



Universiteit  
Leiden  
The Netherlands

## On the efficiency of a VR hand gesture based interface for 3D object manipulations in conceptual design

Alkemade, R.; Verbeek, F.J.; Lukosch, S.

### Citation

Alkemade, R., Verbeek, F. J., & Lukosch, S. (2017). On the efficiency of a VR hand gesture based interface for 3D object manipulations in conceptual design. *International Journal Of Human-Computer Interaction*, 33(11), 882-901. doi:10.1080/10447318.2017.1296074

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/3202542>

**Note:** To cite this publication please use the final published version (if applicable).



## On the Efficiency of a VR Hand Gesture-Based Interface for 3D Object Manipulations in Conceptual Design

Remi Alkemade, Fons J. Verbeek & Stephan G. Lukosch

To cite this article: Remi Alkemade, Fons J. Verbeek & Stephan G. Lukosch (2017) On the Efficiency of a VR Hand Gesture-Based Interface for 3D Object Manipulations in Conceptual Design, International Journal of Human-Computer Interaction, 33:11, 882-901, DOI: [10.1080/10447318.2017.1296074](https://doi.org/10.1080/10447318.2017.1296074)

To link to this article: <https://doi.org/10.1080/10447318.2017.1296074>



Published online: 31 Mar 2017.



Submit your article to this journal [↗](#)



Article views: 972



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 21 View citing articles [↗](#)



# On the Efficiency of a VR Hand Gesture-Based Interface for 3D Object Manipulations in Conceptual Design

Remi Alkemade<sup>a</sup>, Fons J. Verbeek<sup>b</sup>, and Stephan G. Lukosch<sup>c</sup>

<sup>a</sup>Media Technology Programme, Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands; <sup>b</sup>Imagery & Media, Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands; <sup>c</sup>Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

## ABSTRACT

In the early stages of 3D design, sketches are used to quickly conceptualize ideas and gain insight into problems and possible solutions. Computer-aided design tools are widely used for 3D modeling and design, but their required precision and 2D mouse and screen-based interface inhibit the flow of ideas. A study was conducted to explore the efficiency of hand tracking and virtual reality (VR) for 3D object manipulations in conceptual design. Based on existing research on conceptual design and hand gestures, an intuitive hand-based interaction model is proposed. An experiment on basic 3D manipulation shows that participants using a simple VR and hand-tracking interface prototype have similar performance to those using a traditional mouse and screen interface. For the improvement of gestural conceptual design interfaces, the relevant issues are identified.

## 1. Introduction

In early, conceptual stages of 3D design, such as industrial design, 3D art (e.g., in games) and architectural design, designers try to gain insight of a given spatial problem. They start with vague ideas and explore the problem and possible solutions.

The conceptual design phase is crucial to the design process. Fundamental decisions are made in this phase and the impact of design decisions declines steeply as the design matures (Wang, Shen, & Xie, 2002). Important issues that are overlooked in the early stages may be more difficult to overcome in later stages, if not taken into account from the start. During the conceptual design phase, decisions are made that will account for 75% of a product's life cycle cost, including manufacturing, use, repair, and disposal (Nevins & Whitney, 1989).

Design sketches are used not only to generate ideas (*ideation*), for external memory, for association by visual cues, but also as a physical setting in which design thoughts are constructed (Suwa, Gero, & Purcell, 2006). These sketches are low on details and only show the essential elements of the design that are currently under consideration. Designers can use many tools for sketching, the most common being pencil and paper, but also clay, blocks, or other objects and materials can be used.

In 3D conceptual design, different sketching tools have their advantages and disadvantages. Pencil and paper perhaps support the quickest sketches for people who can draw, but the view can be drawn from only one side at a time. Objects and materials can be used for quick visualization of crude

ideas, but the right materials need to be present at the right time.

Computer-aided design (CAD) software is a powerful tool to create complex designs and is often used for detailed modeling, presentation, and production schematics. Digital sketching has advantages over physical sketches, such as modifying existing designs, copying, and histories (undo/redo). However, CAD is mostly used in later stages of design and is less adequate for conceptual design (van Dijk, 1995; Wang et al., 2002). The (required) precision and features nested in menus and modes inhibit the flow of ideas.

For a CAD tool to be effective for conceptual shape design, van Dijk (1995) collected a list of recommendations from literature:

- (1) *Easy data entry*: both entering new data and modifying existing data should be quick and simple.
- (2) *Hand movements*: hand movements in the interaction complement the thinking process.
- (3) *Imprecise data*: exact data requirements inhibit the flow of thoughts of the designer, so the software should accept imprecise data.
- (4) *Switch to details*: the designer should be able to focus on specific features in detail, while temporarily forgetting about others.
- (5) *Review alternatives*: it should be possible to compare different designs.
- (6) *Separate conceptual design system*: the tool for detailed design should be separated from the tool for conceptual design.

- (7) *Clay modeling*: clay modeling provides inherently 3D interactive, free-hand shape (de)formation.

Moreover, conceptual design issues are often interdisciplinary and involve collaboration between customers, designers, and engineers (Wang et al., 2002). For all intended users to be able to learn and use the tool quickly, the tool should also be generally *intuitive*, allowing them to apply knowledge from different contexts to the tool (Blackler, Popovic, & Mahar, 2002).

The most widely used interface for CAD is (and has been for many years) the keyboard, mouse, and screen. There is an inherent inefficiency in this interface that may be possible to improve upon in terms of conceptual design: the user currently works from the 3D real world, through a 2D input, to model a 3D virtual world, which is displayed on a 2D screen (Figure 1). This information flow requires compression from the 3D world (by the user) to the 2D interface and decompression from the interface (by the computer) to the 3D virtual world. This way, an entire spatial dimension of user input is unused by the interface and the user will have to learn how to compress their 3D concept into the 2D interface. In the visual output, the 2D display imposes restrictions on the user's perception of the virtual environment. The monoscopic view impairs depth perception and the view cannot be adjusted by the user naturally (e.g., moving their head); instead, the view has to be adjusted by extra controls.

The development of advanced input technologies and virtual reality (VR) can take away the dimensionality transformations inherent to 2D interfaces for 3D design. VR can create a more natural perception of the virtual world, immersing the designer into the design; 3D input devices can replace the 2D interface for more natural and direct interaction with the virtual world. Moreover, gesture-based interfaces may provide an interface that is more intuitive than hand-held input devices, since humans naturally interact with the physical world using their hands.

In this research, we investigate whether a hand gesture-based interface (combined with VR or screen) can decrease task load of 3D object manipulations and improve time efficiency of such tasks, paving the way for future intuitive and more effective ideation interfaces. We hypothesize that a hand gesture-based VR interface allows 3D object manipulation tasks to be performed faster than conventional mouse and screen based interaction.

The remainder of the article is structured as follows. In the next section, we discuss related work on 3D conceptual design

tools. In Section 3, we discuss hand gesture research to lay a basis for a gesture-based conceptual design interaction, which is then explained in Section 4. In Section 5, we describe the experiment we conducted to test the current<sup>1</sup> feasibility of a hand-based VR interface for conceptual design and we discuss the results in Section 6. Our conclusions are presented in Section 7, along with a consideration of relevant future work.

## 2. Related work

Over the last 15 years, several novel interaction techniques have been proposed to improve the effectiveness of computer-aided conceptual design tools. They can be divided into two categories: improvements to the 2D–3D conversion interface and migration from 2D to 3D interfaces.

### 2.1. 2D–3D conversion

Many conceptual CAD techniques focus on improving the 2D interface to the 3D virtual world, often based on traditional concept sketching. Some detect pen gestures to create primitive shapes and perform actions, as implemented in SKETCH (Zelevnik, Herndon, & Hughes, 2007). GIDeS (Pereira, Jorge, Branco, & Ferreira, 2000) also allows free-hand sketches in top/front/side views that are combined into one 3D reconstruction. Naya, Jorge, Conesa, Contero, and Gomis (2002) proposed a CAD system in which the user could directly draw the axonometric view of an object, which was then converted to a virtual 3D object. While this system assumed the input of straight-edged objects, Teddy (Igarashi, Matsuoka, & Tanaka, 2007) assumes round shapes, which allows the user to quickly free-hand draw organic shapes (such as a teddy bear) that are converted into 3D models.

Benko, Ishak, and Feiner (2005) explored a hybrid interaction model, combining 2D and 3D representations and manipulations of objects. Objects could be not only viewed and manipulated in 2D but also grabbed and pulled into a 3D (augmented reality [AR]) environment using hand gestures (and vice versa). The objects could be parametrically edited in the 2D user interface, with direct feedback in the 3D environment.

Although some of the authors of these systems report user preference for their system over traditional CAD, only GIDeS (Pereira et al., 2000) was subjected to experiments to measure effectiveness. GIDeS outperformed an unspecified commercial CAD system on seven unspecified modeling tasks in task completion time.

## 3. 3D interfaces

Construction of 3D objects based on 2D sketches faces an inherent problem: any algorithm performing this 2D–3D conversion requires assumptions for the third dimension. Replacing the 2D interface of a CAD environment with a 3D interface removes these assumptions and allows direct creation and manipulation in 3D space, which is shown to have usability advantages.

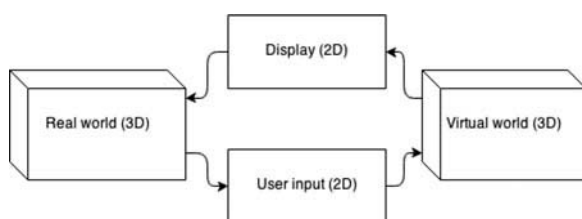


Figure 1. Traditional CAD interface.

<sup>1</sup>Using consumer-grade products.

One type of 3D interface is based on free-hand tracking (e.g., Shen, Ong, and Nee, 2011), and although such technologies can still be much improved, Coelho and Verbeek (2014) found that 3D pointing tasks were already performed more quickly using a 3D input device (Leap Motion) than using a mouse. When using the mouse, the third dimension was controlled with the scroll wheel. They concluded that 3D translation is “less cumbersome when the [third dimension] is provided as input based on real-life movement mappings” (page 84).

Moreover, Buchmann, Violich, Billingham, and Cockburn (2004) found that users of their 3D fingertip-based interface were quickly able to learn the natural manipulative gestures of their system. They also noticed that many of their usability issues were caused by tracking constraints.

A design tool by Encarnação (1999) combines pen gestures and manipulation [similar to those of SKETCH (Zelevnik et al., 2007)] with a 3D AR setup. A virtual table and shutter glasses produce a stereoscopic vision of the scene, while a sketchpad and pen are tracked to interact with the 3D scene. The user can create and manipulate the 2D projection of the 3D scene with the sketchpad using pen gestures. The system was informally tested on several users, who generally found the design very appealing and were able to understand the tools with ease.

A similar approach was taken in Spacedesign (Fiorentino, Amicis, Monno, & Stork, 2002), using infrared cameras to track several tools to interact with a 3D virtual environment. In addition to changing the viewpoint naturally, Spacedesign enables the user to create surfaces from curves that are drawn with a pen directly in 3D space, without the use of a sketchpad. Industrial designers in the automotive industry appreciate the simplicity of the system, because surfaces can be created without having to do the mathematics that are normally required. Because of its intuitiveness and real-time editing, stylist can be active part of the development process.

