IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 3, 2021, accepted April 15, 2021, date of publication April 26, 2021, date of current version May 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3075769

Reducing Arm Fatigue in Virtual Reality by Introducing 3D-Spatial Offset

HASAN IQBAL[1], SEEMAB LATIF[2], (Senior Member, IEEE), YUKANG YAN[1], CHUN YU[1], AND YUANCHUN SHI[1]

[1]Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

[2]School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

Corresponding author: Chun Yu (chunyu@tsinghua.edu.cn)

ABSTRACT Arm fatigue is an important factor affecting user experience in Virtual Reality (VR). In this work, we have proposed ProxyHand and StickHand, virtual hand techniques to address this issue. Using ProxyHand or StickHand, users can flexibly adjust the 3D-spatial offset between the physical hand and its virtual representation. This will allow users to keep their arms in a comfortable posture (vertically down) even when they have to manipulate objects in locations that require lifting of arms using the default interaction method. Proposed ProxyHand and StickHand have a similar Underlying concept that is to introduce a 3D-spatial offset between the physical hand and its virtual representation in VR. However, they respond differently to the user's hand movements because of different working mechanisms. Question arises whether the 3D-spatial offset will negatively impact the hand control ability as the directness of interaction is being violated. To investigate this, we conducted user studies where users were asked to perform object translation, rotation and hybrid tasks. ProxyHand and StickHand are used in combination in some scenarios to maximize positive impact on the user experience in VR. This raises the question to find the best possible combination of these virtual hands to reduce arm fatigue. Firstly, for this purpose, we combined both virtual hands by manually allowing users to switch between ProxyHand and StickHand. Secondly, we used machine learning to automatically switch between both the virtual hands. Results showed that introduction of a 3D-spatial offset largely reduced the arm fatigue while offering equal performance to the default interaction method for all these tasks; translation, rotation and hybrid task. Users preferred using ProxyHand and StickHand to interact in the VR environment for longer periods of time.

INDEX TERMS HMD, VR, ProxyHand, StickHand, 3D-spatial offset.

I. INTRODUCTION

Virtual Reality Head Mounted Displays (HMD) are gaining popularity with the development of head-mounted displays. In future, we can envision people using this platform as part of their daily routine, just like desktop computers.

Even though interaction inside virtual reality is a huge boost for the progress of revolutionizing virtual reality, this interaction gives birth to the arm-fatigue problem also known as "Gorilla arm". This problem is known to harm users as using hand gestures in mid-air for longer periods of time can cause pain and strain in arms, which makes the experience of virtual reality less delightful.

The associate editor coordinating the review of this manuscript and approving it for publication was Eunil Park.

This work attempts to address this fatigue problem by using a combination of feasible techniques, we call "ProxyHand" and "StickHand". These techniques have shown to significantly reduce strain and fatigue in the arms of the VR users. ProxyHand and StickHand enable users to flexibly adjust the 3D spatial offset between the physical hand and its virtual representation in VR. This technique will allow users to keep their arms in a comfortable position even when they have to manipulate objects in locations that requires lifting of arms using the default interaction method. Users can also reset this offset whenever they want by using our technique.

In this work, we tried to answer the question that whether ProxyHand and StickHand can assist users in solving the fatigue problem and if the users can use it with no or minimal coaching. The directness of interaction is violated as there is a 3D-spatial offset between the user's physical hand and its virtual representation. Another question addressed by this research is whether users can perform different tasks accurately and with similar performance as in the default interaction method. Finally we proposed an optimum combination of both of these hands, since ProxyHand and StickHand have their own unique advantages.

Firstly, we designed the ProxyHand to solve the fatigue problem. It enabled the users to work within VR while having their arms in a comfortable position by introducing a 3D-spatial offset between their physical hand and its virtual representation. After analyzing the strengths and drawbacks, we designed StickHand to overcome the scenarios where ProxyHand was not suitable. We combined these techniques to satisfy all the scenarios as none of them was able to overcome these scenarios by itself. For this purpose, we made a manual switch mode in which users could manually switch between ProxyHand and StickHand according to their needs. We designed a user study to compare the performance of the default method and our manual combination of ProxyHand and StickHand. In this user study, we designed translation tasks where users were required to translate objects from one position to another, rotation tasks in which users were required to rotate objects in different orientations, and hybrid tasks where users were required to rotate as well translate objects simultaneously.

Results of this user study showed that there was no significant difference in performance measured by total completion time between the manual combination of ProxyHand/StickHand and the default interaction method. Users could perform the same tasks with much less arm fatigue by using the combination of ProxyHand/StickHand when compared with the default method. The subjective feedback of the user study showed that even though arm fatigue was significantly reduced, but mental engagement had significantly increased by using the combination of ProxyHand and StickHand. We used a machine learning algorithm and developed an automatic switch for changing modes between ProxyHand and StickHand to reduce this mental demand. Automatic switch mode predicted the user's intention to switch between ProxyHand and StickHand. Finally, we conducted a user study to inspect the performance of manual mode and automatic mode to switch between techniques. Both modes had no significant difference in performance. However, mental demand was significantly reduced by automatic mode as compared to manual mode.

Analysis of the user studies showed that ProxyHand and StickHand were able to significantly reduce the arm fatigue levels when compared to the default method while maintaining the comparable performance. The majority of the participants reported that they would prefer to use ProxyHand and StickHand in daily life rather than the default interaction method.

Rest of the paper is organised as follows: section II describes the related work concerning arm fatigue problem.

II. RELATED WORK

We reviewed the fatigue problem as well as indirect pointing and manipulation techniques.

A. FATIGUE PROBLEM IN VR

There are several manipulation techniques for interaction within virtual environments. As described by Poupyrev *et al.* [24], these techniques can be decomposed into *exocentric* and *egocentric* metaphors. They are considered exocentric if the user interacts from outside the virtual environment (third person view). Exocentric techniques include World in Miniature, Automatic Scaling etc. In Egocentric techniques, the user interacts inside the virtual environment. These can be further classified into virtual hand metaphors (Classic Virtual Hands, Go-Go technique, indirect Go-Go, Voodoo Dolls etc) and virtual pointer metaphors (Ray-casting, Flexible pointer, Flashlight, Aperture, Image Plane etc). Nowadays, most VR Applications and Games use mainly Egocentric Techniques i.e. ray-casting and virtual hands. Some of the Ray-casting techniques and almost all virtual hand metaphors require users to lift their arms in the air. Continuously lifting arms in the air for longer periods introduces Arm Fatigue (Gorilla Arm Effect). It has been observed that the arm fatigue problem plays a significant role in user experience in the VR environment [14], [16], [18], [25].

Arm fatigue problem falls in the category of ergonomics. Ergonomics is considered as one of the core constituents of Human-Computer Interaction. It is extensively used to evaluate how humans interact with computer systems and which actions/gestures are of high-cost [10], [26]. Arm Fatigue can be evaluated by using self-report, observing, or direct measurement methods. Self report methods include Likert scale questions [7], NASA TLX Load Index evaluation [9] or Borg CR10 scales [6]. Observational methods include observing fatigue levels using interactive human models like jack [4] or RULA [21] where they surveyed several observed fatigue levels for different postures. Direct measurement methods include evaluating Arm Fatigue by using EMG sensors [3], [11], [29]. These studies also evaluated different arm and body positions, and calculated which orientations and postures accumulate the highest levels of exertion along with fatigue. Higher Fatigue levels not only affect user physically, but they affect the overall user experiences as well. When users performed gestures in the air to interact with devices like Kinect and LeapMotion, the arm fatigue significantly affected the user experience [28], [32]. Based on these results, studies were conducted to optimize the arm posture to reduce the fatigue [16], [18]. Our aim in this paper is to propose a solution for continuously maintaining a comfortable arm posture by the users to avoid the fatigue problem by using ProxyHand and StickHand, the new virtual hands' metaphors.

B. INDIRECT POINTING

The most widely used example of indirect pointing in 2D is mouse control in computers. Other than that, indirect pointing devices and techniques are touchpad, joystick, trackball and ray-casting etc. Many studies have been done to find out the advantages of using indirect pointing techniques as compared to direct pointing techniques [12], [13], [27]. These studies show that direct pointing techniques are better for precision, whereas indirect pointing devices are better in terms of speed. Indirect pointing techniques can alleviate the fatigue problem due to the adjustment of the CD-ratio (Control-Display ratio) [5].

As indirect pointing provides the above-mentioned benefits, it would be of great value to transmit it to the 3D object manipulation in virtual environments. However, in the 3D virtual environment, current object selection is mostly achieved by direct pointing techniques, e.g. using HTC Vive controllers. Previously, researchers also proposed ray-casting techniques to select targets through indirect manner [8], [34]. Users emit a ray from their finger or the controller to point at the target. Multi-touch virtual mouse has been designed to control objects in the 3D virtual environment [30]. The Go-go indirect pointing technique allows the users to reach remote objects with a proxy hand by enlarging the CD-ratio as the hands moves forward [23]. In this paper, we enabled users to customize the distance between the proxy hand and their physical hand to keep their physical hands in a comfortable region. However, the offset between two hands remains static during the interaction process. This provides users with more stable control the proxy hand.

C. TRANSLATION OF INTERACTION SPACE

Fatigue problem can be minimized by translating the interaction space from high fatigue regions to low levels of arm fatigue regions. This translation of interaction space can be done by clutching mechanism, which allows users to relocate control space [2], [17]. Users can grab objects by using this mechanism which they cannot grab normally while standing in the same position. It is primarily used for object manipulation outside around-body interaction space. Work has been done on decoupling the motor space and visual space for flexible manipulation on 2D windows in virtual environments [1]. The interaction region is transformed for easy iteration by the user. In this study, researchers have applied this concept for interaction on 2D applications that were only projected in 3D environments. They have also solved the fatigue problem for ray-casting technique and 2D objects manipulation in 3D environments. This study comprised of the selection tasks only whereas we are solving the fatigue problem for virtual hand metaphor and our experiments tasks comprised of translation, rotation [31] and hybrid tasks. In another study, work has been done to map the visual and physical space geometrically to reduce the arm fatigue for 3D object manipulation in virtual environments [22]. In this work, natural interaction is compromised as the user moves hand in a straight line, the corresponding virtual hand will not move in a straight line because of different geometries of both visual and physical spaces. In our proposed research, we have tried to solve arm fatigue problem while preserving the natural interaction. We believe that our combination of ProxyHand and StickHand is more suitable clutching mechanism in the context of arm fatigue problem.

D. UNCERTAIN INPUT

Uncertain input is a state when there is a possibility of multiple user inputs and the system is unable to distinguish between the user's multiple input intentions requiring some time to recognize it. This concept has been used in the context of pointers and cursors for selection and manipulation. Area cursor [19] and Bubble cursor [15] are single cursors that change according to the need of the situation and user's actions. Multiple cursor techniques, satellite cursor [33] and ninja cursor [20] have also been implemented in the context of uncertain input. Instead of modifying the existing cursors, these techniques make use of replicas of the cursors at different positions. We have incorporated these techniques by assigning two virtual hands and enabling them to move differently when users rotate their hands.

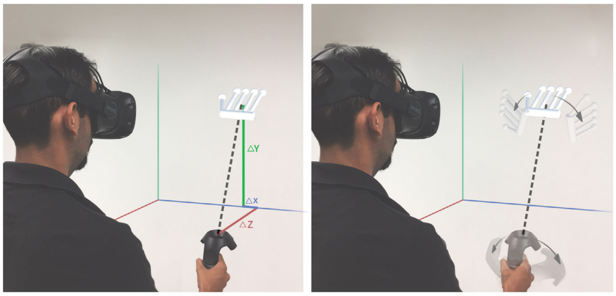
III. ProxyHand AND StickHand DESIGN

A. ProxyHand DESIGN

The default interaction method of virtual reality has enabled the users to interact in virtual environments, which is a great step forward in the advancement of virtual reality. Although, it provides a great user experience but arm fatigue caused by its prolonged use makes the user experience less delightful. The design goal of ProxyHand and StickHand was to enable the users to interact in VR without experiencing arm fatigue while interacting in front-body orientations in virtual environments. Our research goal was to have users arms in comfortable region with noticeable low levels of arm fatigue for different scenarios discussed in the introduction. The object manipulation comprises of selection, translation and rotation tasks. There can be many different egocentric techniques that can be employed for completion of these tasks. For example, ray-casting, Go-Go, virtual hand etc can be used for selection tasks, ray-casting, virtual hand etc can be used for translation tasks, and virtual hand can be used for rotation tasks. The goal of our research was to complete all these tasks simultaneously, while preserving the natural interaction and having similar performance as of physical hand with users keeping their hands in the comfortable region. For this purpose, we used the concept of indirect pointing in the 3D environment. We defined a 3D offset from the absolute position of the controller as shown in figure 2 and called it a ProxyHand.

