Evaluation of Selection Techniques on a Mobile Augmented Reality Game

Leonardo P. Soares, Soraia R. Musse, Márcio S. Pinho
School of Technology - PUCRS
Porto Alegre, Brasil
E-mail: leonardo.pavanatto@acad.pucrs.br,
{soraia.musse, marcio.pinho}@pucrs.br

Jean B. Boussu
École nationale d'ingénieurs de Saint-Étienne
Saint-Étienne, France
E-mail: jean.bertin-boussu@enise.fr

***Abstract*—Selection is one of the most fundamental tasks performed by humans. In Virtual Reality (VR), a number of interaction techniques have been proposed to deal with this type of task. They can be based on different metaphors, that define how the interaction is performed. However, it is still not clear how these different metaphors perform under Augmented Reality (AR). We performed two experiments to evaluate two of those interaction techniques. One was based on a pointing metaphor, and the other on a surface interaction metaphor. To achieve this result, a game prototype was developed, where users had to apply the selection techniques to reach a goal and win. We evaluated the impact of two different types of devices and two types of crowd simulation (with and without flocking behavior) in the user interaction. The scenario allowed the user to easily navigate around the scene from an overview perspective. We performed a two-way ANOVA between the variables and found that the technique based on surface interaction metaphor allowed the users to select multiple objects in the shortest amount of time. However, the technique based on pointing metaphor had a higher accuracy. Moreover, both techniques performed better with a crowd with flocking behavior.**

***Keywords*-3D interaction techniques; crowds behavior; augmented reality; game**

I. Introduction

Augmented Reality (AR) consists of an approach that uses displays, tracking and other technologies to enhance the user's view of a real-world environment with synthetic objects or information [1]. According to Schmalstieg and Höllerer [2], AR holds the promise of creating direct, automatic, and actionable links between the physical world and electronic information. While VR aims to use displays, tracking and other technologies to immerse the user in a synthetic spatial environment [1], AR uses related technologies and concepts to present information that is directly registered over the physical world. The main challenge, then, arises from integrating real and virtual worlds while allowing the user to interact with both of them.

According to Bowman [3], interaction tasks performed by users in 3D environments can be divided into three universal categories: **navigation**, responsible for the movement of the user's viewpoint; **selection and manipulation**, the actions of choosing an object (selection) and then applying translations or rotations (manipulation); and **system control**, responsible for applying changes to the state of the system.

Usually, low-level interaction techniques are developed to solve each of the universal tasks. They are defined as a combination of an input device (hardware) and a user interface (software) [1]. When dealing with complex tasks, such as grabbing an object (selection) and moving it (manipulation) to another place in the environment (navigation), multiple low-level interaction techniques can be composed together to form a high-level interaction technique [1].

Selection and manipulation are one of the most fundamental tasks performed by humans: if an object cannot be manipulated effectively, many specific tasks might just not be performed [1]. These techniques can be split into two categories: the ones that try to act resembling the real-world, related to an **isomorphic view**; or the ones based on "magic", related to a **non-isomorphic view**. The first is used to build a faithful, more natural, representation of the physical world, while the last is mapped and tailored specifically to 3D environments [1], [3].

Mine [4] argues that there are two types of selection techniques: local and at-a-distance. Local techniques allow the user to grab an object that is within its reach, while at-a-distance techniques allow the user to select an object that can't be reached. LaViola et. al [1] suggests a metaphor-based taxonomy with six categories for 3D manipulation: grasping, similar to the local techniques; pointing, similar to the at-a-distance techniques; interacting with a surface, mainly used for touchscreen displays, where dragging or rotating are performed once the user touches the screen and moves his/her finger; indirectly manipulating objects, where the user does not interact directly with the object (no need to navigate to the object's position); bi-manual interactions, where both hands are used to perform the manipulation; and combining metaphors to create hybrid techniques.

We implemented and evaluated two techniques based on different metaphors for the selection task in a mobile AR game. Our main motivation was understanding how users feel about each of the metaphors, and measure aspects of the performance they can obtain. The developed game required users to apply a selection technique successively on individuals of a virtual crowd to win. The scenario was positioned in the real-world by using multiple AR markers aligned in a grid shape over a table. This allowed the user to see his/her surroundings as well as the game environment and interact to the virtual game by moving the camera device around it in an intuitive way.

We conducted two user studies to understand the effects of device size and type of behavior of the virtual crowd on the performance of the selection techniques. We considered two types of crowds: one is self-organized into flock structures as proposed by Reynolds in [5] in his seminal Boids behavioral model, while the other crowd is a set of virtual humans that share the same goal but do not behave in a group.

The main contribution of this paper is the performed evaluation of interaction techniques in a game containing crowds. The organization of the paper is as follows: In Section II we present some works present in the literature, while in Section III we detail the decisions and implementation we propose. Section IV describes the first study, where a comparison between two devices with different sizes is performed. Section V describes the second study, where a comparison between two types of crowd behavior is performed. In addition, Sections VI and VII address some considerations about this work and future improvements.