Besides sketching, some applications are based on clay modeling. VRClay (2014) is a recent tool that uses Oculus Rift VR glasses and two Razer Hydra 6-degrees-of-freedom (6DoF) controllers, which are tracked accurately in any direction in 3D space. The user can hold, rotate, and move an object with one and sculpt it with the other. An important advantage of “virtual clay” over physical clay is that there are no restrictions on the addition of material or shape of parts (e.g., thin legs falling off by gravity). At the time of writing, this application is still being actively developed, but it seems to work very intuitively for organic shapes, as shown in their “sneak peek” video.

Other applications, more or less, follow established CAD principles but provide a 3D interface for more direct interaction. HoloSketch is a VR CAD tool based on simple 2D drawing systems (Deering, 1995). It uses a “Wand” to point and click in 3D space. Menus appear in mid-air and can be clicked with the Wand. From these menus, 3D objects can be selected and placed in the design. Evaluation was done by one artist with experience in traditional CAD software, during 1 month. Overall, she found the interface much more productive than traditional modelers.

A recent and very comprehensive VR CAD tool is MakeVR (Jerald, Mlyniec, Yoganandan, Rubin, Paullus, & Solotko, 2013); it is currently on its way to becoming one of the first VR CAD tools to enter the commercial market. Similar to VRClay, it uses two Razer Hydra controllers and Oculus Rift VR glasses. It is built on top of the ACIS CAD engine and enables operations such as moving objects in all directions in 3D space, boolean operations (e.g., cutting one shape out of the other), and sweeps (e.g., using a sphere to create an irregular-shaped pipe). Although it would be very interesting to see whether this tool would increase ideation effectiveness in conceptual design, experimental research was limited to a Likert scale usability questionnaire.

Many conceptual design interfaces have been proposed; however, only few are thoroughly tested for effectiveness. Rahimian and Ibrahim (2011) investigated the impacts of a 3D sketching tool on designers’ cognitive and collaborative abilities in conceptual collaborative design. A 6DoF haptic device was connected to ClayTools CAD software and employed in a complete conceptual architectural design by three pairs of novice architects. An elaborate protocol study of the design process showed that the design sessions using the 3D sketching tool improved in terms of perceptual actions (more attention to visuospatial features), unexpected discoveries (discoveries from the sketch), coevolution of the problem space (problem finding), and collaborative actions (more proposals and arguments instead of questions and explanations).

### 3.1. Research focus

Generally, 3D conceptual design interfaces are received with enthusiasm by users and professionals when compared to traditional, mouse-based interfaces. Most of the drawbacks and complaints that are mentioned are focused on technical difficulties or implementation choices.

However, very few experiments have been conducted to formally compare 3D interfaces for conceptual design to their 2D counterparts. Moreover, research has been focused on 3D pointing and drawing tools to replace the mouse, while naturally, humans communicate and interact more directly with the world around them by using their hands. Therefore, in our research, we focus on hand-based interaction and experimentally test its effectiveness in basic design tasks.

In this article, we will create a theoretical basis for a gesture-based conceptual design interface, based on research on natural hand gestures. We will then use a limited vocabulary of these gestures in a first experimental step to test their efficiency, keeping the scope small so as to enhance the focus of the measurements. With a full-featured hand gesture-based conceptual design environment at the horizon, this experiment will be the first step to assess its effectiveness. The next section introduces and discusses the theoretical basis for the gesture set.

## 4. Hand gestures

Hand gestures are a natural way of communicating and interacting for humans. A design interface based on these natural gestures is therefore likely to provide intuitive interaction with

a virtual environment. In this section, we explore the domain of natural hand gestures with relation to conceptual design so that we can establish a useful set of gestures for an experimental setup to investigate the efficiency of a gesture-based interface for object manipulations in conceptual design. Whether the described theories of inter-human communication also apply to human-computer interaction remains to be tested. However, they can nevertheless provide a good starting point for an intuitive gesture-based interface.

#### 4.1. Types of gestures

Gestures can assume various roles in human communication. McNeill (1992) listed and described a set of discrete gestures, such as *metaphoric* gestures (describing abstract ideas), *beats* (emphasizing words they co-occur with), and *deictics* (pointing at objects or abstract concepts). Particularly, interesting for our purpose are *iconic* gestures. These have a close similarity to their referents and can depict spatial features, such as size and shape. For example, the concept *up* could be accompanied by pointing a finger in upward direction.

The aforementioned gestures are discrete, communicative arm, or hand movements that can be mapped to a specific meaning through a dictionary. Quek et al. defined such gesturing systems as *semaphoric gestures* and distinguished these from *manipulative gestures*, which are intended “to control some entity by applying a tight relationship between the movement of the gesturing hand/arm with the entity being manipulated” (Quek, McNeill, Bryll, & McCullough, 2002). An important difference between the two types of gestures is that semaphoric gestures are performed and recognized as a whole, while manipulative gestures require (and allow) continuous feedback from the action.

Sturman and Zeltzer (1993) divide gestures into six classes: *direct*, *mapped* and *symbolic* gestures, each further divided into *continuous* and *discrete* gestures. The *continuous* gestures have continuous input to the system, whereas *discrete* gestures (generally postures) are encoded as a single command. *Direct* gestures control kinematically similar actions in the system, *mapped* gestures provide a vocabulary of signals mapped to arbitrary functions, and *symbolic* gestures are interpreted by the system in a wider context and have no one-to-one correspondence with a specific system function.

#### 4.2. Gestures for design

Since gestures are part of the natural interaction of humans, they may provide a valuable interface language between a conceptual design tool and a designer.

Not only are gestures naturally communicative, they also help people think. Cash and Maier (2016) show that designers use gestures to communicate ideas (directed) as well as for themselves (reflective). Other research (described in a review by Alibali, 2005) shows that gestures help speakers activate and remember spatial information (Morsella & Krauss, 2004; Ruiters, 2000; Wesp, Hesse, Keutmann, & Wheaton, 2001); speakers describing images produced more gestures when the images were no longer visible, than when the images were still visible.

An experiment conducted by Hostetter and Alibali (2004) suggests that gestures help conceptualize information. In this experiment, participants were asked describe a dot pattern (audio-only) that was either easy or difficult to conceptualize in terms of geometrical shapes. Dot patterns that were difficult to conceptualize elicited more gestures, even though the participants knew their gestures did not contribute to the communication.

#### Physical manipulation

Perception is closely linked to action. We often need actions to perceive, like moving our head to see and moving our arms to feel. One of the central views in the theory of *embodied cognition* (Gibbs, 2006) is that memory, mental imagery, and problem solving are closely linked to physical interaction: “objects are perceived by imagining how they may be physically manipulated” (page 12).

Ellis and Tucker (2000) found that a range of actions associated with an object is primed upon perception of the object. In their experiment, participants were faster to grab a cylinder in response to the presentation of objects that would be grasped with a similar grip (e.g., hammer, bottle, jar) than in response of objects that would be grasped with a precision grip (e.g., screw, nail, coin).

The way we can interact with objects would thus influence how we perceive those objects and, in a larger context, a problem space. Basing the interface on interaction with the physical world may link the designer’s perception of the virtual world more to that of the physical world, in which humans have a great deal of experience and insight.

#### Iconic gestures

To make proper use of digital capabilities, we cannot solely rely on physical interaction in our conceptual design tool. After all, some interactions (e.g., creation, boolean operations, etc.) are not part of basic interaction with the physical world. For these actions, we can use semaphoric gestures.

Iconic gestures may provide an intuitive interface for these non-physical interactions. They exist in our natural communicative repertoire and, by definition, bear close semantic resemblance to their referent. Doherty (1985) reviewed several studies that showed that iconicity in signs facilitates their acquisition and retention (e.g., Luftig & Lloyd, 1981; Mandel, 1977b).

Marsh and Watt (1998) researched what type of gestures was used by participants that were asked to describe different shapes using only their hands. Following Mandel (1977a), a distinction in iconic gestures was made between *virtual depiction*, in which the hand(s) trace the outline of the object (e.g., tracing a star) and *substitutive depiction*, in which the hand(s) “become” the object (e.g., using the arms to form a bridge). The results showed a strong preference for virtual depiction over substitutive depiction for primitive shapes (e.g., cube, sphere, pyramid). For complex shapes, iconic gestures were still preferred, but occasionally accompanied or replaced by pantemimic (e.g., describing a car by pretending to drive) and other gestures.

Sowa and Wachsmuth (2002) further examined the morphological features of iconic gestures accompanying speech of participants describing several complex shapes. The gestures

used most (86%) were *dimensional* gestures, “representing an object’s outer dimensions via delimiting or enclosing”. Sowa and Wachsmuth also mentioned that these gestures are often subject to *dimensional underspecification*, depicting only one or two dimensions of the object.

More recently, a dataset was constructed by Sadeghipour, Morency, and Kopp (2012), containing data of 29 participants describing 10 primitive and 10 complex shapes using gestures only. The gestures were recorded using a Microsoft Kinect and the dataset is available online as the 3D Iconic Gesture Dataset (3DIG) (Sadeghipour & Morency, 2014). Sadeghipour and Morency built a classifier to recognize the 20 shapes, outperforming human judgment on gestures using a virtual depiction as a drawing technique.

### Combining gesture classes

Mimicking physical manipulation in the virtual environment may provide an intuitive interface for manipulating virtual objects. However, since these are limited to manipulations we can perform in the physical world, we should use other gestures to extend the “unrealistic” functionality, such as object creation. For this purpose, iconic gestures seem very suitable, since they are easy to remember and contain high amounts of spatial information, both due to their close similarity to their referents.

In the next section, we therefore propose a gesture-based interface for conceptual design, based on these principles, including a selection of physical manipulative and iconic gestures. Part of this set we will later use to investigate the efficiency of a hand gesture-based VR interface in basic object manipulations for conceptual design.

## 5. Concept of gesture-based interaction

From the theoretical basis discussed in Section 3, we propose a gesture-based, VR interaction model that supports a wide range of 3D virtual object manipulations for conceptual design. By including a third dimension of input to the interface, there can be a direct coupling between physical and virtual location and this tool should therefore remove some of the restrictions posed by 2D interfaces on 3D conceptual design. We based the interactions on natural human communication and interaction for an intuitive interface with the virtual world, allowing easy, imprecise, and efficient data input using the hands.

We selected this set of gestures to form a theoretically useful interaction model for conceptual design, based on the previously discussed literature. We can then use a subset of these gestures to test the efficiency of a prototype VR hand gesture interface for basic object manipulations in the domain of conceptual design. Figure 3 shows a participatory using our prototype setup.

### 5.1. Interaction

Although interaction mimicking natural physical interaction might be intuitive, some actions that add to conceptual design are not natural (e.g., creating objects from thin air). The virtual world is not equal to our physical world. This,

together with some technological limitations (e.g., accurate tracking), requires us to make a trade-off between intuitiveness and functionality. The interactions described in this section were chosen to form a set of interactions that will support the designs of a large variety of 3D composite shapes, that demonstrates the expressive power of hand gestures in this context, and that we can use to investigate the efficiency of gesture-based interaction in conceptual design. The corresponding gestures are illustrated in Figure 2.

### 5.2. Object creation

Since object creation is not a natural manipulation, we use iconic gestures to create primitive shapes, which can then be combined and manipulated. Since people usually use virtual depiction (Marsh & Watt, 1998) to describe physical shapes using gestures, the program traces the gesture to set the outlines of a new object. To deal with dimensional underspecification (Sowa & Wachsmuth, 2002), we assume the depicted outlines are the two most dominant (i.e., largest) dimensions. The third dimension will be set to an arbitrary minimum size. We also assume the depicted outlines are those of a flat face, perpendicular to the third dimension.

