virtual reality controls	92.21324	3.221368	32.6200232
22	screen and keyboard controls	93.1772	0.770983	30.47915191
22	virtual reality controls	96.16449	1.911942	31.71206089

Heart Rate

For 16 out of 22 participants average heart rate is higher in virtual reality controls version than the screen keyboard version. By applying the binomial test, we get a p-value of 0.0262. From this, we can reject the null hypothesis and our alternate hypothesis is supported.

Skin Conductivity

For 14 out of 22 participants, average skin conductivity is higher in virtual reality controls version than in the screen and keyboard version. By applying the binomial test, we get a p-value of 0.1431. From this we failed to reject null hypothesis.

Skin Temperature

As per the Insko, as stress increases the temperature in the extremities decreases. For 16 out of 22 participants, average skin temperature is less in virtual reality controls version than in the screen and keyboard version. By applying binomial test, we get a p-value of 0.0262 which is marginal probability. From this we can reject null hypothesis.

Experiment 2

The Wilcoxon sum-rank test (Lehmann, 1975) was used to compare the results. This is a non-parametric test for comparing an independent sample of data. It not only allows us to reject the null hypothesis but also to support our alternate hypothesis that in the virtual reality version the user experiences more presence. To prove this, we gave the PQ to two different groups where one group played the virtual reality version and another group played the screen and keyboard version. Now we can compare both the results.

Figure 5 shows the Likert scale of the participants who played the virtual reality version. We have a total of 22 participants, and out of that 11-people played the virtual reality version and 11 people played the screen and keyboard version. People who played the virtual reality controls version agreed more than the people who played on screen and keyboard version. The p-value is significant for "How much did your experiences in the virtual environment seem consistent with your real-world experiences?" because people who played the virtual reality version agreed more strongly than people who played the screen and keyboard version.

When we program Wilcoxon test in “R”, it cannot compute exact p-values for the data where ties occur.

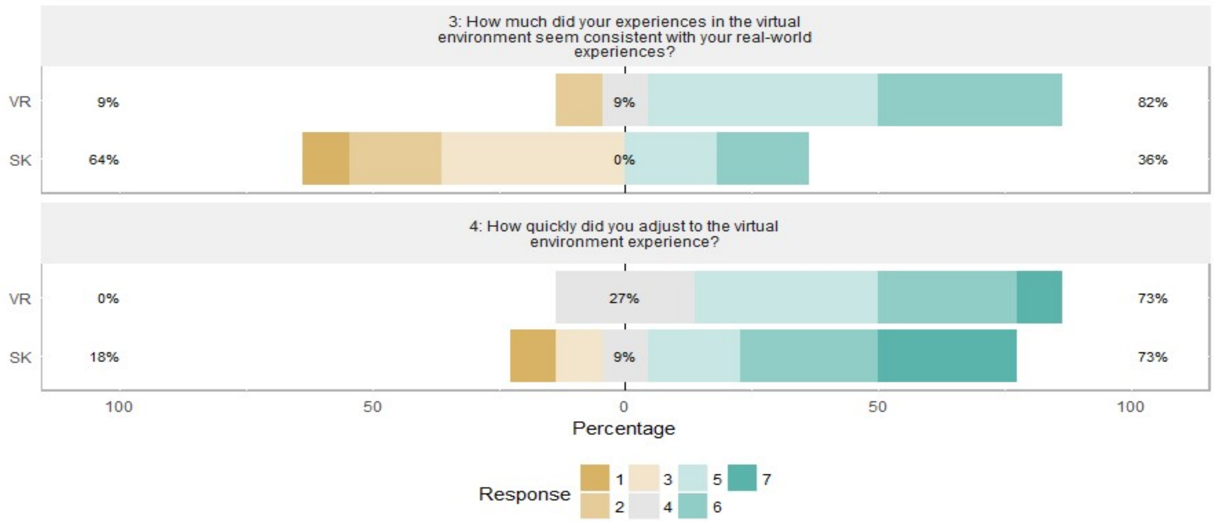

Figure 7: Diverging Stacked Bar Chart for Likert Scale Q3, Q4

Above figure shows the example question and participants responses on that question and the rest of the graphs can be found in the appendix 4.

From appendix 4 figures, we can say that for most of the questions participants who played the virtual reality controls version agreed more than the people who played on SK version.

Also, we test our hypothesis using Wilcoxon sum-rank test.

Table 4: PQ which is positive questions for our hypothesis (Wilcoxon sum-rank test)

Questions	**p-value (SK<VR)**	**Success?**
How natural did your interactions with the environment seem?	0.80	Fail
How natural was the mechanism which controlled movement through the environment?	0.10	Marginal success

How much did your experiences in the virtual environment seem consistent with your real-world experiences?	0.05	Success
How quickly did you adjust to the virtual environment experience?	0.80	Fail
How completely were all your senses engaged?	0.09	Marginal success
How involved were you in the virtual environment experience?	0.10	Marginal success
Did you learn new techniques that enabled you to improve your performance?	0.90	Fail
Were you involved in the experimental task to the extent that you lost track of time?	0.05	Success
How completely were you able to actively survey or search the environment using vision?	0.60	Fail
How much did the visual aspects of the environment involve you?	0.20	Fail
How closely were you able to examine objects?	0.50	Fail
How well could you examine objects from multiple viewpoints?	0.40	Fail
How compelling was your sense of objects moving through space?	0.08	Marginal success
How compelling was your sense of moving around inside the virtual environment?	0.10	Marginal success
How much were you able to control events?	0.30	Fail
How responsive was the environment to actions that you initiated (or performed)?	0.80	Fail
Were you able to anticipate what would happen next in response to the actions that you performed?	0.70	Fail
How much delay did you experience between your actions and expected outcomes?	0.60	Fail
How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	0.60	Fail
How much did the auditory aspects of the environment involve you?	0.20	Fail
How well could you identify sounds?	1.00	Fail
How well could you localize sounds?	0.80	Fail

How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?	0.60	Fail

Table 5:PQ which is negative questions for our hypothesis (Wilcoxon sum-rank test)

Questions	**p-value (SK>VR)**	**Success?**
How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?	0.70	Fail
How much did the control devices interfere with the performance of assigned tasks or with other activities?	0.70	Fail
How inconsistent or disconnected was the information coming from your various senses?	0.80	Fail
How aware were you of events occurring in the real world around you?	0.01	Success

DISCUSSION

As we have discussed in the previous section, we rejected the null hypothesis in most cases, but we failed in some cases. We will discuss the reasons for failure cases.

- "Which environment allowed you to control events more?"

  Participants are used to playing games and simulations using a screen and keyboard, but virtual reality is a new concept. So, participants need to remember some things like they can talk using the touch pad etc.

- "In which environment was it easier to anticipate what would happen next in response to the actions that you performed?"

> In some situations, participants were more conscious of the real environment, for example in the screen and keyboard version, the officer moves forward and backwards by just pressing the up and down keys. However, in virtual reality participants need to walk with their feet. So, as a participant was completely involved in the virtual reality environment, he may have been afraid to move backwards as he might hit some object or wall. In addition, in real time it is difficult to walk backwards.