We added offsets on the X-axis, Y-axis and Z-axis of the virtual hand on the absolute position of the controller as shown in equation one. We maintained the rotation of the virtual hand just like the rotation of the controller as shown in figure 1. The virtual hand appeared at a fixed distance from the absolute position of the controller.

$$(x', y', z') = (x + \Delta x, y + \Delta y, z + \Delta z) \tag{1}$$



(a) (b)

FIGURE 1. **(a) Relative position of virtual hand from absolute position of controller (b) Virtual hand rotates exactly like the rotation of controller.**

Where x', y' and z' represent the coordinates of ProxyHand/StickHand, x, y and z represent the coordinates of the centre position of controller and Δx, Δy and Δz represents the length of the offset for each of their respective coordinates.

We enabled the users to define their 3D spatial offset as shown in figure 1. Users can lift their hands to define the position of their virtual hands in the region of VR environment. Then press and hold the side button of the controller until they drag their hands to a comfortable position with less strain on their arms. When they release the side button, a 3D offset is set between the controller and virtual hand. Users can reset the offset at any time by pressing the same button. Users can set this offset at any position in the around-body interaction space.

Users should be able to easily move an object from one place to another within the front-body orientation space based on our design. However, if users want to move the object beyond their front-body orientation region, (for example, if they want to grab an object in front of them and place it behind them) they would not be able to do so without resetting the 3D spatial offset at least once. We tested this in the pilot study as well.

ProxyHand was designed and implemented using HTC Vive Head Mounted Display alongside two HTC controllers and two base stations. Platforms used for development were SteamVR and Unity3D. However, this technique is not limited to certain hardware or software. It can be implemented by using any other HMDs like Oculus, Sony etc. as well.

Hypothetically, the distance to move an object from one point to another in front-body orientation is the same for the physical hand as well as the ProxyHand. Therefore, the performance of ProxyHand should be similar to the performance of the Physical hand. We believe that this technique of introducing offset between the physical hand and its virtual representation will seem natural to the user providing them a good user experience and quick adaption. We conducted the following pilot study to test assumptions.

B. PILOT STUDY 1

After finalizing the design of ProxyHand, we conducted a pilot study by inviting 5 participants (4 males and 1 female; aged from 22 to 29). We implemented our technique using HTC HMD and its controllers. The prototype of this pilot study was developed using unity 3D and steamVR. There were 3 main objectives for this pilot study. First, to explore the most optimum region where users can perform tasks without any arm fatigue. Second, to find out the possible control space. Third, to learn about the existing limitations of this technique.

We analyzed the most comfortable region by asking participants to identify highest position they wanted their hands to be while their arms were still in their comfortable posture (see figure 3a). Similarly, we asked them about the lowest point they want their hands to be while their arms were in rest position (see figure 3b). We then took the average of these positions for all the participants to find out the typical comfortable region for users which they wanted to use for interaction (see figure 3c).

Next, we asked users to manipulate objects and move them to different around-body interaction regions. We assessed that this technique works perfectly when users set the interaction region just in front of them. Using the current implementation of this technique, it was not possible to move objects from one place to another just by rotating hands as shown in figure 4. To move an object from A to C, users had to define more than one region which was not very convenient. Thus, this technique was most effective when users were manipulating objects in front of them. We designed following user study concerning this limitation. We further leave it as a limitation and discuss possible solutions at the end of this paper.

We asked users to grab an object in the prototype application and move it from one point to another along all the x, y and z axes. Users reported that they could easily manipulate objects along the x-axis and z-axis but faced some difficulties along the y-axis. As our technique was one-to-one mapping, users faced difficulty when they had to manipulate some objects outside the control region. For manipulating the objects along the y-axis (top or bottom), users had to raise/lower their arms outside the comfortable region. Users also reported that the procedure of setting the offset was not so convenient, when they had to reset offset for multiple times required by some tasks (see figure 3).

We examined the strengths and weaknesses of ProxyHand by careful analyzing this pilot study. A major shortcoming of ProxyHand is the inefficiency due to multiple setting of offsets to move the object across the around-body interaction region. To overcome this shortcoming, we came up with a new virtual hand named as StickHand which is described in detail in the next section. The procedure of setting offset was also needed to be changed as users felt uncomfortable raising their hands repeatedly.

C. StickHand–DESIGN

After analyzing the pilot study results, we decided to address the shortcomings of ProxyHand by designing a new technique which we called StickHand. The design goal of StickHand was to enable the users to manipulate the objects across the

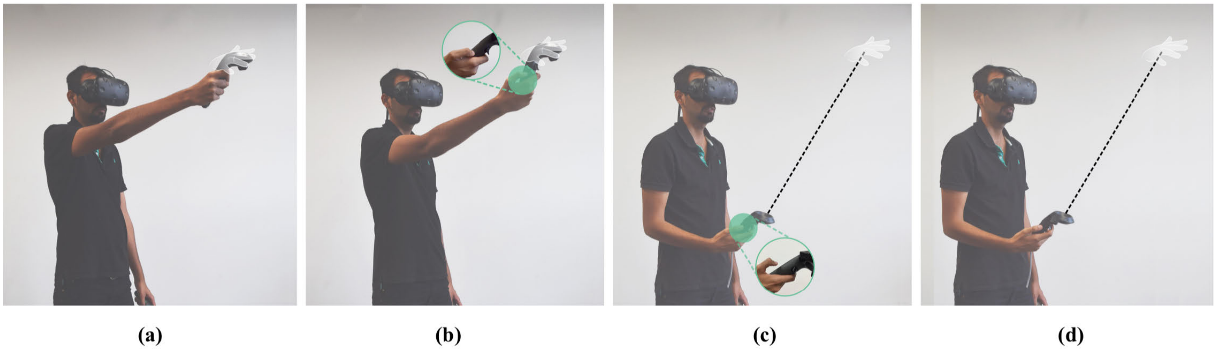


FIGURE 2. A storyboard illustrating a user using the default interaction method and ProxyHand. (a) The user is interacting in VR with his arm stretched using the default method which causes arm fatigue. (b) The user can use ProxyHand by setting a 3D offset between the position of the controller and the virtual hand. For setting offset, the user chooses a position where he wants his hand in VR and then presses the side button of the controller. (c) The user drags down his controller. After adjusting his hand position, he releases the side button which will set the 3D offset. (d) Now, the user can manipulate objects in VR while their arms remain in a comfortable posture.

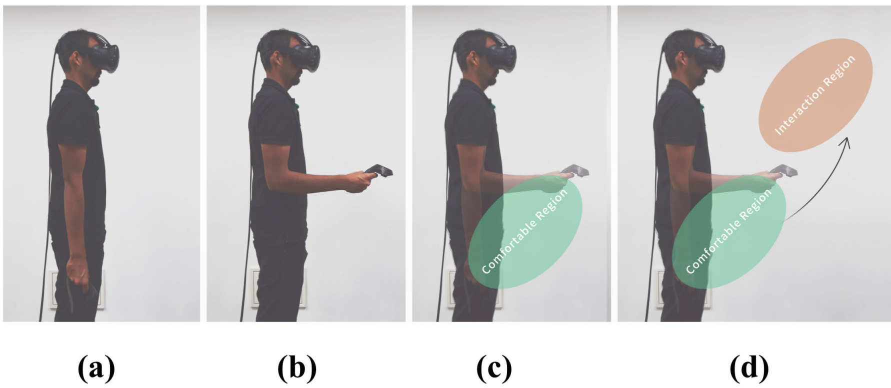


FIGURE 3. (a) lowest point of hand where the user feels his arm is comfortable (b) highest position of hand where the user feels his arm is comfortable (c) comfortable region for interaction (d) One-to-one translation of comfortable region to the interaction region.

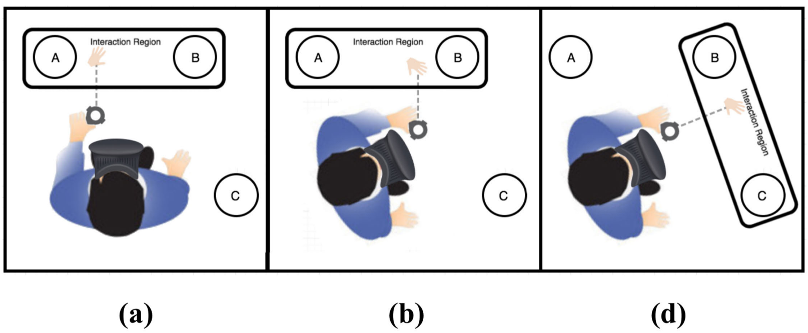


FIGURE 4. (a) User's interaction region where the user can move objects from Point A to B but cannot move objects from Point A or B to point C and vice versa. (b) The interaction region remains the same when the user changes their body orientation (c) Interaction region changes when the user resets offsets. Now the user cannot move objects from B or C to A and vice versa.

around-body interaction space in virtual reality. The essence of the main technique ProxyHand was not disturbed as users were allowed to interact within virtual reality while their hands remained in the comfortable position.

We used the same concept of introducing the 3D spatial offset from the absolute position of the controller. We added an offset on just the Z-axis of the virtual hand from the absolute position of the controller as shown in the equation below. However, we changed the way virtual hands respond to the rotation of the controller as shown in figure 5.

$$(x', y', z') = (x, y, z + \Delta z) \tag{2}$$

where x', y' and z' represent the coordinates of ProxyHand, x, y and z represent the coordinates of the centre position

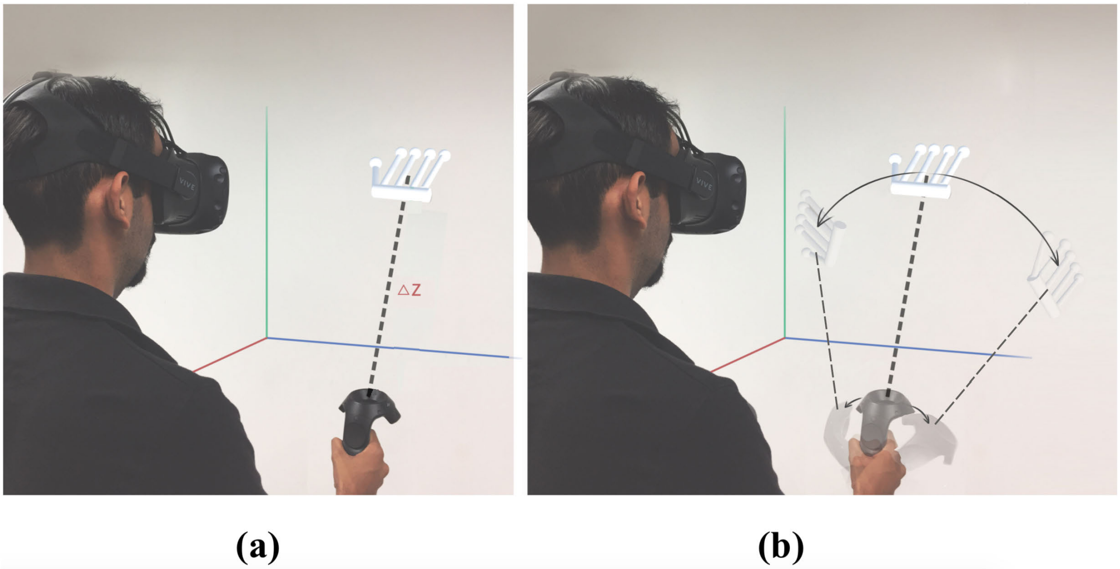


FIGURE 5. (a) Virtual hand is attached to the absolute position of the controller along the z-axis. (b) Virtual hand rotates according to the rotation of controller like a stick in real life.

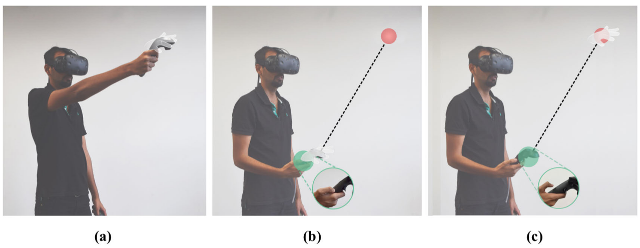


FIGURE 6. (a)User working with default method. (b) User points the laser at the target object. (c)When the button is released, the virtual hand appears at the target position.

of controller and Δz represents the length of the offset for z-coordinate.

We also came up with a new way of setting 3D spatial offset as shown in figure 6. Using this method, users were not required to lift their hands for setting the 3D spatial offset every time. They could set the 3D offset by just pointing the laser in the direction of the object. Moreover, there were two different methods for setting the 3D spatial offset for using this technique and users were given the choice to select one of the methods in the following pilot study for the final design.

