

Article

# “Blurry Touch Finger”: Touch-Based Interaction for Mobile Virtual Reality with Clip-on Lenses

Youngwon Ryan Kim <sup>1</sup>, Suhan Park <sup>2</sup> and Gerard J. Kim <sup>2,\*</sup><sup>1</sup> Korea Electronics Technology Institute, Gwangju 61011, Korea; kyw920@gmail.com<sup>2</sup> Digital Experience Laboratory, Korea University, Seoul 02841, Korea; parksuhan@korea.ac.kr

\* Correspondence: gjkim@korea.ac.kr

Received: 15 October 2020; Accepted: 5 November 2020; Published: 8 November 2020



**Abstract:** In this paper, we propose and explore a touch screen based interaction technique, called the “Blurry Touch Finger” for EasyVR, a mobile VR platform with non-isolating flip-on glasses that allows the fingers accessible to the screen. We demonstrate that, with the proposed technique, the user is able to accurately select virtual objects, seen under the lenses, directly with the fingers even though they are blurred and physically block the target object. This is possible owing to the binocular rivalry that renders the fingertips semi-transparent. We carried out a first stage basic evaluation assessing the object selection performance and general usability of Blurry Touch Finger. The study has revealed that, for objects with the screen space sizes greater than about 0.5 cm, the selection performance and usability of the Blurry Touch Finger, as applied in the EasyVR configuration, was comparable to or higher than those with both the conventional head-directed and hand/controller based ray-casting selection methods. However, for smaller sized objects, much below the size of the fingertip, the touch based selection was both less performing and usable due to the usual fat finger problem and difficulty in stereoscopic focus.

**Keywords:** virtual reality; object selection; pointing; ray casting; task performance; usability

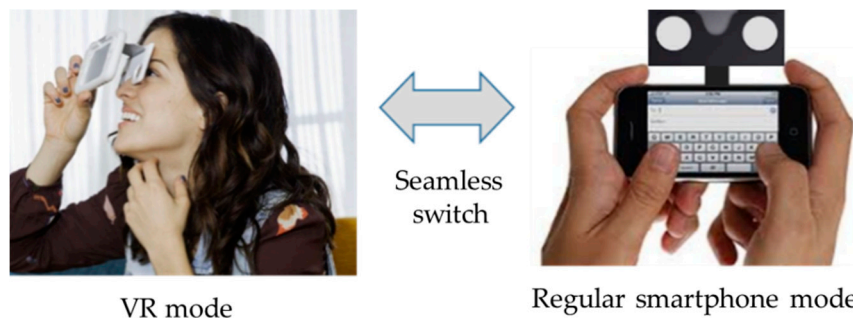
## 1. Introduction

Mobile virtual reality (M-VR) loosely refers to the immersive content platforms that are based on the smartphone [1]. M-VR in its typical form uses a head-set, into which the smartphone is inserted to lend its display (see Figure 1, left). The head-set itself only provides the display optics, simple interaction device (e.g., button or touch pad) and world isolation [2,3]. M-VR is often contrasted with the higher-end (and priced) PC based VR (PC-VR) in which the display integrated head-set is separated from and tethered to a PC [4,5]. M-VR is generally regarded, aside from being mobile and self-contained, more convenient, inexpensive, and targeted for casual use. Nevertheless, M-VR in such a typical form is still not sufficient for a wide adoption by the public mass; the head-set is still somewhat bulky, interaction methods awkward and the ever-needed and must-be-handly smartphone becomes inaccessible during when used for VR (after the insertion).

Recently, a new breed of M-VR, as called “EasyVR,” has appeared in the market in the form of flipping or clipping a cheap non-isolating magnifying lenses on the smartphone (see Figure 1, right). Surprisingly, despite being open and the user’s peripheral view not shut from the outside world, such a display configuration still gives a good level of immersion with the magnified view covering much of the human’s field of view [6]. Given its much-improved convenience and the advantage of quick setting, it serves as a viable alternative configuration for M-VR (there is even a fold-out version built-into the smartphone case [7]). The open configuration also makes the immediate switch to the regular smartphone mode quite possible and vice versa (e.g., switch between VR and texting, see Figure 2).



**Figure 1.** Typical mobile virtual reality with the smartphone inserted and closed into an inexpensive headset (left), and “EasyVR” in which light magnifying lenses are clipped on the smartphone in a non-isolating way.

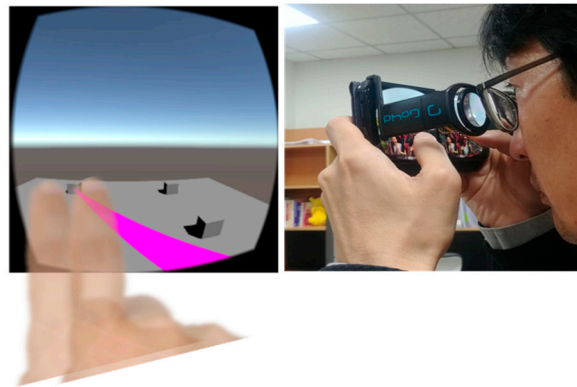


**Figure 2.** Main features of “EasyVR”: (1) quick put-on, (2) no inconvenient mounting on the head, (3) seamless switch between regular smartphone and VR modes, (4) touch screen based interaction possible (and no additional interaction devices—self-contained).

One of the most important elements that makes “EasyVR” easy is perhaps the possibility of using the touch screen for interaction. Conventional M-VR set-ups have offered the head-direction (or gaze) and touch pad (or button) on the side of the head-set for the lack of any better method. The use of the touch screen in EasyVR brings about: (1) the usual and familiar touch based user experience (including the swipe and flick), (2) natural bimanual interaction, and thereby, (3) more direct and efficient interaction, and (4) no need for an additional interaction device.

Even though the open set-up allows the fingers to touch and manipulate objects on the touch screen, they are seen through the magnifying lens. While the virtual objects will be seen in focus by design, the fingers are not and seen blurry by the binocular rivalry. Interestingly, this very feature makes the target object visible and thereby the touch screen interaction feasible. This also partly solves the so called “fat finger” problem in touch based smartphone interaction (see Section 3 for more details) [8].

In this paper, we explore this idea of the touch screen based interaction for EasyVR, called the “Blurry Touch Finger (BTF)” (see Figure 3). We carried out the first stage basic usability evaluation to validate the proposed idea: whether the EasyVR user can in fact select virtual objects accurately using BTF, compared to e.g., the conventional interaction techniques. The evaluation through the performance of object selection, which is one of the most important and basic interactive VR task [9], would provide a good yardstick for the practicality and feasibility of EasyVR and the associated “Blurry Touch Finger” method.



**Figure 3.** Directly selecting a virtual object with the finger on the touch screen. The touching finger is blurred, but because of it, the target object behind it is visible, and reasonably accurate selection is possible. Both two/one hand/finger interactions are possible.

In the following sections, we first review other related research work and describe the details of the BTF method. Then we present the usability experiment and its results. The paper is concluded with a brief summary, discussion and directions for future work.

## 2. Related Work

Object selection in the 3D virtual space has been studied extensively [10,11]. While a variety of methods exist, virtual ray casting [9] is one of the most common and popular approaches, in which a 3D (virtual) ray emanates from the hand (whose position is tracked by a sensor), and the user “shoots” for the object of interest. This method is quite intuitive in the sense that it is a natural abstraction of the real-world finger pointing. A slight variant is the virtual flashlight in which a volumetric cone is formed with the apex at the hand toward the target object [12]. Several different forms of ray casting methods exist depending on how the origin and direction of the ray is determined (and thus where the “cursor” is situated, e.g., using the hand position/orientation (hand directed), hand orientation and head position (head-hand directed), head position/orientation (head directed) [13].

To apply the virtual ray casting or flashlight techniques, the user would need to either employ a hand tracking sensing device that would require a cumbersome process of calibrating its spatial coordinates with that of the view point, or recognize and track the hand from the head mounted camera or sensor which would restrict the hand position to stay within the sensing volume thereby making the ray/cone casting difficult. In the context of EasyVR, the use of hand-worn sensor or controller (a separate module), would be rather prohibitive and go against its very spirit and purpose (i.e., self-containment).