In general, our hypothesis is failed where participants need to interact with the environment, the auditory aspects, or display quality. But we succeed in some cases like virtual reality controls environment seems more natural, mechanism to control the movements, consistency with the real world.

We support our hypothesis using physiological data as participant's average heart rate is higher in virtual reality as compared to screen keyboard version and skin temperature is higher is screen keyboard version than virtual reality. As stress increases skin temperature decreases at the extremities.

FUTURE WORK

In future versions of this kind of simulation new narrative content will be drawn from issues identified by PERF (Police Executive Research Forum, 2012) and from the role-playing scenarios developed by the New Orleans Police Department's newly developed EPIC (Ethical Policing Is Courageous) curriculum (New Orleans Police Department NOPD. EPIC (Ethical Policing Is Courageous) program guide, 2016). These concepts include communicating with the suspect, involving dispatchers, ensuring the safety of civilian, etc.

We can improve this simulation further in multiple ways as follows:

- We will implement this simulation on a larger scale so that participant get more physical space to walk around and to perform different actions.
- In this simulation, we are using a text popup for the talk action, but in the future, we could use speech recognition techniques so that users can directly talk to virtual characters.
- We can add some more actions to the simulation like a drop weapon, pickup weapon, we can add more weapons and some civilians so that user feels it is more natural.
- We could use additional measurement techniques like behavioural measurements to measure presence.

CONCLUSION

In the police use of force simulation the main component we used is the narrative planner and drama manager. The narrative planner provides all the possible states to the drama manager and the dram manager decides the next action for non-player character to perform. In our evaluation, we used two measurement techniques, subjective and physiological. In the subjective measurement, we used PQ and modified version of PQ (Witmer, Bob J; Singer, Michael J, 1998) that asks participants to compare the two experiences. We rejected null hypothesis and supported our hypothesis using physiological data such as heart rate and skin temperature. Overall, this experiment shows that users feel more presence in virtual reality than in a computer simulation.

REFERENCES

Alexander, L Amy; Brunye Tad; Sidman, Jason; Weil, Shawn. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group.*

Dyer, Robert F. (1976). *Questionnaire Construction Manual.* Operations Research Associates Palo Alto.

Farrell, Rachelyn; Ware, Stephen G. (2016). Fast and diverse narrative planning through novelty pruning. *Proceedings of the 12th AAAI International Conference on Artificial Intelligence and Interactive Digital Entertainment*, (pp. 37-43).

Fikes, Richard E.; Hart, Peter E.; Nilsson, Nils J. (1972). Learning and executing generalized robot plans. *Artificial Intelligence, 3*(1972), 251-288.

Freeman, Jonathan; Avons, S E; Meddis, Ray; Pearson, Don E; IJsselsteijn, Wijnand. (2000). Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli. *Presence, 9*, 149-164.

Ginsberg, Matthew; Geddis, Donald F. (1991). Is there any Need for Domain-Dependent Control Information? *Proceedings of the Ninth National Conference on Artificial Intelligence*, (pp. 452-457).

Haslum, Patrik. (2012). Narrative Planning: Compilations to Classical Planning. *Journal of Artificial Intelligence Research, 44*, 383-395.

Hays, Robert T; Singer, Michael J. (1989). *Simulation fidelity in training system design: Bridging the gap between reality and training.*

Herbelin, Bruno; Vexo, Frederic; Thalmann, Daniel. (November 2002). Sense of Presence in Virtual Reality Exposure Therapy. *In proceeding of the 1st International Workshop on Virtual Reality Rehabilitation.*

Hoffmann, J; Nebel, Bernhard. (2001). The FF planning system: fast plan generation through heuristic search. *Journal of Artificial Intelligence Research, 14*, 253-302.

Insko, Breat E. (2003). Measuring presence: Subjective, behavioural and physiological methods. In *Being There: Concepts, effects and measurement of user presence in synthetic environments* (pp. 110-119). IOS Press.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day.

Li, Boyang; Lee-Urban, Stephen; Johnston, George; Riedl, Mark O. (2013). Story Generation with Crowdsourced Plot Graphs. *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, (pp. 598-604).

Lombard, M. and Ditton, T. (1997). At The Heart Of It All: The Concept Of Presence. *Journal of Computer Mediated Communication, 3*(2), 0-0.

Meehan, Michael; Insko, Brent; Whitton, Mary; Brooks Jr, Frederick P. (2002). Physiological Measures of Presence in Stressful Virtual Environments. *ACM Transactions on Graphics (TOG), 21*, 645-652.

Mestre, Daniel R. (2011). Immersion and Presence. *Movement & Perception* (pp. 93-100). CRC Press.

New Orleans Police Department NOPD. EPIC (Ethical Policing Is Courageous) program guide. (2016).

Penberthy, J. Scott; Weld, Daniel S. (1992). UCPOP: A Sound, Complete, Partial Order Planner for ADL. *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning*, *Volume 92*, pp. 103-114.

Police Executive Research Forum. (2012). *An Integrated Approach to De-escalation and Minimizing Use of Force.*

Riedl, Mark; R, Michael Young. (2014). The Importance of Narrative as an Affective Instructional Strategy. *Adaptive Instructional, 2*.

Roberts, David L.; Isbell, Charles L. (2008). A survey and qualitative analysis of recent advances in drama management. *International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning, 4*, 61-75.

Russell, S.J. and Norvig, P. (2016). *Artificial Intelligence: A Modern Approach.* Pearson Education.

Sacerdoit, Earl D. (1977). *Structure of Plans and Behavior.* SRI International Menlo Park CA Aritificial Intelligence Center.

Slater, Mel; Linakis, Vasilis; Usoh, Martin; Kooper, Rob; Street, Gower. (1996). Immersion, Presence, and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess. *ACM Virtual Reality Software and Technology VRST*, (pp. 163-172).

The Police Executive Research Forum PERF. An integrated approach to de-escalation and minimizing use of force. (2012). Critical Issues in Policing Series.

Vere, Steven A. (1983). Planning in Time: Windows and Durations for Activities and Goals. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5*, 246-267.

Wardrip-Fruin; Mateas, Noah; Dow, Michael; Sali, Steven; Serdar. (2009). Agency reconsidered. *Proceedings of the Digital Games Research Association.*

Ware, Stephen G.; Young, R. Michael. (2014). Glaive: a state-space narrative planner supporting intentionality and conflict. *Proceedings of the 10th AAAI International Conference on Artificial Intelligence and Interactive Digital Entertainment*, (pp. 80-86).

Wilkins, David E. (1983). Representation in a Domain-Independent Planner. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, (pp. 733-740).

Witmer, Bob J; Singer, Michael J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Teleoperators and Virtual Environments,*, 25-240.

