STAG: Smart Tools and Applications in Graphics (2018)
M. Livesu, G. Pintore and A. Signoroni (Editors)

A study on Natural 3D Shape Manipulation in VR

E. Cordeiro[1], F. Giannini[1], M. Monti[1], D. Mendes[2], A. Ferreira[2]

[1]CNR-IMATI, Genova, Italy
[2] 2INESC-ID Lisboa, Portugal

Abstract

Current immersive modeling environments use non-natural tools and interfaces to support traditional shape manipulation operations. In the future, we expect the availability of natural methods of interaction with 3D models in immersive environments to become increasingly important in several industrial applications. In this paper, we present a study conducted on a group of potential users with the aim of verifying if there is a common strategy in gestural and vocal interaction in immersive environments when the objective is modifying a 3D shape model. The results indicate that users adopt different strategies to perform the different tasks but in the execution of a specific activity it is possible to identify a set of similar and recurrent gestures. In general, the gestures made are physically plausible. During the experiment, the vocal interaction was used quite rarely and never to express a command to the system but rather to better specify what the user was doing with gestures.

CCS Concepts

•***Human-centered computing*** *→ Human computer interaction (HCI) → Interaction paradigms→ Virtual reality*; •***Computing methodologies*** *→ Computer graphics→Shape modeling;*

1. Introduction

Due to the ability to offer an immersive experience in virtual environments, Virtual Reality (VR) is being used in many application sectors. In the industrial field, it is mainly employed by large companies for the verification, simulation and maintenance of products. In the past, some attempts have been made to investigate its use in 3D shape modelling, but they were strongly limited primarily by the costs and the restricted capabilities of VR tools. The improvement and availability of low cost tools, together with the capabilities for 3D shape acquisition and printing pave the way to new scenarios, which involve new types of users. Current modeling systems can provide many features for creating and modifying 3D shapes, while ensuring efficiency and a high degree of precision. However, effective use of these tools requires specific user training on the tool. Manipulating and modifying the 3D model of an object in VR can provide a more natural interaction and reduce the cognitive complexity of user interfaces of the traditional display and modeling systems. With the aid of these interaction technologies and the development of suitable shape manipulation techniques, it is very likely that end users can be involved in the design and eventual customization of products. Since these techniques should focus on maintaining a low learning curve, as first-time user might be discouraged of using an over-complicated application. Giving the possibility of changing the shape or other key characteristics of a specific object, gives a new meaning to personalization and customization of products. This can also help disabled people to get more appropriate products, as they need objects with different features to do the same actions as a non-disabled person. Likewise, left handed people require to adjust products for their convenience, as not all products have been designed with this in mind.

Furthermore, these advantages also promote collaborative design approach, which can happen in three ways: consumer-consumer, producer-consumer or consumer-producer. Two consumers can exchange their versions of the products between each another in order to gather suggestions or refine their current products. In a producer to consumer scenario, it allows the acquisition of potential costumers' feedback on a specific product, the verification of user acceptability and eventually its correction for a particular consumer. Finally, in the customer to producer scenario, it can allow the customer to directly design a product to be then produced by a 3D printing service and to verify it in the environment in which it must be placed. To make these scenarios possible to non-experts, the shape manipulations and modifications must be as natural as possible.

In this paper, we firstly review related studies on user interaction analysis and in gestural interaction for shape modelling. Then, we present our experiment to study how people would naturally perform some recurrent shape

DOI: 10.2312/stag.20181296

modification operations when no indications and limitations are imposed. With the appropriate questions established, we present the details of our experiment, the applied methodology and the results of the evaluation. Lastly, identified future research steps and opportunities conclude the paper.

2. Related Work

In this section, we present the related works and experiences for analyzing human hand motions and gestures together with efforts towards immersive 3D shape modelling.

2.1. Studies on User Hand Motions and Gestures understanding

Various studies have been carried out to understand which gestures and movements are performed by humans in achieving specific tasks. To this aim, experiments have been designed involving end users whose behavior has then been analyzed to detect commonalities among individuals. In the following, we focus on the characteristics and more importantly, on results and conclusions of experiments considering shape manipulation tasks.