Thus, another possibility is to directly sense (from the self-contained EasyVR unit) the hand movement (and its gestures) using computer vision techniques or employing environment capture or depth sensors from the self-contained VR unit. Small depth cameras (mounted on the head mounted display or on body) have been used to employ 3D hand gesture based interaction [14–21].

While hand based pointing of remote objects by the ray casting is natural and familiar, it can be difficult due to the lack of sufficient proprioceptive feedback [22–26] and incur significant fatigue on the wrist by the offset between the target object/cursor and the hand [27].

On the other hand, current M-VR invariably employs the head-directed (gaze-based) cursor to select an object with the time-based or side-button driven final confirmation. This is primarily because the use of a separate interaction controller is discouraged at first to make M-VR as self-contained as possible. Head-directed target designation is problematic because the tasks of previewing and exploring the world and particular target object selection are coupled. For instance, view-fixed menu system cannot be used if the head-directed cursor is used. Head-directed (or gaze based) interaction is even more problematic with navigation control (although this paper does not deal with it yet) because

the user cannot freely view the world as the moving direction is determined by the gaze direction [28]. Using the head (vs. hand) is more tiring as well [28]. Time-based confirmation is slow, and the use of the side button is less direct. Companies have realized the limitation and have recently introduced separate controllers [29].

In this work, we explore the touch based method, because (1) hand tracking with a single camera is still challenging and prone to errors and latency, (2) depth sensors are still not a common feature on the smartphones yet and (3) the touch based method allows the use of two grabbing hands.

Pierce et al. suggested the “Sticky Finger” as one of their virtual image plane interaction techniques similar to the Blurry Touch Finger, for selecting and manipulating an object in virtual environments [27]. With the Sticky Finger, the user interacts with the 2D projections that 3D objects in the scene make on one’s image plane. Thus, in a monoscopic environment, the method essentially a 2D interaction technique, as the selection is made on the image plane, as also categorized as such by [30]. When using a virtual ray, the “cursor” position (drawn by projecting the user’s fingertip position orthographically) is the point on the image plane collided with the ray [31]. Blurry Touch Finger works similarly except that the selection is made over the “real” image screen space (i.e., touch screen), under the magnifying glass with stereoscopy. Even though the target object is far, it seems as if attached to the finger, and leverages well on the already familiar touch interaction with the smartphone.

There have been few proposals to use the back of the smartphone as the interaction space for M-VR [32,33]. Setting aside the need to install another touch sensor on the back of the M-VR unit, the method of selecting from the “back” is not intuitive especially when the selection is attempted among crowded objects with an occluded view (as seen from the front).

### 3. Blurry Touch Finger

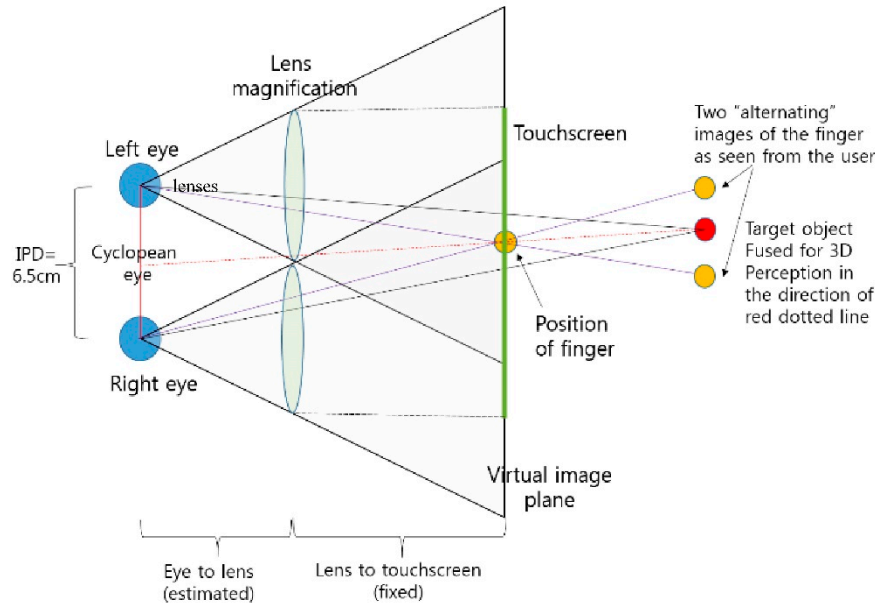
Figure 3 illustrates the concept of Blurry Touch Finger (BTF). First, we assume an “open” EasyVR set-up with a magnifying glass clipped on the smartphone so that the hands/fingers are used to both hold the unit and touch the screen (see Figure 3). Even though the open set-up allows the fingers to touch and manipulate objects on the touch screen, they are seen through the magnifying lens. By design, while the virtual objects will be seen in focus, fingers will be blurred and seen in double vision. Nevertheless, the user is still able to point to and touch the target object near-correctly, similarly to the case of point shooting [34].

Moreover, the target object is visible as if the blurred fingers are transparent due to the binocular rivalry. Binocular rivalry refers to phenomenon of visual perception in which perception alternates between different images presented to each eye [35]. That is, the finger is seen by only one eye (due to the vignette) and the alternative views (one containing the finger and the other not) make the finger effectively semi-transparent, making the imagery behind it completely visible. A similar concept has been applied for real environment object selection in an augmented reality setting [11]. BTF partly solves the so called “fat finger” problem which touch based interaction suffers from for selecting small objects due to the occlusion by the touching finger [8].

Both one-handed and two-handed interactions are possible. For the former, one hand will be used solely for holding the device, while for the latter, both hands are holding the device with fingers extended out to touch the left or right screen. Note that when both hands/fingers are used, the left or right finger is used one at a time (not simultaneously) to select different objects

Figure 4 illustrates the workings of the Blurry Touch Finger. The touch screen is magnified into a larger virtual imagery by the lenses, and the 3D target object (red) is seen in focus, fused for 3D depth perception, in the direction of the purple dotted line (emanating from the mid-point between the two eyes (Cyclopean eye) by a phenomenon called the Allelotropia [36]). Even though this seems to be the line direction with which the user consciously will try to point to the target object and touch the corresponding position on the touch screen, our observation has found that the user naturally makes the touch input on where the target object is rendered on the touchscreen (where the red dotted lines cross the touchscreen either on the left or right, the purple circles in Figure 4). Since the left and right images

are assumed to be fused and not visible separately, this ability seems to be very surprising. Currently, we do not have a clear physiological explanation to this observation. Nevertheless the experiment in this paper demonstrates that such is the case and highly accurate (in par with controller-based ray-casting method) selection is possible.



**Figure 4.** Working of the Blurry Touch Finger. The actual touch is made not along the perceived view direction of 3D target object from the Cyclopean eye, but in the direction from the respective eye depending on where the user touches (on the left or right screen space image).

Note that the red dotted lines are not exactly a straight line due to the refraction by the lenses (but treated straight for now for simplicity and the amount of refraction can be computed). That is, the user can touch either the right or left screen to select the fused target object (when an object only appears in either the right or left imagery, the user will make the touch input in the respective side only). Therefore, a virtual ray emanating from the respective eye (right eye if right half of touch screen is used and vice versa) is used to intersect the target object for selection.

