

A novel user-based gesture vocabulary for conceptual design

Tijana Vuletic^{ax}, Alex Duffy^a, Chris McTeague^a, Laura Hay^a, Ross Brisco^a, Gerard Campbell^{ab}, Madeleine Grealy^b

^a Department of Design Manufacture Engineering Management, University of Strathclyde, Level 7, James Weir Building, 75 Montrose Street, G1 1XJ,

^b Glasgow School of Psychological Sciences & Health, University of Strathclyde, 6th floor, Graham Hills Building, 40 George Street, Glasgow, G1 1QE

Abstract

Research into hand gestures for human computer interaction has been prolific recently, but within it research on hand gestures for conceptual design has either focused on gestures that were defined by the researchers rather than the users, or those that were heavily influenced by what can be achieved using currently available technology. This paper reports on the study performed to identify a user elicited vocabulary of gestures for conceptual design, disassociated from the currently available technology, and its subsequent evaluation. The study included 44 product design engineering students (3rd, 4th year and recent graduates) and identified 1772 gestures that were analysed to build a novel gesture consensus set of vocabulary of hand gestures for conceptual design. This set is then evaluated by 10 other professionals, in order to generalise this set for a wider range of users and possibly reduce the need for training. The evaluation has shown that majority of gestures added to the vocabulary were easy to perform and appropriate for the activities, but that at the implementation stage the vocabulary may need another round of evaluation to account for the technology capabilities. The aim of this work is to create a starting point for a potential future system that could adapt to individual designers and allow them to use non-prescribed gestures that will support rather than inhibit their conceptual design thinking processes, akin to the developments that happened in hand writing recognition or predictive texting.

Keywords: natural interaction, gesture vocabulary, user based, conceptual design, case study

Word count: 15 796 words, including tables, excluding abstract and the reference list.

1. Introduction

Hand gestures are motions performed by humans' hands. Reviews show that they have been extensively researched since the 80s, as a natural and intuitive means of interaction between humans and computerised systems (Pisharady and Saerbeck, 2015, Rautaray and Agrawal, 2015, Santos¹ et al., 2015, Milani et al., 2017, Al-Shamayleh et al., 2018). Hand gestures are explored as a means that allows for touchless interaction providing specific benefits such as hygiene (Lopes et al., 2017) undivided attention focused on the main task (Riener et al., 2013), a more natural fit for interaction with household items (Dinh et al., 2014), large displays (Foehrenbach et al., 2009), or interaction with three-dimensional (3D) objects in Virtual Reality (VR)/Augmented Reality (AR) supported spaces (Kim and Lee, 2016, Memo and Zanuttigh, 2018).

Conceptual design is the initial, fundamental outline of a product during which ideas are explored before the design specification is established (Keinonen and Takala, 2010, pg 17). Hand gestures have also been explored in the field of conceptual design. However, in the applications which have resulted in a prototype the gestures employed are either prescribed, where the authors identify the need for further exploration of more appropriate gestures (Huang et al., 2019), or free-form, which allow natural and efficient but limited types of interactions (Vinayak and Ramani, 2015). Where hand gestures have been explored in the context of conceptual design without practical application, but considering user based input, the work performed has been linked to envisaged Computer Aided Design (CAD) application (Khan and Tunçer, 2019). User based research has been established as an effective gesture elicitation method. It results in a vocabulary of gestures for a specific purpose that is likely to be adopted by the users, and can, if needed, be used to eliminate technological limitations

from consideration (Wobbrock et al., 2009, Wobbrock et al., 2005, Ruiz et al., 2011, Piumsomboon et al., 2013, Khan and Tunçer, 2019). However, use of CAD has been found to interrupt the designers thinking, as instead of thinking about the design they tend to focus on commands and procedures used in CAD for specific shape creation (Huang, 2007).

As the gestures used in interfaces have typically been defined by designers or driven by technological capability, the aim of this paper is to instead identify a vocabulary of gestures that designers would naturally use if there were no limitations. Hand gestures are thus explored for conceptual design in isolation from the technology enabling their implementation, which has not been done in the past. Established user based research approach is applied, disregarding what is possible to achieve with existing technology, in order to identify the most natural and intuitive gestures. Exploring what the designers would naturally do instead of allowing the gesture definition to be driven by what is technologically possible at any given time could be used to identify the requirements that need to be met by a future technology and potentially even drive future implementations. The goal of the research reported in this paper is to explore gestures that are most suitable for the conceptual design process and most likely to support and enhance the design process instead of interrupting it. It is envisaged these gestures would be utilised in a hypothetical future system akin to CAD currently used for detailed design, although at this stage the system itself is not explored and the focus is instead on gestures that may inspire the development of a more appropriate system to support the conceptual design.

This paper reports a user based study of hand gestures for conceptual design, focusing on identification of natural and intuitive hand gestures, addressing the gap identified in the field. The aim of the study is to create a gesture vocabulary for conceptual design, focusing on the user preferences, in terms of how they would use their hands to create, modify and manipulate 3D objects shown on a screen in front of them. The participants, final year design students, were not given any instructions on how to interact with the objects, other than to use their hands. Instead, they are shown different objects on a screen placed on a wall across from the participants, and asked to emulate how they would interact with the objects using their hands. From that, various gestures for creation, manipulation and modification of 3D objects are identified. The gesture vocabulary is evaluated by a different group of other professionals, in order to explore the suitability of the vocabulary for user groups outside of the niche conceptual design application. Figure 1 illustrates the elements of the research reported in this paper, and their key outcomes. Green boxes illustrate user based studies. Two boxes on the left illustrate a user centred gesture study (consisting of two parts) during which a vocabulary/consensus set of gestures for conceptual design is identified, which is considered a primary outcome of this work. Then this vocabulary of gesture is evaluated. Set of green boxes on the right illustrates the evaluation of the consensus set, performed for the full consensus set abstractly, and for the 16 implemented gestures using a VR prototype. Outcome of the evaluation is a variation of the vocabulary, taking into consideration usability and implementation limitations, along with recommendations for future work. Outcomes are represented using blue boxes. Yellow boxes indicate pilot studies that were used to test and refine the approach taken in the user centred gesture study.

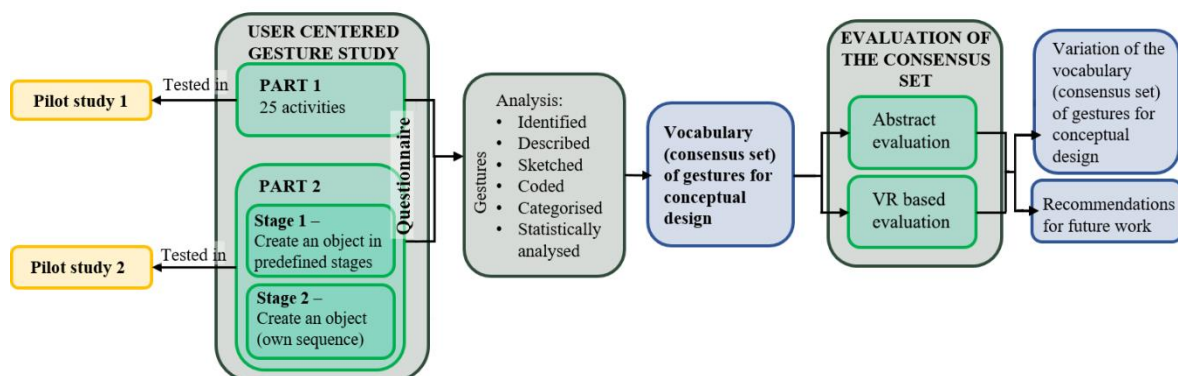


Figure 1 Elements of the research reported in the paper

Section 2 provides a short review of the related work present in the literature. Sections 3 and 4 introduce the methodology used in the user centered gesture study and the analysis approach, respectively. The study results are reported in Section 5, including the key findings and the consensus

set of vocabulary of gestures for conceptual design. Section 6 details the evaluation study performed to test the initial findings and refine the consensus set of vocabulary of gestures for conceptual design. Section 7 provides a discussion of the contributions and explores some potential future research directions. Then the conclusion is given in Section 8.

2. Related work

Gestures are frequently used in novel interaction approaches, primarily because they are habitually used in human communication processes. The use of gestures is even more frequent when the interaction includes physicality of a three-dimensional (3D) object e.g. interaction with a 3D/VR/AR environment (Rautaray and Agrawal, 2015, Koutsabasis and Vogiatzidakis, 2019, Vuletic et al., 2019).

2.1. Gestures in human communication and interaction with objects

Hand gestures are a mode of human expression, both in terms of interpersonal communication and interaction with objects surrounding humans (Zimmerman et al., 1987, Buchmann et al., 2004). When they are used for communication they often accompany speech (McNeill, 1992, Quek, 2004). However, this is not always the case, and they can be used for communication independently of speech. For instance, pantomimic gestures can be used as a sole communication mode e.g. to say “no” a person may wave their finger left to right, or show “thumbs up” to indicate agreement. Depending on the context, hand gestures can be further classified. These sub categories will not be further explored in this paper, as they largely refer to speech-linked gestures, which are not the key focus of this research. At this stage the focus is on gestures only and the identification of activities performed in conceptual design that could be well supported using gestures. In the remainder of this paper, hand gestures will be referred to as gestures.