[CH12] conducted a study to understand human gestures and strategies when performing 3D manipulation tasks from touchscreen input, but some of the key questions are valid when working on other paradigms. The experiment requested a 3D cube manipulation with three points of view for rotations, scaling and translations. They found that most of the observed gestures are continuous and based on physically plausible behaviors. Additionally, several strategies were identified during the experiment; they also noted that a single user uses mainly one of these strategies. [VHR*04] consider that hand motion is a potential input mechanism in computer aided design systems, in particular, in the initial shape design process. They highlighted that integration of such an input mechanism depends on the detection and processing technologies. Their research focused on both the conformity of human motion processing technologies and the optimization of the implementation of their human motion language. It is their understanding that human motion processing technologies could only qualitatively be evaluated for applicability in hand motion based shape conceptualization. In a subsequent study, [VKK*08] analyzed the types of users' gestures used when controlling a media application. They found that the used gesture types demonstrated a pattern, where a position or direction was expressed. They also verified that the use of reverse gestures to signify opposite actions was intuitive. Either participants expressed gestures by hand motions or by hand posture; a clear distinction between these two modes was found for a single person, making recognition possible. Even though the experiment was focused on a media application, participants rarely used computer or DVD player related gestures, suggesting that the gestures identified can be integrated in other applications. It is also observed that the gestures occurred in just two space partitions in front of the participants, and this could allow for a more fine-tuned calibration to the tracking system used. [HGJ*15] explored how a wizard-of-oz experiment could influence users' redefinition of their gestures. They found that users were prone to redefine their gestures when incorrect recognition of previous gestures occurred. They also noted that when one gesture was recognized incorrectly users chose to redefine with similar looking gestures. [CKL*13] also utilize a wizard of-oz approach to define gestures done by children for whole-body interaction. They report that it may be impractical to derive a set of gestures without proper contextual cues. A comprehensive study on user gestures in augmented reality has been conducted in [PCBC13]. They identified forty tasks and a consensus set of forty-four gestures, the most commonly used by the users. The analysis showed that the gestures in the consensus set are, on average, better than those that had been discarded previously, in terms of goodness and ease of execution. Further conclusions have been achieved from the identified gestures, offering important guidelines to the employment of gestures in augmented reality.

2.2. 3D Shape Modification in Virtual and Augmented reality environments

Even with the advances in gesture recognition technologies, defining general gesture grammars that can contain a wide array of gestures that are context-independent has not been achieved yet. In this section, we surveyed previous works on the usage of gesture-based interfaces in the immersive shape modelling context, focusing on the results achieved on interaction evaluation and on the identified open issues.

To achieve 3D immersive modelling superior to desktop modelers, several problems need to be overcome. [DBW*00] state that the use of cordless head and hand tracking are crucial for the unencumbered use of immersive systems. They also identified that the lack of resting positions for the users' arms during the modelling process is one of the primary problems with 3D immersive modelling. Thus, to advance immersive modelling, features like bi-manual interaction, speech recognition, context sensitivity and constraint interactions need to be studied and correctly implemented. [MMS*17b] compared two novel techniques to perform Boolean operations between two 3D objects in virtual reality against each other and a baseline technique. One of the techniques is based on gestures naturally used to manipulate objects, and the other follows a menu-based approach. The use of Myo armbands severely hindered the analysis of the results, and the authors suggest that the use of a more precise tracking system would offer better results. They also describe that having a hands-only representation produced faster performance, when compared to using a full-body representation. [JKW*14] developed AiRSculpt, a wearable augmented reality sculpting system designed to allow users to create and manipulate 3D virtual content with their hands. Results indicate that the system was capable of handling scaling and shape creation tasks, but users struggled when performing rotation tasks. Additionally, the authors referred that there were problems with hand and gesture recognition, and that depth perception during the tests proved difficult for most users. [WS01] describe a sketching system for spline-based free form surfaces on the Responsive Workbench. Using styluses, the system allows drawing curves that can subsequently be connected to create a 3D skeleton of the intended shape. Afterwards, users need to fill the curve network to create the surfaces that constitute the 3D object. They state that the lack of force feedback is one of the problems in immersive modelling applications. However, they also mention that force feedback can constraint the working volume for the hand thus limiting the advantages of VR environments. Surface Drawing [SPS01] also uses the Responsive Workbench as its visualization and interaction medium. Users of this system are equipped with a glove that translates hand motions into geometry. To be able to offer a complete interaction medium, additional operations can be performed by using metaphors of physical tools. A single tong was used for moving objects, and using two tongs simultaneously would scale the object. An eraser tool was also used for removing elements and a magnet tool capable of performing modifications to an already created surface was also employed. The system was tested extensively using informal user studies, artist focused studies and exhibitions. Experienced designers and 3D software users criticized the lack of control, but the overall sentiment was positive. Indeed, the use of tongs was considered very usable as a means of moving the object. The magnet tool on the other hand was considered hard to use. [KRK*06] devised a free-form modelling system using 3D warp brushes. With the standard brushes comprised in the system, it is possible to create new brushes to perform different modelling operations. Having only conducted an informal study, they comment that the stereoscopic vision and 3D input devices are particularly important. Additionally, it is stated that when used on a desktop computer with a mouse and keyboard as input devices, the creative process slows down considerably. [CKF*16] studied two distinct problems. The first is finding common mental features and interaction behaviors when completing 3D modelling and assembly tasks. The second refers to understanding the effect of a virtual hand self-avatar on the user mental models. Three types of tasks were conducted: manipulation, deformation and tool-supported operating tasks in a user experiment. They observe that bi-manual interaction was the most common method during the experiment, but no conclusions were found in regard to the types of gestures used. They also state that the natural gestures for deforming objects can reflect the forces that were exerted in the interaction. Finally, they refer that the hand avatar helped estimate the size of the virtual objects, and that it also aided in planning and visualizing complex process and procedures of certain tool-based tasks. [CFKS16] also conducted a study to determine the efficiency of using a Leap Motion sensor for mid-air interaction in virtual assembly and shape modelling on traditional PC. It also concluded that virtual hand representation had a negative impact on the performance, and that users shifted focus away from the task when the virtual hands were present. With the support gathered from the previous works previously described, a free-hand interaction shape modelling system was created by [CKS16]. It was explored how people perform specific actions to modify and manipulate virtual objects. Participants of the study conducted complimented the learnability and naturalness of the system. The authors also found that their constraint interaction was not widely adopted, participants preferring to use a more natural approach. [KAH*05] created an immersive 3D modelling system capable of recognizing a small set of hand gestures that correspond to specific actions. The actions supported are pause, point, grab, rotate and scale, which can be used to create, manipulate and deform the virtual objects. The system interface is built on top of three components: hand gestures, gizmos as virtual controller objects and textual menus for mode changes. Yet, no user study was conducted to validate if this combination is appropriate.