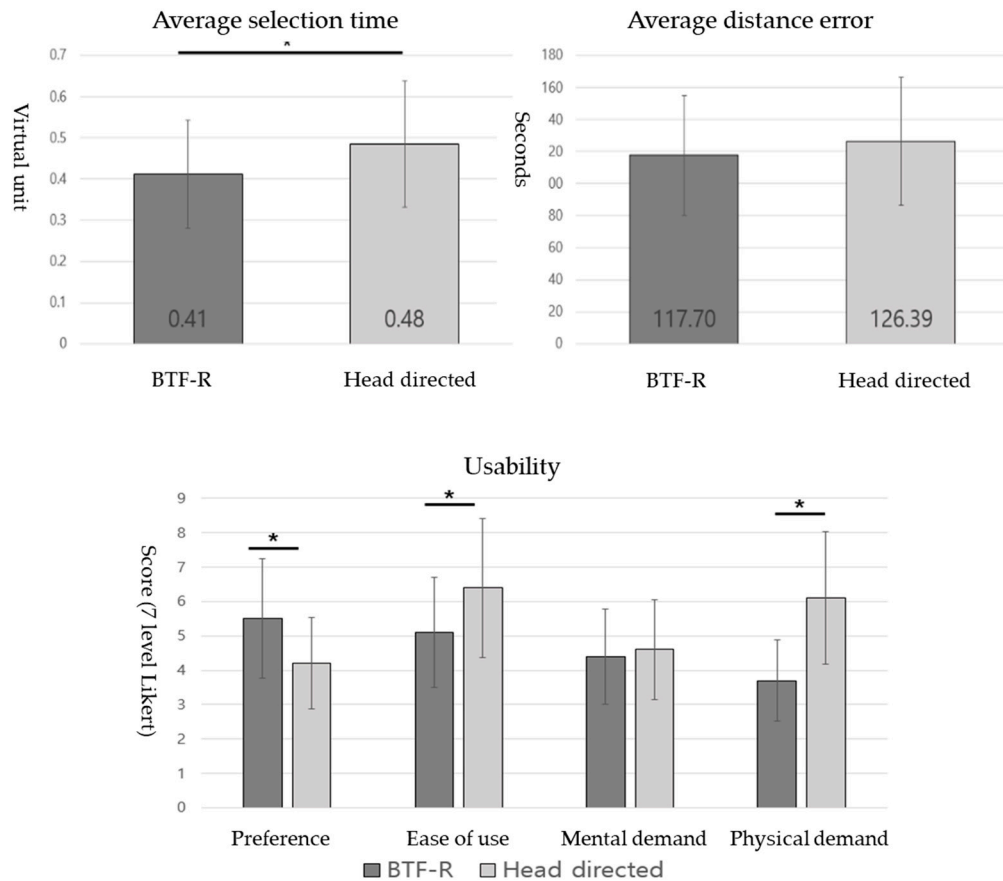
#### 4. Pilot Experiment: BTF-R vs. Head-Directed Selection

##### *Experiment Design*

Thus, we first ran a small pilot study, comparing one form of BTF, namely BTF-R to the head directed method. The BTF is in its bare form, making the object selection upon the first touch (BTF-T). The other variant BTF makes the selection at the release of the finger, after the user makes the initial touch and adjusts the touch finger position for a more accurate final selection (BTF-R).

We only shortly describe the pilot experiment and its results (almost the same experimental procedure was used in the main experiment as described in the following sections). The user was presented with a sequence of 12 objects in one nominal size (about  $0.4 \times 0.4$  cm—48 pixels in the screen space as recommended by the Android user interface guideline [37]) that appeared in a balanced order on the  $4 \times 3$  grid positions. At each appearance of the object, ten subjects were asked to select the object as fast and accurately as possible, using two methods, (1) BTF-R and (2) head-directed. To eliminate any possible biases, both methods were tested using an open configuration, that is, with the flip-on lenses and the final confirmation (for Head-directed) was done using a simple touch on the openly accessible touchscreen (vs. time-based or using a side button). The object selection time, error (the distance between the finger/cursor and target object), and general usability (through a survey) were measured as the dependent variables.

The results in Figure 5 show that the Head-directed method did incur higher error (with statistical significance) and task completion time. The Cohen’s d-value for the distance error (BTF-R vs. Head-directed) was approximately 1.26 (i.e., large effect) and for the selection time, the value was around 0.27 (small to moderate effect).



**Figure 5.** Selection performance (above) and usability (below) between BTF-R and head-directed methods. Note that higher score means more positive toward usability for the categories of preference and ease of use, and vice versa for mental and physical demand. The asterisk mark (\*) indicates statistically significant difference.

The usability survey gave mixed results: e.g., the Head-directed felt to be easier to use, yet BTF was preferred and generally had less mental and physical demand. Note that the pilot test was conducted not using the much more inconvenient closed head-set. Thus, at the least, considering the aforementioned general disadvantages of the Head-directed method (such as the coupling between general viewing and selection), we claim that BTF-R is a viable selection method for EasyVR, and worthy of comparison to the controller-based ray casting method.

### 5. Main Experiment: BTF-T, BTF-R and CONT

The purpose of the main experiment is to validate the idea of the Blurry Touch Finger at the rudimentary level first, particularly to the case of virtual ray casting with a hand-gripped controller, as this is arguably the most popular and stable selection method today (but on PC-VR), and thus serves as a good reference for the level of performance [38]. On the other hand, head-directed method is the most prevalent way of object selection on M-VR, even though we highlighted its well-known disadvantages (see Section 2).

### 5.1. Experimental Design

In this main experiment, we compared the user performances of two styles of the Blurry Touch Finger selection on the EasyVR set-up with regards to the state of the art, controller based ray casting method on the PC-VR as the conventional case (CONT). This would highlight, if any, the potential of BTF (and EasyVR) across different state of the art VR platforms.

The main experiment was designed with two factors (3 selection methods—BTF-T, BTF-R and CONT and 3 object sizes—Small, Medium and Large) within subject repeated measure. We hypothesized that BTF-R would show equal or better performance/usability as compared to CONT. BTF-T is expected to be insufficient for general usage due to its nature to rely solely on the first touch, thus producing significant selection errors due to personal differences (e.g., in one's dominant eye, stereoscopic fusion) and modelling error (e.g., assumed eye position). BTF-R should compensate for them by the additional drag-to-release.

Note that EasyVR with a separate controller was not tested because (1) EasyVR pursues self-containment as a convenient mobile VR platform without any extra devices required other than the smart phone itself (and attached foldable lenses), and (2) holding the smartphone in one hand and controller in the other would be very difficult and unrealistic.

### 5.2. Experimental Task

In the main experiment, the user was similarly presented with a sequence of 12 objects (all in one of three sizes) that appeared randomly on the  $4 \times 3$  grid positions. At each appearance of the object, the subject was asked to select the object as fast and accurately as possible (but with more emphasis on speed to ensure minimum level accuracy). Without such an instruction (e.g., just as fast as possible), we were concerned subjects would just disregard the accuracy aspect. A single selection task was finished upon the initial touch for BTF-T, upon release for BTF-R and upon the trigger button press by the index finger for the CONT. A sufficiently large fixed time-out value was given for the user to select all given objects (although no subjects failed).

Selection of distant 3D objects with three different sizes (screen space size: Small—0.4 cm, Medium—0.5 cm, and Large—0.6 cm) was tested, all situated at a fixed distance (7.3 virtual space unit away from the view point) at 12 different locations laid out in a  $4 \times 3$  grid.

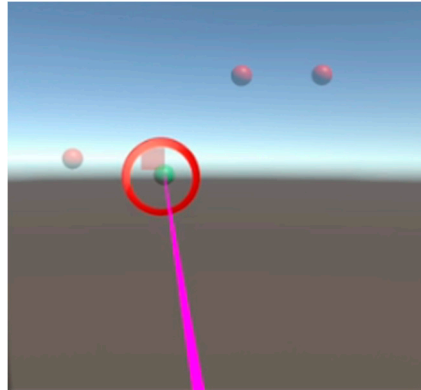
Object selection time (between target object appearance and confirmation events), accuracy (assessed in two ways: the exact planar distance on the screen space to the intended target and number of successful selections) and general usability were measured as the dependent variables. An unsuccessful selection means that the center position of the touch input was not contained within the target object.