Whether communicative or interactive, gestures have a meaning and this meaning needs to be understood in the same manner by all entities involved in the process. In the past, this was achieved through use of prescribed gestures (Isbister and Mueller, 2015, Vuletic et al., 2019). Prescribed gestures are predefined in a vocabulary created prior to their implementation. While set number of gestures may aid the implementation in a system, learning and using predefined gestures is not the most natural and intuitive way of interacting with an interface (Huang et al., 2019, Vuletic et al., 2019). Interaction with the interface discussed in this paper focuses on in-air gestures to create, modify or manipulate 3D objects, using various implementation technologies (AR/VR, holograph, screens or surfaces). Some examples of gestures used for interaction are: in-air sketches, symbols and hand poses recognised as commands for creation of 3D objects (Huang et al., 2019), pulling two hands in a pinch pose apart to zoom when interacting with holographic surfaces (Noor and Aras, 2015), point to select when using virtual pottery (Vinayak et al., 2013), point or pinch to zoom in/out, spread, drag or rotate when manipulating a 3D cardiac model (Riduwan et al., 2013) etc. When prescribed gestures were used, those defined by the users themselves have been shown to be much more memorable than gestures pre-designed by researchers (Wobbrock et al., 2009, Nacenta et al., 2013). Interaction with objects may require natural gestures that do not conform to predefined rules and are instead created by the subject involved in the interaction. They have been found to have higher usability rates than arbitrary predefined gestures, higher sense of control, they were more immersive and were seen to contribute to better interface quality (Bailey et al., 2018).

2.2. Effect of recent advancements in technology on gesture research

Use of hand gestures for interaction is not new, however expansion in numbers of applications incorporating hand gestures seems to coincide with technology developments that have occurred in the past decade (Vuletic et al., 2019). Introduction of sensors such as Kinect (Kinect, 2018) and LEAP (LEAP MOTION INC., 2018) that are relatively cheap, portable and supported by Software Development Kits (SDKs) simplified hand detection and recognition by reducing the need for development of independent algorithms for every step in the process, and provision of shared databases. Technology development is constant, and now both Kinect and LEAP (initially released in 2010 and 2013 respectively) have been replaced with versions with enhanced functionality Azure Kinect DK (released in 2020), Orbbec (released in 2015), Ultraleap (released in 2019), Intel RealSense (released in 2018) etc. These types of contactless sensors, along with traditional cameras and image recognition for detection, have been applied in a variety of fields. For example, hand gesture based interaction is used to reduce the need for user training prior to application use (Buchmann et al., 2004, Kim et al., 2005a, Beyer and Meier, 2011). Sensor based hand detection enables touchless operation guaranteeing sterility or safer interaction in the medical field (Lopes et

al., 2017). Hand gesture or full body gesture interaction supports immersion using Virtual Reality (VR) or Augmented Reality (AR) (Deller et al., 2006). Use of hands aids expression of spatial concepts when externalising ideas (Vinayak et al., 2013). While driving eye contact with the road can be maintained if sensor based hand detection is used to control comfort functions in a car (Riener et al., 2013). Some applications focus on inclusion of older population, by allowing them to control household functions using hand gestures (Bhuiyan and Picking, 2011). A drawback of the applications inspired by technological developments has been that often the gestures used are those that are easily recognisable by a specific system (Schmidt, 2015). To avoid premature focus on an approach that prioritises ease of implementation over needs of the designers, in the first instance we explore the gestures suggested by designers without any technological limitations or consideration of the application. These gestures may evolve over time, once the appropriately advanced technology makes implementation possible.

2.3. Gestures for design

During conceptual design ideas are externalised by designers, and can be vague, incomplete and frequently changing (Liddament, 1999, Varga et al., 2007, Zhong et al., 2011). While this aids idea development and leads to more creative solutions, it does not align well with the methods currently employed in computer based interaction systems, which require fully formed and defined shapes at specific points in the process (Shesh and Chen, 2004, Alcaide-Marzal et al., 2013). These are typically available during the detailed design stage, and well supported by CAD systems (Hartman, 2009). Use of hand gestures for shape creation has been explored in the field of Human Computer Interaction (HCI). In this paper, shape creation was defined as the use of hand gestures to create a new shape in an empty space. Shape modification was defined as interaction with a shape in order to change its geometric characteristics. Shape manipulation was defined as an activity that changes its position in space, translates, rotates or scales it, but does not change it (Vinayak et al., 2013). These activities were explored to different extents in a number of studies, and typically linked to 3D CAD systems. Hand gestures were used for 3D architectural urban planning (Buchmann et al., 2004, Yuan, 2005), cable harness design (Robinson et al., 2007), CAD (Dani and Gadh, 1997, Kim et al., 2005b, Qin et al., 2006, Holz and Wilson, 2011, Kang et al., 2013, Vinayak et al., 2013, Arroyave-Tobón et al., 2015, Huang et al., 2018), or manipulation of already created objects (Chu et al., 1997, Kela et al., 2006, Qin et al., 2006, Bourdot et al., 2010, Kang et al., 2013, Vinayak et al., 2013, Song et al., 2014, Beattie et al., 2015, Noor and Aras, 2015, Xiao and Peng, 2017), and virtual pottery (Dave et al., 2013, Han and Han, 2014, Vinayak and Ramani, 2015). The applications typically used free-form gestures for the creation of splines or surfaces that build up a 3D model (Chu et al., 1997, Buchmann et al., 2004, Kim et al., 2005a, Robinson et al., 2007, Holz and Wilson, 2011, Vinayak et al., 2013, Han and Han, 2014, Arroyave-Tobón et al., 2015, Vinayak and Ramani, 2015). Some intermediate activities, for example selection, were performed using simple prescribed gestures such as pinch or hand grasp. In some applications shapes were created by drawing a profile that is then swept in space along a path, which in itself was a prescribed activity similar to a sweep command in CAD, and were then further modified using free-form and parametric deformation and manipulation (Vinayak et al., 2013, Huang et al., 2018). The gestures were typically chosen by the researchers, not elicited from the population that will ultimately be using them, and the appropriateness of these gestures for the activities they were used for was often not evaluated.

2.4. Existing elicitation studies

While there has been an expansion in numbers of gesture based interfaces, the gestures used in them were typically defined by the researchers developing the studies, often due to ease of application or alignment with the technology used. This has been acknowledged in the field of design (Huang et al., 2018) and HCI in general (Piumsomboon et al., 2013). A study by Jahani and Kavakli (2018) used user elicited gestures for design, which resulted in high level findings that do not explicitly identify specific hand gestures, and where application seemed to diverge from design into car comfort control interfaces.

User-based gesture elicitation has been explored in the field of HCI for AR environments (Piumsomboon et al., 2013), TV control (Dong et al., 2015, Dim et al., 2016), and 3D CAD modelling in conceptual design (Khan and Tunçer, 2019). Approaches followed in them were variations of work by Wobbrock et al. (2009) and Morris et al. (2010), focusing on exploration of the most appropriate gestures for surface technology based interaction. They elicit gestures from inexperienced users, in order to identify the most universally appropriate gestures, and keep the elicitation process itself separate from the technology that may be used during the implementation stage, in order to elicit the most appropriate gestures that are not limited by current capabilities of available technology. There

have been no studies exploring user-elicited gestures for conceptual design in isolation from the technology that may be used to support it and this is the gap the study reported in this paper will address.

3. Developing the user based gesture vocabulary

The key goal of the study was to discover how designers would perform conceptual design if they were not influenced by what was achievable using existing technology and could instead use their hands to create virtual 3D objects in any manner they wished to. Gestures performed were identified, parsed, coded, categorised and analysed for patterns and relations. The most repeated gestures that were not likely to have been repeated by chance then form a gesture vocabulary for conceptual design. The benefit of this approach is that gestures suggested by a larger number of participants are likely intuitive to them, and may in turn be intuitive to other designers (Wobbrock et al., 2009, Nacenta et al., 2013).

3.1. Approach to research design

Elements of established methodologies from the literature exploring user defined gestures were adopted. Typically an effect of the gesture was portrayed in some manner and participants were asked to perform the cause of this effect i.e. gesture causing it (Wobbrock et al., 2009, Piumsomboon et al., 2013). This approach, sometimes referred to as “guessability study methodology”, originates in a study by Wobbrock et al. (2005) aiming to maximise the guessability of symbolic input. In subsequent studies it was expanded to general gesture elicitation from users. In a typical study set up users propose gestures prompted by referents shown to them. Users’ propositions are always acceptable, and “Wizard of Oz” approach is used where the response is emulated by study designer to confirm this acceptance to the users (Lee and Billinghamurst, 2008). This ensures the technology disassociation, and allows elicitation of gestures that are genuinely the most natural to the users. In linguistic terms, the displayed effect of a gesture is a referent, as the gesture performed refers to it (McNeill, 1992). These elements have been implemented in the design of the study reported in this paper. Users were asked to respond to an activity they see enacted on a screen. They suggested a gesture they believed would cause the activity and performed it as they observed the activity unfolding, pretending that they were causing it. This approach was aiming to remove excessive premeditation from the process. They were verbally instructed that all activities they could imagine are possible and achievable. They were told they were in a “magical room” which would instantly recognise the intent of their gestures, to expand their solution space.

The study consisted of two parts. In the first part, participants observed and reacted to a number of predefined activities. The aim was to identify natural and intuitive gestures designers performed for some of the common manipulation and modification activities. In the first part of the study, the activities were discrete, and did not provide information on how participants progressed from one gesture to the other (activities can be seen in Figure 4 in Section 3.4), which would be necessary for the development of any future system. Part two aimed to observe the flow of a conceptual design activity, allow participants to propose both activities and gestures they found appropriate in creation of an object shown to them (activities can be seen in Figure 7 and Figure 8 in Section 3.4).