The works that have been highlighted in this section have tried to understand and develop techniques in various types of interfaces, to arrive at a natural shape modification and manipulation interaction approach. Still, there are open issues that need to be addressed for immersive modelling to become a legitimate alternative to desktop modelers. Regardless the efforts made to understand how humans interact with objects, a natural interface capable of offering CAD-like modelling operations [BCP*10] has not been achieved yet. In addition, despite the advances in gesture recognition technology, human gestures are not precise enough, and some constraints are currently used to offset this issue. However, our final goal is not the creation of a gestural interface for high precision shape modeling functionalities, but rather the creation of an immersive environment in which an untrained user can manipulate and modify the shape of a conceptual model of a product through a manual and vocal interaction in the most natural way possible.

3. Experiment description

To study how humans would naturally interact in a VR scenario while manipulating 3D objects and modifying their shapes, we conducted a user experiment to attain valid, unbiased and minimally constrained results. With this experiment, we aim to study how users would modify and manipulate three-dimensional objects in a virtual reality environment through hand motions and voice commands without any constraints and without using any additional device, i.e. no pointer or controller. Understanding how users would naturally interact with virtual objects is fundamental, and its results can be then applied to a multitude of areas where natural interfaces are required. As described in detail in section 3.2.1, we are only interested in the manipulation of the 3D objects and not in their selection. The selection of objects in virtual environments has already been studied in several domains [AA13]; [MS*17a]; [BOW05]; [PBW*96]; [VGC09].

3.1. Task Description

Participants were asked to complete a set of tasks, which consisted of five different modifications of an object in the virtual environment. We chose these tasks as they represent a subset of operations particularly useful for modifying existing shapes. Some of them have been also applied in [CKF*16]; this permits to verify their experiment conclusions in another context. Table 1 shows the tasks chosen for the user experiment. Each row represents a different task, indicated in the first column; the central column depicts the presented starting object and the goal object is shown in the right column. Figure 1 further illustrates how the two objects were displayed during the experiment using different colours. The first task consisted of positioning a cylinder inside the cylinder-shaped hole of a box, also known as a docking task. It is representative of the modifications of an object requiring the use of an additional one, which has to be correctly positioned on it. In this task, one of the objects is required to be rotated and translated to achieve the correct place. In the second task, users were asked to perform a twisting deformation on the object. The third task required the execution of a bending deformation on the presented object. In the fourth task, the users are asked to change the scale of the object. These three tasks refer to modifications that change the whole object. In the fifth and final task, we asked users to perform a triangle-shaped local depression deformation on the grey cube. This task refers to the insertion of local features on an object. We selected a triangular-shaped feature because it is enough simple to be simulated by the participants and at the same time it does not suggest a too specific gesture thus leaving room to variety of gestures. We did not impose a time limit for the tasks, as we did not want to pressure the participants in performing the actions quickly.

3.2. Methodology

Each user evaluation session followed the same methodology and lasted approximately 30 minutes. The experiment began with a short introduction on the aim of the test and on what was expected from the participant. Each person was then asked to fill out a profiling questionnaire. Afterward, a script was shown with the tasks and the objectives of the experiment that the participant would perform. Participants were given time to adjust themselves to the virtual environment and to the virtual representation of their hands. During this familiarisation time, the users could also practice the selection method in a teapot object (Figure 4). Subsequently, participants were instructed to perform the five tasks described in the above section using any gesture or voice command they preferred. To avoid biased results, the tasks were presented to the user following random order, so that all permutations were attained.