### 5.3. Experimental Set-Up

In the experiment, for the two BTFs, we used the Samsung Galaxy S7 smartphone (152 g, screen size: 5.1 inches, combined screen resolution:  $1440 \times 2560$  pixels [39]). with the Homido-mini VR magnifying glasses (22 g, lens-to-screen distance: 70 mm [40]), giving it a virtual image of approximately 85 degree FOV. As for the CONT, the HTC Vive HMD (555 g, screen resolution:  $2160 \times 1200$  or  $1080 \times 1200$  per eye, FOV: 110 degree) and the two pose-tracking controllers were used (whose experimental content ran on the connected PC) [4].

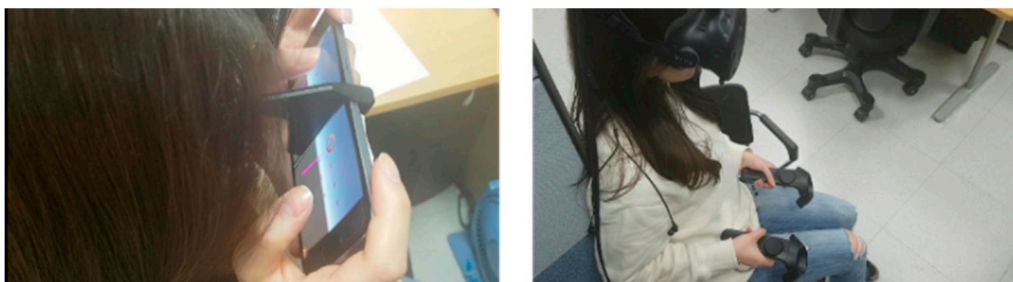
A virtual ray (emanating from the respective eye, e.g., the right eye when right half of the touch screen is touched and vice versa) and a circular cursor were drawn at the end of the ray in stereo. The virtual ray was provided so that the interface looked the same as in the popular controller based ray casting methods (e.g., Tilt Brush/HTC Vive [4]) and give a better understanding of the 3D environment. For this experiment, the eye dominance was asked of but not put into consideration in terms of adjusting the direction of the virtual ray. The cursor and the ray guides the user make the selection, but for the BTF-T, it is of not much a use other than for just providing a confirmation of the selection,

since the selection is made right upon the direct initial touch. Note that the ray/cursor only appears only upon touch in the cases of BTF's (e.g., only one ray/cursor when one finger is touched, and two when both fingers touch the screen), and two rays/cursors always appear in the case of CONT that uses two active controllers (see Figure 6).



**Figure 6.** The ray/cursors(s) appearance when using the BTF's and CONT as a guide for the selection. As for BTF-T, it appears only after the touch is made, and for BTF-R, it remains while the finger stays touched until released. The finger is not shown for clarity.

For the two BTFs, the smartphone based M-VR units were held by both hands and touch interacted with the two thumbs, while for the baseline case, two controllers in each held in the left and right hands were used (see Figure 7).



**Figure 7.** The experimental set-up: (1) for the BTF-T and BTF-R using a smartphone based Easy M-VR (left) and (2) CONT using the PC based VR with the HTC Vive HMD/controllers (right).

For now, the calibration between the real and virtual space, e.g., between the touch screen and virtual image plane (in the case of BTF's) was done by a simple linear scaling transformation. As for the 3D controllers (CONT), the calibration was not necessary since the system already operated with respect to the head-set (default view position).

#### 5.4. Experimental Procedure

Ten paid subjects (eight men and two women between the ages of 23 and 29, mean = 27.9/SD = 5.36) participated in the experiment. After collecting their basic background information, the subjects were briefed about the purpose of the experiment and given instructions for the experimental tasks. As the subjects were the same from the pilot experiment, they were already familiar with the BTF-R, thus an additional short training was given to allow the subjects to become familiar with the VR set-up, two other compared methods (BTF and controller based) and the experimental process. All subjects were right handed, had long experiences in using the smartphone some experiences with the M-VR or PC-VR. Throughout the practice sessions, the subjects mostly tried and adjusted the way they would touch and select the object for higher accuracy.



The subjects sat and held the M-VR unit (smartphone + clip-on glass) with two hands, or wore (strapped to the head) the HMD with controllers in each hand, and looked into the experimental content comfortably. The exact view position was adjusted individually so that the experimental content could be fused in stereo for the proper depth perception. The study included nine different treatments (or conditions), namely, three interfaces  $\times$  three object sizes; in each treatment condition, 12 different target objects laid in the  $4 \times 3$  grid positions were to be selected, comprising of one trial block, and block trials repeated 12 times (a total of  $3 \times 3 \times 12 \times 12 = 1296$  object selections).

Each treatment was presented in a balanced fashion using the Latin-square methodology per subject. After the whole sessions, the subjects were asked to fill out a usability questionnaire (shown in Table 1). We paraphrased the de facto standard usability survey questions from works like the NASA TLX [41] and SUS [42] to fit our purpose. It included standard categories such as ease of use, learnability, fatigue, and general satisfaction. The experiment took about one hour per subject.

**Table 1.** The general usability survey question.

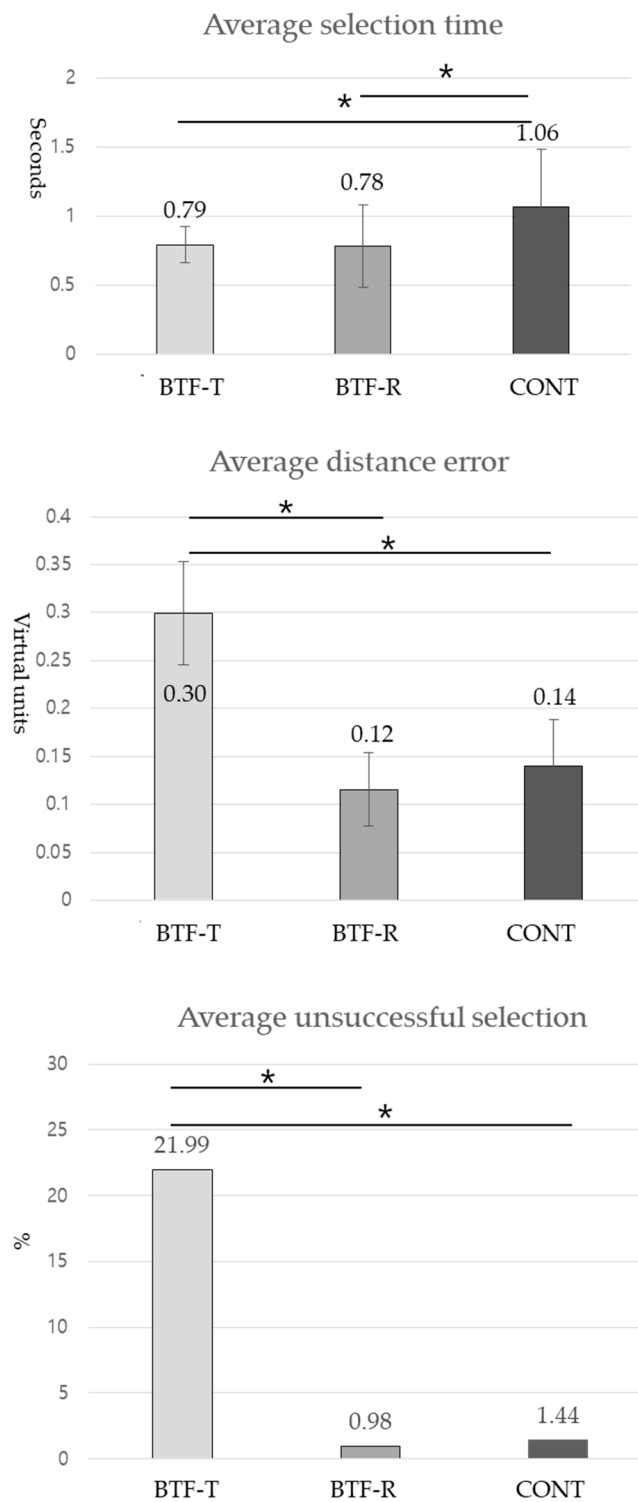
NO./CATEGORY	QUESTIONS
Q1: Ease of use	How easy you felt the interface to be: (1: Very difficult—7: Very easy)
Q2: Naturalness	How natural and intuitive you felt the interface to be: (1: Very contrived—7: Very natural)
Q3: Ease of learning	How easy it was for you to learn the interface: (1: Very difficult—7: Very easy)
Q4: General Satisfaction	Rate the overall level of satisfaction of using the interface: (1: Very unsatisfying—7: Very satisfying)
Q5: Fatigue	How fatigued you were after using the interface: (1: Very fatigued—7: Very fatigued at all)
Q6: Speed	Rate how fast you felt you could accomplish the task: (1: Very slowly—7: Very quickly)
Q7: Accuracy	Rate how accurate you felt you could accomplish the task: (1: Very inaccurately—7: Very accurately)