II. Related Work

A classical approach [3] on how to handle both selection and manipulation is the ray-casting technique: a virtual ray emanates from the user's hand and can select and manipulate objects by pointing to them. It is considered an at-a-distance selection and manipulation technique because the virtual ray can grab objects at any distance. Therefore, it is a fitting technique for mobile devices, where the position and orientation of the device can represent the user's hand, and therefore, the starting point of the ray.

Marzo et al. [6] compared three manipulation techniques for mobile devices: one employing multi-touch interaction, the other using the device position (captured by accelerometers) and the third a combination of both. Their study revolved around 12 participants using an iPod to perform a docking task, where a chair needs to be aligned to a specific position. Their experiment shows that the use of the hybrid method of device positioning and multi-touch yielded the best results. Nonetheless, the use of only device positioning was shown to be interesting, as it appeared to be more intuitive, although it had the worst results overall.

Two novel interaction techniques for mobile devices were proposed by Mossel et al. [7] in 2013. In the first technique, called 3DTouch, the user moves and rotates the objects along the two axes of the plane the device is placed in. In the second technique, named HOMER-S, the user selects an object in the scene and is able to move and rotate it in any direction by performing the desired movement or rotation with the mobile device. This technique provides 6 degrees of freedom (6DOF) for the user's interaction. Their findings show that both methods provide intuitive manipulation with similar performance for traditional positioning or rotating tasks, while in compound tasks of positioning and rotating HOMER-S shown better results. Finally, 3DTouch is better when scaling is involved.

Although many different methods in crowd simulation have been proposed in last years [8] and [9], in this work we decided to only confront the user interaction levels with flocked or non-flocked crowds, as proposed by Reynolds [5]. Indeed, we want to evaluate if the difference in crowd organization, behind the interaction, will change the results.

Reynolds [5] Boid's model allows the simulation of flocks behavior in 3D spaces. The model considers flocks as the aggregate result of the actions of individual birds. Therefore, this approach simulates the behavior of each bird individually. Boid's model is based on the Particle System model [10], where each particle is independent, but are able to represent "fuzzy" objects, with irregular and complex shapes. Boid's model used this concept, but applied the following behavior rules to each bird: **collision avoidance**: steer to avoid crowding local flockmates; **velocity matching**: steer towards the average heading of local flockmates; and **flock centering**: steer to move toward the average position of local flockmates (centroid).

By combining all the rules together, the model is capable of generating a single desired acceleration for each bird, as an emergent phenomenon. This combination is a weighted average of each behavior rule. Using a migratory urge metaphor, the model can also place a global target in the environment (can be a direction or a position) for the birds to go. A herd's model can be derived from the 3D model. To do this, there must be a 2D surface limitation, and an ability for the agent to follow the terrain, which has been implemented in our work using pathfinding over a *NavMesh*, a network of connected 3D planes which represents the navigable area. The *NavMesh* used was the one native of *Unity3D* platform.

III. The Model

This section presents the model developed in this paper. We describe the proposed selection techniques, the game prototype and the AR environment in next sections.

A. Game Prototype

The developed game prototype consists of a post-apocalyptic world, where a group of humans is being attacked by herds of zombies. The scene is composed of a forest, full of zombies, and a refugee camp, full of humans. The game, named *Zeus vs Zombies*, is based on a god view metaphor, where the user can see an overview of the world as if he/she were in the top of *Mount Olympus*. This means that the user is outside of the world, which is also at a smaller scale.

The objective of the game is to eliminate all the zombies in the world in the shortest amount of time possible while reducing the number of human casualties. Zombies are eliminated when they are hit by a Zeus' lightning, which is a selection performed by the user. Humans cannot be selected by the user. Once a human is attacked, he immediately becomes a zombie. If all humans turn into zombies, the game ends and the user loses it. In cases where the number of lightning bolts is limited, the user can also lose the game if their availability gets to zero before all zombies are eliminated.

Figure 1. Game view. *Task was to eliminate zombies (red), and protect humans (green).*

B. Selection Techniques

This work explores two different approaches of selection techniques: a target-based technique, where a target is used to aim at an object and then select it; or a finger-based technique, where the user directly touches over the object that he/she wants to select on a touchscreen with his/her finger.

Using LaViola et al. [1] classification, the target-based technique uses a pointing metaphor, where the position and orientation of the device define what is being pointed. The finger-based technique, on the other hand, uses a surface interaction metaphor since the user can select any object that appears on the screen by simply touching over it.

1) Target-based: This technique places a target in the center of the screen. A ray-casting technique is performed in the forward direction of the target (perpendicular to the screen). As the user moves the mobile device around the AR scene, the target points to different objects present on the scene. If the object being pointed to can be selected, a visual feedback is given to the other: the target's color is modified to show that the object can be selected. At this point, the user can touch anywhere on the screen, and the targeted object is selected.

Our studies considered two variations of this technique: in the first study, the user was allowed to hold his/her finger on the screen and move the device around to select multiple characters faster. On the second study, this was not allowed, because the accuracy of the techniques was also tested, by finding the hits per misses ratio. Therefore, if multiple selection was enabled, the number of misses would be considerably higher.