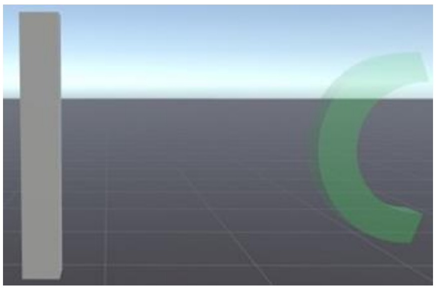


Figure 1: *Example of a starting and target object shown side by side.*

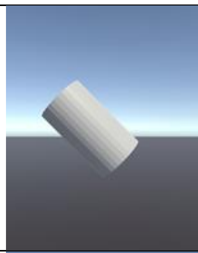
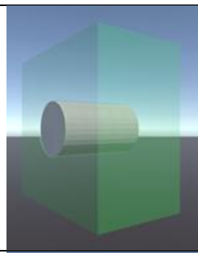
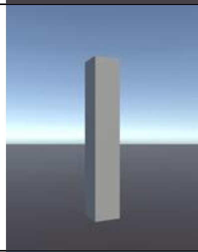
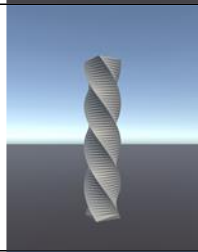
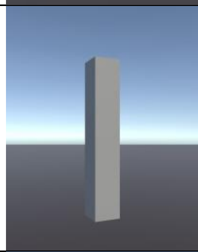
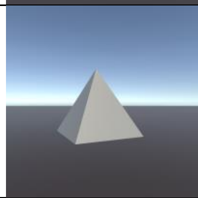
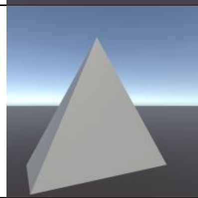
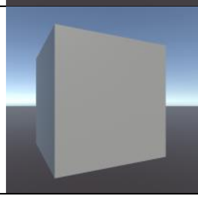
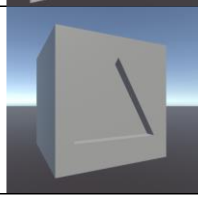
Task	Initial State	Final State
Docking		
Twisting		
Bending		
Scaling		
Local Deformation		

Table 1: *Starting and final configurations of the objects used in the user experiment.*

3.2.1 Wizard of Oz Testing

As stated previously, we conducted a Wizard of Oz user experiment. We chose this approach, as we were interested in understanding how the human motions are performed in a scenario without almost no constraints. In this experiment the user actions are performed by a human operator from behind the scenes, without the user being aware of this. To achieve this we created a set of objects that correspond to different steps of the manipulation and modelling of the objects in each task (Figures 2). During each of the tasks, the operator was able to see on a screen what the user was seeing in head-mounted display, interpreted the user actions and simulated the results of the hand motions and voice commands, making the corresponding changes to the object. To simplify the interaction process we chose to have the operator select the objects for the user. Users were told to place one of their hands inside the object for a brief duration (1-2 seconds) to select the object, and we allowed direct manipulation of the object that was attached to the hand. Resting their hands inside an already selected object would yield the inverse result, a de-selection. The wait period gave the operator time to understand if the user wanted to perform a selection or de-selection, without signalling to the user that he was not actually doing the action.. Since the objective in this study was to understand the user's behavior adopted to perform the assigned task, leaving her/him totally free to choose which gestures and language to use, no wrong behavior was foreseen, i.e. any adopted strategy produced the expected result.

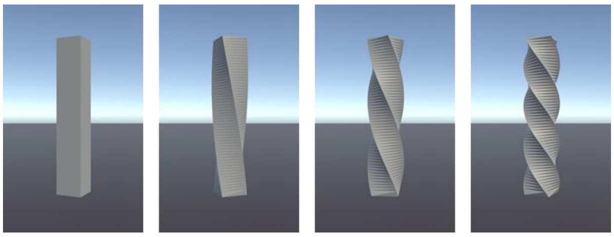


Figure 2: *The different phases of one object during a task.*

3.3 Prototype

To validate our system correctly, we developed a prototype to perform a user evaluation. In this section, we will describe both the hardware components that were involved, and the characteristics of the virtual environment used.

3.3.1. Hardware

The experiment was carried out in our laboratory, a controlled environment, and using the setup depicted in Figure 3. In the figure, the operator is sitting in front of the desktop computer and the participant is wearing the head-mounted display used in the experiment. The fundamental purpose was to echo the user hand motions in the VR environment (i.e. the hand avatars) and to provide a pleasant visualization experience to the participants. To retrieve the pertinent information from the user hand motions we used a Leap Motion sensor. It features two cameras and three infrared LEDs, which are used to track infrared light. For the visualization component, an HTC VIVE head-mounted display was used, in combination with two VIVE cameras to detect the headset in space. This headset is capable of tracking the orientation of the user head using three degrees of freedom. This setting was selected also for its limited cost, which is a strong constraint to effectively achieve a widespread use of VR technology in non-specialised contexts. A standard point-and-shoot video camera was used to record the participants for the later analysis. The video recorded data allowed us to analyze the hand and voice interaction, in order to obtain meaningful results. Additionally, the operator of the tests used a Desktop PC with a common keyboard and mouse input to perform the necessary actions during the user tests.