APPENDIX

Appendix 1

- How natural did your interactions with the environment seem?
- How natural was the mechanism which controlled movement through the environment?
- How much did your experiences in the virtual environment seem consistent with your real-world experiences?
- How quickly did you adjust to the virtual environment experience?
- How completely were all your senses engaged?
- How involved were you in the virtual environment experience?
- Did you learn new techniques that enabled you to improve your performance?
- Were you involved in the experimental task to the extent that you lost track of time?
- How completely were you able to actively survey or search the environment using vision?
- How much did the visual aspects of the environment involve you?
- How closely were you able to examine objects?
- How well could you examine objects from multiple viewpoints?
- How compelling was your sense of objects moving through space?
- How compelling was your sense of moving around inside the virtual environment?
- How much were you able to control events?
- How responsive was the environment to actions that you initiated (or performed)?
- Were you able to anticipate what would happen next in response to the actions that you performed?
- How much delay did you experience between your actions and expected outcomes?
- How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
- How much did the auditory aspects of the environment involve you?
- How well could you identify sounds?
- How well could you localise sounds?
- How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
- How much did the control devices interfere with the performance of assigned tasks or with other activities?
- How inconsistent or disconnected was the information coming from your various senses?
- How aware were you of events occurring in the real world around you?

- How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

Appendix 2

- In which version did your interactions with the environment seem more natural?
- In which version did the mechanism which controlled the movement through the environment seem more natural?
- Which environment seemed more consistent with your real-world experiences?
- In which version, did you adjust to the environment more quickly?
- In which version were all your senses more engaged?
- In which version, did you feel more involved?
- In which version were you better able to learn new techniques that enabled you to improve your performance?
- In which version were you more likely to have lost track of time?
- In which environment was it easier to survey or search the environment using vision?
- Which environment's visual aspects involved you more?
- In which version were you better able to examine objects?
- In which version was it easier to examine the objects from multiple viewpoints?
- In which version was your sense of objects moving through space more compelling?
- In which environment was your sense of moving around more compelling?
- Which environment allowed you to control events more?
- Which environment was more responsive to actions that you initiated?
- In which environment was it easier to anticipate what would happen next in response to the actions that you performed?
- In which environment, did you experience less delay between your actions and expected outcomes?
- At the end of which version did you feel more proficient in moving and interacting with the environment?
- In which version did the auditory aspects of the environment involve you more?
- In which version were you better able to identify sounds?
- In which version were you better able to localise sounds?
- Which version's visual display quality distracted you more from performing assigned tasks or required activities?

- Which version's control devices interfered with the performance of assigned tasks or with other activities more?
- In which version was the information coming from your senses more inconsistent or disconnected?
- In which version were you more aware of events occurring in the real world around you?
- In which version was it easier to concentrate on the assigned task or required activity rather than the mechanism used to perform that task or activity?

Appendix 3

- The participant MUST read and sign an informed consent form.
- Keep the signed consent form.
- Give the participant a second copy of the consent form to take home. It does not need to be signed.
- Assign a unique ID number to the participant and writes it at the top of this page.
- Ask the participant to sit at Computer 1 and put on the headphones.
- Show the participant the introduction video.
- While the video is playing, set up the first survey on Computer 2 and enter the participant’s number on the first screen.
- After the video is finished, strap the Empatica watch to their arm. Press and hold the button for two seconds until the light turns green.
- Ask the participant to take the first survey on Computer 2.
- While the participant is taking the first survey, set up the virtual reality version of the simulation on Computer 1 and confirm that sound is coming through the headphones.
- After the participant finishes the first survey, bring the participant back to Computer 1.
- Show the participant the Vive hand controllers but do not hand them to the participant yet.
- Let the participant know you will tap their shoulder to talk to them.
- Help the participant put on the Vive headset.
- Hand the participant the Vive hand controllers one at a time. Make sure their wrists are through the wrist straps do they don’t drop the controllers.
- Put the headphones on the participant.

- Watch the participant the whole time they are playing. *Focus on the cable and the participant, not the screen, whenever possible*.
- When the participant has finished the tutorial, give them the verbal quiz about the controls.
- When the participant is finished playing, set up the second survey on Computer 2 and enter the participant number on the first page.
- The participant takes the second survey on Computer 2.
- While the participant is taking the second survey, set up the screen and keyboard simulation on Computer 1.
- Make sure the image displaying the controls is visible on the second screen.
- After the participant finishes the second survey, bring the participant back to Computer 1.
- Put the headphones on the participant.
- The participant plays the screen and keyboard version of the simulation.
- When the participant is finished with the tutorial, give them the verbal quiz about the controls.
- While the participant is playing, set up the third survey on Computer 2 and enters the participant number on the first page.
- The participant takes the third survey on Computer 2.
- Press and hold the button on the Empatica watch for 2 seconds until the light turns off. Remove the watch from the participant.