After both parts were completed, the participants were asked to fill in a questionnaire providing further information on their perceptions of activities performed during the study. The questionnaire was designed by the authors. Typically, elicitation studies pose two questions post study, using Likert scale for the answers. For example, how good of a match for purpose and how easy to perform a gesture was (Wobbrock et al., 2009, Piumsomboon et al., 2013), or how easy the task was and how well were participants able to convey their intentions (Khan and Tunçer, 2019). Original questionnaire was designed in part to collect further information about the appropriateness of the study approach, and in part because the goal was to evaluate the gestures by a different group of participants during the evaluation of the consensus set (given in Section 6) instead of having participants evaluate them during the study. Only time limitations were set. Otherwise, there were no limitations imposed, or instructions given to participants on how to perform the gestures, and it was assumed they performed gestures they found appropriate and easy to perform. However, participants were asked to report any difficulties they had with the activities in the questionnaire. To ensure participants are not adopting the gestures for the activities they were asked to perform in the part one, and implementing them in the part two, one half of participants performed part one first, and the second half performed the part two first.

3.2. Participants

Forty-four 3rd year and above Product Design Engineering (PDE) students participated in the study, aged 22.4 on average. Fifteen were female, and 29 were male. Seven were left handed, 33 right handed and one participant was ambidextrous. They had 4.9 years of CAD experience on average (using a variety of CAD software e.g. Solidworks, Creo, Inventor, Catia, Revit, Sketchup, Rhino, AutoCAD, Smartplant3D, Edgecam NX9, ProEngineer, Alias). They also had an average of 1.4 years of design experience in the professional environment, including internships. They have typically spent at least three years of working on student projects or limited projects with industry, but have not worked in professional design environments, rather than short term. Studies in the field typically use non-experienced or non-technical people, in order to avoid influence of previous experience (Piumsomboon et al., 2013) and to ensure the perspective recorded is that of a system user and not a system builder (Wobbrock et al., 2009). However, conceptual design problems are niche and require a level of creativity and mental manipulation of vague concepts that the general public would not necessarily possess. It would be difficult to reliably test the potential participants against this requirement, hence a decision was made to use design students in this study, as by the 3rd year they are considered to have a sufficient grasp of design, have used CAD in their projects, but have still not fully adopted the traditional way of working. They have also displayed the spatial perceptions skills, creativity and concept manipulation throughout their training. They were considered to have good declarative and procedural knowledge of CAD, and medium strategic knowledge. They were not expected to suffer from bounded ideation i.e. would not focus on CAD tools and how to use them rather than ideation. They were also considered less likely to suffer from circumscribed thinking i.e. limit their design steps to those possible in the CAD software they were familiar with, as their knowledge of both CAD tools and design processes was at an intermediate level. They have all successfully completed design projects requiring intermediate level use of CAD in their first three years of studies, i.e. they are highly skilled in use of frequently used CAD commands but may still need guidance in using more advanced tools such as surfacing for example. They can be considered advanced beginners or novice designers that have key characteristics of designers (Liikkanen and Perttula, 2009). However, in order to reduce the influence of design experience on gesture inclusion into the vocabulary, a group of other professionals evaluated the consensus set. More on this process is reported in Section 6.

3.3. Study setup and sequence of activities

The participants were seated at a table, in front of a large 2D screen the animations and presentations containing images of shapes participants were asked to “create” using their hands were shown on. The screen was on the other end of the room, and out of reach of the participants. Two cameras, one positioned under the screen and one on the users’ left hand side recorded them, and an example of camera views can be seen in Figure 2. A pilot study performed to test the setup has shown that the users imagined the objects shown as 3D. A post study questionnaire in this study posed that question explicitly. In the second part of the study, where participants suggested activities freely, think-aloud protocol (Ericsson and Simon, 1984, Jääskeläinen, 2010) was included to record their thought processes.



Figure 2 Screenshot of one of the participants taking part in the study (front view on the left, side view on the right)

In part one of the study the participants were asked to observe an animation of a 3D rendered part being manipulated or modified (this is referred to as an activity throughout the paper), and react to it

by performing a hand gesture they believe would result in the observed activity i.e. imagine they were causing the activity. A flowchart illustrating this process is shown in Figure 3. The approach, including timings and repetition numbers, was successfully tested in a pilot study.

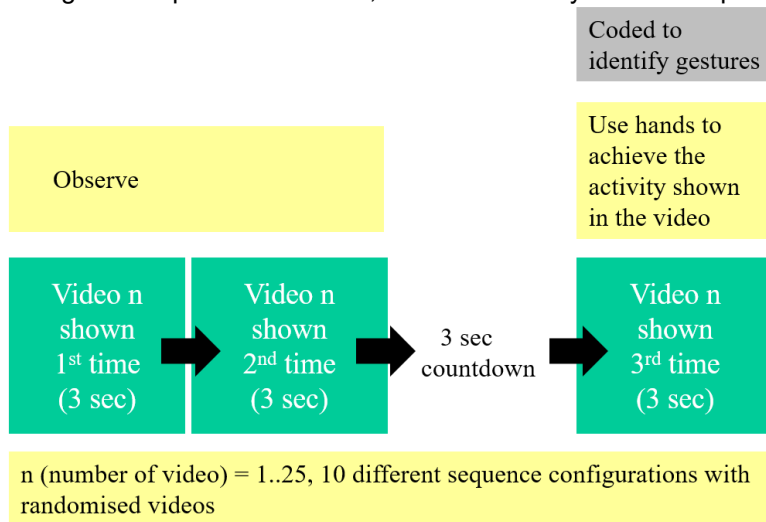


Figure 3 Sequence of activities in Part one

A video containing three repetitions of each activity was played to the participants. A participant watched the first two repetitions, before performing the gesture on the third viewing. Each activity had a three second duration, and between the second a third repetition the screen counted down from three to one in order to give the participants time to get ready to perform the hand gestures concurrently with the third showing of the activity. This time limitation was introduced in order to record the participants' initial reaction, and reduce the likelihood of analogy creation between hand gestures and traditional CAD interaction.

25 activities were shown to the participants, and these are illustrated in Figure 4. They were based on 16 core activities, where some were repeated for different shapes (Figure 4 shows which were repeated): translate up/down/left/right, zoom in/out, rotate clockwise/counter clockwise, select/deselect, extrude cut, extrude cut shallower, extrude up, extrude down. Before that, two or three randomly selected activities were shown to the participants in the training session, in order to ensure they have understood the instructions.

Ten sets of randomized sequences were created, so every tenth participant would perform the same sequence e.g. participants 1, 11, 21, 31, 41 would have performed the activities in the same sequence. Participants were given no further instructions. They were not told what the goal of the study was until both parts were completed. If they asked for more information about how to perceive the object or use their hands e.g. "Is the object on the table in front of me?", they were told to interact with it however they perceive it and use their hands however they wish to. Participants were intentionally not given any instructions on how to use their hands. The goal was to remove as much researcher influence as possible and observe how the participants would choose to interact with the shapes and the space in front of them without constraints imposed on them.

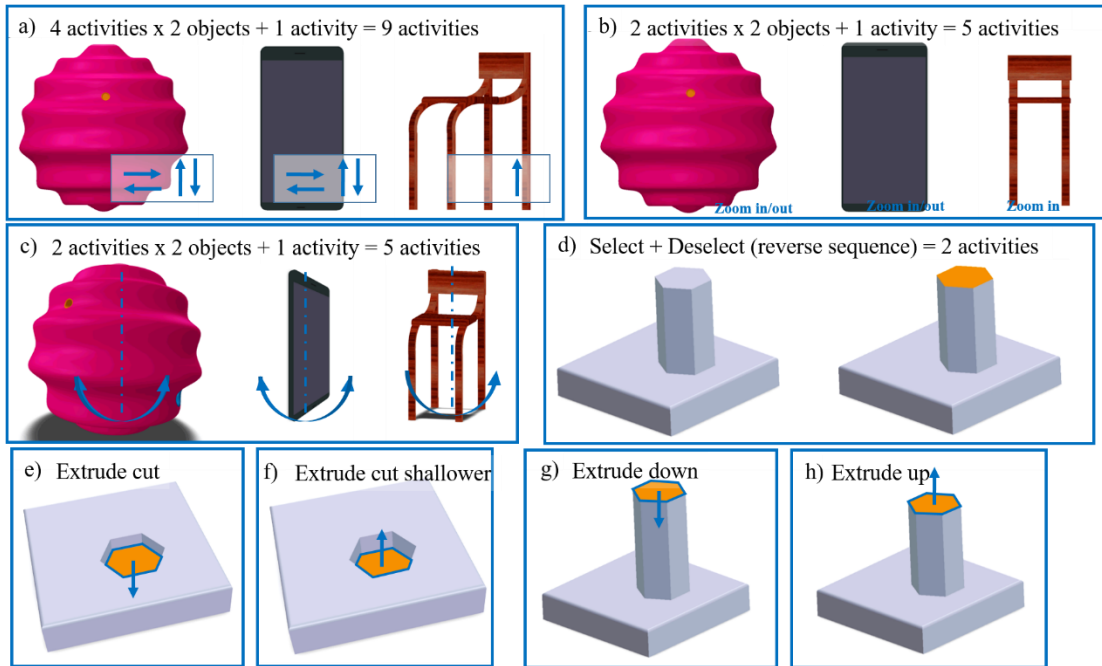


Figure 4 Activities performed in Part one of the full study

The objects animated to incite interaction were either simple and recognisable, with a clearly assigned function e.g. chair and phone, or shapes that were still simple and easy to perceive but without an assigned function e.g. irregular sphere and the console. It was important that the participants can perceive the shape of the object quickly. The variety of object shapes and functions were introduced to explore their effect on the interaction choices.