Figure 3: *The experiment environment*

3.3.2. Virtual Environment

The prototype used in the user tests was developed using the Unity 3D engine. Virtual Reality integration with this engine is transparent and users may enjoy 360° movement of their head and body. The virtual environments was composed of a dark floor and a dark background, the objects chosen for each task and the virtual hands (Figure 4). The dark floor and background were chosen to contrast with the objects, while also keeping the environment uniform and non-distracting. The objects for each task were displayed in front of the user, one in grey that the user could select and manipulate and one in green representing the objective of that task (see Figure 1). To keep the environment as clear as possible, and to minimize distractions, we chose to have only a virtual representation of the user hands instead of a full-body avatar. As expressed in previous research [KTY97] [OSA08], hand-representation is not related to manipulation efficiency, it is only necessary to provide feedback on what the user is doing. To that effect, we used standard hand models included in the Leap Motion asset package.

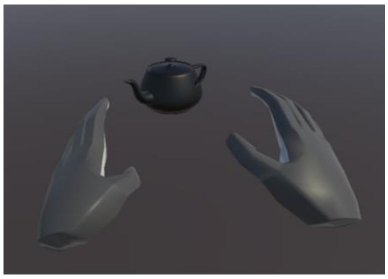


Figure 4: *Virtual environment used in the user experience*

3.4 Participants

The experiment involved 21 participants, 13 being male and 8 females, with ages evenly distributed between 18 and 60. The strong majority held at least a Bachelor degree (90.4%). It is worthy of noting that most of our participants were fellow colleagues with a background in Computer Science or Applied Mathematics, some of which possessed strong knowledge of shape representation and processing, and some had experience of shape modelling tools too. When inquired about their experience with Virtual Reality, 95.2% mentioned that they tried it at most once. This was also the case for Gesture

Detection systems (Leap Motion) usage, as 95.3% responded in the same way. This limited experience in Virtual Reality guarantees that they were not biased in their behaviour by their previous practice.

4. Experiment Results

In this section, we present the results from the user experiment. The data used to obtain these results was gathered via the video recordings, the post-experiment interviews and the profile questionnaire.

4.1 Data Analysis

To classify the interaction mode employed by the participants of our experiment, we followed the concepts established by [GUI87], where human hand gestures can be differentiated in three ways: uni-manual, bi-manual symmetric and bi-manual asymmetric. Uni-manual interaction refers to the use of a single hand to perform a gesture such as drawing, for instance. In a bi-manual symmetric method, there is a use of both hands executing the same motion in parallel, such as rowing a boat. Lastly, bi-manual asymmetric interactions are those in which both hands participate in the action but with different motions, as for example, one hand is holding the object and the other is modifying it. We adopted this taxonomy as we were concerned with identifying what type of interaction was most frequent when performing our tasks.

Figure 5 shows a graph illustrating the different interaction methods used during the experiment for each task. Considering that participants often employed more than one interaction method to complete a single task, we decided to separate the respective modes. For example, if a participant completed one task using both a uni-manual approach and voice, we will attribute a 0.5 value to each of those methods. In the Docking task 100% of participants used an uni-manual approach. For the Twisting task we found that the majority (71.4%) used a symmetrical approach, 19% used only one hand to perform the task and 9.5% employed an asymmetrical style. In the Bending task we found that 57.1% preferred the symmetrical mode, 28.6% chose to use a uni-manual approach and 4.8% adopted an asymmetrical method. Additionally, 2 participants used voice commands to complete this task. We found similar results when analyzing the Scaling task, as 66.7% of participants also chose the symmetrical mode, 38.1% employed a uni-manual mode, 14.3% adopted the symmetrical approach and one participant resorted to voice commands in this task. In the Local Deformation task, the data show that the great majority (81%) adopted the uni-manual mode, while 14.3% chose to use a symmetrical approach.

We also recover data regarding the use of voice interaction during the experiment, which are depicted in Table 2. For the Docking and Twisting task, no participants used voice to express any command or comment. In the Bending task, we found two participants that resorted to voice interaction. One chose to use only voice to complete the task, while the other performed a gesture in combination with a description of what she was performing. In the Scaling task, only a single participant used voice to achieve the required result. Two participants chose to use voice during the Local Deformation, although always accompanying a gestural interaction.