2) Finger-based: This technique does not place a target on the screen. Instead, a ray-casting technique is performed in the forward direction of every point where a touch happens on screen. The user basically taps the object that he/she wants to select. Therefore, the user doesn't need to translate or rotate the device around the scene to select an object, although he/she can. However, there isn't any feedback of objects that can be selected, or if the finger is placed in the correct position. The user only knows if it worked after the selection has already been performed, and some action was triggered by it.

Our studies considered two variations of this technique: in the first study, multi-touch is allowed along with gestures along the screen to select characters faster. On the second study, multi-touch and gestures along the screen were not allowed, because the accuracy of the techniques was also tested, by finding the hits per misses ratio.

C. Crowd Behavior

Herds of zombies were simulated using Reynold's Boids behavioral model [5] with a few adjustments (Figure 2). The original model was developed for 3D movements over the environment, as in a flock of birds and a school of fish. Simulating zombies, however, require the movement of the boid to be constrained to a 2D plane. To achieve that, only 2D information regarding the plane were considered (x and z-axes), with height (y-axis) being disregarded.

Moreover, there are some peculiar actions taken by zombies: they are usually attracted by sound, and they only attack humans when they are very close, given limited capabilities. To implement the concept of sound attraction, the migratory urge feature proposed on the original model was used. This feature allows us to specify a direction or point in space where we want these boids to go, by incrementally changing the boids acceleration towards the specified target. It was based on the concept of bird migrations, where birds move to other places based on climate or specific seasons of the year.

Zombies try to go to places where usually there is noise, such as the center of the camp. Therefore, this model places a migration point in the center of the human's camp. The acceleration towards this point is enforced until the zombie gets close to the center of the point ($d_c = 5$ *meters*). To induce them to attack the humans, a modification of the migratory urge feature was implemented, where the zombies will be strongly attracted towards humans once they are inside a certain radius of distance ($d_z = 3.5$ *meters*). The virtual humans' motion is randomly defined. In addition, the agents are endowed with awareness of zombies presence that comes too close to them ($d_h = 3.5$ *meters*).

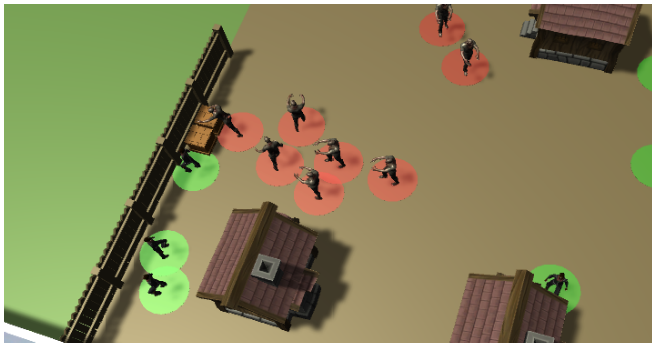

Figure 2. Gameplay. *A herd of zombies is moving towards the humans. One of the zombies (close to the box) is trying to attack the human but is colliding with the box.*

In addition to Reynolds Boids model [5], we also considered the non-flock behavior, i.e. agents do not try to group with others. While in flock behavior the agents, which are part of the same group, share the same goal, in non-flock behavior, agents are individually affected by the rules and have individual goals. The same behaviors and distances applied to flocked agents are used in non-flocked ones, in order to produce comparable results.

D. AR Environment

The game was presented in an AR environment, where a mobile device was used to move around the scene from an overview (God's view) perspective, and to output the AR to the user (Figure 3). From this viewpoint, the user could eliminate a zombie by using the selection techniques being evaluated. Multiple markers were used to set the environment position in the real world, allowing the user to move closer to the environment with at least one of the markers visible. If a setup with only one marker had been used, the marker could be occluded from the camera as the user moves around the environment. The markers page was printed in A2 size, allowing for a space large enough for users to explore, and move around.

IV. First Study: Effects of Device Size on Technique Performance

An experiment through a user study was performed to analyze and compare aspects of the techniques. Given the similarities of the techniques, our aim was to understand how they perform given their specific differences in visual feedback, the need for device manipulation, and the use of a touchscreen. Moreover, the relation between the techniques and the screen size of the device was also investigated.

A. Independent Variables and Hypotheses

This experiment evaluated two independent variables: the interaction technique (TB - Target-Based, FB - Finger-Based), and the device size (smartphone, tablet) used for each trial. Since by combining all variables we obtain 4 different trials, we decided on a within-subjects experiment, where each participant played the game 4 times. The ordering of each trial was counterbalanced between the participants, resulting in 16 different orders. A between-subjects ordering effect was not analyzed due to the reduced size of the sample.

Figure 3. User playing. *User points to the markers and the environment is rendered*

Our main hypotheses were:

- the finger-based technique would perform better than the target-based technique when used on large devices, such as a tablet, since it doesn't require much manipulation of the device;
- the target-based technique would perform better than the finger-based technique when used on small devices, such as a smartphone, since a small device makes it harder to select objects on the screen.