In part two of the study, an uninterrupted design process was observed, and this part consisted of two stages. In both stages, the participants were shown images on the screen and asked to create the objects shown in the images using their hands. Stage 1 contained a number of predefined steps leading the participant to the completion of the final shape. These predefined steps were included to ensure that information on the hand gestures for key design activities were collected. In stage 2, the participants were only shown the final shape and asked to create it any way they saw fit. Compared to part one, the only difference in the instructions given was that the participants were explicitly told to imagine the object shown on the screen as if it was a 3D object suspended in the “virtual space” in front of them. Sequence of activities in part two of the study is shown in Figure 5.

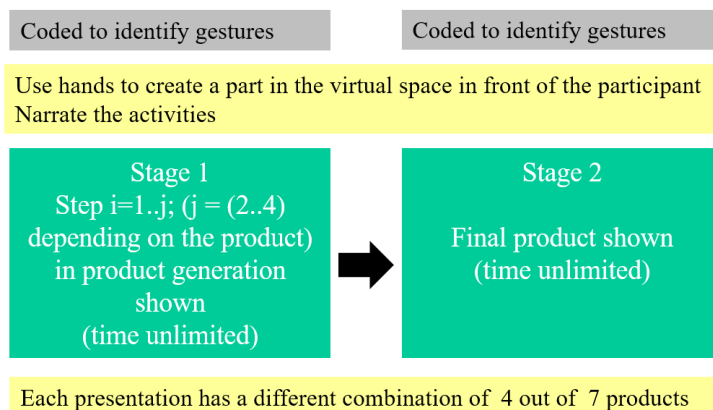


Figure 5 Sequence of activities in Part 2

Stage 1 began by asking the participants to create one of the objects shown in Figure 6.

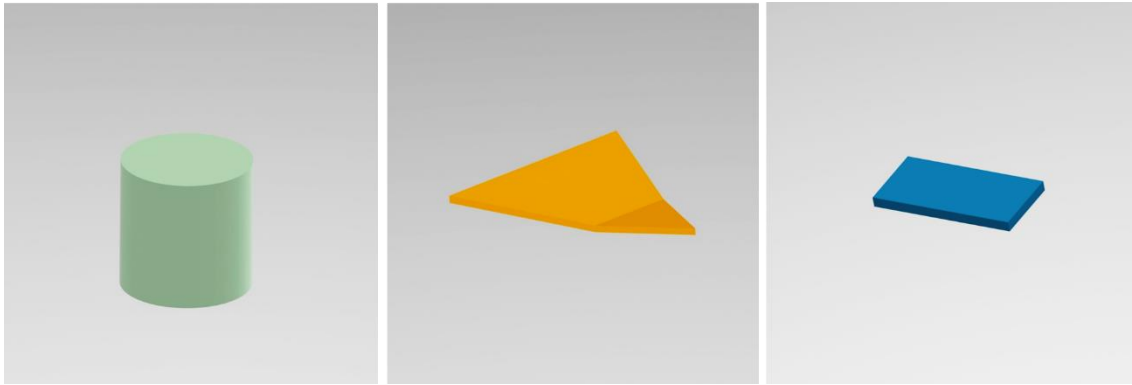


Figure 6 Full study Stage 1, first step for three different parts

Then participants were shown a new slide where a slightly more developed object is placed to the right of the initial object, and asked to further develop the already created object to match the new image. All of the steps are shown in Figure 7. Sequences for a cup and hexagonal plate had two versions each, where the only difference was the last step. The last step for the second version for both is shown to the right of the sequence in Figure 7.

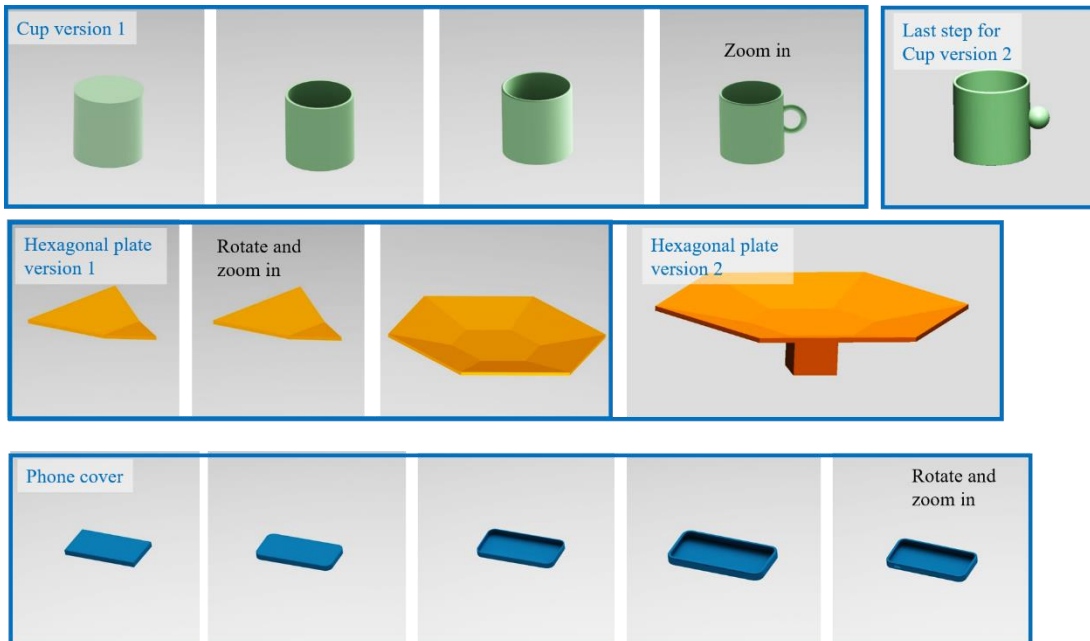


Figure 7 Steps for Stage 1 of Full study for all three parts

In the stage 2, participants were shown only one image displaying the final shape they were asked to create. This final shape is one of the products in the Figure 8.

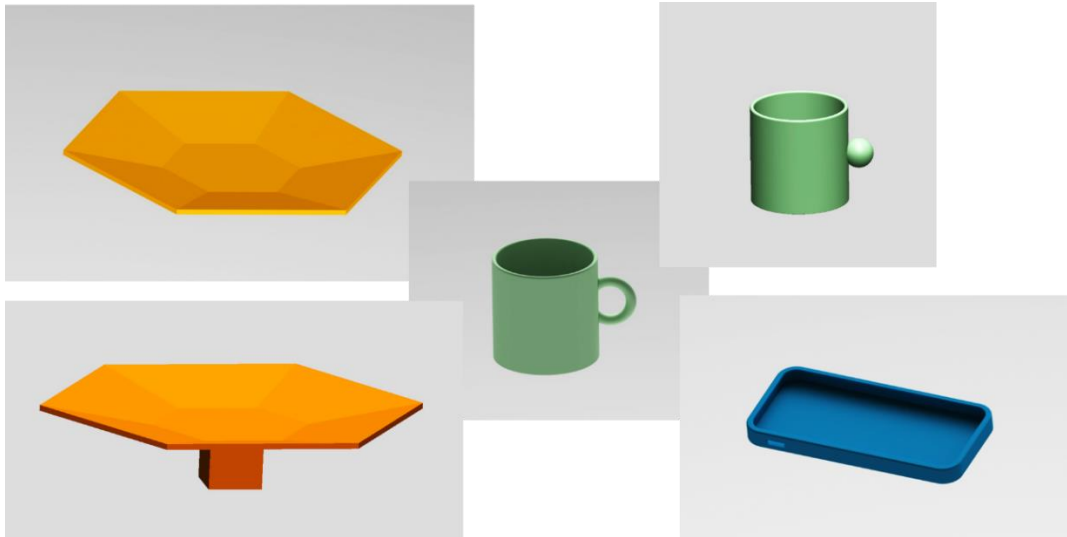


Figure 8 Final products used in Stage 2 of the Full study

Each participant created two products in total, and the combinations of images per presentation shown to participants are given in Table 1.

Table 1 Combinations of objects shown to participants in two stages

Presentation	Stage 1	Stage 2
1	Cup	Hexagonal plate
2	Hexagonal plate	Cup
3	Phone	Cup
4	Cup	Phone
5	Hexagonal plate	Phone
6	Phone	Hexagonal plate
7	Cup version 2	Hexagonal plate version 2
8	Hexagonal plate version 2	Cup version 2

To reduce the effect of potential adoption of the practices from the first stage in the second stage, half of the participants performed the two stages in reverse order.

4. Data analysis

For both parts of the study, video recordings were reviewed and the hand gestures performed in them were first **identified and described** (see Supplementary material, Section 1 for details). Then each gesture was **sketched** on a post-it. Gestures were then **parsed** by grouping the same gestures on post-its together. When the same gesture was performed in different planes, and where a change of a point of view would lead to the gestures being identified as the same they were grouped together. For example, gestures were merged into one when pointer finger tapped in horizontal plane to select a horizontal surface and when the pointer finger tapped in vertical plane to select vertical surface, but the gestures were the same, relative to the surface observed. Once this was completed, the **coding** could start.

The entire gesture coding and categorisation process is illustrated in the flowchart shown in Figure 9.