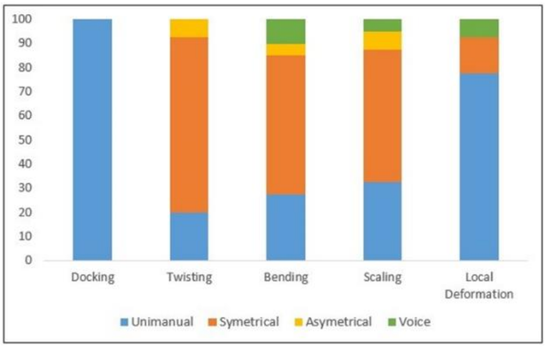


Figure 5: *The percentage of the adopted interaction behaviour in the different tasks*

Task	Hand Motions (%)	Voice (%)
Docking	100,0	0,0
Twisting	100,0	0,0
Bending	92,5	7,5
Scaling	95,0	5,0
Local Deformation	92,5	7,5

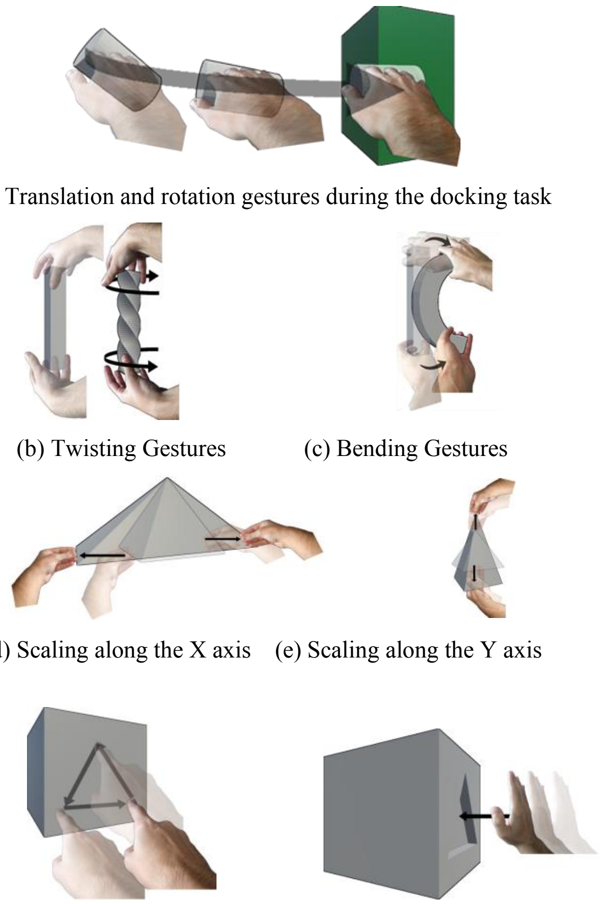
Table 2: *Comparison between the use of hand motions and voice interaction*

4.2 User Strategy

Besides identifying the gestures, which were used during the tests, we are also interested in determining the type of strategy used by the users when completing our tasks. The gestures used in the tasks have been pictured in Figure 6. In the representation of the users' hands, we chose to increase the opacity of the hand to show the progression of the users' motion. As previously pointed out, we found a meaningful preference for a uni-manual approach, which is directly related to the selection method that was chosen for the user tests, as described in section 3.2.1. We observed that after selecting the cylinder, participants naturally moved their hand towards the final position, in order to place it correctly in regard with its proper rotation and position (Figure 6a). We also noticed that the hand used to select the object was not necessarily the dominant one, but the one that would complete the task with a single motion. When reviewing the Twist and Bend tasks we found that most users preferred to use physically plausible movements to accomplish the goal (Figure 6b and 6c). In one instance the participant resorted to the use of voice and drawing a "C" shape. It is also worth noting that there were occasions where participants used only one continuous motion, while others performed quick repetitions of the same motion.

In the Scale task, we also identified a more frequent used interaction method, exemplified in Figure 6d and 6e. This task involved two distinct transformations in different axes. To this aim, nineteen out of twenty participants performed two distinct actions for each axes of transformation, and the remaining participant chose to use voice to finish this task.

When examining users actions in the Local Deformation task we identified that the most frequent gesture - used by sixteen out of the twenty participants - was drawing a triangle on a face of the object (Figure 6f). This action was used either alone, or with a push gesture, indicating a desire to push in the selected region. Indeed, the region selection was the most interesting observation during this task (Figure 6g). Out of all twenty one participants, only four did not perform a region selection. Of these four, two resorted pushing in with their hand directly on the object, and trying to position their hand in accordance to the shape. The remaining two participants used a digging/carving motion to perform this task, alluding to their understanding that the object was made of a material that could be dug through.



(a) Translation and rotation gestures during the docking task

(b) Twisting Gestures (c) Bending Gestures

(d) Scaling along the X axis (e) Scaling along the Y axis

(f) Region specification Gesture (g) Pushing Gesture.

Figure 6: *The most common gestures identified during the user experiment.*

4.3 Discussion

As previously mentioned, we used the [GG87] approach to classify the hand motions performed during our experiment. It is clear from the results that different tasks elicit different ways to interact with the objects. We found that a uni-manual method was preferred in the Docking and Local Deformations tasks. It is important to point out that since in the Docking task it was required that user performs a selection, the chosen selection method had a direct influence on how the users performed this task. On the other hand, in the remaining tasks we observed a greater use of a bi-manual approach. Since these tasks represented a more accurate representation of real world hand motions, they were much more familiar to the participants and elicited behaviors that closely mimicked those used for real objects. Additionally, the use of symmetrical hand motions was much more prevalent than asymmetric, although this can be attributed to the nature of the tasks itself.