To allow the user to also create some important primitives that do not satisfy these assumptions, we created a vocabulary of iconic gestures to create a sphere, cylinder or cube. For this vocabulary we used the most popular iconic gestures used to describe these primitives in the 3DIG Dataset (Sadeghipour & Morency, 2014).

### 5.3. Grabbing

For grabbing objects, we mimic physical manipulation: an object is grabbed by closing a hand around it. The interaction is designed without force feedback, so hand and fingers can freely pass through virtual objects. A more realistic, physics-based grasping technique was proposed by Borst and Indugula (2006), using a virtual spring model of a hand that would apply force on the boundaries of the object and be restricted by it. The visual feedback would be the virtual hand grasping the virtual object, while the real fingers would pass through the virtual object. However, in their user tests, this created some practical difficulties, such as delayed release because of a deep grasp and difficult precision manipulations (for small objects) due to imperfections in tracking and hand model geometry.

We propose a pinching gesture for object grabbing consistent for all objects. When thumb and index finger touch, the pinching gesture is activated at the location of the touch. If the pinching gesture is activated within an object, the object is grabbed. This pinching gesture provides a single grabbing location, which simplifies the interaction model in contrast to full-hand grabbing. Pinching is also more easily recognized by hand-tracking devices such as the Leap Motion.

Grabbing objects enable different actions to be performed on that object, such as movement, scaling, and (un)grouping.

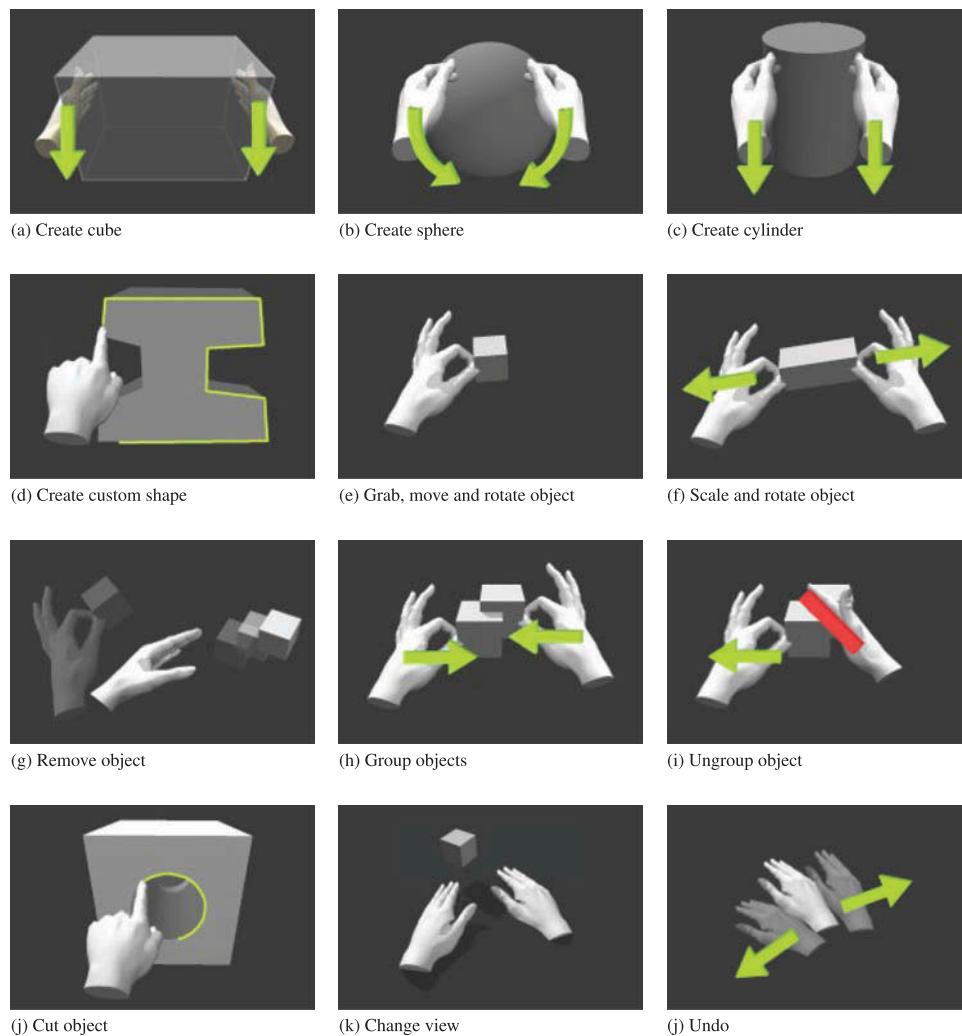


Figure 2. Gestures for conceptual design.

#### 5.4. Moving and rotating

When an object is grabbed, moving the grabbing hand will move the object along with it. The 3D hand tracking allows a 1:1 translation of the hand position to the virtual object's position. Furthermore, the relative rotation of the hand, while the object is grabbed, can be applied directly to the object. By grabbing an object, the user can therefore rotate and move the object directly in 6DoF.

This direct translation, being based on physical manipulation, should be intuitive and draw on the 3D eye-hand coordination that is learned from young age.

#### 5.5. Scaling

Scaling is performed by an interaction mimicking stretching materials. The user can grab an object on two sides and increase or decrease the distance between the hands. The object boundaries will follow this hand movement, causing it to scale with that distance. Since the boundaries of the object “stick” to the user's hands, it can still be moved and rotated during scaling.

As with the moving and rotating interaction, scaling draws on the learned coordination of the user with physical materials.

#### 5.6. Removing

Removing an object from the virtual scene mimics the human behavior of throwing something away. Grabbing the object, moving it quickly, and releasing it during the movement will remove the object, after it has flown out of the screen.

With this action, it does not matter where the object will land. Therefore, the movement can be fast and uncontrolled. The user is likely to be familiar with throwing away objects.

#### 5.7. Grouping

In CAD software, it is possible to group and select multiple objects for simultaneous movement. This way, the designer can create complex from several smaller ones, regarding the group as one larger object. In 2D interfaces, multiple objects can be selected by drawing a rectangle around the shapes.

In 3D, however, Oh, Stuerzlinger, and Dadgari (2006) proposed a hierarchical grouping technique that, in a small user test, showed potential to make group selection more efficient. This grouping technique was based either on gravity or connectivity. In the first, objects that were positioned on top of each other would be grouped hierarchically such that moving the lower object would simultaneously move all objects on top of it. In the latter, objects that touched were automatically grouped.

A gravity-based grouping mechanism would assume the scene contains mostly loose objects stacked on top of each other. Due to the restrictions this assumption poses, we consider the connectivity grouping more practical for conceptual design. This would mimic physical manipulation based on sticky objects, making objects stick when pushed together. Grouping would therefore be as simple as moving one object partly into another, making them “stick” together. When, afterward, one of the objects is grabbed and moved or scaled, any grouped object will transform with it.

Ungrouping would similarly be based on the physical manipulation of ripping objects apart. In this interaction, there is a possible conflict with scaling. Grabbing two objects in the same group and moving them in different directions could release the grouping and enable independent movement, but the interaction already refers to scaling the object (group). We therefore propose an open hand pose for one of the hands, mimicking the action of holding back the other objects, while “ripping” the grabbed object apart from them.

When ungrouping two objects that are connected indirectly through other objects within the group, a historical hierarchy is maintained, separating the group at the link that was most recently made. This interaction assumes that components are first modeled independently before being attached to each other. The “weakest” link in the two connected components would then be the connection that was made last, joining the components together. Should this not be the case, the user can still separate the group at the desired link by grabbing the two objects directly involved in that link. That link will then be renewed and become the weakest link in the group. This allows smaller “modules” to be made and then combined, after which they can easily be detached from the others.

### 5.8. Cutting

The shape of objects can be controlled in more detail by a cutting operation. Shapes can be cut out of objects in a similar way as shapes can be created, by sketching its outlines with the index finger. Instead of sketching in the air, however, a cutting shape is sketched on the surface of an object. A tunnel is then created in the shape of the hole, all the way to the other side of the object. The hole can then be manipulated like any other object (e.g., moved around within the object, rotated, and scaled). Once again, this gesture draws on the natural use of iconic gestures.

### 5.9. Changing view

Using VR with head tracking brings the advantage of natural view changing. The user can simply move their head to look

from a different angle, the same way they would do in the real world. However, if this is the only way of changing the view, it may put the designer in uncomfortable positions when viewing the model from certain angles for prolonged periods of time.

For this reason, we enable the user to move and rotate the workspace in two dimensions (left–right and forward–backward). The view can also be “zoomed” by enlarging or reducing the workspace and all objects in it. When both hands touch the workspace surface (which appears as a solid, flat cube, as if it were a table top), the workspace is moved, rotated, and scaled so as to keep the touching points at the hands. This interaction (except for the scaling) is similar to moving around a physical model on a tabletop. To zoom in or out, the user would place the hands on the surface and move them apart or together, respectively.

### 5.10. Undo

Undoing a previous action is the only action in this set that does not require accurate spatial coordination with an object. Therefore, we can use a relatively uncoordinated gesture so that it is easily distinguishable from the other gestures (both for the user to learn and for the system to recognize). The undo action has neither dimensions nor a “physical” virtual representation, so we cannot apply physical manipulation or iconic gestures. We therefore chose a simple “waving” gesture for the undo action. Unlike the other gestures, this does not need to be coordinated and targeted at a specific object or location. It is also easy to perform and yet is not easily triggered by mistake.

### 5.11. Selection

This set of interactive gestures was assembled based on the theoretical basis of natural gestures and conceptual design. Alternative selections of gestures are possible; however, this set of gestures is a theoretically suitable interaction model for conceptual design and will allow us to investigate the efficiency of gesture-based VR object manipulations for conceptual design.

## 6. Experiment

In an experimental setup, we compared the usability, task load, and performance of two gesture-based interfaces (with screen or VR) and a traditional mouse and screen interface, with regards to basic object manipulation. For this, we investigated a basic subset of design interactions, including movement, rotation, scaling, and undo. Creation and deletion of objects were left out in order to limit the complexity of the experiment.

We implemented a prototype consisting of off-the-shelf consumer components, to investigate whether this novel interface could be achieved with current, mainstream technology. The hardware behind the prototype consisted of a laptop (running the graphics smoothly), the Oculus Rift (Development Kit 1) VR headset, and the Leap Motion (version 2) hand-tracking device. The Oculus Rift has a

wide angle view and head-orientation tracking and the Leap Motion is a small optical device that performs full-hand tracking in an area of about 60 cm × 60 cm × 60 cm. Since the focus of this experiment was on interaction and not on feature richness, we implemented the prototype in the Unity 4.6 game engine on Windows 7, rather than a CAD tool, since the first provides more tools for customization of interaction than the latter. Both Oculus and Leap Motion provide assets (plugins) for Unity to program with the hardware.

In our prototype, we implemented the interactions as described in Figure 2: movement (e), rotation (e/f), scaling (f), and undo (j). Virtual hands were visible, following the movement and pose of the user's hands. When thumb and index finger touched, the touch point was used as the location for a grabbing gesture (for movement, rotation and scaling). That point on the object would then follow the touch point. The undo gesture was implemented as three direction changes (above a threshold magnitude) of a hand in 0.5 s. This would undo the transformations made on the current object.