B. Apparatus and Software

This experiment used two devices: a smartphone; and a tablet. The chosen smartphone was an iPhone 6©. It has the dimensions of 138.1 x 67.0 x 6.9 mm (h x w x d), and weights 129 grams. A 4.7-inch screen, with a resolution of 1334-by-750-pixel resolution (326 ppi). It uses an A8 chip with 64-bit architecture and an M8 motion coprocessor. It has an 8-megapixel camera, with f/2.2 aperture. The chosen tablet was an iPad Mini 2©. It has the dimensions of 200 x 134.7 x 7.5 mm (h x w x d), and weights 331 grams. A 7.9-inch screen, with a resolution of 2048-by-1536 resolution (326 ppi). It uses an A7 chip with 64-bit architecture and an M7 motion coprocessor. It has a 5-megapixel camera, with f/2.4 aperture.

The software prototype was developed using the Unity Engine, version 5.5.5p1, a cross-platform game engine that is extensively used in the field of VR due to the flexibility and ease of use. The ARToolkit 5.3.2 framework was also used to perform the tracking of the AR markers and render the game scenario over them. A multi-marker configuration was used, allowing the system to still track correctly even if some marks are not visible or occluded.

C. Task and Dependent Variables

The task consisted of solving the game presented in the previous section. The participants had to rapidly eliminate all the zombies while minimizing the collateral damage (humans turned into zombies). In this study, the number of times the user could try to select a zombie was infinite for both techniques. Moreover, the target-based technique allowed for holding the touch and moving the device around to select multiple targets, while the finger-based technique allowed for multiple touches on different points of the screen at the same time. There was no penalization mechanism for the user for missing a selection.

The following dependent variables were evaluated:

- **time** to finish the game: all zombies were eliminated, or zombies eliminated all humans;
- and a **score**: based on the user performance in the game. Every human that turns into a zombie results in a reduction of 10 points in the score. Every zombie eliminated by the user results in an addition of 10 points in the score. However, if the user eliminated more than a single zombie within a 1-second window, it would add a double amount of points (20).

D. Participants

16 participants (15 males and 1 female) from the university undergraduate and graduate population were recruited on a voluntary basis for our study. 10 participants were undergraduate students, 4 were graduate students 2 were professionals. The participants ranged in age from 19 to 44 years (median age 27 years). We excluded participants under the age of 18, due to legal matters.

E. Procedure

Participants had to sign a consent form, which briefly explained the study and guaranteed their rights during the experiment. These rights included anonymity of the generated data, access to the results of the research, possibility of stopping the test at any time for any reason, and the right to disagree with any of the previous terms. The consent form also authorized the collection of data from the trial and publishing of it. After that, they filled a background questionnaire, with questions regarding the characterization of participants (age, gender, and education).

They were given an explanation about the study. It was explained that they would test two selection techniques in two different devices while playing a game. The game was presented, explaining the task they would have to perform. They were explicitly told to eliminate the zombies in the minimum amount of time possible while reducing the number of human casualties.

Each participant performed 4 trials: TB with the smartphone, TB with the tablet, FB with the smartphone and FB with the tablet. The ordering was counterbalanced, leading to 16 different orders. This within-subjects design and its ordering were defined to minimize any potential learning factor. Since the techniques were simple, no training time was allowed. They would start the first trial with only the information about the game and a brief description of the techniques.

After the tests were completed, each participant completed a post-experiment questionnaire with specific questions about their experience using the system for each of the devices. They were asked to rate their perception of the techniques for easy to use, results, comfort using a 7-point Likert Scale. They were also asked to choose their preferred technique for that device.

F. Results

The experiment was evaluated by two methods: obtaining the metrics from each trial's simulation, comprised of the score made by the user and the time it took his/her to finish the game; and a questionnaire, with characterization of participants (age, gender, education), and their perception about the techniques.

A two-way ANOVA was performed between the trials being analyzed for each of the metrics. First, the main effects on the score metric were analyzed. The comparison between the techniques showed significant differences for score ($F_{1,30} = 1000.50, p < .001$). The comparison between devices showed no significant differences ($p = .964$). No interaction effect was observed between both variables ($p = 0.690$). A comparison of score results can be seen in Figure 4, where the error bars indicate standard deviation.

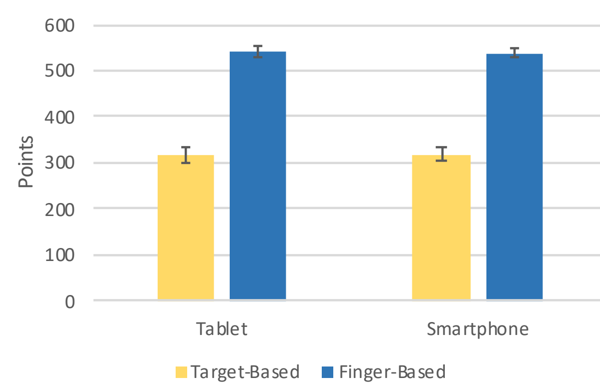

Figure 4. Score comparison between techniques and devices.

The main effects on the time metric were also analyzed. The comparison between the techniques showed no significant differences for time ($p = .074$). The comparison between devices showed no significant differences ($p = .696$).

No interaction effect was observed between both variables ($p = 0.698$).