The StickHand was designed to manipulate objects across the around-body interaction space by users rotating controllers in particular directions. Hypothetically, translating object from one place to another required the rotation of the wrist as compared to dragging the whole hand from one place to another in case of default method or ProxyHand. However, it was evident that since the position of this virtual hand depended on the rotation of the controller in StickHand which might affects its accuracy. These effects had to be checked on users in the pilot study so that the subsequent user study could be designed accordingly. StickHand was designed and implemented using the same devices and platforms which were used for ProxyHand.

D. PILOT STUDY 2

After finalizing the design of the StickHand, we conducted a pilot study on ProxyHand and StickHand prototype. The main purpose of this study was to evaluate the strengths and weaknesses of both techniques and to ask users which method they would prefer for setting 3D spatial offset, either setting offset by uplifting hands or by pointing the laser onto the target object.

We invited 5 participants from campus (3 males and 2 females; aged from 21 to 27) having experience of the virtual environment. To assess both techniques thoroughly, we designed several scenarios in our prototype. Firstly, we asked users to pick one method for selecting a 3D spatial offset. Then we asked the users to use both techniques one by one and complete given scenarios. After users became familiar with both techniques, we asked them which hand they would prefer to use in the following scenarios.

1. *Where accuracy is required?*
2. *Where rotation is required?*
3. *Where translation is required?*
4. *Where short translation (less than 40cm) is required?*
5. *Where long translation (more than 40cm) is required?*
6. *Where long translation and high accuracy is required?*
7. *Where long translation and low accuracy is required?*

All the participant chose ProxyHand for the first and second scenario. All the participants chose StickHand for the third scenario. Everyone except one participant selected ProxyHand for the fourth scenario. All the participants chose StickHand for the fifth scenario. All the participants answers were very interesting for the sixth scenario, as everyone wanted to combine both the techniques i.e., first translate using StickHand and then use ProxyHand for high accuracy. All participants chose StickHand for the seventh scenario. They also gave their feedback on both techniques.

We assessed strengths and weaknesses of both the techniques by careful observation of the participants' actions during completion of scenarios and analysis of their subjective feedback. ProxyHand was suitable for scenarios where higher accuracy was required i.e. users had to place an object carefully and where users had to rotate an object because the rotation of the virtual hand was exactly mimicking the rotation of the controller. However, it was not suitable for long translation of objects as users had to drag their hands. Users also could not translate an object across the whole around-body interaction space by setting only one offset. On the other hand, StickHand was suitable for long translation as users just needed to rotate their controller for translating an object from one place to another and they did not need to drag their hands. Users also could move an object around-body interaction space using StickHand. However, it was not suitable for rotating objects as the virtual hand moved instead of rotating when users moved their controllers. It was also not feasible for scenarios where high accuracy was needed because of the same reason. Users mentioned in their feedback that both methods have distinct benefits. Both of these techniques complement each other in such a way that the advantages of one technique counter the shortcomings of the other technique and vice versa.

As it was evident that both of these techniques have their benefits and purpose, it was required to combine them. For this purpose, we came up with a manual switch mode that switch between ProxyHand and StickHand mode based on the user requirement. We assigned a separate button on the controller for this manual switch mode. For example, if users were using StickHand, they could change the mode of virtual hand to ProxyHand by clicking that mode switch button and vice versa. We conducted the following user study on combining both of these techniques using Manual Switch.

IV. USER STUDY 1

The goal of this study was to compare the performance of the default interaction method in virtual reality and the combination of ProxyHand and StickHand using a manual mode switch.

A. PARTICIPANTS

We recruited sixteen participants from the campus (11 males and 5 females; aged 21-36 years, Mean= 25.375). Their heights ranged from 160cm to 191cm with an average of 175cm. Their arm lengths ranged from 59cm to 87cm with an average of 73cm. Ten participants were familiar with virtual reality experiences whereas six of them had no or very little virtual reality experience.

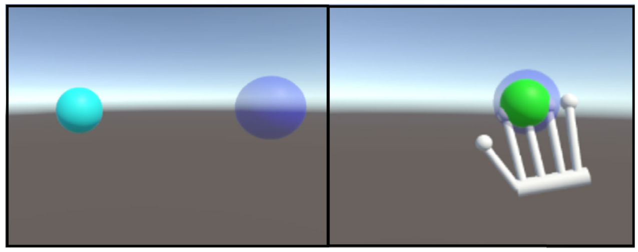


FIGURE 7. (a) two spheres at distance (b) Sphere turns green when its center is in certain accuracy threshold from the center of target sphere.

B. APPARATUS

This user study was designed and implemented using the same devices and platforms which were used for designing ProxyHand and StickHand in section 3. We implemented the combination of both techniques by assigning the hair-trigger of the controller for grabbing objects, the side grip for setting offset and the top button for switching mode between ProxyHand and StickHand.

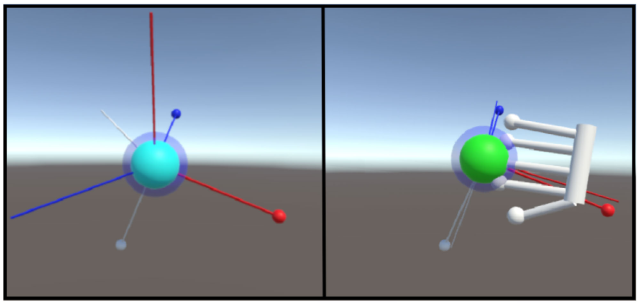
C. DESIGN

We designed translation, rotation and hybrid task (combination of translation and rotation) to calculate the completion times concerning both techniques; default method and combination of ProxyHand and StickHand using manual switch. We used a within-subjects experiment design for all the three tasks.

1) TRANSLATION TASK

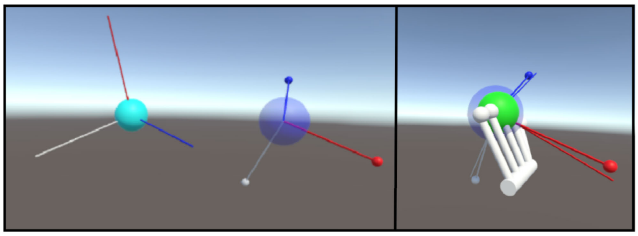
For the translation task, independent variables were Distance Accuracy and Distances whereas the dependent variable was Time. Users were required to grab the main sphere, move it to the target sphere and drop it inside the centre of the target sphere. The main sphere turned green whenever it was inside the target sphere and the distance between their centres was under a set threshold as shown in figure 7. The next scenario was shown when the user released spheres after they turned green. For distinction, the target sphere was transparent and relatively bigger than the main sphere.

In the previous user study, we observed that regions had no significant effect on the way techniques performed. Both the default method and ProxyHand had similar performances. As users highlighted that tasks were long in the previous study, we left regions parameter out to reduce the completion time of the task. We used the same distance accuracies as last time i.e. 1cm, 2cm, 3cm. However, this time, we set three distances as 30cm, 60cm and 90cm. We generated spheres at random positions at around-body interaction space of users. The order of each scenario was also randomized. However, each participant performed both techniques on the same randomly generated scenarios. We set the following number of scenarios: 2 Techniques $\times$ 3 Distance Accuracies $\times$ 3 Distances $\times$ 5 number of trials = 90.



(a) (b)

FIGURE 8. (a) Two spheres with fixed centers, having different colored spikes along x, y and z-axes (b) Sphere turns green when the angles between all the axes are in certain threshold.



(a) (b)

FIGURE 9. (a) Two spheres with spikes at distance (b) Sphere turns green when its center is at certain distance accuracy threshold from the center of target sphere and the angles between all the axes are in certain angular accuracy threshold.

2) ROTATION TASK

For the rotation task, the independent variable was angular accuracy and the dependent variable was time. For this task, spheres were designed with coloured spikes on x, y and z-axes. Centres of both spheres were fixed and users needed to match all axes within a certain threshold. The main sphere turned green when angles between all axes were within the allowed level of threshold as shown in figure 8. Small spheres were fixed to differentiate axes of the main sphere and target sphere.

We set three angular accuracies; 8 degrees, 12 degrees and 16 degrees. The scenarios were randomized as discussed in the translation task. We set the following number of scenarios for this task: 2 Techniques $\times$ 3 Angular Accuracies $\times$ 5 number of trials $=$ 30.

3) HYBRID TASK

The hybrid task was the combination of translation and rotation tasks as users were required to translate and rotate an object simultaneously as shown in figure 9. In this task, independent variables were distance, distance accuracy and angular accuracy and the dependent variable was time.

In this task, users were required to place the main sphere inside the target sphere and also align the axes of both the spheres. There were two levels of distance accuracies; 1cm and 3cm, two levels of distances; 30cm and 90cm and two levels of angular accuracies; 8 degrees and 16 degrees. We set the following number of scenarios for this task: 2 techniques $\times$ 2 distance accuracies $\times$ 2 distances $\times$ 2 angular accuracies $\times$ 5 number of trials $=$ 80.

D. PROCEDURE

We asked each participant to get familiar with the default method and ProxyHand/StickHand manual mode. Users did this warm up on both of these tasks randomly. Then we gathered information from them such as their age, height, arm length, experience in VR and sickness with 3D environment.

Before the actual trial, we described them what they are required to do. Then we gave them translation, rotations and hybrid tasks to perform with both the techniques, default method and ProxyHand/StickHand manual mode. The order of these tasks was random. Users were given the rest of 5-10 minutes between every task.

We calculated the time taken by users for doing individual scenarios of all these tasks. We also used NASA Task Load Index along with three more added questions which were: Arm Fatigue level, Perceived Speed and Perceived Accuracy. There were five levels of each question; very low, low, medium, high and very high. We gave these questionnaires at the end of each task. After users performed all of these tasks, we asked them about the advantages and disadvantages of the default method, ProxyHand and StickHand. We also asked which technique they would prefer to use in daily life. The whole test for one user took around 60-80 minutes.

E. RESULTS

For all of these tasks, there was no significant effect of the technique on completion time.

1) TRANSLATION TASK

The average time taken by all users to complete the translation task with the default method was 1952 ms whereas it was 1944 milliseconds with ProxyHand/StickHand manual switch mode as shown in table 1. RM-ANOVA results show that the difference of completion times of both techniques for this task was not significant ($F_{1,15} = 0.007$, $p = 0.936$). This non-significant difference showed that both techniques gave similar performance overall for this task. The average time taken by each participant for switching between modes in manual switch mode was 148ms for each scenario.

Results showed significant difference between completion times for different distance accuracies ($F_{2,30} = 46.5$, $p < 0.001$). Total time taken for each respective level of distance accuracy is given in table 1 and is shown in figure 10. This result was consistent with the result of the translation task in the previous user study. Although both techniques did not have a significant difference in completion times at each respective distance accuracy level ($F_{2,30} = 3.158$, $p = 0.057$), but their mean times had different trends as shown in figure 11. Default method meantime for all the users was less than manual mode at 1cm accuracy level. However, manual mode had less meantime than the default method at 2cm and 3cm.

TABLE 1. Time taken for different scenarios in Translation Task using both techniques (Default and Manual switch mode) in milliseconds. The 1st column represents the time taken considering all scenarios. Next three columns represent three different levels of Distance Accuracies (DA). Last three columns represent three different levels of Distances (D).

Method	Overall	1cm-DA	2cm-DA	3cm-DA	30cm-D	60cm-D	90cm-D
Default	1952	2552	1766	1538	1454	2065	2337
Manual	1944	2719	1634	1480	1612	1997	2224

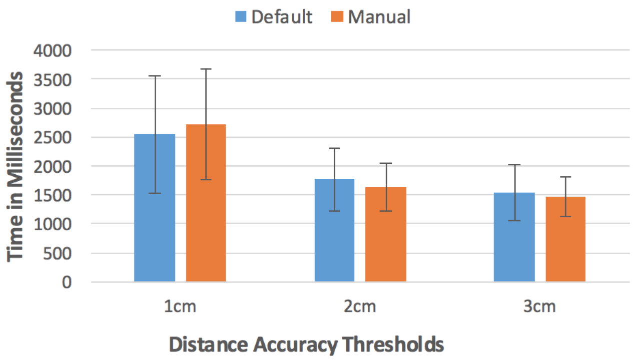


FIGURE 10. Comparison between different distance accuracy thresholds of translation task using default method and manual mode.