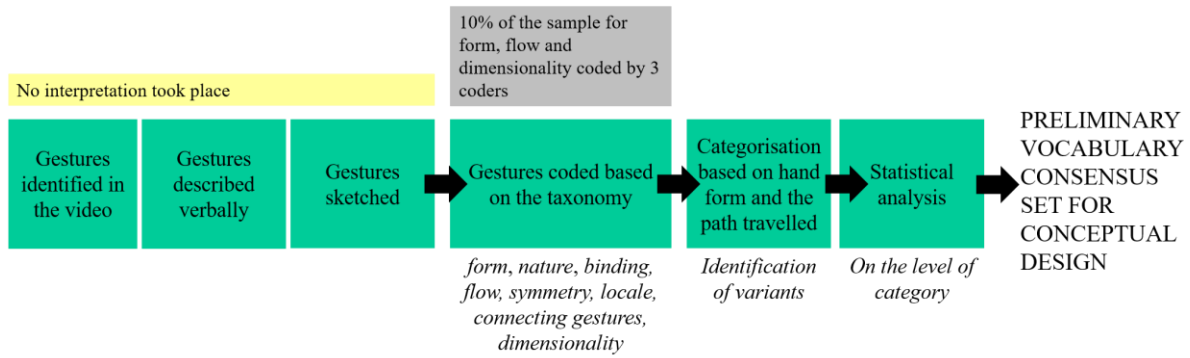


Figure 9 Gesture coding and categorisation process

At the end of the classification and categorisation process, a consensus set of gesture vocabulary for conceptual design was compiled.

4.1. Coding taxonomy

Taxonomy used in the coding, was based on the participatory design technique (Schuler and Namioka, 1993) and extended from Wobbrock et al. (2009), Morris et al. (2010) and Piumsomboon et al. (2013). Wobbrock et al. (2005) explored the gestures performed when using surface computing and Piumsomboon et al. (2013) explored hand gestures for AR interaction. Similar approach was used by Ruiz et al. (2011) to identify gestures for mobile interaction. Wobbrock et al. (2009) classify gestures based on their **form**, **nature**, **binding**, and **flow**. Piumsomboon et al. (2013) add **symmetry** and **locale**. In this study the taxonomy has been extended with **connecting gestures** (activity performed between the two gestures, in order to identify the connecting motions), and **dimensionality** of gestures (2D or 3D). The taxonomy is summarised in Table 2, and described in the remainder of this section.

Table 2 Summary of the taxonomy of the gesture coding

Classification type	Sub-classes
Form	Static pose Dynamic pose Static pose and path Dynamic pose and path
Nature	Pantomimic (physical) Metaphorical pantomimic Iconic
Binding	Object-centric (location defined wrt object features), World-dependent (location defined wrt world features), World-independent (location can ignore world features), and Mixed dependencies (world-independent plus another).
Flow	Discrete Continuous
Symmetry	Unimanual, Bimanual symmetric mirrored, Bimanual symmetric copied, Bimanual asymmetric.
Locale	On the surface In-air
Connecting gestures	Hands remain in previous position Hands resting on the table Open palms vertical in air
Dimensionality	2D 3D

4.1.1. Form

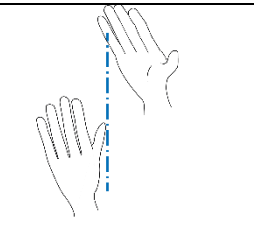
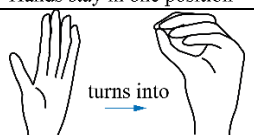
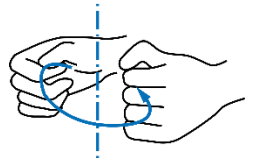
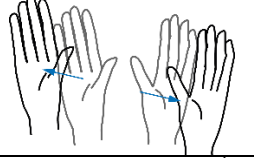
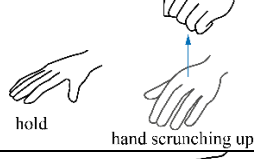

Based on the form gestures can perform a:

- Static pose – Hand and fingers are static. If both hands are used, if at least one of the hands is moving the pose is considered to be dynamic.
- Dynamic pose – Hand does not move along a path, but fingers do move along their individual paths.
- Static pose and path – Hand and fingers assume a static shape and move along a path
- Dynamic pose and path – Hand and fingers change shape while moving along a path.

Wobbrock et al. (2009) also included one-point touch and one-point path, as they were using tablets, however in the 3D environment one point touch and one point path are obsolete (Piumsomboon et al., 2013).

Examples for each of the form codes are shown in Table 3.

Table 3 Examples of form coding

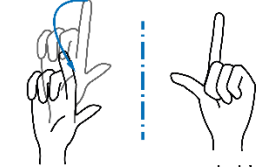
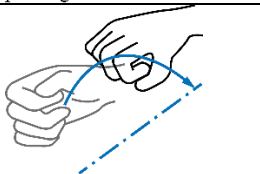
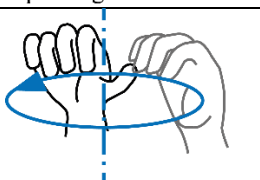

Gesture	Code and explanation
 <p>Hands stay in one position</p>	Static pose. To rotate both hands stay motionless in one pose.
	Dynamic pose. To zoom out hand turns from open palm into a full hand pinch, but stays in the same spot.
	Static pose and path (example 1). To rotate both hands follow a circular path, but the hand shape doesn't change.
	Static pose and path (example 2). To zoom in both hands follow separate paths.
	Dynamic pose and path (example 1). To undo right hand holds, but left hand changes shape from open hand with palm facing up into a fist while moving upwards in the process.
	Dynamic pose and path (example 2). To zoom out, right hand moves forward, and transforms in the process from a fist into an open hand with the palm facing down

4.1.2. Nature

Wobbrock et al. (2009) classified gestures, based on their nature, as symbolic, physical, metaphoric or abstract. In-air free form 3D gestures observed in this study required a modification of this classification. The meaning of the gesture would be less relevant for the eventual implementation in comparison to the information it provides and ability of others to understand/interpret the gesture without having to learn it prior to using a system. Some gestures were abstract, some might have had a symbolic meaning, but the common thread with both was that they would have had to be learnt and did not fully describe the motion, unless additional information was provided. If gestures with abstract

or symbolic meaning were observed without additional verbal explanation, an observer would not have been able to tell what the performed gesture was aiming to achieve, and in these cases the gesture was classified as iconic. It should be noted that gravity was not taken into account i.e. object being interacted with is considered to be a weightless virtual shape suspended in the air. Gestures classified as physical in the similar studies were classified as pantomimic in this paper. It was argued that the hand gestures performed to interact with the 3D object in the imagined virtual space were the same that would have been performed if the same interaction was performed in the physical world i.e. physical activity was pantomimed in the virtual world. In some cases, the hands were performing the motions that did pantomime an activity which can be performed in a physical world, but one that had a metaphoric meaning and if observed literally would indicate a different activity. In these cases the gestures were classified as metaphorical pantomimic gestures. Examples for each of the nature code types are illustrated in Table 4.

Table 4 Examples of nature coding

Gesture	Code and explanation
 <p data-bbox="204 824 464 869">repeat circular pushing motion hold</p>	<p data-bbox="491 651 1390 869">Pantomimic. To rotate left hand replaces an axis, and right hand spins the object around like a globe, or a basketball.</p>
 <p data-bbox="204 1055 464 1086">"pouring water"</p>	<p data-bbox="491 869 1390 1086">Metaphorical pantomimic. To lift the bottom of the cut higher up, water is poured so the level of "water" rises. It was never indicated that the water was in the cut, therefore additional meaning seems to have been added to the visual by the user.</p>
	<p data-bbox="491 1086 1390 1265">Iconic (example 1). To fillet an edge a hand traces it. While this may be considered to be a pantomimic gesture as you may form an edge of a sculpted cup that way, it may as well be just tracing a shape and to pick one of the two more information is needed.</p>
	<p data-bbox="491 1265 1390 1480">Iconic (example 2). To translate an object down hand is moved downwards, parallel to the ground with an open palm. This may be understood as translate down, but it may also indicate change of height, compressing something, moving only one surface down etc. Without more information it is hard to say which option is more likely.</p>

4.1.3. Binding

Wobbrock et al. (2009) and Piumsomboon et al. (2013) provide four classification options for binding classification of objects:

- object-centric (location defined with regards to object features),
- world-dependent (location defined with regards to world features),
- world-independent (location can ignore world features), and
- mixed dependencies (world-independent plus another).

In this study all gestures were object-centric as the guidance given to participants was to focus on the object only.

4.1.4. Flow

In the established studies, in terms of flow, gestures were classified as discrete, where response occurred after the user acted, or continuous, where response occurred while the user acted. Again, due to study set up and instructions given to participants to perform gestures while viewing the animation and pretending they were causing the activity, gestures were continuous in all instances.

4.1.5. Symmetry

In terms of symmetry Piumsomboon et al. (2013) classified gestures as unimanual (and then further as dominant or non-dominant depending on the participants' handedness) and bimanual (and then further as symmetric or asymmetric). While differences in use of dominant and non-dominant hands were noticed during the study reported in this paper, hand dominance has not been further explored, as it was expected that left or right handedness would not be a technical issue in the foreseeable future, and that handedness was not a priority at this, early, phase of gesture exploration. For example, LEAP, a low cost consumer recognition system, can detect both hands equally well and does not intrinsically differentiate between them. Symmetric gestures were further classified in this study as symmetric mirrored (both hands have the same form and follow the same path, mirrored around a central axis) or symmetric copied (both hands have the same form and follow the same path).