We compared the new interactive system with a traditional setup, using a mouse and a screen. The Unity Editor itself provided typical manipulation tools for object movement, rotation, scaling, and undo (Figure 4), using the mouse, keyboard, and screen. At the top of the screen, there were buttons for three main object manipulation modes: moving, rotating, and scaling. Clicking an object in either of these modes would show the handles that could be dragged to manipulate the object in that mode. The camera could be zoomed using the scroll wheel, rotated by pressing Alt while dragging the mouse, and moved using the "hand" tool or pressing Alt Gr while dragging the mouse. All buttons and features of the Unity Editor that were irrelevant for the basic object manipulations were hidden. These editor tools allowed us to test the different interfaces on similar object manipulation tasks in the same environment.

Each participant was randomly assigned to one of three conditions: (1) Mouse & Screen, (2) Hands & Screen, and (3) Hands & VR. The second condition was added to gain extra insight on the influence of the VR component on the modeling tasks.

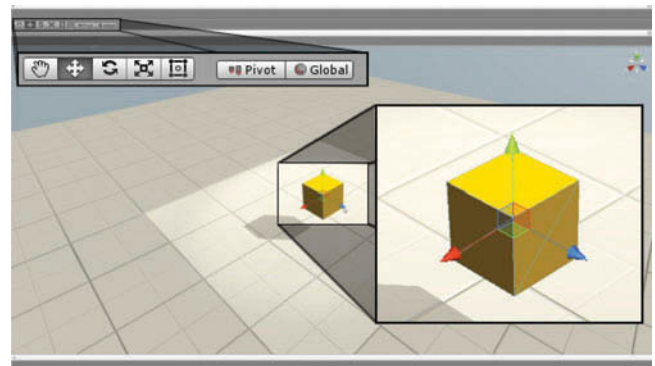


Figure 4. Object manipulation in Unity Editor. Buttons top-left switch modes (e.g., move/scale/rotate), handles on the object can be dragged to control the mode's dimensions. Spacebar could be used to undo the current transformations. All other interfaces of the editor were hidden from the participant.

### 6.1. Tasks

Each participant was asked to go through one *trial*. Each trial consisted of six to nine *tasks*, in which a yellow *shape* appeared and had to be transformed (moved, scaled and/or rotated) to match a red, transparent *target* (Figure 5). Considering the sketching nature of conceptual design, the target did not require 100% overlap with the shape. Instead, the target was completed when at least 40% of the target and shape overlapped, which, in our view, is a reasonable fit for a 3D sketch. After each task, the previous shape would turn transparent gray and the next shape-target set appeared. The order of the tasks within each trial was randomized to control for order effects. Finally, the correctly transformed shapes would together form a model (e.g., a house or table and chair), which was transparently visible throughout the trial.

To find differences between transformation types, tasks were chosen to be in four different categories: Move, MoveRotate, MoveScale, and MoveRotateScale. These categories we call *task types* and they encode what kind of transformations has to be performed on the shape to fit the target (for instance, shapes in the task category MoveRotate should be moved in and rotated). Example tasks for each task type are illustrated in Figure 6.

The objective was to complete the trial as quickly as possible. The trial was presented as a game and high scores with the other participants were displayed at the end of the trial, to

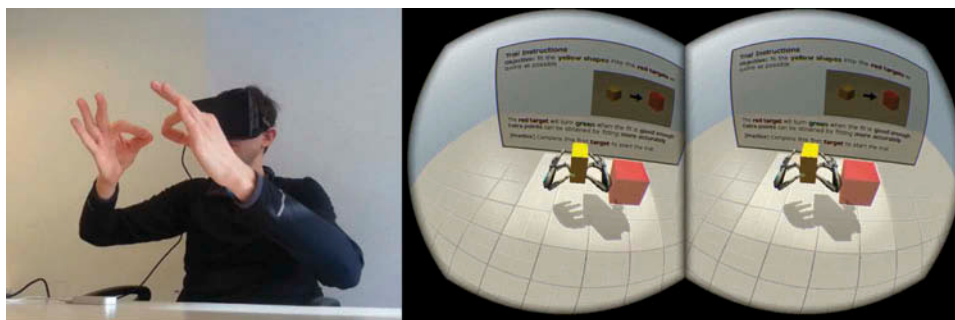
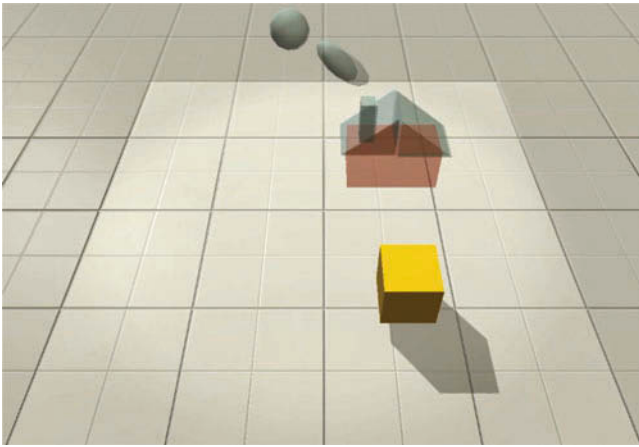


Figure 3. Participant resizing a shape during experiment. Left: The participant wearing an Oculus Rift VR headset, his hands being tracked by the Leap Motion. Right: The stereoscopic image the participant could see through the display.



**Figure 5.** Experiment task. The yellow shape must be fitted into the red target shape by moving, scaling, and/or rotating. When the target is completed (40% overlap of shape and target), the shape turns gray, the next target is highlighted red and the corresponding shape appears at its start position. In this example, the yellow shape must be moved forward and scaled horizontally to fit in the red target. The gray shapes are (inactive) previous and next tasks.

motivate participants to complete it as quickly as possible, instead of taking the time to try out the technology. When, during the trial, it was clear to the experimenter that a participant had misunderstood or forgotten about the rules or controls, they were explained again. This was done to focus the experiment on the possible proficiency with the used medium instead of the retainment of knowledge about the specific interface.

### 6.2. Procedure

Before each trial, the participant was introduced to the interactions with the CAD software, during a 5-min “Sandbox”

mode, in which they could freely manipulate a single cube in the virtual space. Since the Leap Motion hand tracking showed inaccuracies in usage during preliminary testing, some extra time was taken in the Hands & Screen and Hands & VR conditions, to instruct participants about the optimal hand poses and interactions with the Leap Motion tracker. Since participants generally already know how to use a mouse, but not the Leap Motion, we consider this a fair instruction.

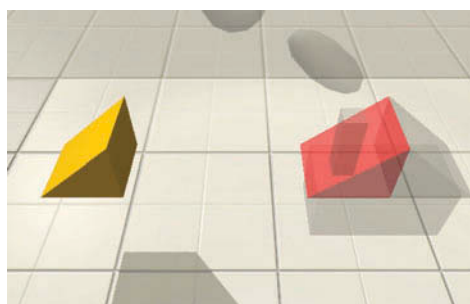
After the Sandbox mode, a practice task was presented to the participant, containing trial instructions on the screen. The measured trial started as soon as the participant finished the practice task. Since each participant needed some training time, six to nine tasks were presented per participant in order to gather a reasonable amount of data. This is taken into account in the analysis of the results.

### 6.3. Methods

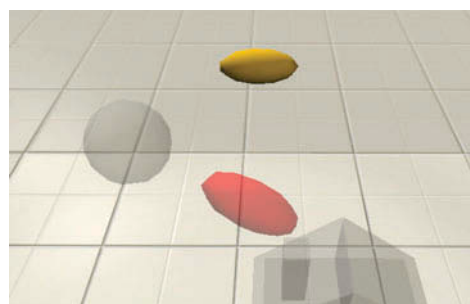
During the trial, duration and accuracy of each task was recorded automatically, as well as the type of transformation (moving/rotating/scaling). Although the tasks were randomized to minimize any learning effects, the task number (in order of appearance) was recorded to find any learning effect, if present.

At the end of each trial, the participant was asked to fill out a questionnaire with demographic and experience data, standardized System Usability Scale (SUS) questions (Brooke, 1996), a modified version of NASA Task Load Index (TLX) questions (Hart & Staveland, 1988) and specific questions on the interactions. For all questionnaires, a 5-point Likert scale was used. The used questionnaires are shown in Appendix A.

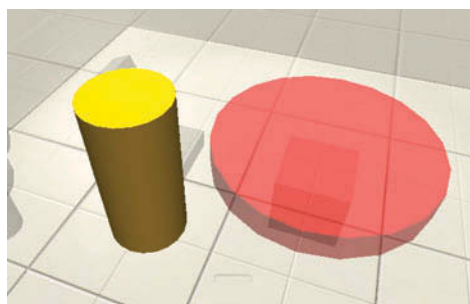
Finally, observations and user comments were noted by the experimenter, for qualitative analysis.



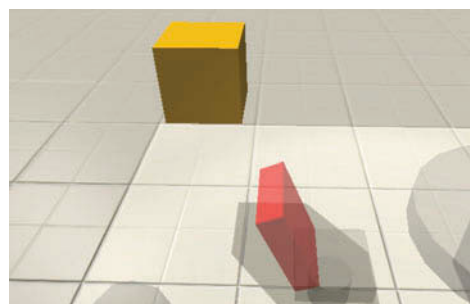
(c) Move



(d) Move and rotate



(e) Move and scale



(d) Move, scale and rotate

**Figure 6.** Task types.

## 7. Results

In the experiments, 33 participants performed a total of 243 tasks, in 3 (randomly assigned) groups of 11 participants per condition (Mouse & Screen, Hands & Screen, Hands & VR). All participants were volunteers, 24 of whom were undergraduates in the Media Technology program at Leiden University in the Netherlands. Most of the participants (23) did not have any prior experience with Leap Motion and 11 of them had no prior experience with CAD. There were 23 males and 10 females between 22 and 31 years old, the majority (21) of Dutch nationality. The complete experiment took about 20 min per participant including the instructions. The recorded trials took an average of 11 min, with a standard deviation of 7 min.

### 7.1. Task performance

The three interfaces (conditions) were compared on the users' ability to quickly transform shapes to match a target's position, rotation, and/or size. Due to the experimental design with multiple tasks per participant, a mixed model ANOVA with random effects was run in R (R Core Team, 2014), package nlme (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2014) to compare the differences of the means between the three interfaces. In the mixed model, task duration was the dependent variable with condition and task type as fixed effects. Since a different interface might perform differently for different task types, an interaction effect was modeled between condition and task type. To control for having multiple samples per participant, the participant factor was added as a random effect.

There was a statistically significant difference of the mean duration over both the conditions ( $F(2, 30) = 6.85$ ,  $p = 0.004$ ) and task types ( $F(7, 189) = 12.6$ ,  $p < 0.0001$ ). Also the interaction between condition and task type was significant ( $F(14, 189) = 2.05$ ,  $p = 0.02$ ). A Tukey *post-hoc* test using the multcomp package (Hothorn, Bretz, & Westfall, 2008) revealed a significant difference between conditions Hands & Screen and Mouse & Screen ( $p = 0.02$ ), as well as between Hands & Screen and Hands & VR ( $p = 0.02$ ).