A post-experiment questionnaire was applied to verify the participants' perception of the techniques on each device. Using the smartphone, 56.25% reported preferring the target-based technique. Using the tablet, 56.25% of the participants reported preferring the finger-based interaction technique. Only 25% of the participants preferred different methods of interaction for each device. From the ones who always preferred the same, 50% always preferred the finger-based technique, and the other 50% preferred the target-based technique. A comparison of results can be seen in Figure 5.

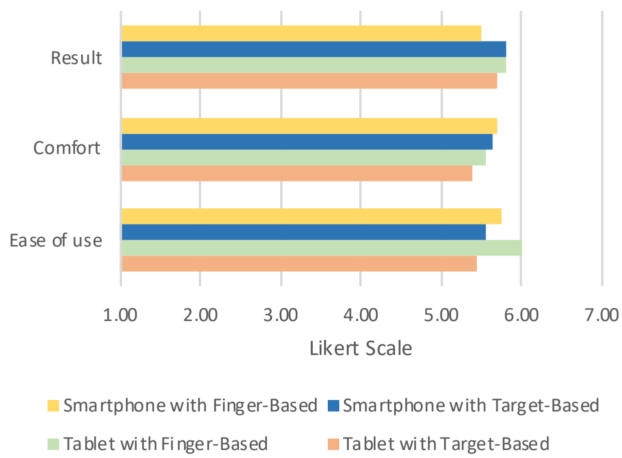

Figure 5. Qualitative score attributed by the participants to the different combinations of platform and techniques.

V. Second Study: Effects of Crowd Behavior on Technique Performance

As the dependent variable time in the first study has not shown a statistically significant difference, we decided to perform a second study to further investigate the differences between the techniques.

We speculate that the game prototype was too easy to win, therefore not allowing the differences to surface. To achieve such objective, two different types of zombies behavior were compared against the same previous techniques: a flocked behavior, which follows the boids model present in the first study, and a non-flocked behavior, where each zombie would try to reach its own goal, individually.

A. Independent Variables and Hypotheses

This experiment also evaluated two independent variables: the interaction techniques (TB - Target-Based, FB - Finger-Based), and the type of crowd (flocked or non-flocked). Since by combining all variables we obtain 4 different trials, we decided on a within-subjects experiment, where each subject played the game 4 times. The ordering of each trial was counterbalanced between-subjects, resulting in 24 different orders. A between-subjects ordering effect was not analyzed due to the reduced size of the sample.

Our main hypotheses on this study were:

- the target-based technique performs equal or better than the finger-based technique when a crowd is flocked since it requires less physical handling of the device by the user;
- the finger-based technique performs better than the target-based technique when a crowd is non-flocked since it is faster to touch on another point of the screen than moving the device;

B. Apparatus and Software

This experiment used only the iPhone 6©smartphone device described in the previous study.

The software prototype was slightly modified from the first study. During the course of the investigation, the framework used for tracking fiducial AR markers had to be replaced. The Vuforia 7 framework was used instead. The same markers were kept, although differences in tracking capabilities between frameworks might be present. As a result, the project was also ported to Unity Engine version 2017.3.1p4, which has native support of the framework.

C. Task and Dependent Variables

The same task was used. However, this time, a few new constraints were introduced. In this study, the number of times the user could try to select a zombie was finite for both techniques. Holding touch and multitouch was disabled. Because of this changes, a penalization mechanism for the user for missing a selection could be created. The participant starts the game with 30 selections available. Every time he/she tries to select, this number is decreased. The number of hits and misses is also counted, allowing for new accuracy metrics. Also, a new possibility of game end was added. If the number of selection available reached 0, the participant would lose the game.

The following dependent variables were evaluated:

- **time** to finish the game: all zombies were eliminated, or zombies eliminated all humans;
- **score**: based on the user performance in the game. Every human that turns into a zombie results in a reduction of 10 points in the score. Every zombie eliminated by the user results in an addition of 10 points in the score. However, if the user eliminated more than a single zombie within a 1-second window, it would add a double amount of points (20).
- **humans left**: number of humans that survived when the game was completed.
- **zombies left**: number of zombies that kept existing when the game was completed.
- and **hits/misses ratio**: a ratio between hits and misses. Misses were incremented by one to avoid division by zero, in cases where the participant did not miss.

D. Participants

48 participants (39 males and 9 female) from the university undergraduate and graduate population were recruited on a voluntary basis for our study. 23 participants were undergraduate students, 11 were graduate students 14 were professionals. The participants ranged in age from 18 to 58 years (median age 24 years). We excluded participants under the age of 18, due to legal matters. 20% reported to play video-games every day, 12% reported to play often, 13% reported to sometimes, and 3% reported to never play.

E. Procedure

A procedure similar to the first study was performed, but considering the differences in the independent variables. Each participant performed four trials: TB with flocked behavior, FB with flocked behavior, TB with non-flocked behavior and FB with non-flocked behavior.

In this study, after each trial, the participant completed a questionnaire with specific questions about that trial, which included a 7-point Likert scale on affirmations regarding the techniques and the behavior of the zombies. The participant was allowed to write a paragraph about his/her perception and feedback on the trial.