Appendix 4

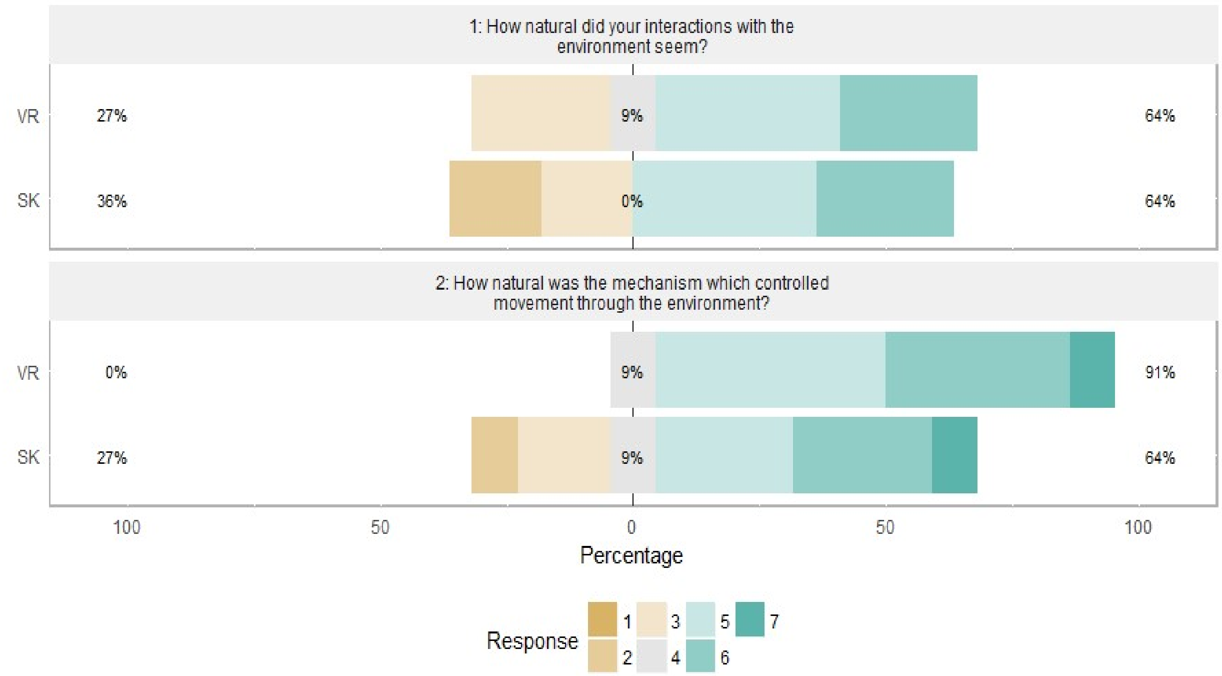

Figure 8: Diverging Stacked Bar Chart for Likert Scale Q1, Q2

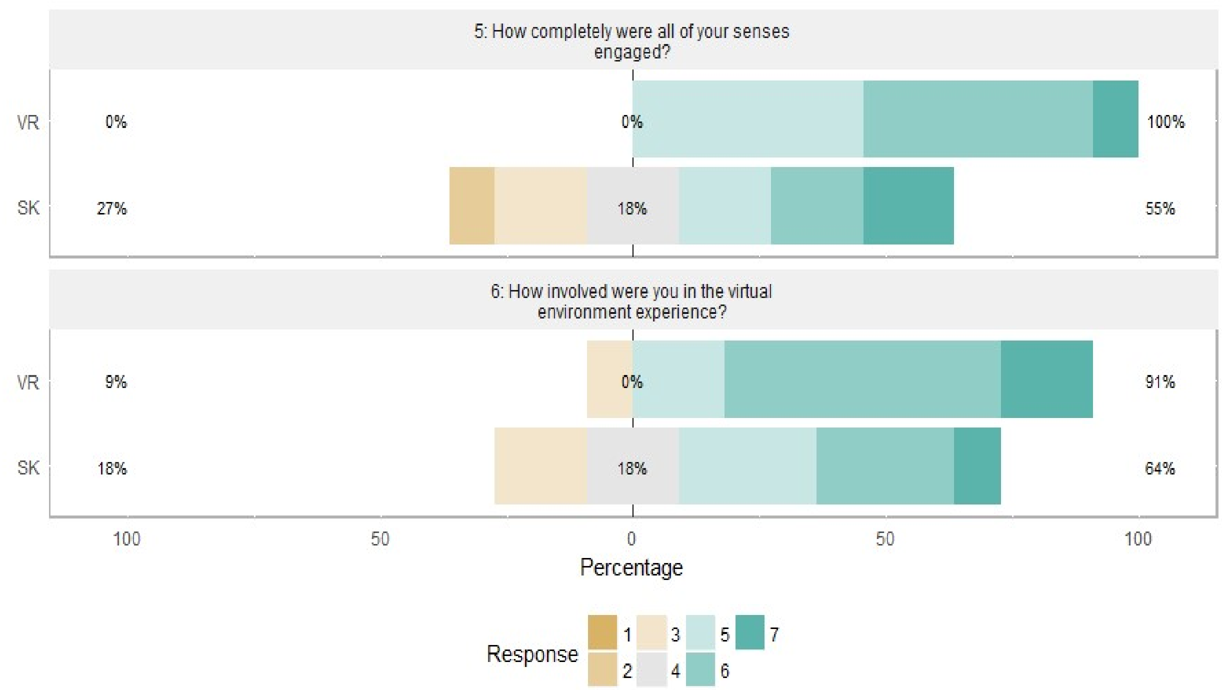

Figure 9: Diverging Stacked Bar Chart for Likert Scale Q5, Q6

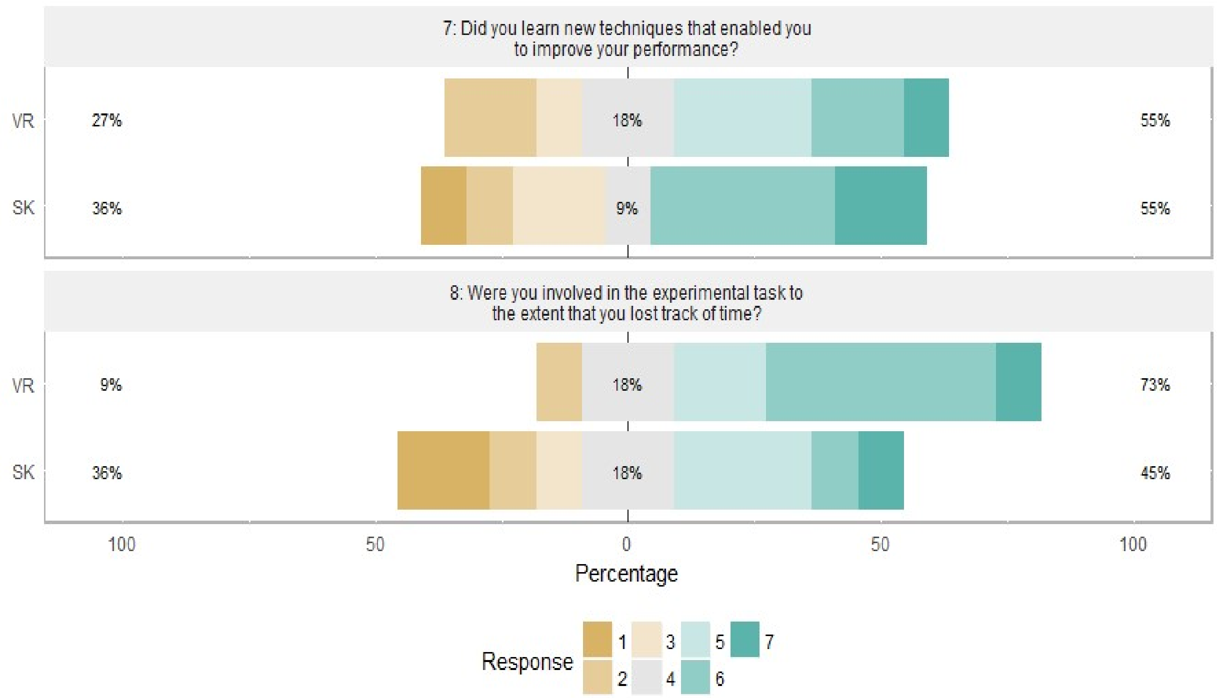

Figure 10: Diverging Stacked Bar Chart for Likert Scale Q7, Q8

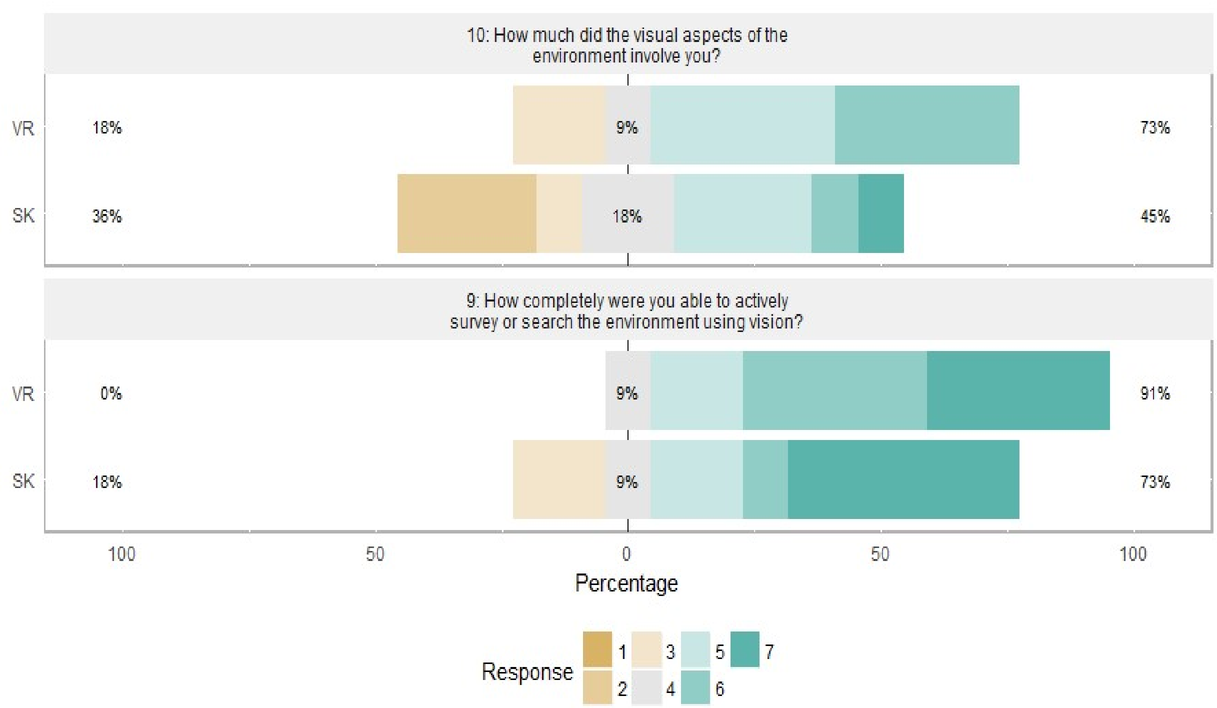