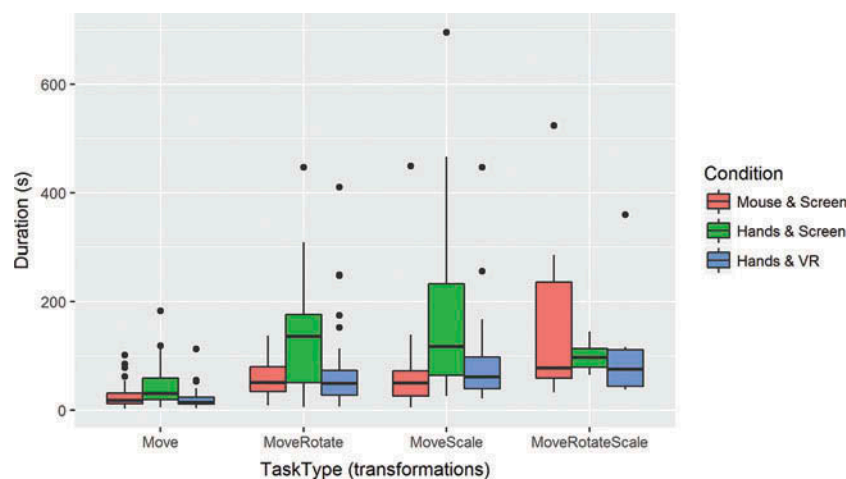


Figure 8. Tukey boxplots of distributions of task duration per condition and task type.

Condition	Mean duration (seconds)	Std.Dev.
Mouse & Screen	57	80
Hands & Screen	109	121
Hands & VR	64	87

Figure 7. Means, standard-deviations, and coefficient of variation of task duration in the different interface conditions.

However, no significant difference was found between Mouse & Screen and Hands & VR. The mean duration and standard deviation of each task are displayed in Figure 7. Participants in the Mouse & Screen condition were fastest, however, not significantly faster than those in the Hands & VR condition. On the other hand, participants the Hands & Screen condition required nearly twice as much time as those in the Mouse & Screen condition.

The mean duration of each task type per condition is displayed in Figure 8. In nearly all task types, participants in the Hands & Screen condition were the slowest, while the other two conditions did not seem to differ much. The exception seems to be the task type MoveRotateScale, in which the condition Mouse & Screen appears to take more time than the other conditions, and with a larger variance.

### 7.2. Questionnaires

The questionnaires at the end of each trial contained SUS questions, NASA TLX questions, and extra questions specifically about these tasks. The distribution of answers to these questions per condition is visualized in Figures B1, B2, and B3 (Appendix B), respectively.

The SUS questions and TLX questions were converted to scale numbers (0–4) and summed up per participant. For negatively posed questions, contributions were calculated as 4–score, as described by Brooke (1996). For the TLX questionnaires, the weights of the questions in the sums were set equal. Finally, the sums were normalized to a range between 0 and 100. The distribution of scores is displayed in Figure 9.

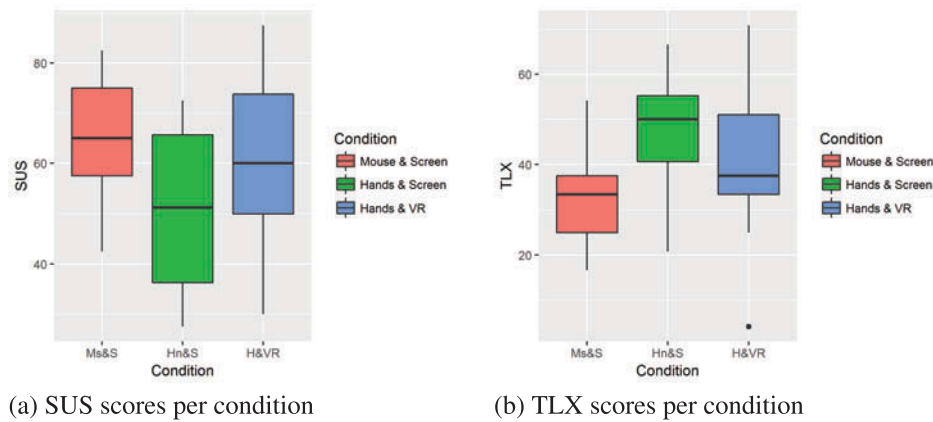


Figure 9. SUS and TLX scores per condition.

Both SUS and TLX scores were statistically analyzed using a pairwise *t*-test in R (R Core Team, 2014), with a Bonferroni correction for multiple comparisons. In the SUS score analysis, none of the groups were significantly different from the others, as shown in Figure 10. In the TLX score analysis, a significant difference was found only between the conditions Hands & Screen and Mouse & Screen, as shown in Figure 11. Participants in the Mouse & Screen condition reported a lower perceived task load than those in the Hands & Screen condition.

All questions were tested using a Kruskal–Wallis test, which is suitable for a Likert-style questionnaire because it deals with ordinal data, does not require normal distributions, and allows more than two groups to be tested. When a significant difference was found between the conditions on Likert-style questions, the contrasts between conditions were tested using a *post-hoc* Dunn test, which is appropriate to follow after the Kruskal–Wallis test, again with a Bonferroni adjustment to control for multiple comparisons. The, per question computed, *p*-values are reported in Appendix C,

Conditions	Hands & Screen	Hands & VR
Mouse & Screen	$p = 0.15$	$p = 1.00$
	$t = 2.05$	$t = 0.54$
Hands & VR	$p = 0.42$	
	$t = 1.51$	

Figure 10. *p* and *t*-values of differences between pairs of conditions on SUS scores, as a result of a pairwise *t*-test with Bonferroni correction for multiple comparisons.

Conditions	Hands & Screen	Hands & VR
Mouse & Screen	$p = 0.04^*$	$p = 0.64$
	$t = -2.61$	$t = -1.27$
Hands & VR	$p = 0.57$	
	$t = -1.33$	

Figure 11. *p* and *t*-values of differences between pairs of conditions on TLX scores, as a result of a pairwise *t*-test with Bonferroni correction for multiple comparisons. Marked with an asterisk (\*) is a significant difference ( $p < 0.05$ ).

Figure C1, C2, and C3 for SUS, TLX and extra questions, respectively.

In the SUS questions, a significant difference was found in Kruskal–Wallis test on the statement that the participant required support of a technical person ( $\chi^2(2, N = 36) = 8.16$ ,  $p = 0.02$ ). The Dunn-test showed that participants in the Hands & Screen condition agreed significantly more with this statement than those in the Hands & VR condition ( $p = 0.02$ ) and than those in the Hands & VR condition ( $p = 0.02$ ). Another significant difference was found on the statement that the interface was easy to use ( $\chi^2(2, N = 36) = 9.56$ ,  $p = 0.01$ ), where participants in the Mouse & Screen interface agreed more than in the Hands & Screen interface ( $p < 0.01$ ). A third significant difference was found on the statement that the system was unnecessarily complex ( $\chi^2(2, N = 36) = 6.94$ ,  $p = 0.03$ ), where participants in the Hands & Screen condition agreed significantly more than those in the Hands & VR condition ( $p = 0.01$ ). The contrasts between conditions that were not mentioned in these three questions were statistically insignificant.

All conditions were statistically indistinguishable in the (dis) agreement of participants on the SUS statements that they would use the system frequently ( $\chi^2(2, N = 36) = 0.122$ ,  $p = 0.94$ ), the various functions were well integrated ( $\chi^2(2, N = 36) = 0.72$ ,  $p = 0.70$ ), there was too much inconsistency ( $\chi^2(2, N = 36) = 0.72$ ,  $p = 0.70$ ), the system was very cumbersome to use ( $\chi^2(2, N = 36) = 0.63$ ,  $p = 0.73$ ), participants felt confident using the system ( $\chi^2(2, N = 36) = 5.88$ ,  $p = 0.05$ ), they needed to learn a lot before they could get going ( $\chi^2(2, N = 36) = 5.52$ ,  $p = 0.06$ ), and that they would imagine most people would learn to use the system quickly ( $\chi^2(2, N = 36) = 0.07$ ,  $p = 0.97$ ).

In the TLX questions, a significant difference was found in the statement that the tasks were physically demanding ( $\chi^2(2, N = 36) = 8.21$ ,  $p = 0.02$ ), where participants in the Hands & VR condition agreed more than those in the Mouse & Screen condition ( $p = 0.01$ ). Also, the statement that the participant found they were successful in accomplishing their tasks showed a significant difference ( $\chi^2(2, N = 36) = 9.74$ ,  $p = 0.01$ ): participants in the Mouse & Screen condition agreed more than those in the Hands &

Screen condition ( $p = 0.01$ ). The last significant difference in TLX questionnaire was found in the statement that the participant had to work hard to accomplish their level of performance ( $\chi^2(2, N = 36) = 11.22, p < 0.01$ ), where participants in the Hands & Screen condition agreed significantly more than those in the Mouse & Screen condition ( $p < 0.01$ ).

No significant differences were found in the reported (dis) agreement of participants on the statements that the tasks were mentally demanding ( $\chi^2(2, N = 36) = 0.91, p = 0.63$ ), the tasks were rushed or hurried ( $\chi^2(2, N = 36) = 1.36, p = 0.51$ ), the participants were insecure, discouraged, irritated, stressed, or annoyed ( $\chi^2(2, N = 36) = 0.70, p = 0.71$ ).

In our extra questions, there was a significant difference between conditions in the agreements to the statement that the participant could easily rotate objects to their desired state ( $\chi^2(2, N = 36) = 7.83, p = 0.02$ ), where participants in the Mouse & Screen state agreed more than those in the Hands & VR state ( $p = 0.01$ ). A significant difference was also found in the statement that the input device worked according to their intentions ( $\chi^2(2, N = 36) = 6.04, p = 0.05$ ), where participants in the Mouse & Screen condition agreed more than in those in the Hands & VR condition ( $p = 0.02$ ). The final significant difference was found in the statement that the participant had trouble perceiving 3D locations of objects ( $\chi^2(2, N = 36) = 0.98, p = 0.01$ ), where participants in the Hands & VR condition agreed significantly less than those in the Mouse & Screen condition ( $p = 0.01$ ) and the Hands & Screen condition ( $p = 0.02$ ). There were no significant differences found between conditions on the statements regarding ease of moving shapes ( $\chi^2(2, N = 36) = 2.29, p = 0.32$ ), ease of scaling shapes ( $\chi^2(2, N = 36) = 5.46, p = 0.07$ ), amount of spatial overview ( $\chi^2(2, N = 36) = 2.55, p = 0.28$ ), enjoyment of working with the tool ( $\chi^2(2, N = 36) = 0.19, p = 0.91$ ), whether the tool could help them gain insight into spatial problems or designs ( $\chi^2(2, N = 36) = 0.24, p = 0.89$ ), and naturalness of the interface ( $\chi^2(2, N = 36) = 4.70, p = 0.10$ ).

Over all conditions, 66% of participants reported a preference for using the hands over a mouse and 10% preferred using a mouse over the hands (24% had no preference). Furthermore, 59% reported a preference for using VR over a screen and 14% preferred using a screen over VR (27% had no preference). There was no statistically significant difference in preference between conditions ( $\chi^2(2, N = 36) = 0.16, p = 0.92$  and  $\chi^2(2, N = 36) = 0.09, p = 0.96$ , respectively), meaning these preferences were regardless of the interface the participant used.

### 7.3. Observations and user comments

During the experiments, observations were made about the interactions of the participants with the interface. Together with the open questions in the questionnaire, they might shed some light on some issues that are difficult to quantify.

First of all, participants generally had to get familiar with the tracking behavior of the Leap Motion. Not all hand poses are equally easy to recognize for the tracking device. For

example, at all times during interaction, the camera inside the device should have a clear line of sight to the user's index finger and thumb (which were the main fingers used in this experiment). Besides the hand pose, participants needed to get used to the limited working area. The Leap Motion has a tracking area of roughly 60 cm × 60 cm × 60 cm. To interact anywhere outside that area, the user must move the view toward that point.

A more technology independent issue is letting go of an object. As discussed in Section 4.1, it was observed that the point at which the virtual object is released is critical for accurate placement of the object. Many users experienced virtual objects "sticking" to their thumb while they were already retracting their hand.