To better understand the obtained results, due to the similar nature of [CH12] work, we answer some key questions identified in their work:

1. Is there a common behavior to manipulate and modify 3D objects?
2. Are user gestures based on physically plausible movements, or on a gesture language they create?
3. Do users always use the same strategy?
4. Can we gather a recurrent set of gestures?

From the data gathered, we identified that the majority of participants chose actions based on physically plausible behaviors (Q1). It is clear that users employ different strategies, but the use of physically plausible movements are the most prevalent (Q2). When a strategy is chosen, it persists across all tasks (Q3).

In addition, we identified a clear set of gestures across the majority of participants (Q4).

We observed that the strategies used in the experiments remained consistent when considering the tasks individually. This, coupled with the prevalent use of physically plausible movements indicates that users prefer to employ as natural as possible metaphors to interact with the virtual objects. Although, it was also clear that some users felt confused when trying to understand what were the possibilities and limitations of the system. In one instance, after re-affirming that the user could do the same as they would in the real world, they took immediate action and the goal was achieved

It is also important to understand voice interaction during the experiment, or rather the absence of it. Only six users resorted to using voice commands, in three out of the five tasks, as shown before in Table 2. We found that voice commands were used in two ways: to determine what the system was capable of, and as a means to describe what they wanted to do or were doing with their hands. In the first case, the approach was to use command-like words or sentences (”bend right” in the case of the Bend task) to perform the action. On the other case, users simply wanted to explain what they were doing, as a helping mechanism, e.g. ’’Now I draw a triangle’’ (while performing the corresponding motion). With this in mind, we argue that without imposing directly the use of voice the majority of users will prefer the use of gestures.

Important information regarding the user-experience was obtained in the post-experiment interview and while observing the participants throughout the experiment. Given that most participants had never experimented with virtual reality, we found that some had difficulties understanding how they could interact with the virtual objects. A user commented that the perception that the object is not real is strange at the start. Having said this, the great majority of users had no difficulty interacting with the objects, and praised the visual feedback that was given to their gestures. It is important to note that no participant realized that their actions were being reproduced by the operator, and most claimed that their gestures worked like in the real world. We also observed that participants did not change the position or the rotation of the objects during the tasks (excluding the Docking task) to perform them more easily, instead, they chose to move themselves or tilt their heads in order to check the correctness of the operation. This last consideration indicates that immersive virtual reality can definitively be preferred to flat displays.

Overall, we retrieved important information that will be essential in continuing research into immersive shape modelling..Above all, the prevalent use of physically plausible gestures indicates that it should be possible to automatically identify similar gestures, independent of the user cultural background, which can be associated to a specific task. This can pave the way to the creation of natural and understandable shape modification actions in immersive environment, very easy to remember, which is one of the key aspects in designing user interfaces as highlighted by Norton in [Nor10].

It should be emphasized that there is a problem of precision not only in the acquisition of the signal but also deriving from the difficulty of the user to perform accurate gestures on a virtual shape without having a physical feedback; therefore mid-air gestures can be valid for rough changes in the study of alternative shape solutions. The availability of tangible interaction devices capable of providing the sense of touch would greatly improve the ease and quality of the operation.

5. Conclusions

Despite the recent technological advances in virtual reality, there is not an appropriate natural interface for immersive shape modification. In this paper, we presented a study that advances the understanding of hand and voice interaction for shape modifications in mid-air within immersive environments. A user experiment was conducted using a Wizard of Oz approach as a way to gather the natural user interaction when performing shape modification tasks. The results indicate that the strategy used by the participants was distinct from task to task, and that physically plausible hand motions were the most used. We found that region selection was the primary way of interaction when performing a local deformation task. In addition, the use of one or two hands to perform shape modifications is task dependent. In our experiment, voice interaction was very rarely used, and when it occurred it served as a helping mechanism and not to express a command to the system.

5.1 Future Work

Following the results obtained, we intend to continue researching natural interfaces for 3D shape modification in immersive environments. To continue the exploration of this research area with the aim to build fully natural interfaces for 3D modelling, the next iteration of our work will focus on shape modification operations that will be actually executed by user in real time, instead of using a wizard of oz approach. In addition, we are interested in further investigating other shape modification tasks and how speech can be incorporated. The communication with the system has deep implications on how it is used, and the most common use currently is to express an already known command.

Acknowledgments

The authors would like to express their gratitude to all the participants of the user study, for the time granted and for the feedback provided too. A special thank goes to Martina Paccini and Luigi Rocca for their help and valuable input.

References

[AA13] Argelaguet F., Andujar C.: A survey of 3D object selection techniques for virtual environments. Computers & Graphics (2013);37(3):121–36.

[BCP*10] Bourdot P., Convard T., Picon F., Ammi M., Touraine D., Vézien J.M.: VR–CAD integration: Multimodal immersive interaction and advanced haptic paradigms for implicit edition of cad models. Computer-Aided Design (2010);42(5):445–61.