The standard two-way ANOVA was conducted that examined the effect of the two factors (1) interface type (BTF-F, BTF-R, CONT) and (2) object size (large, medium, small) on user performances (selection time, distance error, number of unsuccessful selection—numerical data). The Tukey’s Student Test was used for the pairwise comparisons. For the usability analysis, the Mann-Whitney U Test was used since the questionnaires were rather subjective in nature although scored in the 7 level Likert scale.

### 5.5. Results and Discussion

#### 5.5.1. Interface Type

The analysis showed a significant main effect for the interface type (Figure 8) to all three user performance categories (selection time:  $[F(2,81) = 8.417, p < 0.001^*]$ , distance error:  $[F(2,81) = 147, p < 0.001]$ , number of unsuccessful selections:  $[F(2,81) = 139.99, p < 0.001^*]$ ).



**Figure 8.** Quantitative performance measures (average selection time, average distance error and average number of unsuccessful selections in percentage) among the three interface types. The asterisk mark (\*) indicates statistically significant difference.

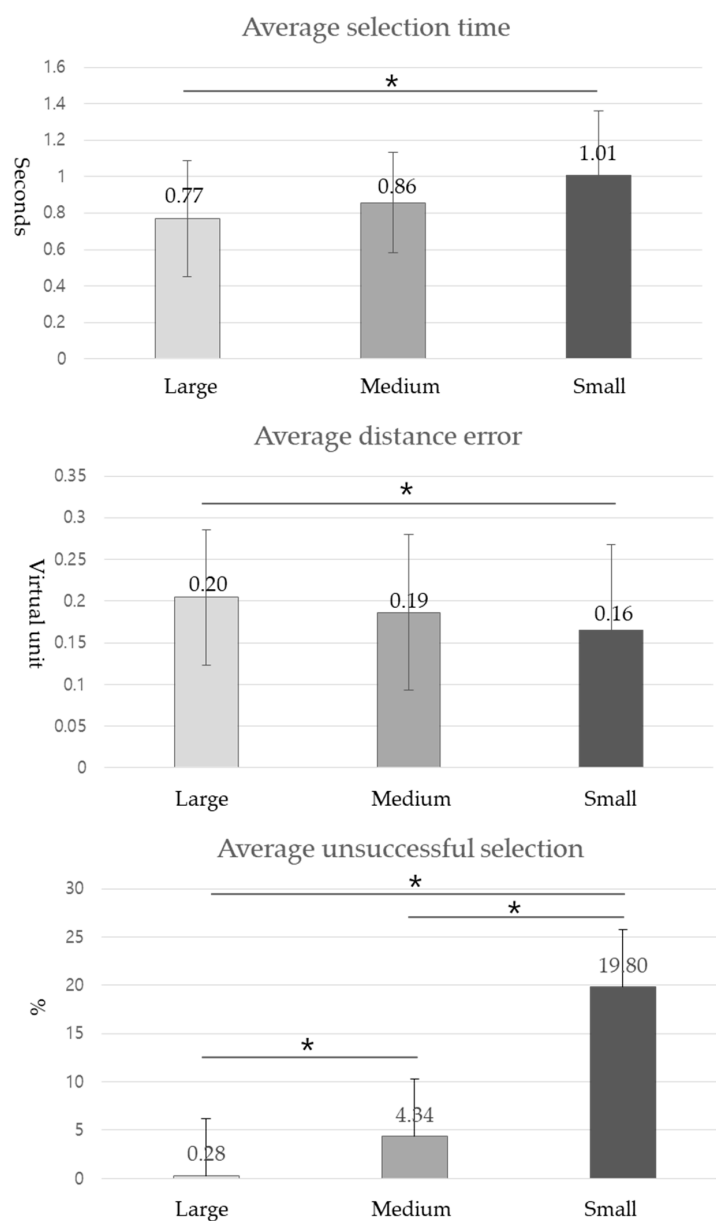
Using the post hoc Tukey’s Pairwise Comparisons, both BTF-T and BTF-R showed faster performance than the CONT due to its directness (BTF-T vs. BTF-R:  $p = 0.991$ ; BTF-T vs. CONT:  $p = 0.002$  \*; BTF-R vs. CONT:  $p = 0.001$  \*).

As for the distance error, BTF-R and CONT showed a similar level of accuracy with significant difference to the much lower that of BTF-T (BTF-T vs. BTF-R:  $p < 0.001$  \*; BTF-T vs. CONT:  $p < 0.001$  \*; BTF-R vs. CONT:  $p = 0.092$ ).

For the number of unsuccessful object selection, BTF-R showed the best performance among the three (but statistically no difference from CONT). Similarly with the distance error, BTF-T clearly showed an unusable level of object selection performance (BTF-T vs. BTF-R:  $p < 0.001$  \*; BTF-T vs. CONT:  $p < 0.001$  \*; BTF-R vs. CONT:  $p = 0.946$ ).

### 5.5.2. Object Size

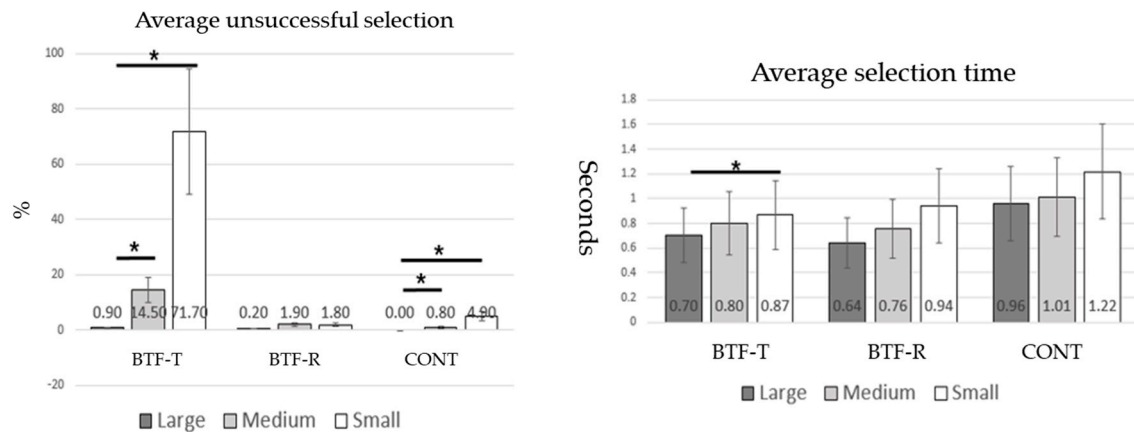
ANOVA showed a significant main effect for target object size (Figure 9) for all (selection time:  $[F(2,81) = 4.894, p = 0.010$  \*], distance error:  $[F(2,81) = 5.939, p = 0.004$  \*, number of unsuccessful selections:  $[F(2,81) = 103.10, p < 0.001$  \*]).



**Figure 9.** Quantitative performance measures (average selection time, average distance error and average number of unsuccessful selections in percentage) among the three target object sizes. The asterisk mark (\*) indicates statistically significant difference.

Using the post hoc Tukey’s Pairwise Comparisons, the selection were significantly faster with large targets (Large vs. Small:  $p = 0.008 *$ ). Yet, the error distance when selecting large targets were significantly higher than when selecting small targets (Large vs. Small:  $p = 0.003 *$ ).