Another note to be made is that the system suggests a higher level of naturalness and realism than it can support. When users first see their "own" hand as a virtual hand, moving the same way as they feel it moving, many of them seem to expect it to behave much the same as their real hand would, including interaction with objects. Without instructions, the first response of many users to a virtual cube is to grab it with the whole hand, while that is actually one of the hardest hand poses to track.

Some participants also commented about a lack of feedback during the interactions. Some mentioned haptic feedback as a way to make the interaction feel more natural, others suggested more visual feedback as a way to communicate the internal state of the software. For example, a (clearer) visual signal when the virtual hand entered its "pinching" state.

### 7.4. Discussion

Our hypothesis that a hand-based VR interface would allow faster 3D object manipulation was not yet completely supported by the results. Our prototype was generally received with enthusiasm and participants' performance could not be distinguished between the new media (VR and hand tracking) and the traditional mouse and screen; however, some issues should be addressed if this novel interaction technique is to outperform the latter.

The results and observations of our experiment suggest that there are some issues yet to overcome before the proposed type of interaction could exceed the accuracy, user training, and feature-richness of the established mouse-and-screen interface.

Tracking errors were one issue that seemed to inhibit task efficiency. The limited working area and accuracy of hand pose tracking did not allow the users to interact as freely and naturally as is necessary for a productive ideation process.

Another issue was found in the interaction design: since the virtual hands look very natural and follow the user's real hand quite accurately up to some point, the interface suggested to the users that the interaction would be very natural as well. Instead, the interactions and effective hand poses were limited. This caused a mismatch between the user's gestures and the gestures that could be recognized by the system and the tracker. In terms of the *Gulfs of execution and evaluation*

of Norman (1986), we attempted to bridge the Gulf of Execution by bringing the interaction closer to the user's natural intentions. However, by suggesting interaction the system cannot support, the user's expectations of the system state as a result of their actions are not met. Therefore, the Gulf of Evaluation should be bridged to a greater extent, by providing the user with feedback that accurately reflects the system's complexity and capabilities. It should be investigated which type of feedback is most effective in enabling the user to make this mental mapping.

The results in our experiment give an indication of the comparison between hand-tracking interfaces in VR and conventional interfaces, with regards to 3D object manipulations.

An interesting result was the significant difference between the hands and screen interface, compared to the hands and VR interface. The VR component of our interface seems to have a positive effect on the visual perception of 3D space, allowing us to better perceive distance toward the objects we are interacting with. This is backed up by the questionnaire results, in which users reported having less trouble perceiving depth in this Hands & VR condition. A similar conclusion was drawn by Loup-Escande, Jamet, Ragot, Erhel, and Michinov (2017), after an experiment in which participants were more successful in a spatial learning task using stereoscopic display than those using monoscopic display.

The results from the questionnaires confirm that our proposed Hands & VR interface performs similarly to the traditional Mouse & Screen interface on most measured aspects, with a few differences.

Apparently, object rotation was found easier using a mouse and screen, which may be due to restricted wrist movement in the hand-tracking interface. While we estimate a wrist to be able to rotate an object for 100° at most before requiring a new grasp, the mouse-based interface allows infinite rotation by using 2D handles.

Furthermore, the mouse was considered more reliable as an input device than the hand-tracking Leap Motion. As consumer-grade hand-tracking technology is still in its early phases, we hope this difference will soon disappear. We expect the ease of manipulation (moving, rotating, scaling, etc.) to improve with it.

According to the task load results, the third dimension provided by hand tracking comes at the cost of more physical demand. The arms cannot rest on a desk while gesturing, as they can while using a mouse. This would make the interface less suitable for extended use. It would be better suited for short, early conceptual design sessions.

Two final notes on the experiment should be considered. First, as a representative of the mouse-based interaction, the Unity3D Editor was chosen. Although this allows an equal comparison between hand-based interaction and mouse-based interaction using the same software in the same environment, the Unity3D Editor, as a game engine, may or may not be an accurate representative of the mouse-based CAD. We have not tested how the Unity Editor controls compare to other CAD interfaces. However, the interface had similar features to other CAD software, such as draggable handles on the object and axis-locked transformations and 2D transformations. In any case, it allowed us to measure, to some

extent, the coordination of the participants during basic object manipulations.

Second, the tasks, in which participants were asked to transform one shape into another, were created by the experimenter. For each task type, several tasks were created manually to fit within that task type. The 243 recorded tasks therefore consist of a multiple of approximately 15 unique tasks (6–9 per participant). This allowed for more equal comparison between conditions; however, the possibility of the tasks not being representative for the task types was not specifically probed.

Our hand tracking, VR interface did not yet outperform the mouse and screen interface; however, it also did not perform worse. The mouse has been developed over many years to become the leading pointing device in human-computer interaction. It has an integrated role in our society: most people learn to use one already at a young age. In comparison, most participants in this study used the hand-tracking input device for the first time. That being said, (consumer) hand-tracking technology is still in its infancy and will probably receive a large development boost in the next few years. Performance may improve with user practice, along with many configurable parameters in hand tracking technology, which is similar to the conclusion by Canare, Chaparro, and He (2015) after a comparative study between hand/gaze tracking and a mouse for 2D pointing tasks.

With further improvements in tracking accuracy and research in hand-based interaction models, along with the development of VR, a 3D hand-based VR conceptual design tool will have a strong reason for existence in the near future.

## 8. Conclusions and future work

CAD software is suitable for later stages of design but often less practical for the conceptual design phase. The required detail of the input inhibits the flow of ideas and the traditional 2D mouse and screen based interface requires 2D–3D conversion of input and output, restricting the natural mapping from the physical to the virtual world. While many researchers replace the 2D mouse with a 3D pointing device, we investigated the use of hand gestures and VR as a natural, intuitive 3D interface.

We proposed a gestural interaction model for conceptual design in VR. The gestures are based on natural manipulative and communicative gestures that stimulate thought, insight and conceptualization.

Participants using the VR and hand-tracking interface were, in general, not detectably slower nor faster than those using the mouse and screen interface on basic 3D object manipulations. For tasks that included moving and rotating as well as scaling of objects, participants with hand tracking and VR seemed to outperform those with the mouse and screen in efficiency, and with a lower variance. However, this difference was not yet found to be statistically significant.

Participants using VR and hands outperformed those using only hands and a screen. They also reported having less trouble with depth perception, compared to both screen conditions, suggesting the VR headset has a positive effect on

eye-hand coordination, improving the efficiency at which users can manually interact with objects in virtual space.

Improvements could be made in both the interaction design and tracking technology. The interface should clearly reflect which gestures are (not) supported by the system. A model of a physical hand can therefore be confusing if it does not support the perceived affordance. The used tracking technology did not support all user intended interactions during the basic object manipulation tasks. Improved tracking technology could therefore also contribute toward a more effective conceptual design interface.

In the experiment, we tested three conditions: Mouse & Screen, Hands & Screen, and Hands & VR. One (possibly very interesting) condition could also be explored: Mouse & VR. To measure the impact of the VR display independently from the hand-tracking interface, it would be very interesting to include this condition in an experiment. However, this would bring an important issue to take into account: the combination of a 2D input device with a 3D display would require a new interaction paradigm. The mouse pointer (which is able to move in 2D) needs to be displayed in 3D space. Therefore, a  $z$ -value must be chosen for the  $x, y$ -location of the cursor. Furthermore, on a 2D screen, object selection can be detected simply by comparing  $x$  and  $y$  values. In 3D space on the 2D screen, a “ray” is cast from the screen into 3D space and the first hit object is selected. In a 3D (stereoscopic) interface, the origin of the ray cannot be chosen in such a way that, for both eyes, the cursor occludes the point clicked in 3D space, as long as the cursor has a fixed  $z$  value from the camera.

This research focused on an interface using free hand-gesture interaction in VR; however, other setups could still be explored. For instance, an AR setup in which virtual objects are overlaid on the image of the user’s hands would allow the use of real objects as “tools” to manipulate virtual objects (Datcu, Lukosch, & Brazier, 2015). Furthermore, allowing users to specify their own set of gestures (Bordegoni, 1994) might take more time to setup the system, but for those users could, in the long run, improve efficiency.

For further exploration of the potential of hand-based interaction in 3D conceptual design, learning rate could be measured in a longitudinal study over an extended period of time. Moreover, measuring distance in manipulation tasks would give insight into the performance at different scales.

The focus of this experiment was on using off-the-shelf components, to investigate the feasibility of the proposed interaction model using current mainstream technology. For insight in the future potential of the interaction model, the experiment could be repeated using state-of-the-art components, such as high-accuracy data gloves and a high resolution VR headset. Such systems may have a reduced error rate compared to the interfaces used in this experiment and even outperform mouse and screen interface. With the presented data as a baseline, the development of new technologies can be monitored over time in longitudinal follow-up studies. The availability of benchmark data for new technologies would help to shape improvements.

Our experiment provides insight on the usability of a gesture-based VR interface in virtual object manipulations; however, further research is required to evaluate its

effectiveness in conceptual design. Therefore, a complete prototype of the proposed interaction model should be compared to traditional tools in real world conceptual design tasks, using metrics based on models of creativity and ideation effectiveness (e.g., Shah, Smith, & Vargas-Hernandez, 2003; Suwa et al., 2006).

## References

- Alibali, M. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, 5(4), 307–331. doi:10.1207/s15427633scc0504
- Benko, H., Ishak, E., & Feiner, S. (2005). Cross-dimensional gestural interaction techniques for hybrid immersive environments. *IEEE Proceedings. VR 2005. Virtual Reality, 2005*, 209–217. doi:10.1109/VR.2005.1492776
- Blackler, A. L., Popovic, V., & Mahar, D. P. (2002). Intuitive Use of Products. In *Common Ground. Design Research Society International Conference 2002* (pp. 120–134). London, UK: Design Research Society. Retrieved from <http://eprints.qut.edu.au/1879/>
- Bordegoni, M. (1994). Parallel use of hand gestures and force-input device for interacting with 3d and virtual reality environments. *International Journal of Human-Computer Interaction*, 6(4), 391–413.
- Borst, C. W., & Indugula, A. P. (2006). A spring model for whole-hand virtual grasping. *Presence*, 15(1), 47–61.
- Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability Evaluation in Industry*, 189, 4–7.
- Buchmann, V., Vjolic, S., Billingham, M., & Cockburn, A. (2004). FingARtips Å S, Gesture based direct manipulation in augmented reality. *Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, 1(212), 212–221.
- Canare, D., Chaparro, B., & He, J. (2015). A comparison of gaze-based and gesture-based input for a point-and-click task. *International Conference on Universal Access in Human-Computer Interaction*, 9176, 15–24. doi:10.1007/978-3-319-20681-3
- Cash, P., & Maier, A. (2016). Prototyping with your hands: The many roles of gesture in the communication of design concepts. *Journal of Engineering Design*, 27(1–3), 118–145. doi:10.1080/09544828.2015.1126702
- Coelho, J., & Verbeek, F. (2014). Pointing Task Evaluation of Leap Motion Controller in 3D Virtual Environment. In *Creating the Difference: Proceedings of the Chi Sparks 2014 Conference*. (p. 78). The Hague, The Netherlands: The Hague University of Applied Sciences and Chi Nederland. Retrieved from <http://chi-sparks.nl/2014/proceedings/>
- R Core Team. (2014). *R: A language and environment for statistical computing* (Computer Software Manual). Vienna, Austria.
- Datcu, D., Lukosch, S., & Brazier, F. (2015). On the usability and effectiveness of different interaction types in augmented reality. *International Journal of Human-Computer Interaction*, 31 (3), 193–209.
- Deering, M. F. (1995, September). HoloSketch: A virtual reality sketching/animation tool. *ACM Transactions on Computer-Human Interaction*, 2(3), 220–238. doi:10.1145/210079.210087
- de Ruiter, J. P. (2000). The production of gesture and speech. *Language and Gesture*, 2, 284.
- Doherty, J. (1985, January). The effects of sign characteristics on sign acquisition and retention: An integrative review of the literature. *Augmentative and Alternative Communication*, 1(3), 108–121. doi:10.1080/07434618512331273601
- Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, 91, 451–471.
- Encarnação, L. (1999). A translucent sketchpad for the virtual table exploring motion-based gesture recognition. *Computer Graphics Forum*, 18 (3), 277–286.
- Fiorentino, M., de Amicis, R., Monno, G., & Stork, A. (2002). Spacedesign: A Mixed Reality Workspace for Aesthetic Industrial Design. In *Proceedings of the 1st International Symposium on Mixed*