Overall, with regards to symmetry and number of hands used, gestures were classified as:

- Unimanual,
- Bimanual symmetric mirrored,
- Bimanual symmetric copied,
- Bimanual asymmetric.

4.1.6. Locale

In terms of locale, gestures could be performed on the surface or in the air. As in-air gestures were explored, majority of the gestures were performed fully in-the-air. However, at times participants used the table in front of them as an aid, with one or both hands. When table is interacted with for the full duration of the activity, gestures were classified as on-the-surface.

4.1.7. Connecting gestures

If a gesture vocabulary consensus set identified via this study was to become a new interaction paradigm for 3D modelling, the transitions between the gestures would become more important. Thus, a code was added describing the hand activity between two distinct gestures, in part two of the study only, as it included uninterrupted design process, where transitions would be identifiable.

Two different activities took place:

- Hands remained in the previous position or
- Hands were resting on the table.

4.1.8. Dimensionality

Participants viewed the animations on a 2D screen, and performed the gestures in a 3D space. After the activity completion, participants were explicitly asked if they perceived the object in the video as a 3D object suspended in front of them, and 86% of them confirmed that this was the case. Five participants (14% of the sample) did not agree. Two participants stated they did not know how they perceived the objects, one disagreed and said the gestures they performed were not in the 3D space and two strongly disagreed. Gestures these participants performed were coded for dimensionality. Gesture was considered to be 3D if hands "break" the plane i.e. use more planes than the vertical plane the image is shown in.

Gestures were coded as 2D if:

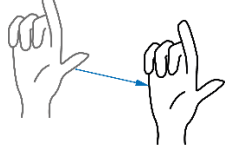
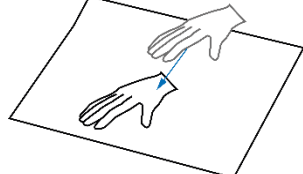
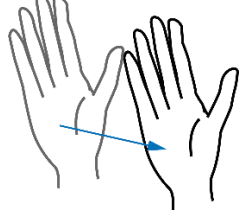
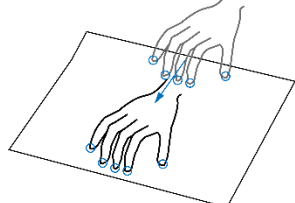
- All of the motions were performed in one plane that matches the plane gestures were shown in (vertical plane of the wall the screen was on), and users appeared to interact with a touch screen.
- All of the motions were performed on the table (e.g. participant pushed the imaginary object forward with their palm touching the table).

Gestures were coded as 3D if:

- Participants seemed to interact with an object suspended in the air in front of them and used multiple planes, while at least one part of the hand was used for the interaction with the object.
- Gestures were performed as if the imagined object is located on the table but was three-dimensional (e.g. when a participant held the object's imaginary vertical axis and "rotated" the object by touching the "sides of the object").

Illustrations of examples for each of the dimensionality codes are shown in Table 5.

Table 5 Examples of dimensionality coding

Gesture	Code and explanation
 <p data-bbox="233 483 376 506">1, 2 or 3 fingers</p>	<p>2D (example 1). To translate an object left, hand is indication it is touching the vertical plane with one, two or three fingers. Gesture appears like it is interacting with a touch screen and it is not giving the object any depth.</p>
	<p>2D (example 2). To translate up, gesture is sliding an imaginary object forwards, but is fully touching the table, not allowing for any other dimension for the object other than the front face in the vertical plane</p>
	<p>3D (example 1). To translate an object left, right “side” of the object is pushed by the open palm, giving it depth.</p>
	<p>3D (example 2). Gesture is sliding an imaginary object backwards to “translate down” in one plane, but the gesture is grasping the object in a plane perpendicular to it, using a plane that did not exist in the visual representation of the activity.</p>

4.2. Inter-coder reliability

To assess the coder reliability Krippendorff’s alpha measure of reliability was used. It is a general measure that can be used regardless of “the number of observers, levels of measurement, sample sizes, and presence or absence of missing data” (Hayes and Krippendorff, 2007). The first author coded 100% of the gestures recorded. Two additional coders were asked to code randomly selected 10% of the sample for the part of the coding in where more interpretation was required (form, nature and dimensionality), to ensure that a consensus was reached among the coders. Either 50 units or 10% of the sample is considered to be an appropriate size of the sample (Lombard et al., 2002). The coders were research assistants currently in their PhD write up year. They have both previously completed an MSc in the field of design and their research field is product design. The elements where interpretation occurred more were: dimensionality coding for the 18% of participants that did not state performed gestures were 3D, and hand path coding and gesture type coding for all participants.

For the form Krippendorff’s Alpha reliability estimate of 0.8158 was calculated, for the nature of gestures it was 0.7458, and for the dimensionality it was 0.8468. Agreement of $\alpha \geq 0.8$ is customarily required, with value of $\alpha \geq 0.667$ being the lowest required value where tentative conclusions are acceptable (Krippendorff, 2004). This means that for the form and dimensionality the agreement was at a reliable level, whereas for the nature of gestures it fell slightly short of the required value, but it was still above the lowest required values for tentative results. Due to the discrepancies between the gesture definitions and their applicability to gestures for design this is considered acceptable by the authors, particularly since the nature of the gestures does not play a crucial role in the gesture categorisation primarily used to form the consensus set (reported in Section 5.3).

4.3. Categorisation

Once all the gestures were coded, for both parts of the study, they were **categorised and analysed for patterns and relations**. Sketches were then assigned unique identifiers. The coders were trained to apply the same approach to interpretation of gestures. Gestures that performed same activity

following the same path were grouped in the same category. However, there were variants within the category, based on the form the hands took. These variants were given their own sub-codes, expressed by the decimal value. Each unique activity was given a unique code and number, and the numbers have the form of nn.n, e.g. TL01.1. The list of codes and example of a coding process can be seen in the Supplementary material (Section 2), along with the Images containing sketches of all gestures performed.

4.4. Statistical analysis

Once the gestures were grouped, the Agreement Rate (AR), Chi Square and Fleiss Kappa were calculated for them. This was done to increase the robustness of the study and ensure that the gestures added to the vocabulary are those most frequently used, that showed statistical significance and were not likely to have been repeated by chance, instead of relying on an arbitrary number of repetitions as a criteria.

Agreement Rate (AR), calculation derived by Findlater et al. (2012) and adopted by Vatavu and Wobbrock (2015), was calculated for each of the activities and the categories within it. AR rate measures the homogeneity for nominal data.

Agreement rate is widely used, but not universally accepted as a measurement for selection of appropriate gestures for the inclusion in the consensus set. Tsandilas (2018) suggests that an additional measure, either Fleiss' κ or Krippendorff's α , should be used to chance-correct the coefficients and specific agreement. For this study Fleiss' κ was calculated to correct for chance of agreement.

AR is used to discover if there is sufficient agreement between the participants proposing gestures from a theoretically infinite set of gestures. Fleiss' κ determines if the agreement is statistically significant. AR and Fleiss' κ do not provide information on what number of repetitions would be required for a specific gesture to be included in the consensus set. To determine this, Chi square Goodness of Fit analysis was performed to determine if the number of repetitions for different categories within each activity was likely to happen by chance, and what number of repetitions was expected to happen by chance for each category.

The calculations were performed using SPSS. Chi-Square goodness of fit calculations also allowed for observation of the repetitions for each shape, where same activities were performed for different shapes.

5. Results of the user centered gesture study

This section reports the results of the data analysis approach outlined in Section 4, for 1785 gestures collected during the were collected during the user centered gesture study.

5.1. Gesture parsing

In part one 1100 gestures were planned to be collected, but 1083 gestures were performed successfully by the participants. Five gestures were skipped in part one of the study e.g. participants did not understand what happened on the screen when a surface was deselected and did not perform a gesture with their hands. One gesture was eliminated from the analysis, as the gesture performed was not a hand gesture i.e. the participant moved their head closer to the object to zoom in. In part two participants had more freedom to determine the activities they would perform, and this meant that the number of gestures they performed was variable and 702 gestures were recorded. Due to the camera failure, four gestures for participant four and seven gestures for participant 20 were not recorded in part one of the study, and the gestures performed during part two of the study for the participant four were not recorded for part two. This left 1785 gestures for the full analysis for both parts of the study combined.

5.2. Gesture coding and taxonomy

In terms of **form** vast majority of gestures performed static pose and path, 1711 gestures, or 96% of the performed gestures. Gestures using dynamic pose were performed by only thirteen participants, and were used to *fill*, *draw* or *extrude*. Eleven different participants used static pose, and they used it for *shape creation*, *plane selection*, *loft activity* or *rotation*. On their own, these static gestures would not be able to convey the activity, unless they were predefined and linked to a specific activity prior to their use. Eleven participants performed gestures using dynamic pose and path. They were used to *extrude*, *zoom in or out*, *undo*, *draw*, or *extrude cut* a shape, and were generally similar to the static pose and path gestures performed for the same activities. The difference was that the moving hand

changing shape provided more indication of what the activity was e.g. for *zoom in* fingers of the moving hand move apart to indicate the increase of the size of the object being zoomed into. It is interesting that, aside from *zoom in* or *out* activity, dynamic pose and path type of gestures were mostly found in the second part of the study, where participants were free to propose their own activities.