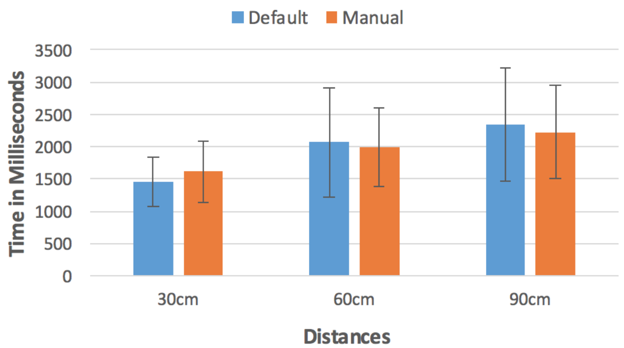


FIGURE 12. Comparison between different distances of translation task using default method and manual mode.

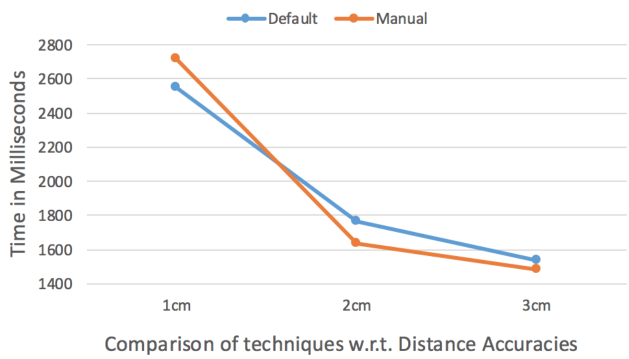


FIGURE 11. Comparison of Both techniques (default and manual mode switch method) with respect to distance accuracies in translation task.

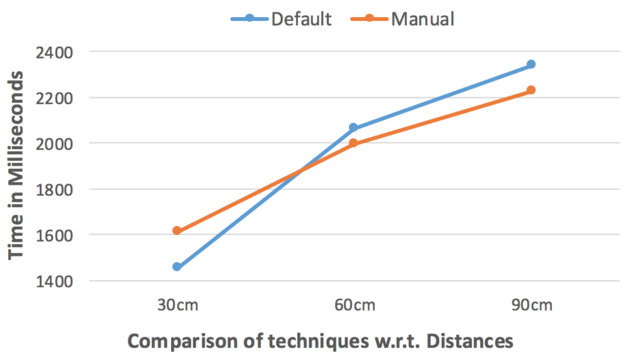


FIGURE 13. Comparison of Both techniques (default and manual mode switch method) with respect to distances in translation task.

Results showed significant difference between completion times for different distances ($F_{2,30}$= 27.17, p < 0.00). Total time taken for each respective level of distances are given in table 1 and are shown in figure 12. This result was different from the result of the translation task in the previous user study. This might be because of the reason that the difference between each parameter was 30cm instead of 10cm, as was in the previous study. Although both techniques did not have a significant difference in completion times at each respective distances ($F_{2,30}$ = 1.104, p = 0.345), but their mean times had different trends as shown in figure 13. Default method meantime for all users was less than manual mode at 30cm distance. However, manual mode had less meantime than the default method at 60cm and 90cm. This trend was similar to that of distance accuracy discussed above.

We analyzed subjective feedback of users by using Wilcoxon test (see figure 14). There were no significant differences between temporal demand (Z= −0.758, p=0.448), performance (Z= −1.19, p=0.234), effort(Z= −1.925, p=0.054), frustration (Z= −0.879, p=0.38), perceived speed (Z= −1.941, p=0.52) and perceived accuracy (Z= −1.069, p=0.285). However, there were significant differences between mental demand (Z= −2.069, p=0.039), physical demand (Z= −3.0, p = 0.003) and arm fatigue (Z= −3.110, p=0.002).

Subjective feedback results showed that both factors like physical demand and arm fatigue were greater in the default method than the manual switch mode because users got tired using the default method. However, mental demand was higher in manual switch mode because users had to remember lots of buttons for each action. They got confused sometimes.

2) ROTATION TASK

The average time taken by all the users to complete the rotation task with the default method was 6819ms and it was 6426ms with ProxyHand/StickHand manual switch mode as shown in table 2. RM-ANOVA results showed that the time difference of both techniques of rotation task was not significant($F_{1,15}$ = 2.259, p = 0.154). The average time taken

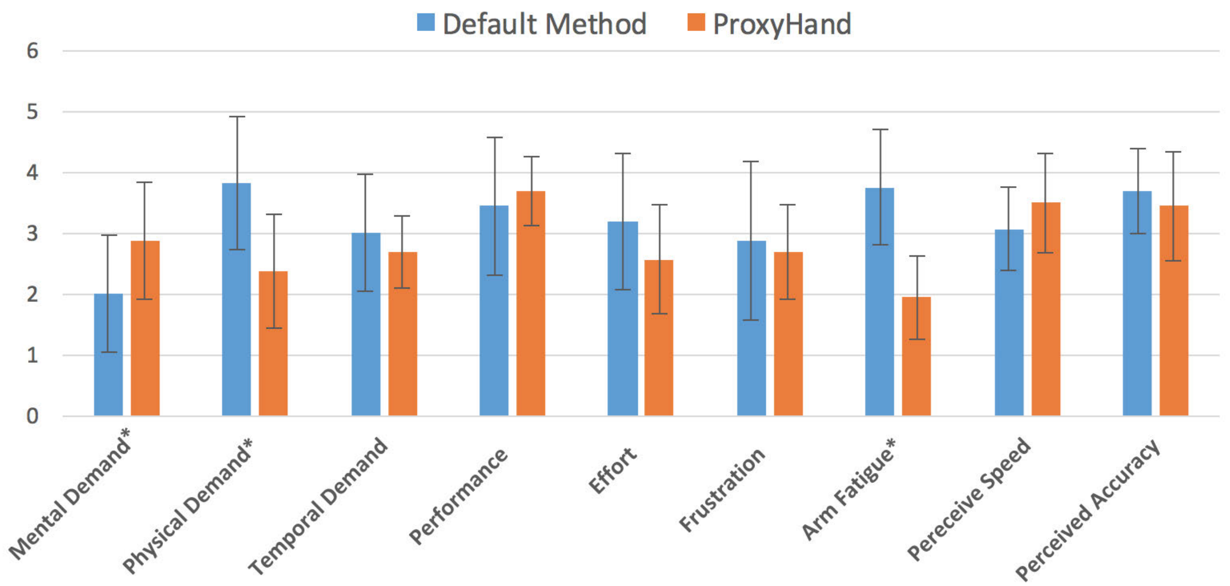


FIGURE 14. Subjective feedback of participants over both default and manual mode techniques (from 1 – very low to 5 – very high) in Translation task. Significant factors are denoted by * in the end.

TABLE 2. Time taken for different scenarios in Rotation task using both techniques (Default and Manual switch mode) in milliseconds. The 1st column represents the time taken considering all scenarios. Next three columns represent three different levels of Angular Accuracies.

Method	Overall	8 °-AA	12 °-AA	16 °-AA
Default	6819	7344	6631	6480
Manual	6426	6953	6684	5640

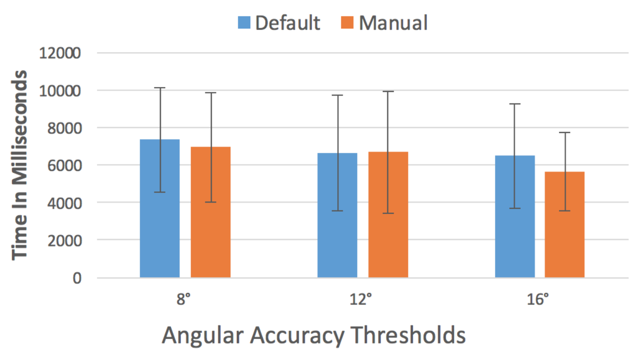


FIGURE 15. Comparison between different angular thresholds of rotation task using default method and manual switch mode.

for switching between modes was 4ms. The reason behind this small switch time was that users switched to ProxyHand and then did not switch back to StickHand as only ProxyHand was sufficient for this rotation task.

There was also a non-significant difference between three different levels of angular accuracy i.e. 8, 12 and 16 degree angular accuracies ($F_{2,30} = 2.591$, $p = 0.092$) as shown in figure 15. This result was odd as compared to the angles task of the previous user study. One reason could be that rotational complexity was not controlled in the given scenarios which might have affected this result. However, there was a significant difference between 8 degrees and 16 degrees ($p = 0.48$).

For subjective feedback (see figure 16), there were no significant differences between mental demand ($Z= -0.513$, $p=0.608$), temporal demand ($Z= -1.412$, $p=0.158$), performance ($Z= -1.512$, $p=0.131$), effort($Z= -1.209$, $p=0.227$), frustration ($Z= -1.548$, $p=0.122$), perceived speed ($Z= -0.264$, $p=0.792$) and perceived accuracy ($Z= -0.832$, $p=0.405$). However, there were significant differences between physical demand ($Z= -2.698$, $p = 0.007$) and arm fatigue ($Z= -2.796$, $p=0.005$). These results were similar to the translation task of this user study except for mental demand. We believe that it was non-significant because users did not switch mode from ProxyHand to StickHand for rotation. Thus the mental demand was low as they did not need to remember which button was assigned for switching mode.

3) HYBRID TASK

For the hybrid task, the mean of the total time taken for each technique is shown in table 3. RM-ANOVA results showed that the time difference for both techniques on the hybrid task was not significant ($F_{1,15} = 0.747$, $p = 0.401$). The average time taken by each participant for switching between hands in manual switch mode was 243ms.

There was a significant difference between different levels of distance accuracies ($F_{1,15} = 19.299$, $p = 0.001$). Participants took more time when higher distance accuracy was required and less time for lower accuracy demand. (see figure 17).

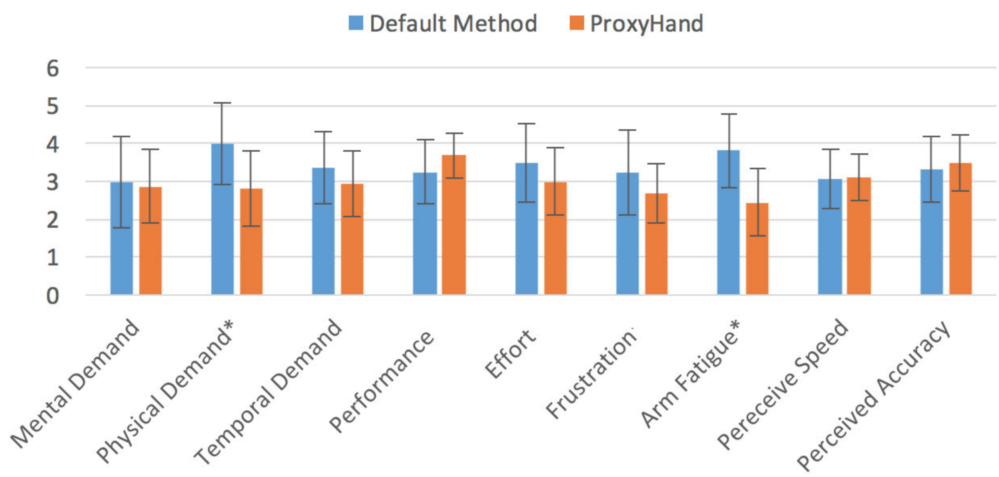


FIGURE 16. Subjective feedback of participants for rotation task over both default and manual switch mode technique (from 1 - very low to 5 - very high). Significant factors are denoted by * in the end.

TABLE 3. Time taken for different scenarios in the Hybrid task using both techniques (Default and manual switch mode) in milliseconds. The 1st column represents the time taken considering all scenarios. Next two columns represent two different distance accuracies. Next two columns represent two different levels of distances (D). Last two columns represent two levels of angular accuracies.

Method	Overall	1cm-DA	3cm-DA	30cm-D	90cm-D	8 °-AA	16 °-AA
Default	7314	8059	6570	7043	7585	8041	6587
Manual	7515	8139	6891	6831	8198	7866	7163

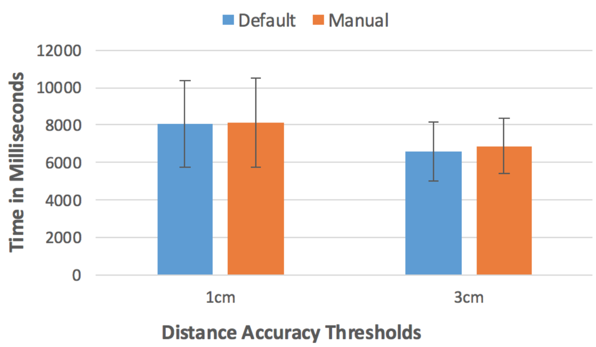


FIGURE 17. Comparison between different distance accuracy thresholds of hybrid task using default method and manual switch mode.