Figure 11: Diverging Stacked Bar Chart for Likert Scale Q9, Q10

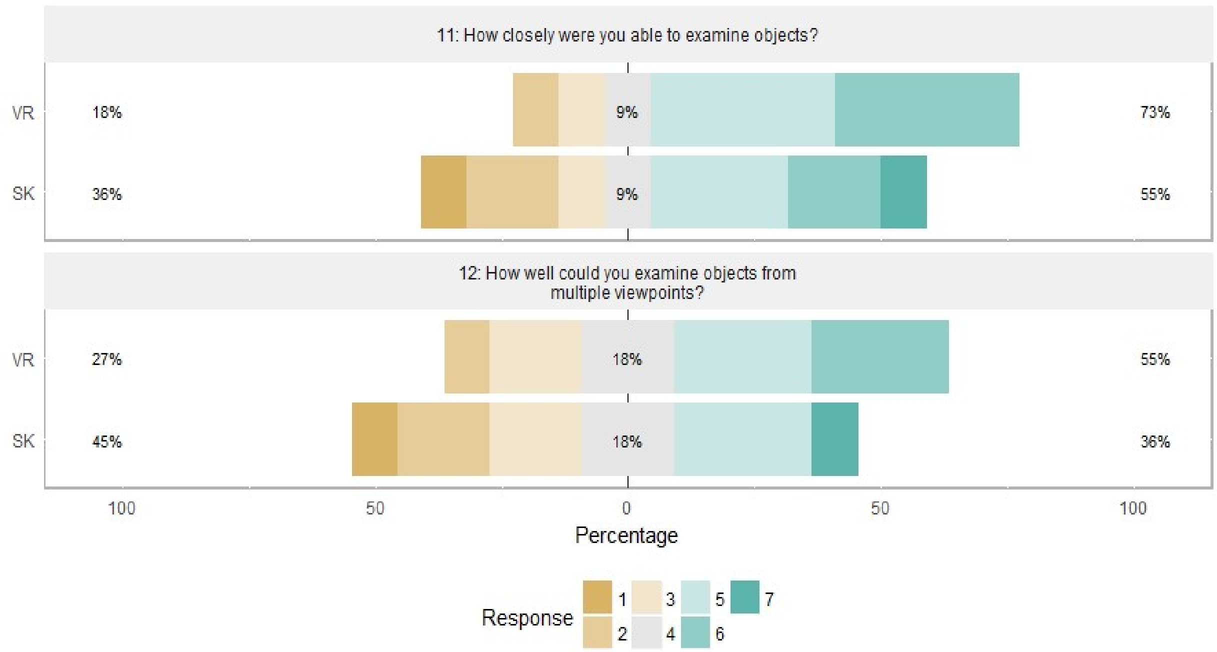

Figure 12: Diverging Stacked Bar Chart for Likert Scale Q11, Q12

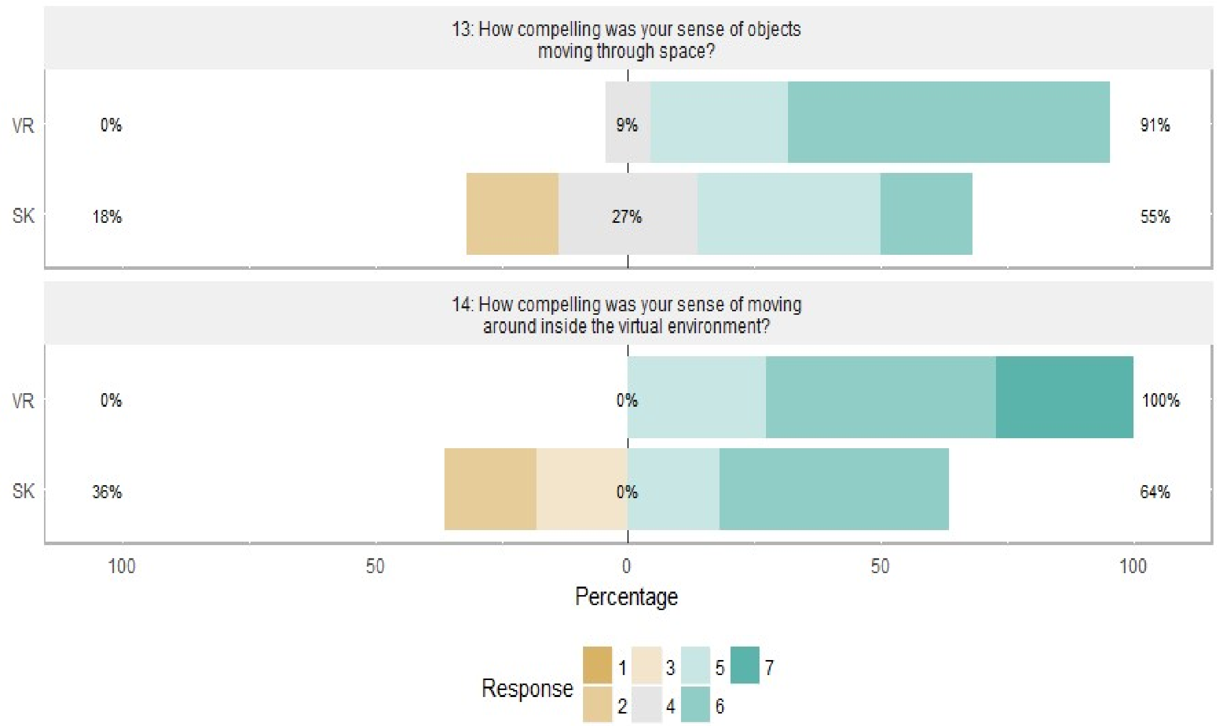

Figure 13: Diverging Stacked Bar Chart for Likert Scale Q13, Q14

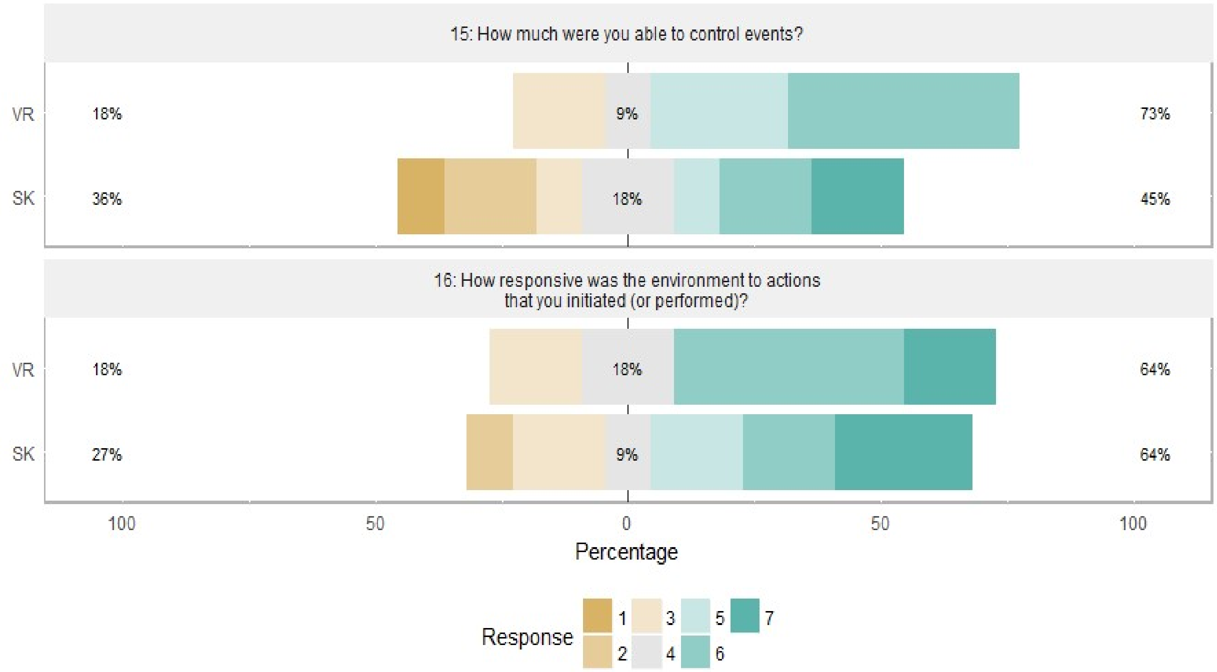