For the rate of unsuccessful selection, all sizes had significant differences (Large vs. Medium:  $p = 0.016 *$ , Large vs. Small:  $p < 0.001 *$ , Medium vs. Small:  $p < 0.001 *$ ). However, for just BTF-R, there was no significant difference between the varied sizes (they existed only for BTF-T and CONT) (Figure 10). This is meaningful because there were no significant differences in the selection time between the BTF-R and the CONT condition. This indicates that if the target is small and the user needs to act precisely, BTF-R could be the way to go.



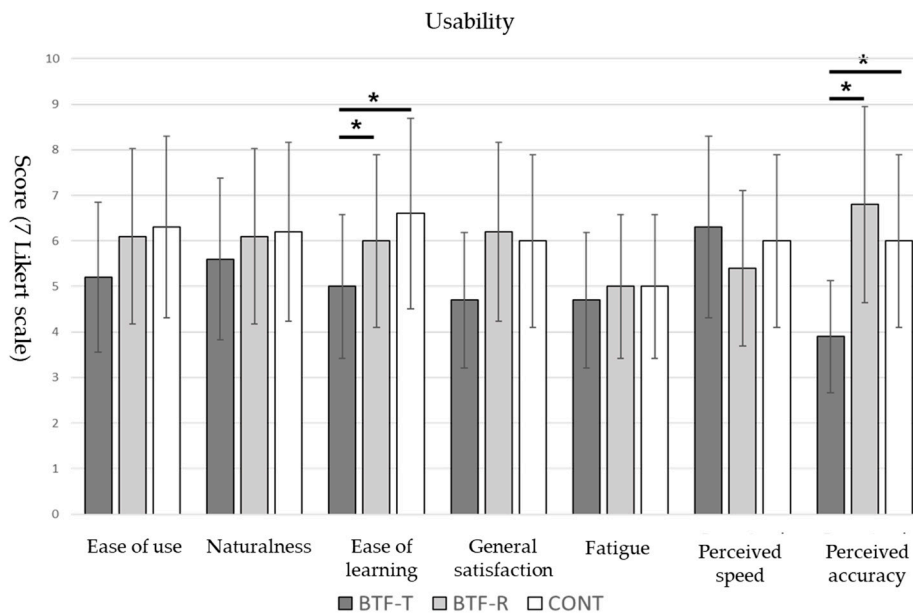
**Figure 10.** BTF-T is fast and accurate for large target objects. The asterisk mark (\*) indicates statistically significant difference.

Another interesting fact is that when using BTF-T, if the target objects were large, BTF-T was good enough to compete with the other two interfaces—equally accurate than BTF-R or CONT and faster than CONT but slightly slower than BTF-R (Figure 10). The Cohen’s d values were around 0.34 for BTF-T vs. BTF-R and 0.76 for BTF-T vs. CONT for the selection time. Considering the fact that the size of the large targets in the experiment matches the smallest size required from the Android Design Guidelines, any size that is bigger than the “large” size in this experiment can benefit from using BTF-T.

The results follow the usual speed-accuracy trade-off with respect to object size. The smaller the object is, more time was spent to select it, but with nearly equal accuracy for BTF-T (Figure 10) and better accuracy with BTF-R, by the visible target (and guiding cursor) and less fat finger problem.

### 5.5.3. Usability

The responses to the survey generally showed a similar level of usability among the three interface types (Figure 11). Statistically significant differences were found only on the categories of ease of learning (Mann-Whitney:  $p$ -value = 0.002 \*) and perceived accuracy (Mann-Whitney:  $p$ -value = 0.001 \*). BTF-T was regarded the least easy to learn while also clearly faster (although not with a statistical significance) and less accurate. The overall trend was that BTF-T was rated marginally less usable (with partial statistical support) than the other two, most probably due to the lack of method to correct the initial touch position.



**Figure 11.** Responses to the survey which generally showed a similar level of usability among the three interface types. BTF-T was the least easy to learn, and perceived to be fast but less accurate. The asterisk mark (\*) indicates statistically significant difference.

## 6. Supplement Experiment: Smaller Objects at Different Depths

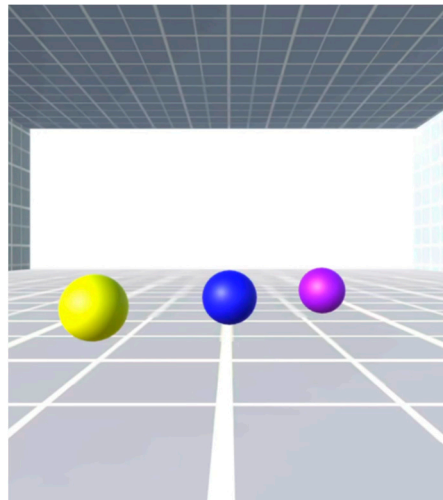
The pilot and main experiment were conducted with the user selecting (relatively large) objects of different sizes at a fixed depth. For further validation, a supplement experiment was performed for selecting even smaller objects presented at different depths. The smallest object size tested in the main experiment was about  $0.4 \text{ cm} \times 0.4 \text{ cm}$  which was about the smallest recommended object size for mobile application [37]. However, in VR applications (that often operate at different depths) such a guideline may not be meaningful.

### 6.1. Experimental Procedure

The experimental procedure was mostly same as the main experiment except that that equal sized target objects were placed at three depths such that the closest appeared with the screen space size of about 0.4 cm, second closest, 0.3 cm and farthest 0.2. These sizes were not only significantly below the size of an average fingertip ( $1.5 \text{ cm} \times 1.5 \text{ cm}$  for the index and  $2 \text{ cm} \times 2 \text{ cm}$  for the thumb), but also smaller than the object sizes tested in the main experiment. Only BTF-R (who exhibited the best performance in the main experiment) and CONT were compared with regards to the selection performance. Thus, the supplement experiment was designed with two factors (2 selection methods—BTF-R and CONT and 3 object sizes (screen space)—0.2 cm, 0.3 cm, and 0.4 cm) within subject repeated measure. We hypothesized that CONT would show better performance as compared to BTF-R, as the fat finger problem is starting to become a major performance deterrent. Only BTF-R and CONT were compared with regards to the selection performance. Figure 12 shows the test virtual environment.

As in the main experiment, the user was similarly presented with a sequence of 12 objects (all in one of three sizes) that appeared randomly at three different depths. At each appearance of the object, the subject was asked to select the object as fast and accurately as possible. Object selection time and general usability were measured as the dependent variables. Thirteen paid subjects (eight men and two women between the ages of 23 and 29, mean = 27.9/SD = 5.36) participated in the experiment. The subjects were replaced with a different group from the main/pilot, because the same original could not be summoned again, and in addition, the new subject group could reaffirm the previous results with less bias. The study involved six different treatments (or conditions), namely, two interfaces  $\times$

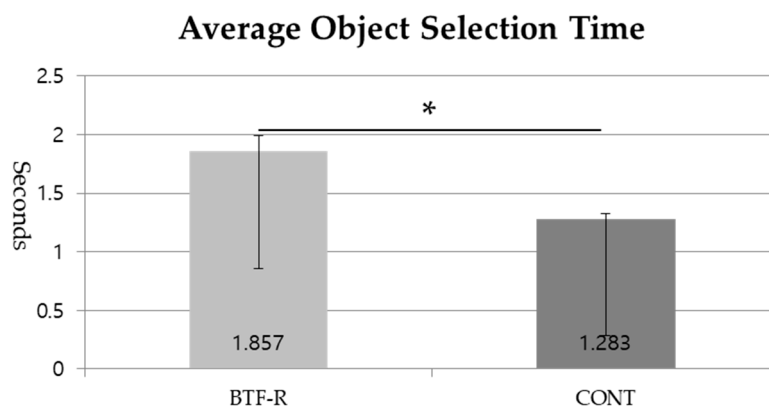
three object sizes; in each treatment condition, 12 different target objects were to be selected, comprising of one trial block, and block trials repeated 12 times (a total of  $2 \times 3 \times 12 \times 12 = 864$  object selections). Each treatment was presented in a balanced fashion using the Latin-square methodology per subject. After the whole sessions, the subjects were asked to fill out the same usability questionnaire. The same analysis methods were applied as in the main experiment.