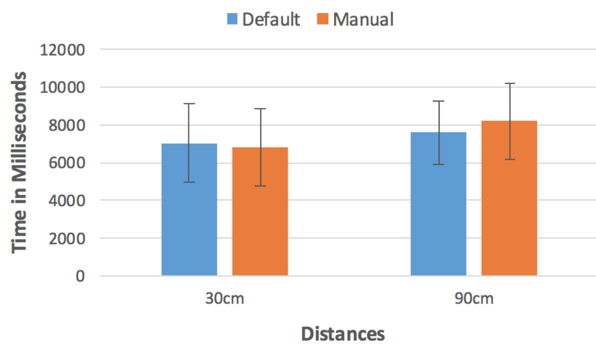


FIGURE 18. Comparison between different distances of hybrid task using default method and manual switch mode.

There was a significant difference between different distances ($F_{1,15} = 13.521$, $p = 0.002$). Participants took more time when the distance was large and less time when the distance was small (see figure 18). There was also a significant difference between different levels of angular accuracies ($F_{1,15} = 13.612$, $p = 0.002$). Participants took more time where angular accuracy demand was higher i.e. 8 degrees and took less time when demand was lower i.e. 16 degrees (see figure 19).

For this task, there were no significant differences between temporal demand ($Z = -1.29$, $p=0.197$), perceived speed ($Z = -1.1897$, $p=0.058$) and perceived accuracy ($Z = -0.796$, $p=0.426$). There were significant differences between mental demand ($Z = -2.684$, $p=0.007$), physical demand ($Z = -2.517$, $p=0.012$), performance ($Z = -2.309$, $p=0.021$), effort($Z = -2.886$, $p=0.004$), frustration ($Z = -2.375$, $p=0.018$), arm fatigue ($Z = -2.632$, $p=0.008$). Physical demand, effort, frustration and arm fatigue were higher for the default method whereas mental demand and performance were higher for ProxyHand/StickHand manual switch mode. (see figure 20). This difference in physical demand, effort, frustration and arm fatigue was obvious because of the increased fatigue levels of participants. However, users felt more mental demand in ProxyHand/StickHand manual switch mode because they had to switch again and again between two modes and it confused them.

F. DISCUSSION

We summarized the questionnaire of all the users where we asked them about the advantages, disadvantages of Default

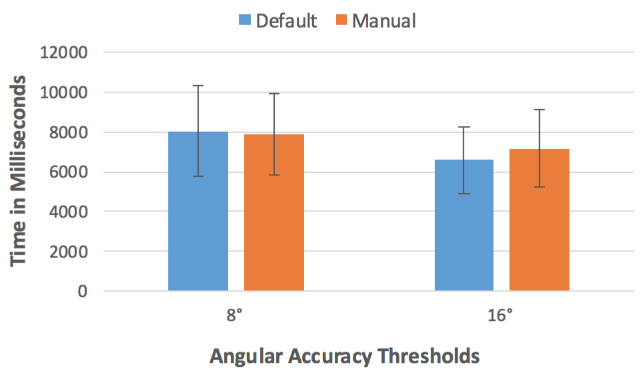


FIGURE 19. Comparison between different angular accuracies of hybrid task using default method and manual switch mode.

Method, ProxyHand/StickHand and their preference. All the participant agreed that the main advantage of the default method is its natural way of interaction and they did not have to learn lots of buttons. According to the participants, arm fatigue was the most common disadvantage. Most participants agreed that ProxyHand had reduced the arm fatigue and it was suitable for and rotation high accuracy tasks. According to the participants, ProxyHand was not suitable for long translations tasks and object manipulation across whole around-body interaction space. All the participants agreed that StickHand was very convenient for long translation tasks as it was efficient as compared to ProxyHand and the default method. StickHand was not suitable for rotation and accuracy tasks. Regarding preference, 13 of the participants preferred the combination of ProxyHand and StickHand whereas, 2 preferred ProxyHand and only 1 participant preferred the default method.

Following are some of the participants' general comments:

> *"I would like to use default method for objects which are too near to body, ProxyHand for rotation and StickHand for long distance translation" - Participant 3.*
>
> *"I would prefer to use mixture all three methods" - Participant 5.*
>
> *"ProxyHand and StickHand definitely reduces arm Fatigue. However, I would like to have simpler method for switching modes as I got confused because of lots of buttons." - Participant 9.*
>
> *"Proxy and StickHand are something I would use for daily life because they have very low fatigue levels in my opinion." - Participant 11.*
>
> *"I prefer to use ProxyHand only as I felt overwhelmed during switching mode from StickHand to ProxyHand or vice versa", Participant 15*

After careful analysis of the objective and subjective feedback of all the participants, we concluded that the combination of ProxyHand/StickHand with manual switch mode was better than using only ProxyHand because StickHand satisfied all the shortcomings of ProxyHand. However, objective feedback showed that users spent a lot of time switching the mode during task completion which wasted a considerable amount of time. According to the subjective feedback of the participants, mental demand was significantly higher than the default method as there were so many controls that they had to remember while performing the task. Participants were overwhelmed because of remembering many controls and felt confused sometimes while performing the task. This showed the trade-off between Physical Demand and Mental Demand i.e. in the default method, there was more physical demand but less mental demand and in our technique, there is less physical demand but more mental demand. This motivated us to design the automatic mode switch between ProxyHand and StickHand so that users do not need to remember the controls and the mental demand problem for this combination of ProxyHand/StickHand gets addressed.

V. AUTOMATIC SWITCH USING MACHINE LEARNING

After the comprehensive user study carried out in the previous section, we came to know that the combination of ProxyHand and StickHand by manual switch reduced the arm fatigue as well as addressed the problems encountered in the user study discussed in section 3. However, subjective feedback of user study conducted in section 4 showed that mental demand for ProxyHand/Stick hand manual switch was higher than default method because participants had to remember many controls while performing tasks. This reduction of arm fatigue was at the cost of increased mental demand. To reduce increased mental demand caused by the combination of ProxyHand/StickHand manual switch, we designed an automatic switch for changing modes using machine learning.

A. DESIGN

There were many options for switching between ProxyHand and StickHand from manual mode to automatic mode. After careful consideration of all the available options, we opted to use machine learning for this particular problem.

To use machine learning, we had to use the concept of uncertain input i.e. initially the system is uncertain that the user is using which hand (ProxyHand or StickHand). But after a specific amount of time, it decides by examining the user's way of using controllers that if they are using ProxyHand or StickHand. For this purpose, we showed ProxyHand and StickHand together in a scene as shown in figure 21. When the users set the offset using a laser pointer and grabbed the object, the virtual hand got split into two hands (ProxyHand and StickHand) and the object also got split into two, one attached with ProxyHand and the other with StickHand. After a specific amount of time, the user's intention was identified i.e. which hand the user was using and that hand was retained whereas the other hand was hidden. Machine learning was used for user's intention classification.

B. TRAINING OF THE MODEL

Correctly predicting the user's intention was a binary classification problem, since there were only two virtual hands, ProxyHand and StickHand. We chose to use logistic

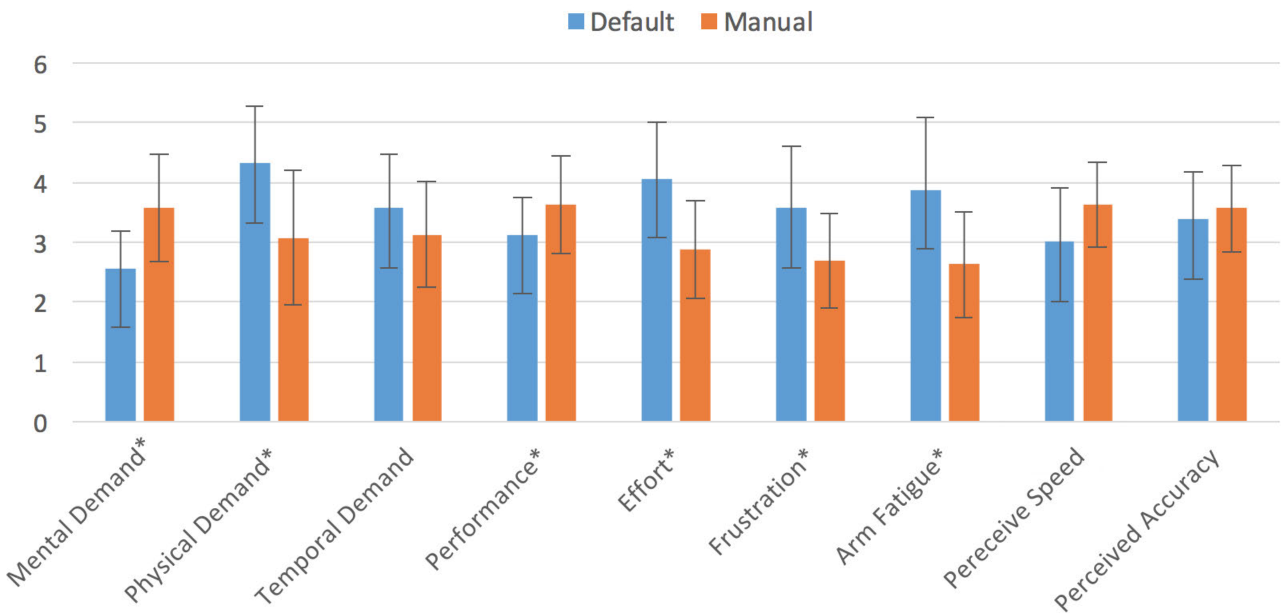


FIGURE 20. Subjective feedback of participants for the hybrid task over both default and manual switch mode technique (from 1 - very low to 5 - very high). Significant factors are denoted by * in the end.

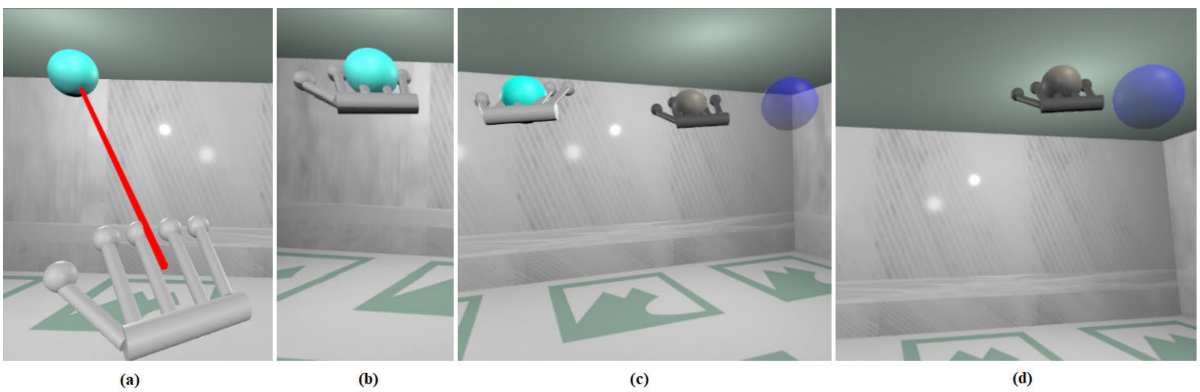


FIGURE 21. (a) User sets the offset using a laser. (b) Users grabs the objects using ProxyHand (c) Virtual hand splits into two i.e. ProxyHand and StickHand. Object is also split into two objects. (d) After a specific amount of time, a decision is made using logistic regression that the user is using ProxyHand or StickHand and the other hand disappears.

regression which is one of the widely used binary classification model in machine learning. To train the model, we used the position, rotation and other data of HMD and controllers from manual switch mode during user study in section 4. The data comprised of 16 participants and each participant did the translation task on average of 1944ms for 45 scenarios, rotation task on 6425ms for 15 scenarios and hybrid task on average of 7515 for 40 scenarios. So the total data we had was of 7751280ms ((1944ms $\times$ 40 + 6425ms $\times$ 15 + 7515ms $\times$ 45) $\times$ 16 participants). Since in current HMDs, one second (1000ms) has 90 frames and 11.1ms roughly equals 1 frame. So we had 7751280/11.1 = 698313 frames data.

The challenging part of this task was to choose the best parameters for feeding into the logistic regression classifier. From the data, we had X, Y, Z positions of controller and HMD and X, Y, Z rotations of controller and HMD of each frame. As uncertain input design was used for automatic switch mode, gaze data was also an important factor. Since there were two hands in the starting scene of automatic switch mode, users would be focusing and seeing the hand which they want to use. Thus, we created two extra variables named ProxyHand-seen and StickHand-seen. If users were using ProxyHand in the user study conducted in section 4, ProxyHand-seen was 1 for the frame in which ProxyHand was in the middle of the sight of the user calculated by the angle of HMD and 0 otherwise. For calculating StickHand-seen if users were using ProxyHand, we calculated the arbitrary position of StickHand manually using equation 2 and the position and rotation of the controller. StickHand-seen was assigned 1 for the frame in which arbitrarily calculated

StickHand was in the middle of the sight of the user calculated by the angle of HMD and 0 otherwise. If users were using StickHand, we calculated StickHand-seen and ProxyHand-Seen by applying the same logic mentioned above in reverse.