Figure 14: Diverging Stacked Bar Chart for Likert Scale Q15, Q16

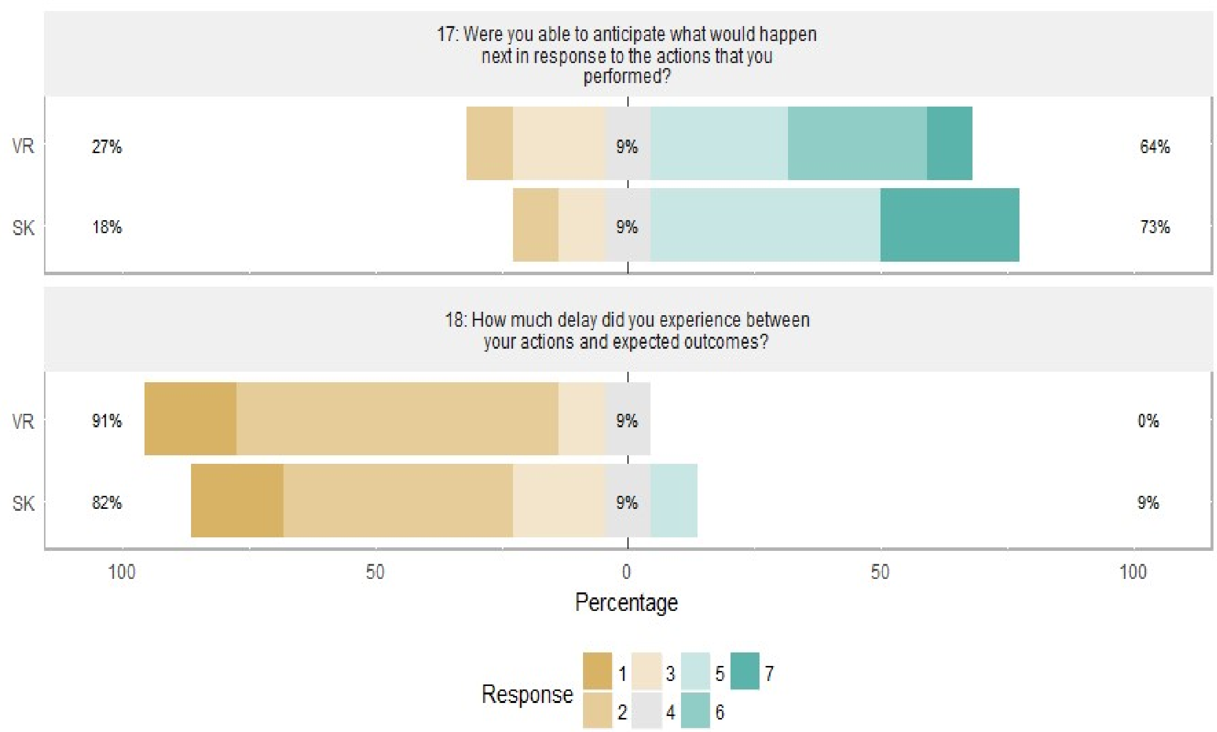

Figure 15: Diverging Stacked Bar Chart for Likert Scale Q17, Q18

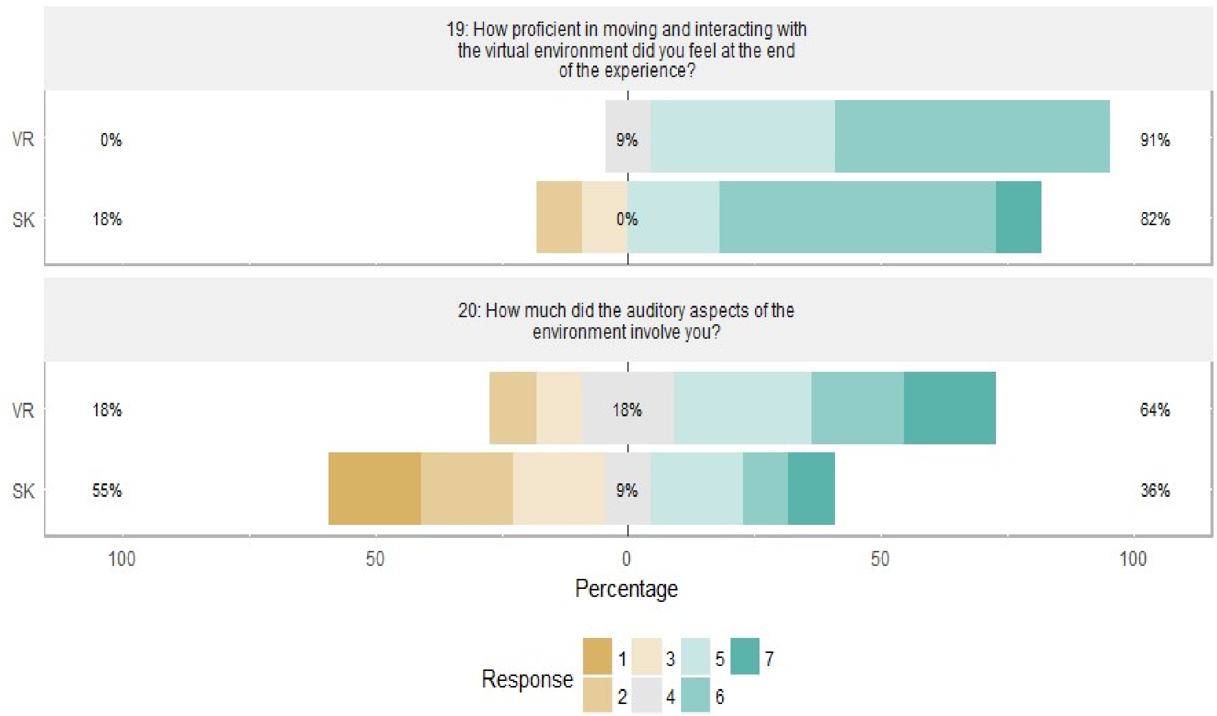

Figure 16: Diverging Stacked Bar Chart for Likert Scale Q19, Q20

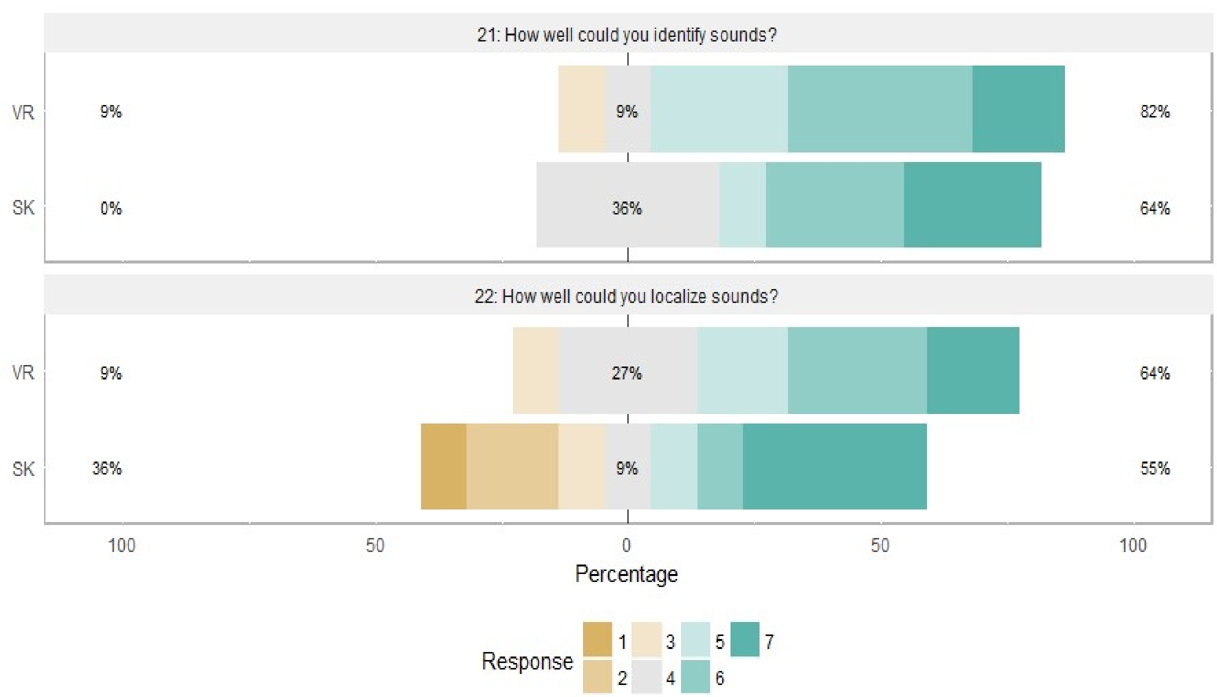

Figure 17: Diverging Stacked Bar Chart for Likert Scale Q21, Q22