- and *Augmented Reality* (p. 86). Washington, DC, USA: IEEE Computer Society. Retrieved from <http://dl.acm.org/citation.cfm?id=850976.854976>
- Gibbs, R. W. J. (2006). *Embodiment and cognitive science*. Cambridge, UK: Cambridge University Press.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX: Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183.
- Hostetter, A. B., & Alibali, M. W. (2004). On the tip of the mind: Gesture as a key to conceptualization. In *Proceedings of the twenty-sixth annual conference of the cognitive science society* (Vol. 26). Chicago, IL: Cognitive Science Society.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363.
- Igarashi, T., Matsuoka, S., & Tanaka, H. (2007). Teddy: A sketching interface for 3D freeform design. In *ACM SIGGRAPH 2007 Courses*. New York, NY: ACM.
- Jerald, J., Mlyniec, P., Yoganandan, A., Rubin, A., Paullus, D., & Solotko, S. (2013). Makevr: A 3d world-building interface. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on* (pp. 197–198). Orlando, FL.
- Loup-Escande, E., Jamet, E., Ragot, M., Erhel, S., & Michinov, N. (2017). Effects of stereoscopic display on learning and user experience in an educational virtual environment. *International Journal of Human-Computer Interaction*, 33 (2), 115–122.
- Luftig, R. L., & Lloyd, L. L. (1981). Manual sign translucency and referential concreteness in the learning of signs. *Sign Language Studies*, 30(1), 49–60.
- Mandel, M. (1977a). Iconic devices in American sign language. *On the Other Hand: New Perspectives on American Sign Language*, 1, 57–107.
- Mandel, M. (1977b). Iconicity of signs and their learnability by non-signers. In *Proceedings of the first national symposium on sign language research and teaching* (pp. 259–266). Chicago, IL.
- Marsh, T., & Watt, A. (1998). Shape your imagination: iconic gestural-based interaction. In *Virtual Reality Annual International Symposium, 1998. Proceedings, IEEE 1998* (pp. 122–125). <http://doi.org/10.1109/VRAIS.1998.658465>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, 117, 411–424.
- Naya, F., Jorge, J., Conesa, J., Contero, M., & Gomis, J. M. (2002). Direct modeling: from sketches to 3D models. In *Proceedings of the 1st Ibero-American Symposium in Computer Graphics SIACG* (pp. 109–117). Guimarães, Portugal
- Nevins, J. L., & Whitney, D. E. (1989). *Concurrent design of products and processes: A strategy for the next generation in manufacturing*. New York, NY: McGraw-Hill Companies.
- Norman, D. A. (1986). Cognitive engineering. *User Centered System Design: New Perspectives on Human-Computer Interaction*, 3161, 31–61.
- Oh, J.-Y., Stuerzlinger, W., & Dadgari, D. (2006). Group selection techniques for efficient 3D modeling. In *3D User Interfaces, 2006. 3DUI 2006. IEEE Symposium on* (pp. 95–102). Alexandria, VA.
- Pereira, J., Jorge, J., Branco, V., & Ferreira, F. (2000). Towards calligraphic interfaces: sketching 3D scenes with gestures and context icons. In *WSCG '2000: Conference proceeding: The 8th International Conference in Central Europe on Computers Graphics, Visualization and Interaktive Digital Media '2000 in cooperation with EUROGRAPHICS and IFIP WG 5.10* (pp. 314–321). Plzen, Czech Republic: University of West Bohemia.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2014). *{nlme}: Linear and Nonlinear Mixed Effects Models. R Foundation for Statistical Computing*. Retrieved from <https://cran.r-project.org/package=nlme>.
- Quek, F., Mcneill, D., Bryll, R., & Mccullough, K. E. (2002). Multimodal human discourse: Gesture and speech university of illinois at Chicago. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(3), 171–193.
- Rahimian, F. P., & Ibrahim, R. (2011, May). Impacts of VR 3D sketching on novice designers: A Z' spatial cognition in collaborative conceptual architectural design. *Design Studies*, 32(3), 255–291. doi:10.1016/j.destud.2010.10.003
- Ruiter, J. D. (2000). The production of gesture and speech. *Language and Gesture*.
- Sadeghipour, A., & Morency, L.-P. (2014). *3DIG - 3D Iconic gesture dataset*.
- Sadeghipour, A., Morency, L.-P., & Kopp, S. (2012). *Gesture-based object recognition using histograms of guiding strokes*. In *Proceedings of the British Machine Vision Conference* (p. 44.1–44.11). Durham, UK: BMVA Press.
- Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2003, March). Metrics for measuring ideation effectiveness. *Design Studies*, 24(2), 111–134. doi:10.1016/S0142-694X(02)00034-0
- Shen, Y., Ong, S.-K., & Nee, A. Y. (2011). Vision-based hand interaction in augmented reality environment. *International Journal of Human-Computer Interaction*, 27(6), 523–544.
- Sowa, T., & Wachsmuth, I. (2002). Interpretation of shape-related iconic gestures in virtual environments. In I. Wachsmuth & T. Sowa (Eds.), *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop, GW 2001 London, UK, April 18&20, 2001 Revised Papers* (pp. 21–33). Berlin, Heidelberg, Germany: Springer Berlin Heidelberg.
- Sturman, D. J., & Zeltzer, D. (1993). A Design Method for “Whole-hand” Human-computer Interaction. *ACM Trans. Inf. Syst.*, 11(3), 219–238.
- Suwa, M., Gero, J., & Purcell, T. (2006). *Unexpected discoveries and s-inventions of design requirements: A key to creative designs* (Computational Models of Creative Design IV, Key Centre of Design Computing and Cognition). Sydney, Australia: University of Sydney.
- van Dijk, C. G. (1995, January). New insights in computer-aided conceptual design. *Design Studies*, 16(1), 62–80. doi:10.1016/0142-694X(95)90647-X
- VRClay. (2014). *VRClay - Sculpting in virtual reality*. Retrieved from <http://vrclay.com/>
- Wang, L., Shen, W., & Xie, H. (2002). Collaborative conceptual design: A state of the art and future trends. *Computer-Aided Design*, 34(13), 981–996.
- Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, 114, 591.
- Zelevnik, R., Herndon, K., & Hughes, J. (2007). SKETCH: An Interface for Sketching 3D Scenes. In *ACM SIGGRAPH 2007 Courses*. New York, NY, USA: ACM.

## About the Authors

**Remi Alkemade** ([remi@remialkemade.nl](mailto:remi@remialkemade.nl)) is a computer scientist with an interest in new interaction modalities and artificial intelligence; he is an independent researcher who graduated in early 2015 from the Media Technology master's program at the Leiden University in the Netherlands.

**Fons J. Verbeek** ([f.j.verbeek@liacs.leidenuniv.nl](mailto:f.j.verbeek@liacs.leidenuniv.nl), <http://liacs.leidenuniv.nl/~verbeekfj>) is head of the Imagery & Media cluster of Leiden Institute of Advanced Computer Science; he heads the section Imaging & BioInformatics. Apart from data analysis in images, he is specifically interested in image interaction and interactivity in data spaces. Hence, new interaction devices are studied in his research group in correspondence projects in data analysis.

**Stephan G. Lukosch** ([s.g.lukosch@tudelft.nl](mailto:s.g.lukosch@tudelft.nl), <http://www.tudelft.nl/sglukosch>) is a computer scientist with an interest in virtual co-location using augmented reality; he is an Associate Professor in the Faculty of Technology, Policy, and Management at Delft University of Technology.

## Appendix A. QUESTIONNAIRES

The following are the questions from the questionnaires given to participants after the trials. Participants could rate each question on a 5-point Likert scale (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree).

### SUS questions

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought that the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found that the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

### TLX questions

1. The tasks were mentally demanding.
2. The tasks were physically demanding.
3. The tasks were very rushed or hurried.
4. I was successful in accomplishing what I was asked to do.
5. I had to work hard to accomplish my level of performance.
6. I was insecure, discouraged, irritated, stressed, or annoyed.

### Extra questions

1. I could easily move the shapes to their desired position.
2. I could easily rotate the shapes to their desired state.
3. I could easily scale the shapes to their desired size.
4. I had a good spatial overview while performing the tasks.
5. This tool could help me gain insight into a spatial problem/design.
6. The input device worked according to my intentions.
7. I had trouble perceiving the 3D locations of objects.
8. The interface felt natural.
9. I enjoyed working with this tool.
10. I would prefer designing with my hands over using a mouse (assuming they both flawlessly).
11. I would prefer designing in virtual reality over using a screen.

## Appendix B. QUESTIONNAIRE RESPONSE DISTRIBUTIONS

The figures below illustrate the results of the questionnaires for each answer, grouped by condition.