Instead of using x, y, z positions and rotations of controllers and HMD as it is, we decided to use only magnitudes of change of position and rotation data of the controller alongside ProxyHand-seen and StickHand-seen. We calculated magnitudes of position and rotation of controller for each frame as follows:

$$\Delta Rotation = \sqrt{(xr - xr')^2 + (yr - yr')^2 + (zr - zr')^2}$$

$$\Delta Position = \sqrt{(xp - xp')^2 + (yp - yp')^2 + (zp - zp')^2}$$

where ΔRotation is the total change of rotations of the controller between current and previous frame, xr, yr and zr are the x, y and z rotations of the controller of current frame whereas xr', yr' and zr' are the x, y and z rotations of the controller of the previous frame. ΔPosition is the total change of positions of the controller between current and previous frame, xp, yp and zp are the x, y and z positions of the controller of the current frame whereas xp', yp', zp' are the x, y and z positions of the controller of the previous frame.

Now, there were 4 variables, ΔRotation, ΔPosition, ProxyHand-seen and StickHand-seen for each frame. As mentioned above, current HMDs have 90 frames per second. We needed to use as few frames as we could so that the result was inferred as soon as possible to ensure a smooth user experience. For this purpose, we tried using 10, 20, 30, 40, 50 and 60 frames. Instead of using all the frames individually, we concatenated the data of all the frames into one single row. Each row was treated as one training example. Each training example had the following parameters: $x1$, $x2$, $x3$, $x4$ and y, where $x1$, $x2$, $x3$, and $x4$ were input parameters ΔRotation, ΔPosition, ProxyHand-seen and StickHand-seen respectively, and y was the output parameter i.e. ProxyHand or StickHand.

After feeding data of 10 frames in Logistic Regression, only 23% accuracy was achieved. For 20, 30, 40, 50 and 60 frames, accuracies of the model were 56%, 92%, 92%, 93% and 93% respectively. The possibility of using 10 and 20 frames was discarded as the accuracy obtained was too low. Since there was no huge difference of accuracies between 30 frames and 60 frames data and the time difference was almost double i.e. for 30 frames around 333.3ms and 60 frames around 666.6ms, we decided to use 30 frames for training the model. So that whenever the user grabs an object, the first 30 frames will be used to decide which hand the user wants to use.

We took 30 rows of ProxyHand data one by one, modified them into one row and labelled them ProxyHand. We did the same procedure for StickHand Data. Since we had a total of 698314 frames, when we modified 30 frames into 1 frame, we got 23277 training examples where 14232 were labelled as ProxyHand and 9045 were labelled as StickHand. So, m = 14232 where y = 1(ProxyHand) and m = 9045 where y = 0(StickHand). For the training model, we used the Scikit-Learn python tool. Firstly, we normalized all the data since ΔRotation had values between 0 and 1350, Δ Position had values between 0 and 1.2 and ProxyHand-seen and StickHand-seen had values between 0 and 30. After normalization of data, it was put into the logistic regression model using the Scikit-Learn. Training data had 18622 input examples chosen randomly whereas the rest 4655 examples were used as test data. The total number of interactions, N were 100000.

TABLE 4. Confusion matrix of test set data after applying Logistic Regression.

	StickHand-Predicted	ProxyHand-Predicted
StickHand-Actual	1537	259
ProxyHand-Actual	113	2746

The confusion table (see table 4) shows all the correctly identified StickHand and ProxyHand labels by model and also shows the false positives and false negatives. The precision of the model was 91.3% and recall was 96%. Accuracy was 92% and F1 score was 0.935 (93.5%). Weights for each parameter which were extracted from the model were $w1 = -5.42$, $w2 = 5.21$, $w3 = 5.299$, $w4 = -5.11$ whereas intercept was −0.03f. When the user grabbed the object, data of the first 30 frames were collected and added together to form 1 row with 4 parameters. We applied weights discussed above to these parameters and intercepts in sigmoid function $(1/(1+e^{-x})$ where $x = x1*w1 + x2*w2 + x3*w3 + x4*w4 + intercept)$. After applying all this, when the probability was above 0.5, we assigned ProxyHand and StickHand otherwise.

VI. USER STUDY 2

The goal of this study was to compare the performance between manual mode switch and automatic mode switch and to observe if automatic mode switch reduced mental demand as compared to manual mode switch.

A. PARTICIPANTS

For this study, we recruited sixteen participants from the campus (10 males and 6 females; aged 21-30 years, Mean= 24.5) who had no prior knowledge or practice of the technique. Their heights were from 162cm to 189m with an average of 174cm. Their arm lengths were from 61cm to 86cm with an average of 72cm. 11 participants were familiar with virtual reality experiences and 5 of them had no or very little virtual reality experience.

B. APPARATUS, DESIGN, AND PROCEDURE

For this user study, we used the same apparatus and tasks which we used in the user study 1 (see section 4). For the translation task, the total number of scenarios were: 2 techniques × 3 distance accuracies × 3 distances × 5 no. of trials = 90. The total number of scenarios in the hybrid task were: 2 techniques × 2 distance accuracies × 2 distances × 2 angular accuracies × 5 number of trials = 80. The same procedure which was adopted in the user study 1 was adopted

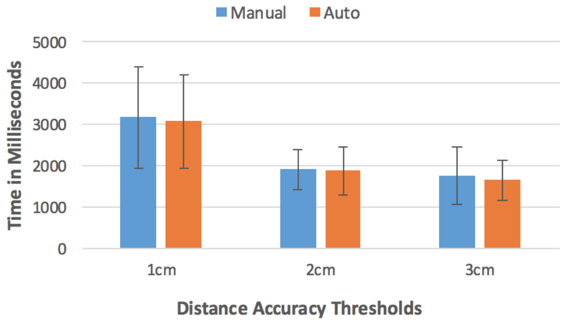


FIGURE 22. **Comparison of Both modes (manual and automatic mode) with respect to distance accuracies in translation task.**

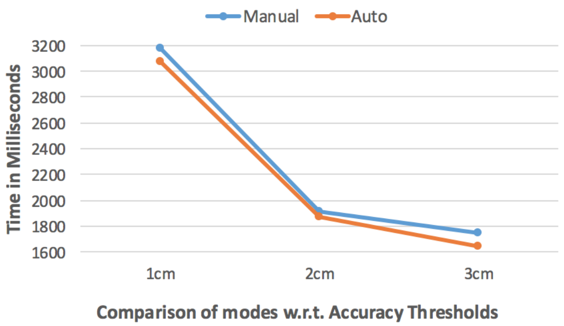


FIGURE 23. **Comparison of Both modes (manual and automatic mode switch) with respect to distance accuracies in translation task.**

here. Since, there was no rotation task in this user study, for one user, it took around 45-60 minutes for completing the whole test.

C. RESULTS

For both of the tasks, translation and hybrid, there was no significant effect of mode on completion time.

1) TRANSLATION TASK

The average time taken by all of the users to finish the translation task with the manual switch was 2277 ms and with the automatic switch, it was 2196 ms as shown in table 5. Repeated Measures ANOVA results showed that the time difference between both modes was not significant ($F_{1,15} = 1.053$, $p = 0.321$). This indicated that the time taken for completing the task with both modes was similar and the performance of both modes was also similar.

There was a significant difference in completion times for different distance accuracies ($F_{2,30} = 51.088$, $p < 0.0001$). Total time taken for each respective level of distance accuracy is given in table 5 and shown in figure 22. This result was consistent with the translation task of both user studies in section 3 and section 4 of this work. The reason for this significant difference was also the same that users spent more time when accuracy demand was high and spent less time when accuracy demand was low. Even though there was no significant difference in completion times between both modes at each respective distance accuracies ($F_{2,30} = 0.103$, $p < .0903$), the meantime of automatic mode for all users was less at each level than manual mode as shown in figure 23. The main reason for this trend can be that users spent a little amount of time on switching the mode manually whereas, in the automatic switch that time was saved.

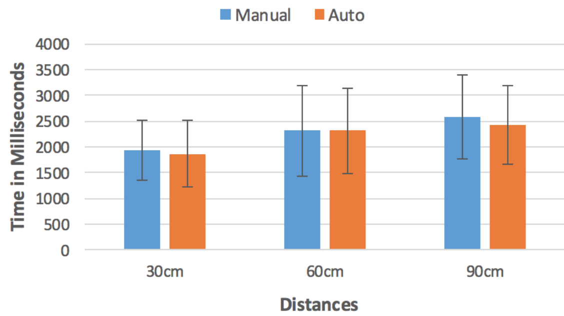


FIGURE 24. **Comparison between different distances of translation task using manual mode and automatic mode.**

There was a significant difference in completion times for different levels of distances ($F_{2,30} = 37.271$, $p < 0.0001$). Total time taken for each respective distance is given in table 5 (see figure 24). This result was consistent with the result of the translation task of the user study in section 4. The reason for the difference was the same that users took more time when the distance was long and less time when the distance was short.

Subjective feedback of participants was analyzed by using the Wilcoxon test (see figure 25). There were no significant differences between temporal Demand ($Z = -0.832$, $p=0.405$), arm Fatigue ($Z = -1.134$, $p=0.257$), perceived Speed ($Z = -1.633$, $p=0.102$) and perceived Accuracy ($Z = -0.577$, $p=0.564$). However, there were significant difference between mental demand ($Z = -3.331$, $p=0.001$), physical demand ($Z = -2.456$, $p = 0.014$), performance ($Z = -2.598$, $p=0.009$), effort ($Z = -2.585$, $p=0.010$) and frustration ($Z = -2.697$, $p=0.007$). Mental demand, physical demand, effort and frustration were higher for manual mode than automatic mode whereas performance was lower in manual mode than automatic mode. These results were the evidence that the mental demand of users was brought down using automatic mode and this affected physical demand, effort, frustration and performance of users as well.

2) ROTATION TASK

The average time taken by all the users to complete the rotation task with manual mode was 6084ms and it was 5928ms with automatic mode as shown in table 6. RM-ANOVA results showed that time difference of both techniques of rotation task was not significant($F_{1,15} = 0.285$, $p = 0.602$)

There was a significant difference between 3 different levels of angular accuracy i.e. 8, 12 and 16 degree angular accuracies ($F_{2,30} = 5.601$ $p = 0.009$) as shown in figure 26. This result was obvious as participants took more time to rotate and adjust when angular accuracy demand was higher and took less time when it was lower.

TABLE 5. Time taken for different scenarios in Translation Task using both modes (Manual and Automatic switch mode) in milliseconds. The 1st column represents the time taken considering all scenarios. Next three columns represent three different levels of Distance Accuracies (DA). Last three columns represent three different levels of Distances (D).

Method	Overall	1cm-DA	2cm-DA	3cm-DA	30cm-D	60cm-D	90cm-D
Manual	2277	3172	1911	1748	1935	2310	2587
Auto	2196	3072	1871	1644	1857	2312	2418

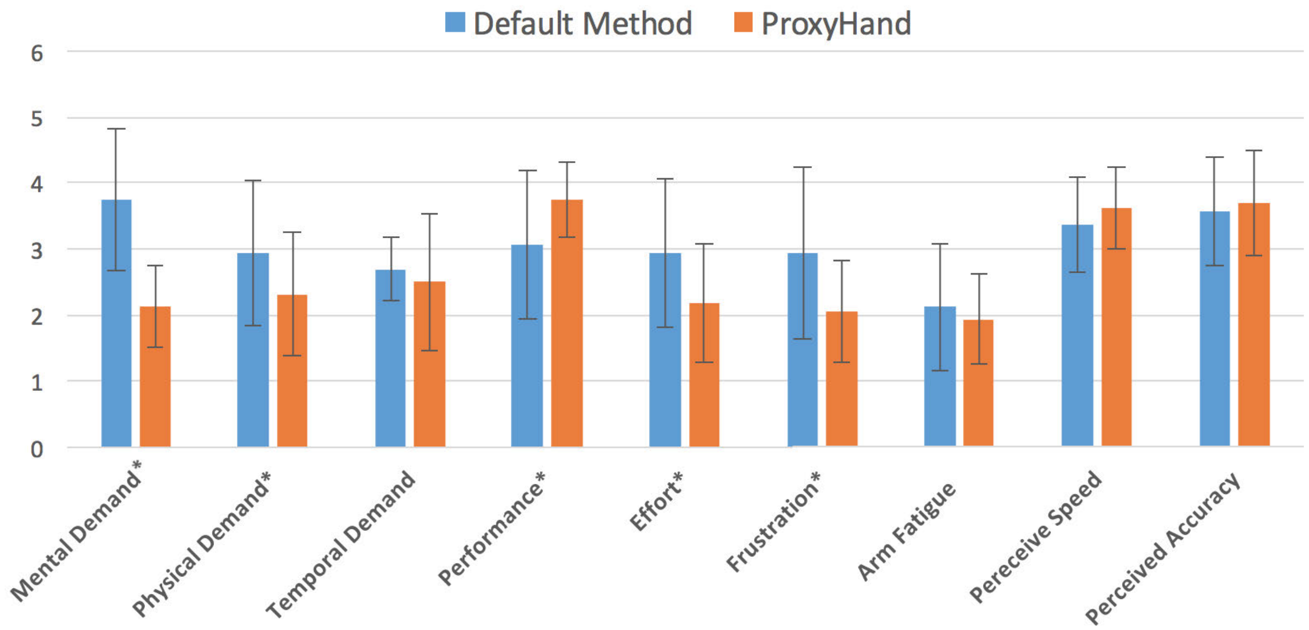


FIGURE 25. Subjective feedback of participants over both manual and automatic mode (from 1 – very low to 5 – very high) in translation task. Significant factors are denoted by * in the end.