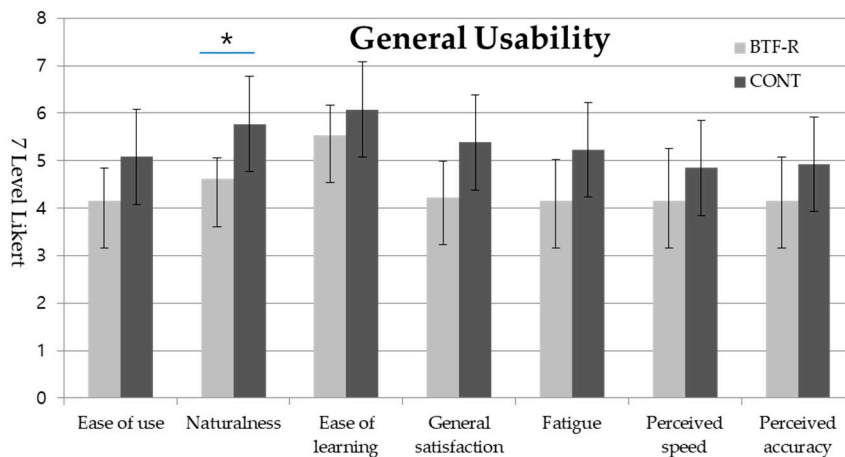
**Figure 12.** Selecting target objects at different depths whose screen spaces were smaller than in the main experiment and the average fingertip size.

6.2. Results and Discussion

Figure 13 shows the average selection time between BTF-R and CONT for the smaller object selection collectively. ANOVA revealed a statistically significant difference ( $p < 0$ ,  $F = 3.333$ ). The performance for each individual size followed the similar trend (specific results not given here). As the object size went well below the smallest recommended size for mobile interaction, the fat finger problem have affected the touch based performance negatively, and likewise the usability as illustrated in Figure 14. In all categories, the CONT exhibited higher usability, with statistical significance for Q2 and Q3.



**Figure 13.** The average selection time between BTF-F and CONT with objects with sizes below 0.4 cm. The asterisk mark (\*) indicates statistically significant difference.



**Figure 14.** Responses to the usability survey for selecting objects with sizes below 0.4 cm between BTF-R and CONT. The asterisk mark (\*) indicates statistically significant difference.

More seriously, our observation and post-briefing revealed that the subjects had quite a difficulty in fusing the stereoscopic imagery with the EasyVR set-up, contrast to the main experiment in which no such phenomenon occurred. For some reason, subjects had trouble being able to focus on the small objects as laid out at different depth. This goes to show that additional interface would be needed to resolve this problem.

## 7. Conclusions

In this paper, we proposed, Blurry Touch Finger, a touch screen based interaction technique an open mobile VR platform with fingers accessible to the screen. We emphasize that the main objective was to show effectiveness of BTF on EasyVR (not EasyVR itself as a framework), as a viable alternative to existing selection method like the gaze/button and even controller based ray-casting methods. Our work partly demonstrated the utility of the proposed method; the object selection performance was as good as when using head-directed and the hand-driven ray casting methods.

Binocular rivalry is usually a phenomenon that affects the visual perception in a negative way. The blurriness caused by the rivalry incidentally made it possible for the user to look through the finger, and added by the sense of proprioception, and accurately select the wanted object. In addition, such touch based selection is already a quite familiar task through the usual interaction with the smartphones.

However, even with the after-release technique and the see-through nature of the “blurry” touch selection, for objects fairly below the size of the fingertip, the touch based performance suffered from the usual fat finger problem and difficulty in focus. Additional interactional provision can help alleviate this problem, e.g., by relying on techniques such as the “scaled” grab [43], in which the potential target can be effectively be enlarged with respect to the user for easier selection.

Although the EasyVR platform (or the use of open clip-on lenses) is not as popular as Google Cardboard or regular head-set based VR at this time, we believe that users and developers will now understand that EasyVR offers reasonable immersion despite being open (refer to [6] by the same authors) and can possibly allow similar user experience to regular non-VR (e.g., touch based smartphone) mobile gaming. Thus, we believe there is a big potential in the future for further popularizing VR and smearing its usage into our everyday lives.

Furthermore, there are many other basic VR tasks (e.g., object manipulation and rotation, navigation, menu selection) with which BTF can be extended and applied to. Note that any interaction methods for EasyVR is bound to involve touch based selection, e.g., of interface buttons. In fact, we have already begun the study of navigation methods in EasyVR [44] in which it was observed that no users have any difficulty in interacting with the touch screen buttons and interfaces using the Blurry

Touch Finger. Even the BTF selection method itself can be realized in a variety of forms (e.g., look of cursor, generalizing into a cone-shaped flashlight metaphor). Continued studies e.g., with respect to the Fitts' law with a larger number of participants are needed to fully validate and characterize the interaction performance and explore the design space of the Blurry Touch Finger as the main interface for EasyVR.

In addition, we assumed that EasyVR was in general more convenient to use than closed M-VR (such as Cardboard/GearVR). However, there may be negative effects e.g., to the level of immersion/presence due to peripheral distraction and touch based interaction. This is subject of another on-going work [6]. Touch based interaction is generally more suited for direct interaction with the target objects, however, there may be instances where indirect interaction is more appropriate in VR. Accidental touches can also be problematic.

Nevertheless, our study has given a promising glimpse into the potential of the touch based interaction for M-VR, especially in the context of bringing EasyVR to a larger user base. We envision EasyVR/BTF can be used for causal touch-driven games and quick reviewing of 360 videos. It can also work to provide dual modes in serious games and social networking situation, allowing the user to seamlessly switch between regular smart phone and immersive viewing with the same touch interaction.

**Author Contributions:** Conceptualization, Y.R.K. and G.J.K.; methodology, Y.R.K. and G.J.K.; software, Y.R.K., S.P.; validation, Y.R.K., S.P.; formal analysis, Y.R.K., S.P. and G.J.K.; investigation, Y.R.K. and G.J.K.; resources, Y.R.K. and G.J.K.; data curation, Y.R.K., S.P.; writing—original draft preparation, Y.R.K. and G.J.K.; writing—review and editing, Y.R.K. and G.J.K.; visualization, Y.R.K., S.P.; supervision, G.J.K.; project administration, G.J.K.; funding acquisition, G.J.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2016-0-00312) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lewis, P. Daydream VR vs. HTC Vive & Oculus Rift|Mobile VR vs. PC. Available online: <http://www.pcadvisor.co.uk/review/wearable-tech/daydream-vr-vs-htc-vive-oculus-rift-mobile-vr-vs-pc-vr-3647738/> (accessed on 8 October 2020).
2. Google Cardboard. Available online: <http://vr.google.com/cardboard/> (accessed on 8 October 2020).
3. Samsung. Samsung Gear VR. Available online: <http://www.samsung.com/global/galaxy/gear-vr/> (accessed on 8 October 2020).
4. HTC. HTC Vive. Available online: <http://www.vive.com/kr/> (accessed on 8 October 2020).
5. Oculus. Oculus VR. Available online: <http://www.oculus.com/> (accessed on 8 October 2020).
6. Kim, Y.R.; Kim, G.J. Presence and immersion of “easy” mobile VR with open flip-on lenses. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, Gothenburg, Sweden, 8–10 November 2017; pp. 1–7.
7. Fignant VR, Fignant VR. Available online: <http://www.fignantvr.com/> (accessed on 8 October 2020).
8. Siek, K.A.; Rogers, Y.; Connelly, K.H. Fat finger worries: How older and younger users physically interact with PDAs. In Proceedings of the 10th IFIP TC13 International Conference on Human-Computer Interaction, Rome, Italy, 12–16 September 2005; pp. 267–280.
9. Bowman, D.A.; Hodges, L.F. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In Proceedings of the 1997 Symposium on Interactive 3D Graphics, Providence, RI, USA, 27–30 April 1997; pp. 35–ff.
10. Bowman, D.A.; Kruijff, E.; Laviola, J.J.; Poupyrev, I. *3D User Interfaces: Theory and Practice*; Addison-Wesley Professional: Redwood City, CA, USA, 2004.
11. Yu, J.E.; Kim, G.J. Blurry (sticky) finger: Proprioceptive pointing and selection of distant objects for optical see-through based augmented reality. In Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Merida, Mexico, 19–23 September 2016; pp. 336–337.