After all trials were completed, each participant completed a post-experiment questionnaire with specific questions about their experience with the techniques and zombies behaviors. They chose their preferred technique, if they preferred different techniques for different crowd behaviors, and if the crowd behavior changed the way they played the game.

F. Results

The experiment was evaluated by two methods: obtaining the metrics from each trial's simulation, comprised of the score made by the user, the time it took to the user to finish the game, the number of humans and zombies that still exist at the end of the game, and the number of hits and misses; and a questionnaire, with characterization of participants (age, gender, education, frequency that plays video-game), and their perception about the techniques and crowd behavior.

A two-way ANOVA was performed between the trials being analyzed for each of the metrics. First, the effects on the score metric were analyzed. The comparison between the techniques showed significant differences for score ($F_{1,188} = 7.53, p = 0.006$). The comparison between types of crowd behavior showed significant differences ($F_{1,188} = 3.86, p = 0.050$). No interaction effect was observed between both variables ($p = 0.994$). A comparison of the results can be seen in Figure 6, where the error bars indicate standard deviation.

The effects on the time taken to finish the game were also analyzed. The comparison between the techniques showed significant differences for time ($F_{1,188} = 12.53, p < 0.001$).

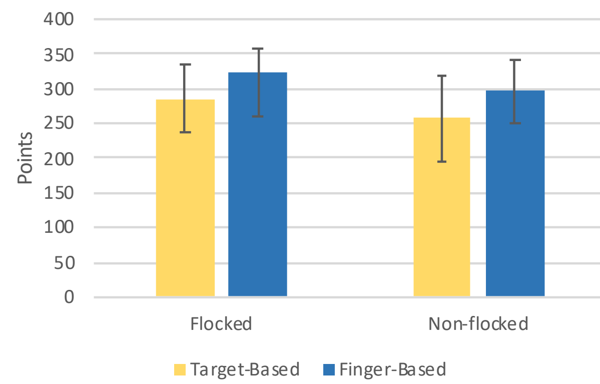

Figure 6. Score comparison between techniques and devices.

The comparison between types of crowd behavior showed no significant differences ($p = .587$). No interaction effect was observed between both variables ($p = 0.476$). A comparison of the results can be seen in Figure 7, where the error bars indicate standard deviation.

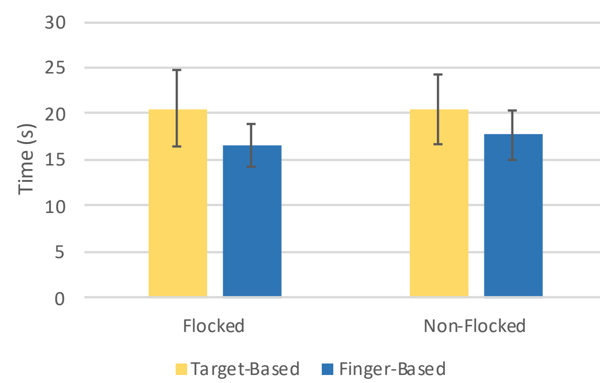

Figure 7. Time comparison between techniques and devices.

The effects on the number of humans that survived were also analyzed. The comparison between the techniques showed significant differences for number of survivors ($F_{1,188} = 8.72, p = 0.003$). The comparison between types of crowd behavior showed significant differences ($F_{1,188} = 5.49, p = 0.020$). No interaction effect was observed between both variables ($p = 0.608$). A comparison of the results can be seen in Figure 8, where the error bars indicate standard deviation.

The effects on the number of zombies that were not eliminated were also analyzed. The comparison between the techniques showed no significant differences for the number of survivors ($p = 0.065$). The comparison between types of crowd behavior showed significant differences ($F_{1,188} = 5.03, p = 0.025$). No interaction effect was observed between both variables ($p = 0.813$).

The effects on the hits per misses ratio were also analyzed.

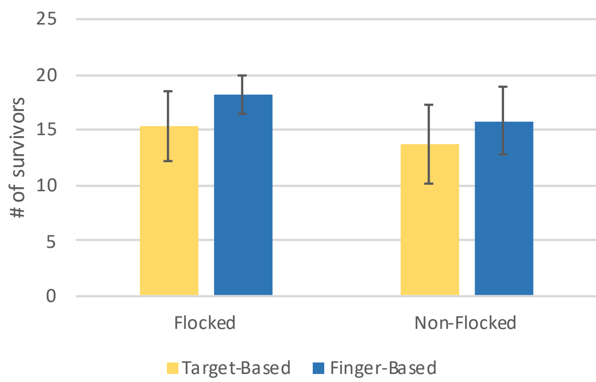

Figure 8. Number of humans who survived comparison between techniques and devices.

The comparison between the techniques showed significant differences for hits per misses ratio ($F_{1,188} = 23.19, p < 0.001$). The comparison between types of crowd behavior showed no significant differences ($p = 0.393$). No interaction effect was observed between both variables ($p = 0.147$). A comparison of the results can be seen in Figure 9, where the error bars indicate standard deviation.