TABLE 6. Time taken for different scenarios in the Rotation task using both modes (Manual and Automatic) in milliseconds. The 1st column represents the time taken considering all scenarios. Next three columns represent three different levels of Angular Accuracies (AA).

Method	Overall	8 °-AA	12 °-AA	16 °-AA
Manual	6088	6658	6327	5627
Automatic	5928	5952	5823	4509

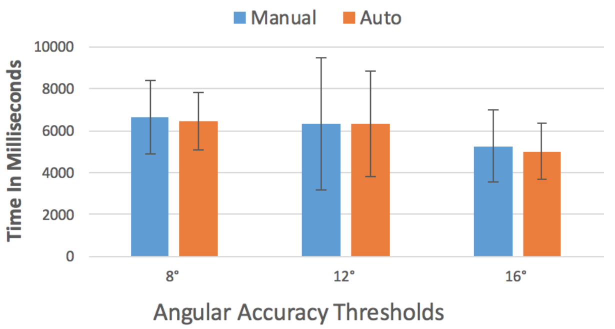


FIGURE 26. Comparison between different angular thresholds of rotation task using manual and automatic mode.

For subjective feedback (see figure 27), there were no significant differences between mental demand (Z= −0.879, p=0.380), physical demand(Z= −0.306, p=0.760), temporal demand (Z= −0.743, p=0.457), performance (Z= −1.134, p=0.257), effort(Z= −1.508, p=0.132), arm fatigue(Z= −0.570, p=0.569), perceived speed (Z= −1.897, p=0.058) and perceived accuracy (Z= −0.632, p=0.527). However, there was a significant difference between frustration (Z= −2.565, p=0.01). Most of these factors did not show significant differences as users did not switch in manual mode as it only required rotation and all the users switched to ProxyHand.

3) HYBRID TASK

For the hybrid task, the mean of the total time taken for each mode is shown in table 7. Repeated Measures ANOVA results showed that the time difference for both the techniques on the hybrid task was not significant ($F_{1,15} = 0.083$, $p = 0.777$).

Different levels of distance accuracies had a significant difference in completion times ($F_{1,15} = 17.073$, $p = 0.001$). The reason was the same that participants took more time when distance accuracy demand was high and took less time when the accuracy demand was low (see figure 28).

Different levels of distances had a significant difference in completion times ($F_{1,15} = 16.447$, $p = 0.001$). Reason was the same that participants took more time when the distance was large and less time when the distance was small (see figure 29).

TABLE 7. Time taken for different scenarios in the hybrid task using both modes (manual and automatic mode) in milliseconds. The 1st column represents the time taken considering all scenarios. Next two columns represent two different distance accuracies (DA). Next two columns represent two different levels of distances (D). Last two columns represent two levels of angular accuracies (AA).

Method	Overall	1cm-DA	3cm-DA	30cm-D	90cm-D	8°-AA	16°-AA
Default	8275	8783	7768	7805	8745	8891	7660
Manual	8376	8830	7922	7806	8946	8982	7770

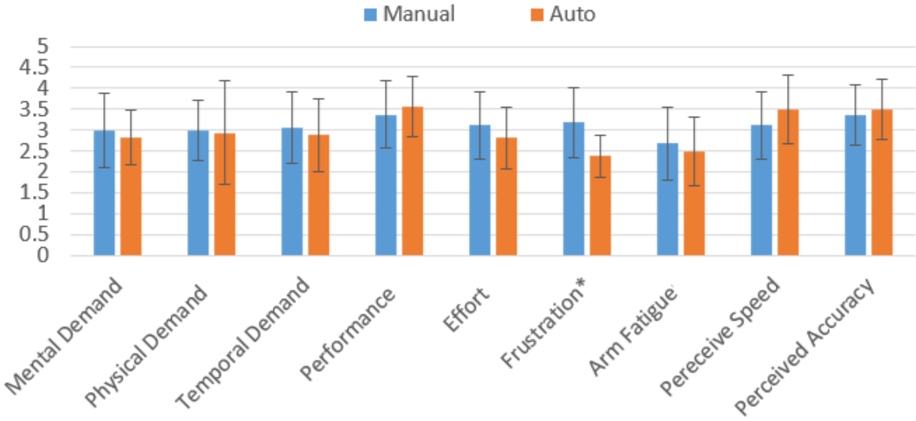


FIGURE 27. Subjective feedback of participants for the rotation task over both manual and automatic modes (from 1 very low to 5 very high). Significant factors are denoted by * in the end.

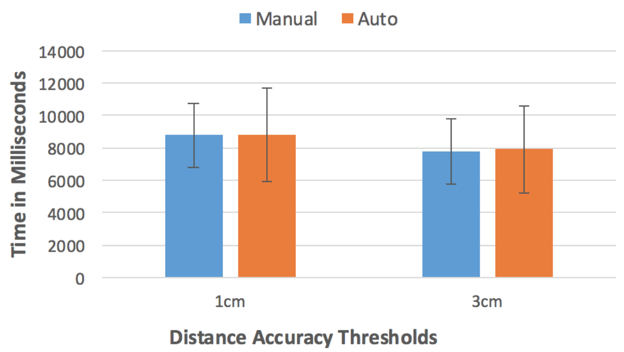


FIGURE 28. Comparison between different distance accuracy thresholds of hybrid task using manual mode and automatic mode.

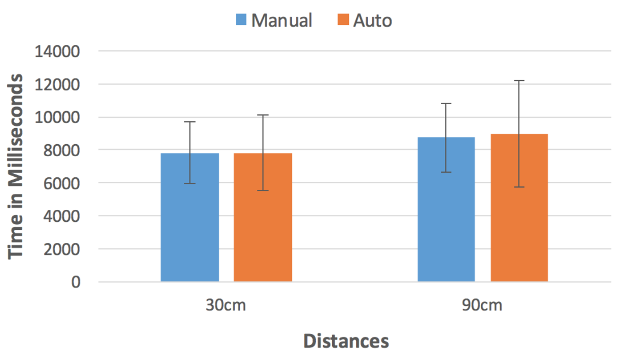


FIGURE 29. Comparison between different distances of the hybrid task using manual mode and automatic mode.

There was also a significant difference between different levels of angular accuracies ($F_{1,15} = 21.760$, $p < 0.0001$). The reason behind it was also the same that

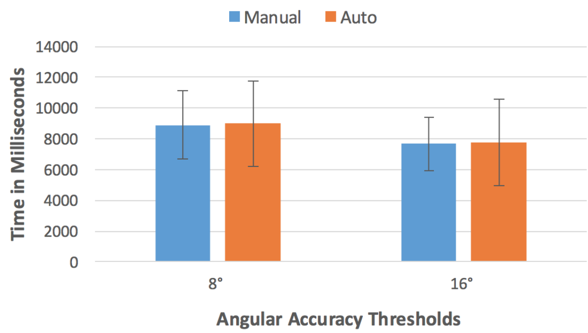


FIGURE 30. Comparison between different angular accuracies of the hybrid task using manual mode and automatic mode.

when angular accuracy demand was high i.e. 8-degrees, participants took more time and when demand was lower i.e. 16 degrees, they took less time to complete the task (see figure 30).

For this task, there were no significant differences between temporal demand ($Z = -1.29$, $p=0.197$), performance ($Z = -0.378$, $p=0.705$), arm fatigue($Z = -1.0$, $p=0.317$) and perceived accuracy ($Z = -0.0$, $p=1$). There were significant differences between mental demand ($Z = -3.573$, $p<0.0001$), physical demand ($Z = -3.051$, $p=0.002$), effort($Z = -2.696$, $p=0.009$), frustration ($Z = -2.586$, $p=0.01$), and perceived speed ($Z = -2.653$, $p=0.008$). Mental demand, physical demand, effort, frustration were higher for manual mode whereas mental demand and perceived was higher for automatic mode. (see figure 31). The reason behind increased mental demand, physical demand, effort and frustration was that users had to remember more controls in manual mode which made them confused.

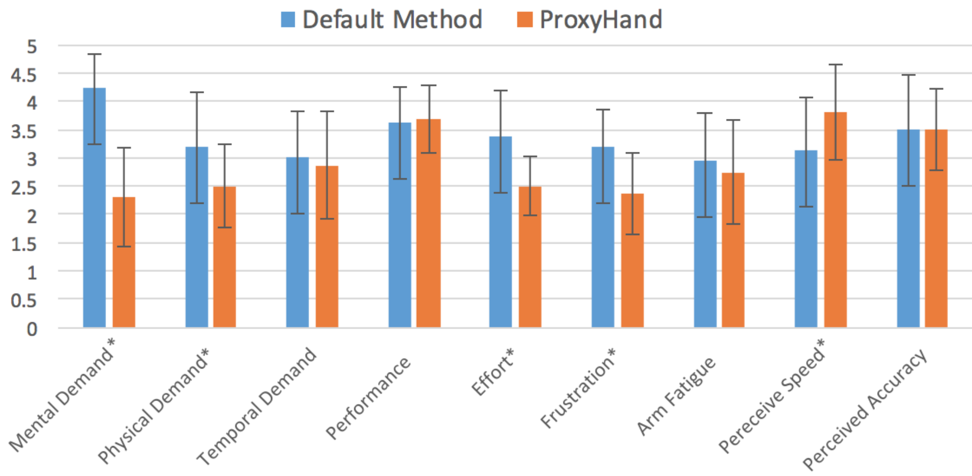


FIGURE 31. Subjective feedback of participants for the hybrid task over both manual mode and automatic mode (from 1 very low to 5 very high). Significant factors are denoted by * in the end.

D. DISCUSSION

At the end of the user study, we summarized the questionnaire which was given at the end of all tasks to participants in which we asked them about the advantages and disadvantages of both manual mode and automatic mode. About the advantages of manual mode, the majority of participants said that the hands were always changed according to their needs as they had full control of the switch. Because in automatic mode, sometimes the wrong mode was selected and it was irritating for them. About disadvantages, they said that they felt overwhelmed because of remembering so many controls. They forgot to switch mode sometimes or pressed the wrong button and got frustrated eventually. According to participants, the advantages of the automatic mode were that it was efficient and they did not have to remember controls as it switched modes automatically. About the disadvantages of automatic mode, they reported that two hands and objects were a distraction at the start. However, they got used to it after some practice. About 15 of them preferred automatic mode whereas only 1 preferred manual mode.

Following are some of their general comments:

> *"Automatic mode is way better than manual mode in a sense that I don't feel frustrated because of remembering so many controls as in case of manual mode." - Participant 2.*

> *"Automatic mode was little distracting in start as two hands confused. But after sometime I got used to of it. I prefer automatic mode over manual mode anyday." - Participant 7.*

> *"I would like to see combination of automatic and manual mode as even though I prefer auto mode, I would like to use manual mode as well." - Participant 8.*

> *"I am truly amazed to see how well automatic mode recognises my intentions. It just recognized wrongly only once for me." - Participant 12.*

Assessment of objective and subjective feedback of users indicated that automatic mode provides equal performance to manual mode in terms of task completion time. However, it significantly reduced mental demand as compared to manual mode and thus also decreased frustration. It was observed that automatic mode misjudged the intentions of participants a few times but still, they were happy to use automatic mode over manual mode. An automatic switch, thus, was the better solution for combining both ProxyHand and StickHand than a manual switch.