12. Bolt, R.A. "Put-that-there" Voice and gesture at the graphics interface. In Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, Seattle, WA, USA, 14–18 July 1980; pp. 262–270.
13. Liang, J.; Green, M. JDCAD: A highly interactive 3D modeling system. *Comput. Graph.* **1994**, *18*, 499–506. [[CrossRef](#)]
14. Colaço, A.; Kirmani, A.; Yang, H.S.; Gong, N.W.; Schmandt, C.; Goyal, V.K. Mime: Compact, low power 3D gesture sensing for interaction with head mounted displays. In Proceedings of the 26th Annual ACM Symposium on User interface Software and Technology, St. Andrews, UK, 8–11 October 2013; pp. 227–236.
15. Bowman, D.A.; Hodges, L.F. Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *J. Vis. Lang. Comput.* **1999**, *10*, 37–53. [[CrossRef](#)]
16. Lee, G.A.; Bai, H.; Billinghamurst, M. Automatic zooming interface for tangible augmented reality applications. In Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry, Singapore, 2–4 December 2012; pp. 9–12.
17. Wachs, J.P.; Kölsch, M.; Stern, H.; Edan, Y. Vision-based hand-gesture applications. *Commun. ACM* **2011**, *54*, 60–71. [[CrossRef](#)]
18. Mistry, P.; Maes, P. SixthSense: A wearable gestural interface. In Proceedings of the SIGGRAPH Asia 2009 Art Gallery and Emerging Technologies: Adaptation, Yokohama, Japan, 16–19 December 2009; pp. 85–85.
19. Ha, T.; Woo, W. Bare hand interface for interaction in the video see-through HMD based wearable AR environment. In Proceedings of the International Conference on Entertainment Computing, Cambridge, UK, 20–22 September 2006; pp. 354–357.
20. Piumsomboon, T.; Clark, A.; Umakatsu, A.; Billinghamurst, M. Poster: Physically-based natural hand and tangible AR interaction for face-to-face collaboration on a tabletop. In Proceedings of the 2012 IEEE Symposium on 3D User Interfaces (3DUI), Costa Mesa, CA, USA, 4–5 March 2012; pp. 155–156.
21. Prisacariu, V.A.; Reid, I. 3D hand tracking for human computer interaction. *Image Vis. Comput.* **2012**, *30*, 236–250. [[CrossRef](#)]
22. Gielen, C.; van den Heuvel, P.; van Gisbergen, J. Coordination of fast eye and arm movements in a tracking task. *Exp. Brain Res.* **1984**, *56*, 154–161. [[CrossRef](#)] [[PubMed](#)]
23. Sarchilli, K.; Vercher, J.L. The oculomanual coordination control center takes into account the mechanical properties of the arm. *Exp. Brain Res.* **1999**, *124*, 42–52. [[CrossRef](#)] [[PubMed](#)]
24. Bernier, P.M.; Chua, R.; Franks, I.M. Is proprioception calibrated during visually guided movements? *Exp. Brain Res.* **2005**, *167*, 292–296. [[CrossRef](#)] [[PubMed](#)]
25. Helsen, W.F.; Elliott, D.; Starkes, J.L.; Ricker, K.L. Coupling of eye, finger, elbow, and shoulder movements during manual aiming. *J. Mot. Behav.* **2000**, *32*, 241–248. [[CrossRef](#)] [[PubMed](#)]
26. Li, Z.; Jarvis, R. Real time hand gesture recognition using a range camera. In Proceedings of the Australasian Conference on Robotics and Automation (ACRA), Sydney, Australia, 2–4 December 2009; pp. 21–27.
27. Pierce, J.S.; Forsberg, A.S.; Conway, M.J.; Hong, S.; Zeleznik, R.C.; Mine, M.R. Image plane interaction techniques in 3D immersive environments. In Proceedings of the 1997 Symposium on Interactive 3D Graphics, Providence, RI, USA, 27–30 April 1997; p. 39.
28. Bowman, D.A.; Koller, D.; Hodges, L.F. A methodology for the evaluation of travel techniques for immersive virtual environments. *Virtual Real.* **1998**, *3*, 120–131. [[CrossRef](#)]
29. Samsung. Samsung GearVR With Controller. Available online: <http://www.samsung.com/us/mobile/virtual-reality/gear-vr/gear-vr-with-controller-sm-r324nzaaxar/> (accessed on 8 October 2020).
30. Poupyrev, I.; Ichikawa, T.; Weghorst, S.; Billinghamurst, M. Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. *Comput. Graph. Forum* **1998**, *17*, 41–52. [[CrossRef](#)]
31. Lee, S.; Seo, J.; Kim, G.J.; Park, C.M. Evaluation of pointing techniques for ray casting selection in virtual environments. In Proceedings of the International Conference on Virtual Reality and its Application in Industry, Hangzhou, China, 9–12 April 2002; Volume 4756, pp. 38–44.
32. Gugenheimer, J.; Dobbstein, D.; Winkler, C.; Haas, G.; Rukzio, E. Facetouch: Enabling touch interaction in display fixed uis for mobile virtual reality. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, Japan, 16–19 October 2016; pp. 49–60.
33. Lee, J.; Kim, B.; Suh, B.; Koh, E. Exploring the front touch interface for virtual reality headsets. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 2585–2591.

34. Point Shooting. Available online: [https://en.wikipedia.org/wiki/Point\\_shooting](https://en.wikipedia.org/wiki/Point_shooting) (accessed on 8 October 2020).
35. Alais, D.; Blake, R. *Binocular Rivalry*; MIT Press: Cambridge, MA, USA, 2005.
36. Allelotropia. Available online: [https://en.wikipedia.org/wiki/Binocular\\_vision](https://en.wikipedia.org/wiki/Binocular_vision) (accessed on 8 October 2020).
37. Android User Interface Guideline. Available online: [https://developer.android.com/guide/practices/ui\\_guidelines/index.html](https://developer.android.com/guide/practices/ui_guidelines/index.html) (accessed on 8 October 2020).
38. Cournia, N.; Smith, J.D.; Duchowski, A.T. Gaze-vs. hand-based pointing in virtual environments. In Proceedings of the CHI'03 Extended Abstracts on Human Factors in Computing Systems, Ft. Lauderdale, FL, USA, 5–10 April 2003; pp. 772–773.
39. Samsung. Samsung Galaxy S7. Available online: <http://www.samsung.com/global/galaxy/galaxy-s7/> (accessed on 8 October 2020).
40. Homido. Homido Mini VR Glasses. Available online: <http://homido.com/en/mini> (accessed on 8 October 2020).
41. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv. Psychol.* **1988**, *52*, 139–183.
42. Brooke, J. SUS: A “quick and dirty” usability. *Usability Eval. Ind.* **1996**, *189*, 4–7.
43. Mine, M.R.; Brooks, F.P., Jr.; Sequin, C.H. Moving objects in space: Exploiting proprioception in virtual-environment interaction. In Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 3–8 August 1997; pp. 19–26.
44. Kim, Y.R.; Choi, H.; Chang, M.; Kim, G.J. Applying touchscreen based navigation techniques to mobile virtual reality with open clip-on lenses. *Electronics* **2020**, *9*, 1448. [[CrossRef](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).