[Bow05] Bowman D.A.: Design and evaluation of 3D multiple object selection techniques. In: Virginia Polytechnic Institute and State University. Citeseer; 2005. .

[CH12] Cohé A., Hachet M.: Understanding user gestures for manipulating 3D objects from touchscreen inputs. In: proceedings of Graphics Interface (2012) Canadian Information Processing Society; 2012. pp. 157–64.

[CKLP13] Connell S., Kuo P.Y., Liu L., Piper A.M.: A wizard-of-oz elicitation study examining child-defined gestures with a whole-body interface. In: Proceedings of the 12th International Conference on Interaction Design and Children. ACM; (2013) pp. 277–80.

[CFKS16] Cui J., Fellner D.W., Kuijper A., Sourin A.: Mid-air gestures for virtual modeling with leap motion. In: International Conference on Distributed, Ambient, and Pervasive Interactions. Springer; (2016). pp. 221–30.

[CFKS16] Cui J., Kuijper A., Fellner D.W., Sourin A.: Understanding people's mental models of mid-air interaction for virtual assembly and shape modeling. In: Proceedings of the 29th International Conference on Computer Animation and Social Agents. ACM; (2016) pp. 139–46.

[CKS16] Cui J., Kuijper A., Sourin A.: Exploration of natural free-hand interaction for shape modeling using leap motion controller. In: Cyberworlds (CW), (2016) International Conference on. IEEE; (2016) pp. 41–8. 38

[DBW*] Deisinger J., Blach R., Wesche G., Breining R., Simon A.: Towards immersive modeling-challenges and recommendations: A workshop analyzing the needs of designers. In: EGVE. 2000. (2000) pp. 145–56.

[Gui87] Guiard Y. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. Journal of motor behavior (1987);19(4):486–517.

[HGJ*15] Henschke M., Gedeon T., Jones R.: Touchless gestural interaction with wizard-of-oz: Analysing user behaviour. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction. ACM; 2015. (2015) pp. 207–11.

[jkw*14] Jang S.A., Kim H., Woo W., Wakefield G.: Airsculpt: A wearable augmented reality 3d sculpting system. In: International Conference on Distributed, Ambient, and Pervasive Interactions. Springer; 2014.(2014) pp. 130–41. 13

[KRK*06] Kil Y.J., Renzulli P., Kreylos O., Hamann B., Monno G., Staadt O.G.: 3D warp brush modeling. Computers & Graphics (2006);30(4):610–8. 17

[KAH*05] Kim H., Albuquerque G., Havemann S., Fellner D.W.: Tangible 3D: Hand gesture interaction for immersive 3D modeling. In: IPT/EGVE. 2005. (2005) pp. 191–9.

[KTY97] Kiyokawa K., Takemura H., Yokoya N.: Manipulation aid for two-handed 3-d designing within a shared virtual environment. In: HCI (2). 1997. p. 937–40. 24

[MMS*17a] Mendes D., Medeiros D., Sousa M., Cordeiro E., Ferreira A., Jorge J.A.: Design and evaluation of a novel out-of-reach selection technique for VR using iterative refinement. Computers & Graphics (2017);67:95–102.

[MMS*17b] Mendes D., Medeiros D., Sousa M., Ferreira R., Raposo A., Ferreira A., Jorge J.: Mid-air modeling with boolean operations in VR. In: 3D User Interfaces (3DUI), 2017 IEEE Symposium on. IEEE; (2017). pp. 154–7.

[Nor10] Norman. D.A.: Natural user interfaces are not natural. Interactions 17, 3 (May 2010), pp. 6-10.

[Osa08] Osawa N.: Two-handed and one-handed techniques for precise and efficient manipulation in immersive virtual environments. Advances in Visual Computing (2008):987–97.

[PCB*13] Piumsomboon T., Clark A., Billinghurst Cockburn A.: User-defined gestures for augmented reality. In: CHI'13 Extended Abstracts on Human Factors in Computing Systems. ACM; (2013) pp. 955–60.

[PBW*96] Poupyrev I., Billinghurst M., Weghorst S., Ichikawa T.; The go-go interaction technique: non-linear mapping for direct manipulation in vr. In: Proceedings of the 9th annual ACM symposium on User interface software and technology. ACM; (1996) pp. 79–80.

[SPS01] Schkolne S., Pruett M., Schroder P.: Surface drawing: creating organic 3D shapes with the hand and tangible tools. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM; 2001 (2001) pp. 261–268.

[VGC09] Vanacken L., Grossman T., Coninx K.: Multimodal selection techniques for dense and occluded 3D virtual environments. International Journal of Human-Computer Studies (2009) 67(3):237–55.

[VHR*04] Varga E., Horvath I., Rusak Z., De Smit B., Broek H.: Survey and investigation of hand motion processing technologies for compliance with shape conceptualization. In: Proceedings of DETC. 2004 volume 4; (2004) pp. 1–14.