VII. CONCLUSION AND FUTURE WORK

The arm fatigue problem, famously known as "Gorilla Arm", is a huge problem for user experience in virtual environments. In this work, users can set their customizable 3D offset. Users can do all tasks while keeping their arms in rest posture (arms vertically down). Now they can even do tasks in such regions while keeping their arms in rest posture where they would have needed to stretch their arms at or above shoulder level using the default interaction mechanism. Users got used to ProxyHand and StickHand by minimal training. This work proves that the combination of ProxyHand and StickHand is an optimal solution to arm fatigue problem in VR. This work considerably reduces the fatigue levels of arms during the interaction. The performance of this work for doing tasks in VR is the same as the performance of the default method. This technique is mostly applicable where users have to interact in VR continuously for long time.

The main objective of this work was to analyze if arm fatigue while interaction in virtual reality can be reduced while maintaining the same performance as compared to the default method. After a series of different user studies and various comparisons between the default method, ProxyHand, StickHand, a combination of ProxyHand and StickHand using a manual switch and using an automatic switch, it can be inferred that this work tremendously decreases arm fatigue levels while providing similar performance to the default method.

The validity of all user studies' designs can be examined by checking the results of these user studies. Results of all tasks were following Fitt's law that is when the distance was large, users took more time to complete the task and took less time when the distance was small. Users took more time when the requirement of distance accuracy was high and less time when it was low. Moreover, Fitt's law was not just holding for linear tasks, but it was also holding on to the angular task i.e. users took more time when more angular accuracy was required and took less time when less angular accuracy was needed.

This work has few areas where we see opportunities for future work. The accuracy of the model used in automatic mode can be increased using different parameters or using more complex models. The two hands shown in automatic mode at the start confuse users a little bit. This can be improved by carefully transitioning the modes by making hands equally transparent at first and then changing their transparencies gradually as probabilities of both hands, ProxyHand and StickHand, changes. One of the positive unexpected outcomes of this work was that it can manipulate objects which were beyond the reach of users. Thus this technique can also be used to manipulate distant objects and in future, works can be done to compare the performance of this work with the work which is specifically designed for manipulating distant objects like Go-Go techniques [8] etc.

So, this work concludes that ProxyHand and StickHand are great alternatives for reducing Arm fatigue in virtual reality while offering equal performance to the default interaction method. This claim is supported by the extensive series of user studies.

REFERENCES

[1] C. Andujar and F. Argelaguet, "Virtual pads: Decoupling motor space and visual space for flexible manipulation of 2D windows within VEs," in *Proc. IEEE Symp. 3D User Interface*, Mar. 2007, pp. 99–106.

[2] F. Argelaguet and C. Andujar, "A survey of 3D object selection techniques for virtual environments," *Comput. Graph.*, vol. 37, no. 3, pp. 121–136, May 2013.

[3] M. Bachynskyi, G. Palmas, A. Oulasvirta, and T. Weinkauf, "Informing the design of novel input methods with muscle coactivation clustering," *ACM Trans. Comput.-Hum. Interact.*, vol. 21, no. 6, p. 30, 2015.

[4] N. I. Badler, C. B. Phillips, and B. L. Webber, *Simulating Humans: Computer Graphics Animation and Control*. London, U.K.: Oxford Univ. Press, 1993.

[5] R. Blanch, Y. Guiard, and M. Beaudouin-Lafon, "Semantic pointing: Improving target acquisition with control-display ratio adaptation," in *Proc. Conf. Hum. Factors Comput. Syst. (CHI)*, 2004, pp. 519–526.

[6] G. Borg, *Borg's Perceived Exertion Pain Scales*. Champaign, IL, USA: Human Kinetics, 1998.

[7] S. Boring, M. Jurmu, and A. Butz, "Scroll, tilt or move it: Using mobile phones to continuously control pointers on large public displays," in *Proc. 21st Annu. Conf. Austral. Comput.-Hum. Interact. Special Interest Group Design: Open 24/7 (OZCHI)*, 2009, pp. 161–168.

[8] D. A. Bowman and L. F. Hodges, "An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments," in *Proc. Symp. Interact. 3D Graph. (SI3D)*, 1997, p. 35.

[9] E. A. Bustamante and R. D. Spain, "Measurement invariance of the Nasa TLX," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 52. Sage, CA, USA: Los Angeles, 2008, pp. 1522–1526.

[10] J. M. Carroll, *Designing Interaction: Psychology at the Human Computer Interface*, vol. 4. Cambridge, U.K.: CUP Archive, 1991.

[11] M. Cifrek, V. Medved, S. Tonković, and S. Ostojić, "Surface EMG based muscle fatigue evaluation in biomechanics," *Clin. Biomech.*, vol. 24, no. 4, pp. 327–340, May 2009.

[12] L. Findlater, K. Moffatt, J. E. Froehlich, M. Malu, and J. Zhang, "Comparing touchscreen and mouse input performance by people with and without upper body motor impairments," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2017, pp. 6056–6061.

[13] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan, "Direct-touch vs. Mouse input for tabletop displays," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2007, pp. 647–656.

[14] E. S. Goh, M. S. Sunar, and A. W. Ismail, "3D object manipulation techniques in handheld mobile augmented reality interface: A review," *IEEE Access*, vol. 7, pp. 40581–40601, 2019.

[15] T. Grossman and R. Balakrishnan, "The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area," in *Proc. SIGCHI Conf. Human Factors Comput. Syst. (CHI)*, 2005, pp. 281–290.

[16] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, "Consumed endurance: A metric to quantify arm fatigue of mid-air interactions," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2014, pp. 1063–1072.

[17] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell, "A survey of design issues in spatial input," in *Proc. 7th Annu. ACM Symp. User Interface Softw. Technol. (UIST)*, 1994, pp. 213–222.

[18] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani, "Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2017, pp. 3328–3339.

[19] P. Kabbash and A. S. W. Buxton, "The 'prince technique': Fitts' law and selection using area cursors," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI)*, New York, NY, USA: ACM Press/Addison-Wesley, 1995, pp. 273–279.

[20] M. Kobayashi and T. Igarashi, "Ninja cursors: Using multiple cursors to assist target acquisition on large screens," in *Proc. 26th Annu. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2008, pp. 949–958.

[21] L. McAtamney and E. Nigel Corlett, "RULA: A survey method for the investigation of work-related upper limb disorders," *Appl. Ergonom.*, vol. 24, no. 2, pp. 91–99, Apr. 1993.

[22] R. A. Montano Murillo, S. Subramanian, and D. Martinez Plasencia, "Erg-O: Ergonomic optimization of immersive virtual environments," in *Proc. 30th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2017, pp. 759–771.

[23] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa, "The go-go interaction technique: Non-linear mapping for direct manipulation in VR," in *Proc. 9th Annu. ACM Symp. User Interface Softw. Technol. (UIST)*, 1996, pp. 79–80.

[24] I. Poupyrev and T. Ichikawa, "Manipulating objects in virtual worlds: Categorization and empirical evaluation of interaction techniques," *J. Vis. Lang. Comput.*, vol. 10, no. 1, pp. 19–35, Feb. 1999.

[25] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Intelligent human-computer interaction based on surface EMG gesture recognition," *IEEE Access*, vol. 7, pp. 61378–61387, 2019.

[26] G. Salvendy, *Handbook of Human Factors and Ergonomics*. Hoboken, NJ, USA: Wiley, 2012.

[27] K. A. Satriadi, B. Ens, M. Cordeil, B. Jenny, T. Czauderna, and W. Willett, "Augmented reality map navigation with freehand gestures," in *Proc. IEEE Conf. Virtual Reality 3D User Interface (VR)*, Mar. 2019, pp. 593–603.

[28] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu, "A handle bar metaphor for virtual object manipulation with mid-air interaction," in *Proc. ACM Annu. Conf. Hum. Factors Comput. Syst. (CHI)*, 2012, pp. 1297–1306.

[29] A. Troiano, F. Naddeo, E. Sosso, G. Camarota, R. Merletti, and L. Mesin, "Assessment of force and fatigue in isometric contractions of the upper trapezius muscle by surface EMG signal and perceived exertion scale," *Gait Posture*, vol. 28, no. 2, pp. 179–186, Aug. 2008.

[30] L. Vlaming, C. Collins, M. Hancock, M. Nacenta, T. Isenberg, and S. Carpendale, "Integrating 2D mouse emulation with 3D manipulation for visualizations on a multi-touch table," in *Proc. ACM Int. Conf. Interact. Tabletops Surf. (ITS)*, 2010, pp. 221–230.

[31] V. Vuibert, W. Stuerzlinger, and J. R. Cooperstock, "Evaluation of docking task performance using mid-air interaction techniques," in *Proc. 3rd ACM Symp. Spatial User Interact.*, Aug. 2015, pp. 44–52.

[32] R. Walter, G. Bailly, and J. Müller, "StrikeAPose: Revealing mid-air gestures on public displays," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2013, pp. 841–850.

[33] C. Yu, Y. Shi, R. Balakrishnan, X. Meng, Y. Suo, M. Fan, and Y. Qin, "The satellite cursor: Achieving MAGIC pointing without gaze tracking using multiple cursors," in *Proc. 23nd Annu. ACM Symp. User Interface Softw. Technol. (UIST)*, 2010, pp. 163–172.

[34] K. J. Zuiderveld, A. H. Koning, and M. A. Viergever, "Acceleration of ray-casting using 3-D distance transforms," *Proc. SPIE*, vol. 1808, pp. 324–336, Sep. 1992.

HASAN IQBAL received the bachelor's degree in software engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2011, and the master's degree in advanced computing from Tsinghua University, China, in 2018. He is currently an Algorithm Engineer at the Research and Development Centre in Shanghai, China. He has working experience of full-time engineer as well as an internee in multiple companies, nationally and internationally. His research interests include machine learning, deep learning, computer vision, and human–computer interaction (AR/VR). He was a recipient of the Rector's Gold Medal for best final year project and thesis at NUST.

SEEMAB LATIF (Senior Member, IEEE) received the Ph.D. degree from The University of Manchester, U.K. She is currently an Assistant Professor and a Researcher with the National University of Sciences and Technology (NUST), Pakistan. Her research interests include artificial intelligence, machine learning, data mining, and NLP. Her professional services include Industry Consultations, Conference Chair, Technical Program Committee Member, and a Reviewer for several international journals and conferences. In the last three years, she has established research collaborations with national and international universities and institutes. She has also secured research grants from the National ICT R&D Grass-Root Initiative, the Higher Education Commission Technology Development Fund, and the U.K. ILM Ideas. She has received the School Best Teacher Award, in 2016, and the University Best Innovator Award, in 2020. She is also the founder of NUST spin off company, Aawaz AI Tech.

YUKANG YAN is currently a Postdoctoral Researcher at the HCI Group, Department of Computer Science and Technology, Tsinghua University. His research area includes human–computer interaction (HCI), in which the focus is on developing multi-modality input techniques, gesture interaction methods, and computational behavior models to improve the user experience in VR/AR, with wearable devices and voice user interfaces. He has developed skills in signal processing, gesture recognition, VR/AR development and user study design, and statistical analysis. His goal is to realize the "perfect" MR experience in the near future, where multi-modality input and feedback are enabled and enhanced, and digital and real objects integrate seamlessly to facilitate information exchange.

CHUN YU received the Ph.D. degree from Tsinghua University, in 2012. He is currently an Associate Professor with Tsinghua University. He is keen to research computational models and algorithms that can facilitate the development of high-efficiency multi-modal user interface, such as mobile phones, large displays, and VR/AR headsets. He has published more than 30 papers in prestigious conferences and journals in human–computer interaction, including CHI, UIST, CSCW, and *IJHCS*. His recent research won the ACM CHI Honorable Mention Award, in 2013, 2015, 2016, and 2017, the ChineseCHI 2015 Best Paper Award, the IEEE UIC 2014 Best Paper Award, and the ACM MobileHCI 2013 Honorable Mention Award.

YUANCHUN SHI received the B.Sc., M.Sc., and Ph.D. degrees from Tsinghua University, China. She joined the faculty of Tsinghua University, in 1993, where she is currently a Changjiang Distinguished Professor with the Department of Computer Science and Technology. She has published more than 100 articles in top conferences and journals (CS Ranking, HCI), including over 30 CHI, UIST, Ubicomp, TOCHI, and *IJHCS* papers over the most recent five years alone. Her research interests include human–computer interaction, ubiquitous computing, and multimedia, which has won many academic awards, including two State Science and Technology Progress Awards, Faculty Awards, and many best paper and honorable mention awards from conferences, such as CHI and MobileHCI. Her recent HCI contributions include user behavior modeling, intelligent interaction techniques for mobile phones, large displays, and VR/AR headsets.