Observing the **nature** of the gestures over 70% of gestures performed were iconic. These gestures would be required to be learnt prior to use. Although they often resembled physical gestures that would have been used to manipulate the object, they were lacking some elements or information that would fully describe the activity performed to observers that had no prior knowledge of the goal of the activity. Less than 30% of the gestures performed were pantomimic. Three gestures were metaphoric pantomimic gestures. For example, to *zoom in* two participants have “pulled a rope”. To raise the height of an *extruded cut* surface one participant had “poured water” into it. The observations in this paragraph can only be considered tentative because of the lower alpha value calculated for the agreement between the coders for the nature of the gestures.

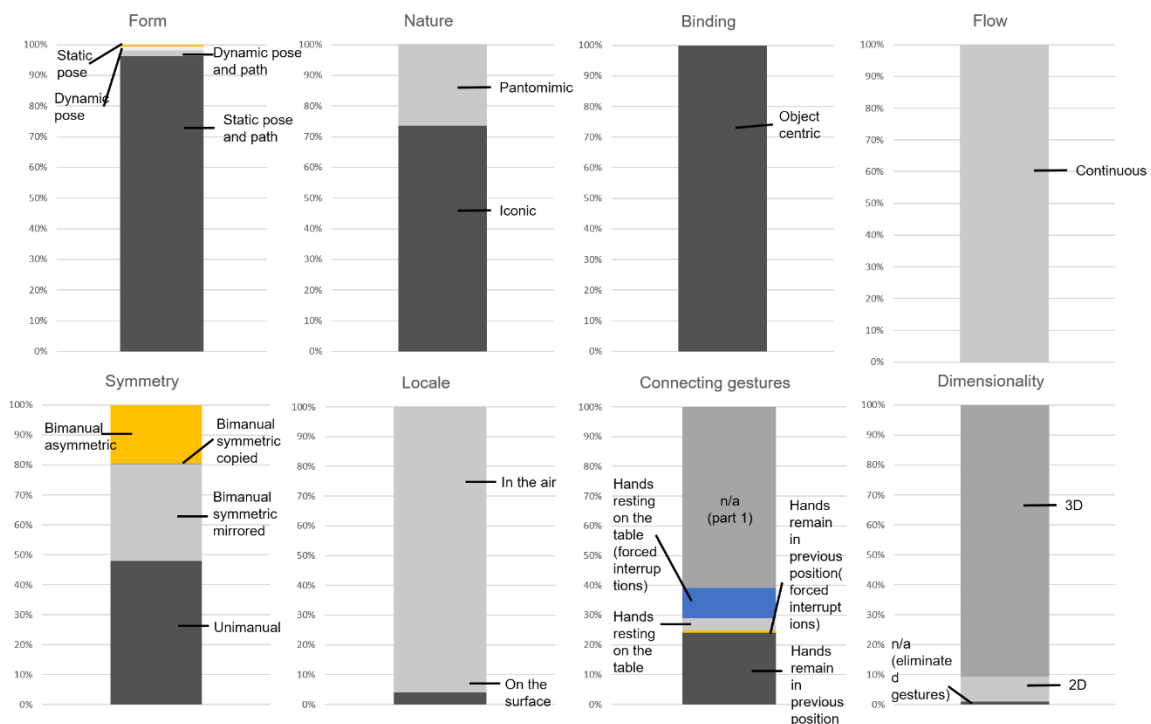


Figure 10 Gestures coded by form, nature, binding flow, symmetry, locale, connecting gestures and dimensionality

Classification based on **binding** criteria showed that all gestures were object-centric. This was predetermined by the study design and participants were explicitly asked to manipulate the object they saw using their hands.

Similarly, study design required the participants to perform gestures as if they were controlling the activity shown on the screen while it was happening, which meant that for part one of the study in terms of **flow** gestures were continuous by design. In part two participants were not instructed to perform continuous gestures, but due to the nature of the activities, they all still were continuous. Participants started by imagining they were creating a part in the empty space in front of them, and the activity was complete when the imaginary 3D model suspended in the space in front of them has been “created”. Each step contributed to creation of a piece of the geometry and imagining it was appearing as the gestures progressed. Naturally, continuous gestures corresponded better to the activities than the discrete gestures.

Split between unimanual and bimanual gestures was close to 50/50, with unimanual gestures accounting for 47.8% of the performed gestures. Bimanual asymmetric gestures were 19.4% of the total sample, bimanual symmetric mirrored gestures were 32.4% of the total sample, and only seven gestures (0.4% of the sample) were bimanual symmetric copied gestures. Some unimanual and bimanual gestures were identical, if only the moving hand was observed. In the bi-manual variant, the static hand was used to hold the object being modified in place. In this study these variations were

classified as different gestures, however in the future work, if the vocabulary was to reach the implementation stage, these gestures may merge into a single gesture. For example if the object would be assumed to be stationary unless indicated otherwise.

In terms of **locale**, majority of gestures (96%) were performed in-air. As participants were seated at a table, some used the table to rest one of their hands while performing the gestures. As the other hand was in all cases still used in-air, and the resting hand did not an active role in the gesture performance, these gestures were classified as in-air. Gestures were classified as on-the-surface if they were in contact with the table for the entirety of the activity, without the in-air element. For example, in one of the gestures used to translate left shown in Figure 11, coded with TL06, hand slides to the left along the table, and it was classified as on-the-surface. Gesture used to pattern a shape around, MulPat07, in some instances touches the table. However, the motion “moving” the shape to the next position is fully in-air, conveying a key element of the activity, and without it the gesture would be incomplete. Hence, this gesture was classified as in-air, as shown in the image in Figure 11. The table was a part of the study setup primarily to provide a more comfortable environment for the participants. The added benefit of its presence was the ability to observe the extent to which the participants needed to use it. It was encouraging to see that 96% of participants performed in-air gestures instead.

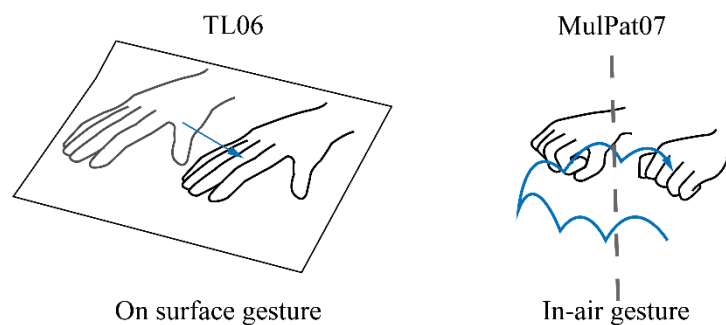


Figure 11 Example of on surface and in-air classification

Connecting gestures category was not predefined, but instead identified from the gestures performed in the part two of the study, as that is where uninterrupted design took place. Only two distinct activities took place in the full study. Either hands remained in the form they took at the end of the previous gesture, while the participant was thinking about the next step (85%), or they were resting on the table (15%). In the stage 1 of part 2 of the full study there were interruptions between the predefined stages. The connecting gestures between predefined stages were not included in the sample as the participants may not have been thinking about the design activities during these. For majority the hands were resting on the table (92%).

Observing **dimensionality** for the participants who stated they did not perceive the objects as 3D or did not know if they perceived them as 3D, it was found that 90% of the gestures were performed as if the object was perceived in 3D. Majority of 2D gestures were used for translation, presumably emulating the tablet interaction paradigms.

5.3. Statistical analysis and gesture categorisation

Agreement rates were first calculated for both parts of the study and they are illustrated in Figure 12. For the Part one all agreement rates are above 0.1, indicating at least medium agreement. AR for Rotate clockwise, Rotate counter clockwise and Extrude cut are above 0.3, indicating high agreement. For the Part two, AR for 11 activities indicate low agreement, and AR for nine activities are above 0.1, indicating at least medium agreement. AR for Resize is above 0.3, indicating high agreement.

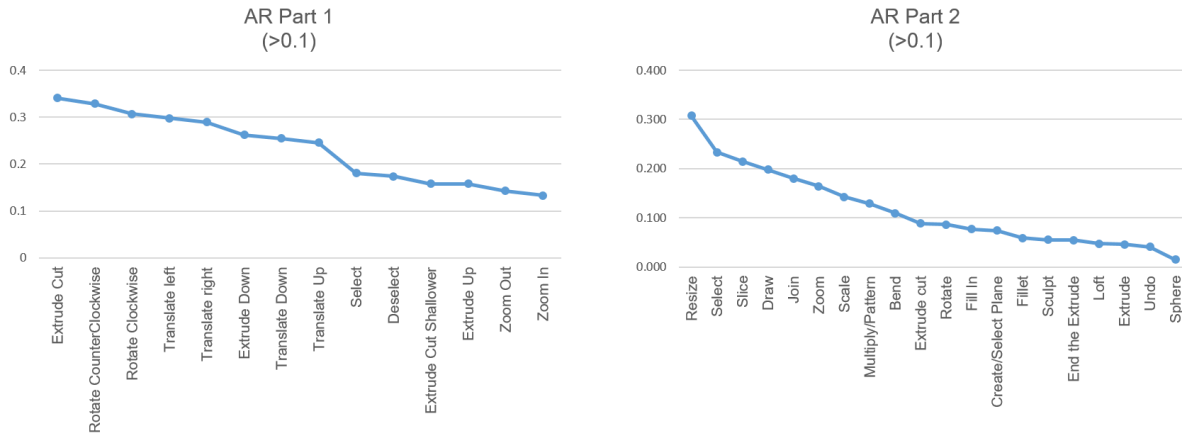


Figure 12 Agreement Rates for Part 1

To chance-correct, the coefficients and specific agreement Fleiss Kappa values were calculated and κ values for different activities are given in Table 6, along with the AR values.