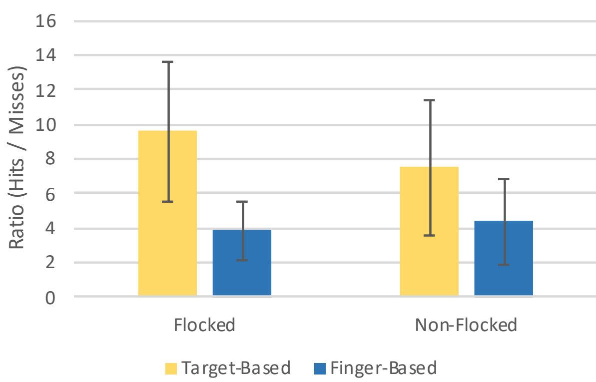

Figure 9. Hits/Misses comparison between techniques and devices.

A questionnaire was applied to verify the participants' perception of the techniques and crowd behaviors after every trial. According to the questionnaire, participants felt that the finger-based technique prevented them from selecting the correct character more than the target-based and made it harder to visualize the game, but it was fast to perform the selection. Regarding ease of use, the finger-based technique was considered better, while they also reported that the target-based technique also required more experience with video-games, and required more effort. Both techniques performed similarly regarding limitations, comfort and being a good solution to the problem. No influence of the type of crowd on technique was observed on this questionnaire. Figure 10 shows the results for each variable.

Figure 10. Average of 7-point likert-scale on sentences regarding the technique.

Regarding the crowd behavior, the perception that they were working in group, and were intelligent change between the types of crowd, with the flocked behavior performing higher. Regarding the speed of the zombies, the user's reported a higher score while using the target-based technique than using the finger-based.

After all trials were completed, a final questionnaire was answered by the participants. 56% of the participants considered the non-flocked behavior harder. 64.5% of the participants always preferred the target-based technique, while 22.9% preferred the finger-based technique. 2.5% preferred both at different moments. One participant argued that the finger-based was better for non-flocked and the target-based for flocked behavior.

VI. Discussion

The first study results found that the used devices had not influenced the performance of each technique. As we compare results from each technique between the devices, they had similar results. Therefore, our hypotheses for the first experiment, which both linked the techniques to a

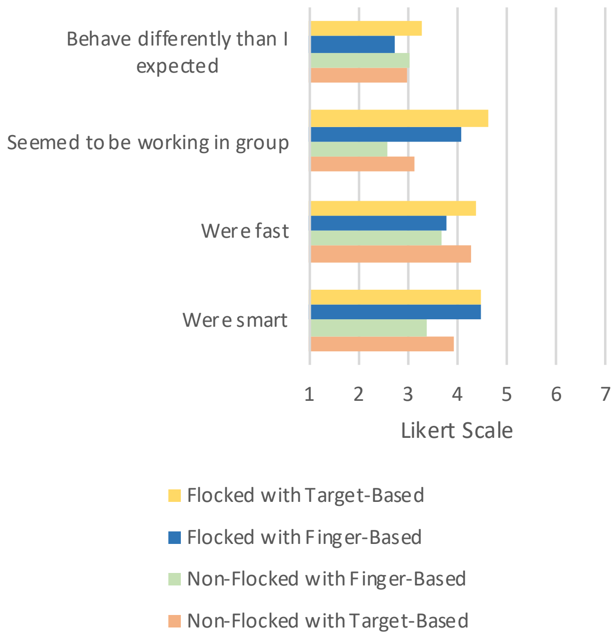

Figure 11. Average of 7-point likert-scale on sentences regarding the zombies behavior.

category of device being used at first were not true. The finger-based technique always performed better in terms of score.

In both studies, the finger-based technique achieved higher scores than the target-based technique. The distribution of points in the game is based on a human turning into a zombie (-10), a zombie being eliminated (+10), and a zombie being eliminated within a 1-second window of another zombie being eliminated (+20). Therefore, the fact that the technique performed better suggests that it can select a higher number of objects over an interval of time. A possible explanation for this is that it is faster to touch on different points of the screen than to move the device to align a target.

In the first study this did not imply in a smaller time to complete the task, suggesting that, although it is easier to select multiple objects faster, the user still needed to perform some movements on the device to locate zombies. In the second study, however, the time to complete the task was smaller for finger-based than target-based technique. A possible explanation is that, in the second study, the "multiple selections" and "hold to keep selecting" options were disabled, leading to a more difficult game, and therefore making the differences clearer.

In the second study, the number of humans who survived showed significant differences between both techniques and both types of crowd. The finger-based technique yields an end of the game with more survivors in both types of crowds. This can indicate that the finger-based was more efficient in preventing attacks, implying that the user could click on the screen quicker. Still regarding the number of survivors, for both techniques, the flocked behavior also yield to more survivors. This indicates that it is quicker to select objects that are close together.

In addition, the number of zombies who were not destroyed at the end of the game was smaller for the finger-based technique. This agrees with the idea that this technique was more effective in preventing attacks, and that the user could click on the screen quicker.

The hits/misses results help to understand the accuracy difference between the techniques. The target-based technique performed better in both types of crowds, which could mean that the visual feedback provided by the target gave user the option to only try to select after knowing it was in the correct position. In the finger-based technique, the user could only learn if it was a hit after he tried to select.