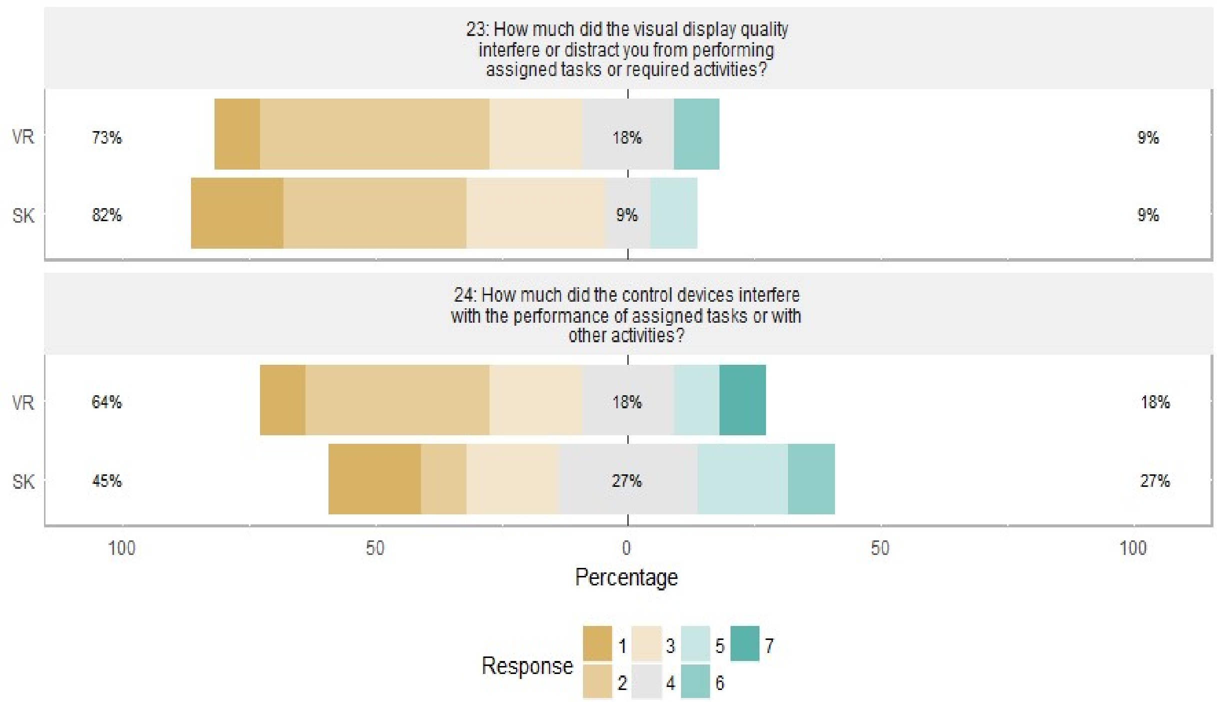

Figure 18: Diverging Stacked Bar Chart for Likert Scale Q23, Q24

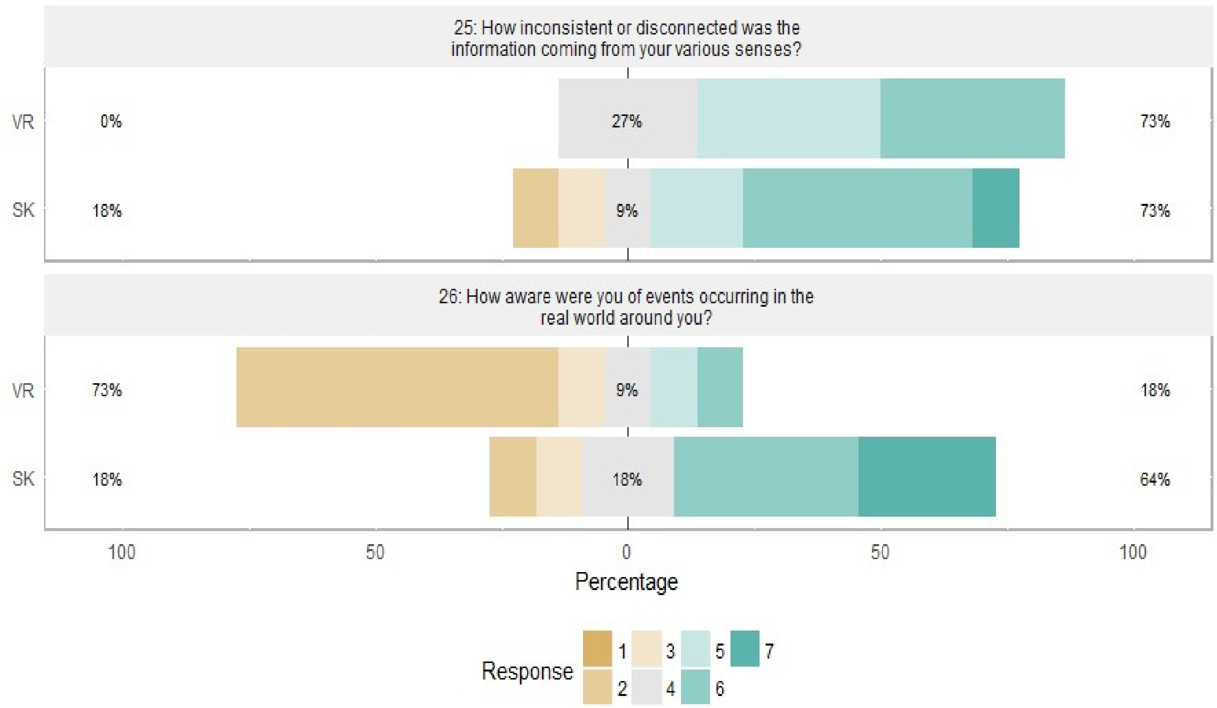

Figure 19: Diverging Stacked Bar Chart for Likert Scale Q25, Q26

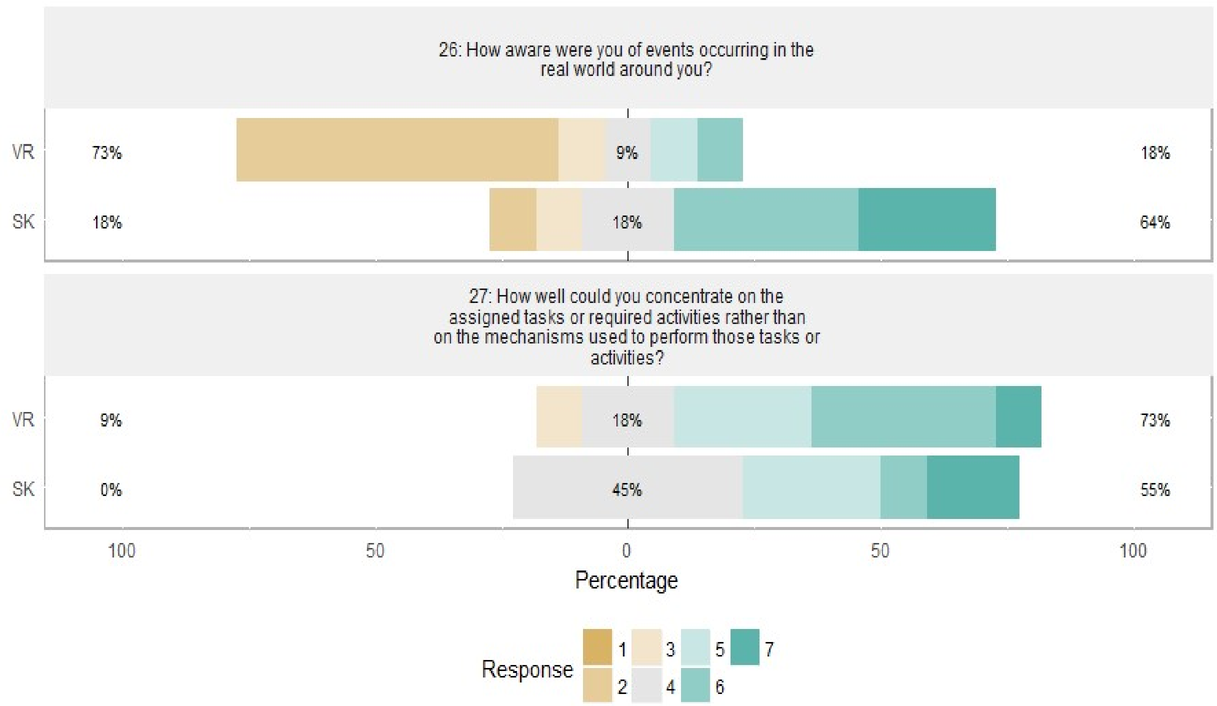

Figure 20: Diverging Stacked Bar Chart for Likert Scale Q26, Q27

VITA

The author Dharmesh Desai was born in Nasik, Maharashtra, India. He obtained his Bachelor's degree in electronics and telecommunication from Pune University in 2011. He joined the University of New Orleans computer science graduate program to pursue an MS degree with a concentration in Artificial Intelligence. This research work was done under the supervision of Dr Stephen Ware in 2017.