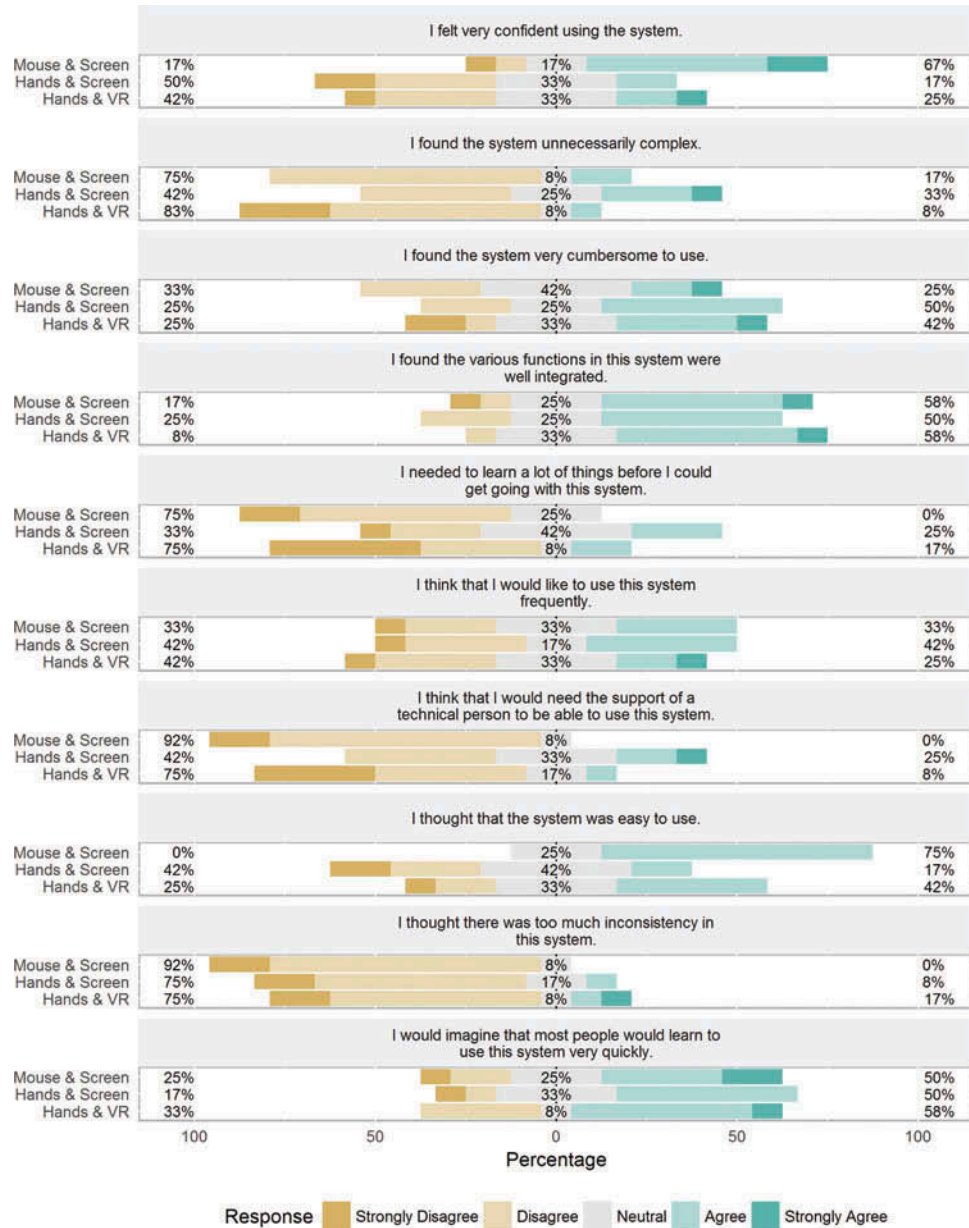


Figure B1. Distribution of answers to SUS questionnaire, grouped per condition.

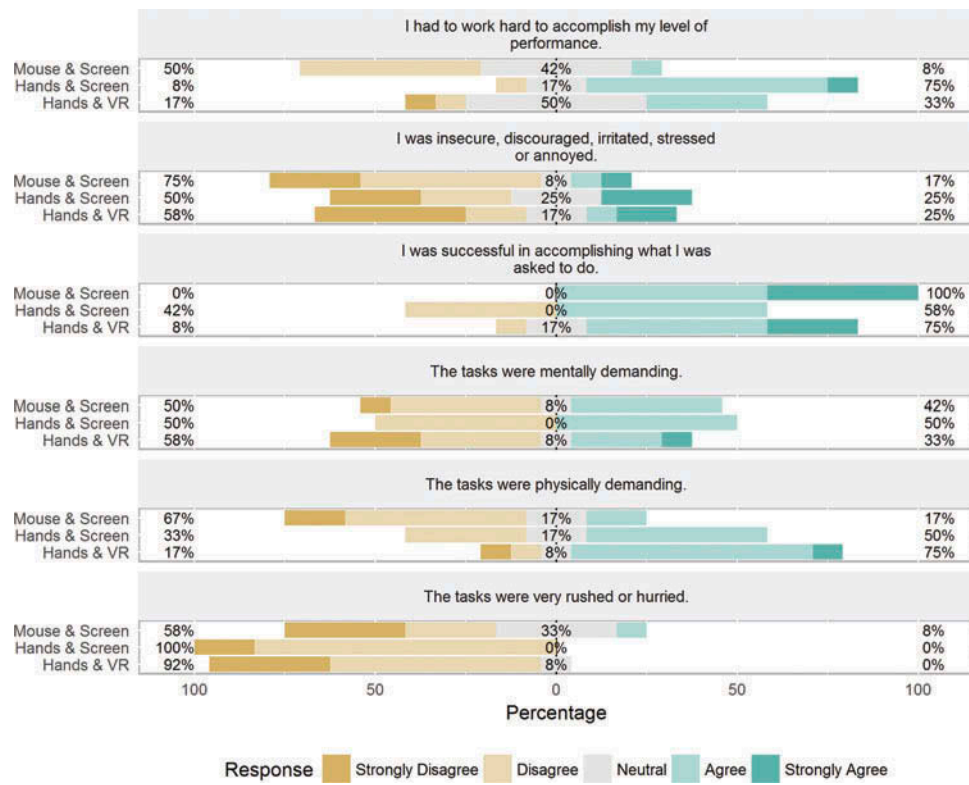


Figure B2. Distribution of answers to TLX questionnaire, grouped per condition.

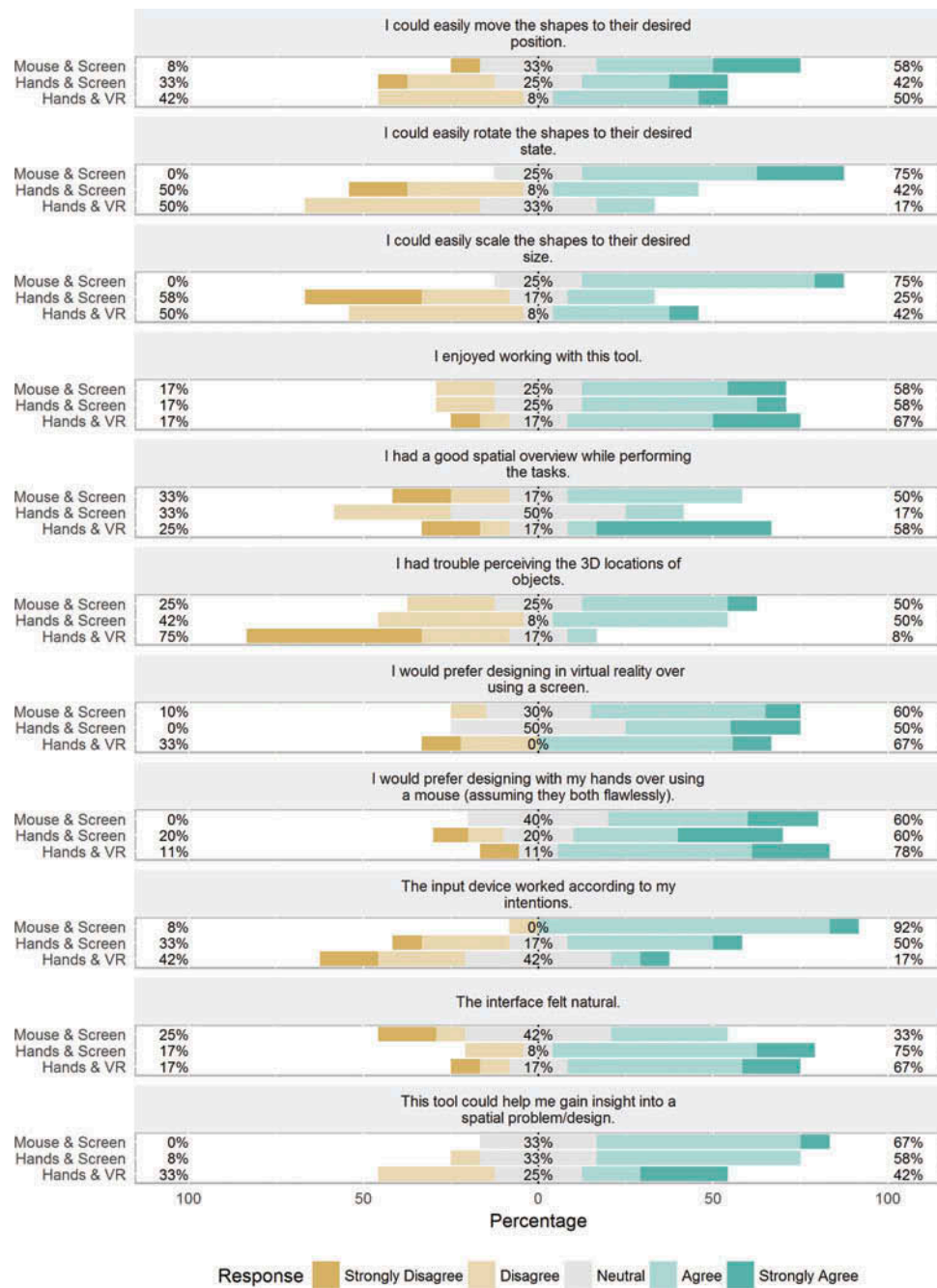


Figure B3. Distribution of answers to extra questions, grouped per condition.

## Appendix C. QUESTIONNAIRE SIGNIFICANCE VALUES

The tables show the results of the Kruskal–Wallis test by ranks and the Dunn *post-hoc* test for group comparisons. Asterisks (\*) indicate significant *p*-values (at significance level  $p < 0.05$ ). Group comparisons are only considered significant if the grouped *p*-value is significant.

Question	$\chi^2$ (2, $N = 36$ )	<i>p</i> Grouped	<i>p</i> M & S - H & S	<i>p</i> M & S - H & V	<i>p</i> H & S - H & V
1	0.12	0.94	1.00	1.00	1.00
2	6.94	0.03*	0.22	0.35	0.01*
3	9.56	0.01*	0.00*	0.11	0.29
4	8.16	0.02*	0.02*	1.00	0.02*
5	0.72	0.70	0.82	1.00	0.62
6	0.72	0.70	0.75	0.65	1.00
7	0.07	0.97	1.00	1.00	1.00
8	0.63	0.73	0.64	0.94	1.00
9	5.88	0.05	0.03	0.13	0.77
10	5.52	0.06	0.10	1.00	0.04

**Figure C1.** SUS Questionnaire Dunn and Kruskal–Wallis test results. M is for Mouse, S is for Screen, H is for Hands, and V is for VR.

Question	$\chi^2$ (2, $N = 36$ )	<i>p</i> Grouped	<i>p</i> M & S - H & S	<i>p</i> M & S - H & V	<i>p</i> H & S - H & V
1	0.91	0.63	1.00	0.84	0.52
2	8.21	0.02*	0.10	0.01*	0.47
3	1.36	0.51	0.66	0.38	1.00
4	9.74	0.01*	0.00*	0.24	0.13
5	11.2	0.00*	0.00*	0.18	0.11
6	0.70	0.71	0.68	1.00	0.73

**Figure C2.** TLX Questionnaire Dunn and Kruskal–Wallis test results. M is for Mouse, S is for Screen, H is for Hands, and V is for VR.

Question	$\chi^2$ (2, $N = 36$ )	<i>p</i> Grouped	<i>p</i> M & S - H & S	<i>p</i> M & S - H & V	<i>p</i> H & S - H & V
1	2.29	0.32	0.55	0.20	0.79
2	7.83	0.02*	0.05	0.01*	0.86
3	5.46	0.07	0.03	0.56	0.26
4	2.55	0.28	1.00	0.24	0.25
5	0.24	0.89	1.00	1.00	0.99
6	6.04	0.05	0.48	0.02	0.21
7	9.76	0.01*	0.95	0.01*	0.02*
8	4.70	0.10	0.07	0.13	1.00
9	0.19	0.91	1.00	1.00	1.00
10	0.16	0.92	1.00	1.00	1.00
11	0.09	0.96	1.00	1.00	1.00

**Figure C3.** Extra Questionnaire Dunn and Kruskal–Wallis test results. M is for Mouse, S is for Screen, H is for Hands, and V is for VR.