Regarding the second experiment, we hypothesized that the two types of crowd behavior would influence the performance of the techniques individually. We found out, however, that both techniques perform better when the crowd is flocked, but keeping a similar improvement between them. With the target-based technique performing better than the finger-based regarding the ratio of hits/misses, and the finger-based technique performing better for score and time.

The qualitative data on both studies show that the preference of the user's interaction for each method is well distributed, with most users preferring the same interaction method regardless of the device they are using or crowd type. However, it should be noted that most users preferred using the finger-based technique with the tablet and the target-based technique with the smartphone, suggesting that, although no performance correlation was found, these techniques can be more suitable to a given device based on the user's perception.

The qualitative data from the second study show some interesting insights into the techniques and crowd behavior. On technique, there is a clear distinction between them, with the target-based being regarded as more novel, but requiring more effort and experience. The finger-based, however, seems to allow a fast selection, but a lower accuracy, with the user not always being able to select correctly.

VII. Final Remarks and Future Work

This paper presented an evaluation of two selection techniques for interaction in AR environments, each based on a different metaphor. A game was developed in order to test the performance of the techniques. This game implemented a modified version of the Boids' model [5], with a herd of zombies being simulated. The techniques were also compared between two different types of devices, to find if there was any relation between the techniques and the devices.

Two experiments were performed. In the first study, participants tested both techniques with two different devices, having their order counterbalanced to minimize any learning effect. The effects of two independent variables were evaluated within-subject: the interaction technique (TB - Target-Based, FB - Finger-Based), and the device (smartphone, tablet). The dependent variables were score and time to complete the task of eliminating zombies by selecting them. In the second study, they tested both techniques with two different crowd behaviors. The effects of two independent variables were evaluated within-subject: the interaction technique (TB - Target-Based, FB - Finger-Based), and the crowd behavior (flocked, non-flocked). The dependent variables were the score, time, surviving humans and and hits per misses.

Overall, results show that the finger-based technique yields a better performance, while the target-based yield a better accuracy. Regarding devices, the users preferred the finger-based technique with the tablet, possibly because of the bigger size of the screen, and target-based technique with the smartphone, perhaps because the smallest screen could make it harder to select the zombies with a finger, while it should be easier to manipulate around the environment, but the dependent variables did not support any distinction between devices. Regarding crowds, both techniques performed better when the flock model was being used. However, none of the techniques performed better because of a specific crowd model, the finger-based always performed better in terms of time and score, and the target-based always performed better in terms of hits per misses ratio.

In future works, a broader research may be made in order to draw deeper relations between the variables. The two compared devices were somewhat similar, and differences could surface when using devices with more expressive differences. Regarding crowds, it would be interesting to experiment with different sizes of crowds, as a way of managing the difficulty of the task. As the game get's harder, it could modify the performance and accuracy of the techniques. Experiment-wise, each user performed each of the 4 trials once. A longer experiment, with users repeating the same trials multiple times, and in different levels of difficulty could be an important factor.

VIII. Acknowledgments

Our research is partially funded by (a) the National Institute of Science and Technology in Medicine Assisted by Scientific Computing (Grant CNPq 181813/2010-6) and (b) the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior – Brasil (CAPES) – Finance Code 001.

References

[1] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev, *3D user interfaces: Theory and practice*. Addison-Wesley Professional, 2017.

[2] D. Schmalstieg and T. Hollerer, *Augmented reality: principles and practice*. Addison-Wesley Professional, 2016.

[3] D. A. Bowman, E. Kruijff, J. Joseph J. LaViola, and I. Poupyrev, "An introduction to 3-d user interface design," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 1, pp. 96–108, 2001.

[4] M. R. Mine, "Virtual environment interaction techniques," *UNC Chapel Hill CS Dept*, 1995.

[5] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 25–34, Aug. 1987. [Online]. Available: http://doi.acm.org/10.1145/37402.37406

[6] A. Marzo, B. Bossavit, and M. Hachet, "Combining multi-touch input and device movement for 3d manipulations in mobile augmented reality environments," in *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction*, ser. SUI '14. New York, NY, USA: ACM, 2014, pp. 13–16. [Online]. Available: http://doi.acm.org/10.1145/2659766.2659775

[7] A. Mossel, B. Venditti, and H. Kaufmann, "3dtouch and homer-s: Intuitive manipulation techniques for one-handed handheld augmented reality," in *Proceedings of the Virtual Reality International Conference: Laval Virtual*, ser. VRIC '13. New York, NY, USA: ACM, 2013, pp. 12:1–12:10. [Online]. Available: http://doi.acm.org/10.1145/2466816.2466829

[8] D. Thalmann and S. R. Musse, *Crowd Simulation, Second Edition.* Springer, 2013.

[9] M. K. N. I. B. Nuria Pelechano, Jan M. Allbeck, *Simulating Heterogeneous Crowds with Interactive Behaviors*. CRC Press, 2017.

[10] W. T. Reeves, "Particle systems—a technique for modeling a class of fuzzy objects," *ACM Trans. Graph.*, vol. 2, no. 2, pp. 91–108, Apr. 1983. [Online]. Available: http://doi.acm.org/10.1145/357318.357320