Table 6 AR, Fleiss' kappa values for activities in Part 1 and Part 2 of the study

Part 1			Part 2		
Gesture	AR (>0.1)	Fleiss κ	Gesture	AR (>0.1)	Fleiss κ
Extrude Cut	0.341	0.897	Resize	0.308	0.163
Rotate Counter Clockwise	0.328	0.898	Select	0.233	0.632
Rotate Clockwise	0.307	0.923	Slice	0.214	0.457
Translate left	0.297	0.989	Draw	0.198	0.927
Translate right	0.289	0.971	Join	0.180	0.679
Extrude Down	0.262	0.879	Zoom	0.165	0.35
Translate Down	0.254	0.974	Scale	0.143	-1.021
Translate Up	0.245	0.963	Multiply/Pattern	0.129	0.43
Select	0.180	0.828	Bend	0.110	0.316
Deselect	0.174	0.831	Extrude cut	0.089	0.684
Extrude Cut Shallower	0.158	0.775	Rotate	0.087	0.509
Extrude Up	0.158	0.748	Fill In	0.077	-0.172
Zoom Out	0.142	0.789	Create/Select Plane	0.074	0.605
Zoom In	0.132	0.715	Fillet	0.059	0.016
			Sculpt	0.055	0.059
			End the Extrude	0.055	-1.33
			Loft	0.048	-6
			Extrude	0.046	0.606
			Undo	0.041	0.218
			Sphere	0.015	-3.713

Gestures were added to the consensus set only if AR was above 0.1, and if Fleiss' κ was showing at least fair agreement (>0.21). All activities performed in the part one had representatives in the consensus set. In part two only eight activities had both AR and Fleiss' κ that were higher than the required value: *Resize*, *Select*, *Slice*, *Draw*, *Join*, *Zoom*, *Multiply/Pattern* and *Bend*. These are represented in bold font in Table 6.

Consensus set was then further refined by adding information from the Chi-Square analysis, in order to explore the likelihood of repetition happening by chance. Gestures where number of repetitions for a category of gestures was lower than would have happened by chance would be removed from the consensus set. Values for different activities and objects interacted with in part one of the study can be found in Table 7. While all p values are significantly lower than the Bonferroni corrected p value, indicating that the repetition of gestures was unlikely to happen by chance, for a number of gestures expected number of repetitions was lower than 5, meaning that Chi-Square could not provide definitive conclusions. These cases were recalculated using the Exact one calculation, and those values were reported instead.

Table 7 Chi-Square results for Part 1 of the study

Gesture	Bonferroni corrected p value	Chi Sq Phone	Chi Sq Sphere	Chi Sq Chair	Chi Sq Console	Expected number of repetitions
Extrude Cut	0.050				0.000	6.1 (console)
Rotate CounterClockwise	0.025	0.000	0.000			8.8 (phone); 7.0 (sphere)
Rotate Clockwise	0.017	0.000	0.000	0.000		8.8 (phone); 6.1 (sphere); 8.6 (chair)
Translate left	0.025	0.000	0.000			6.1 (phone); 7.3 (sphere)
Translate right	0.025	0.000	0.000			6.3 (phone); 7.2 (sphere)
Extrude Down	0.050				0.000	5.4 (console)
Translate Down	0.025	0.000	0.000			5.5 (phone); 6.3 (sphere)
Translate Up	0.017	0.000	0.000	0.000		6.1 (phone); 5.8 (sphere); 5.5 (chair)
Select	0.050				0.000	2.6 (console)
Deselect	0.050				0.000	2.8 (console)
Extrude Cut Shallower	0.050				0.000	3.9 (console)
Extrude Up	0.050				0.000	4.3 (console)
Zoom Out	0.017	0.000	0.000	0.000		3.3 (phone); 3.4 (sphere); 4.4 (chair)
Zoom In	0.025	0.000	0.000			2.8 (phone); 2.9 (sphere)

Values for different activities and objects interacted with in part two of the study can be found in Table 8. Since part two did not have a controlled number of repetitions i.e. participants chose their own activities during object creation, for the majority of the gestures and objects the number of expected repetitions was not high enough. This also meant there were variations in the purpose of activity if the specific activity was performed for a different shape. For example to create a shape shown in Figure 7 titled “Hexagonal plate version 2”, some participants started by drawing a hexagon (marked by Hex-Hex in the table), others a triangle that they then rotated to create a hexagon (marked by Hex-Tri in the table), and some started by drawing a square first (marked by Hex-Square). Additionally, the number of participants that performed specific gestures varied. Only two gestures were shown to have happened more frequently than expected, and only when exact calculation was applied to them. These were *Draw* and *Select* and they were added to the main consensus set. The remainder of gestures were retained for descriptive purposes in a separate set (provided in the supplementary

material). In essence, this means that it has not been proven that the number of repetitions for consensus set for part two did not happen by chance for majority of the activities.

Table 8 Chi-Square results for Part 2 of the study

Gesture	Bonferroni corrected p	Chi sq															Expected number of repetitions (relative to column number)
		Phone	Phone-shell	Phone-mic	Cup	Hexagon	Hex-Hex	Hex-Tri	Hex-Square	Hex-bend tri	Hex-bend hex	Hex-pattern with rot	Hex-copy	Hex-undo pat	Hex-undo square	Sphere	
Resize	0.010	0.04															1.8;
Select	0.017	0.96			0.01	0.01											1.9;2;2.2
Slice	0.017					0.34											1.6
Draw	0.017	0.00			0.00		0.01	0.01	0.14								3.6; 4.7;3.1;3.3
Join	0.017				0.43	0.02											2.4;4..3
Zoom	0.025	0.32			0.68												2;1.4
Scale	0.010	0.90			1.00												1.3;1
Multiply/Pattern	0.025											0.39	0.49				2.3;1.7
Bend	0.017									0.74	0.91						2.6;1.5
Extrude cut	0.010		0.47	0.04	0.18	0.18	0.18										2.5;2.3;2.9;2.9
Rotate	0.013	0.20			0.39	0.71											2;2;3.5
Fill In	0.025				0.68												1.4
Create/Select Plane	0.017	1.00			0.57	1.00											2;2.5;1.4
Fillet	0.017	0.80			0.43												1.8;2.5
Sculpt	0.025	Too many different shapes to have statistical significance															-
End the Extrude	0.013				0.98	1.00											1.2;1
Loft	0.050					1.00											1
Extrude	0.017	0.65			0.05		0.01	0.90	0.58								2.1;2.8;1.9;1.4;1.6
Undo	0.050	1.00			0.78									0.65	0.90		1.1;1.6;1.3;1.3
Sphere	0.025															1.00	1.1

5.4. User defined set of gestures/consensus set

Following the data analysis the most repeated gestures that did not happen by chance and were statistically significant formed a consensus set of gesture vocabulary for conceptual design. The consensus set is shown in Figure 14 and Figure 15. The excluded set is given in the supplementary material. Sketches used to illustrate the gestures followed the gesture representation framework developed by McAweeney et al. (2018), further modified for the needs of this study. The framework is illustrated in Figure 13.

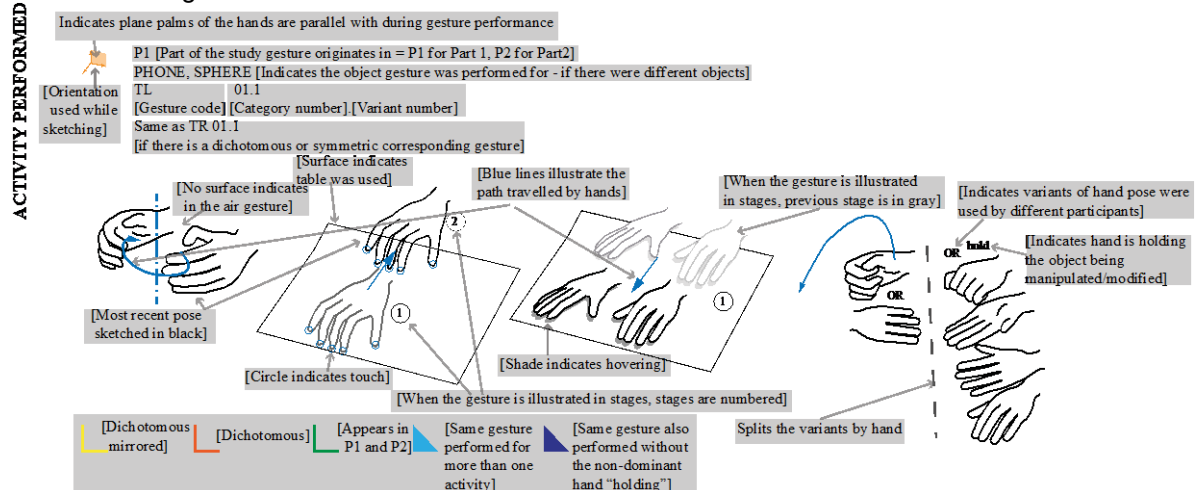


Figure 13 Framework for gesture sketches

Gestures were represented in the isometric view. Text above the gesture sketch indicated which part of the study the gesture appeared in, which objects it was used for, and what the gesture code, category and variant number were. Where the gestures had more than one sequence, the sequence stages were numbered. In-air gestures were shown without a surface below them. Where the table was used surface was represented. If the table was touched this was indicated with circled touchpoints in blue. When the hand was hovering over the surface of the table shadow was added to indicate this. Motion paths were indicated by blue arrows and paths. Coordinate system in the top left corner indicated the directions of the axis and planes of the space the isometric view was set in, and highlighted planes indicated parallelism with the palms of the hands shown in the specific gesture sketches. Where multiple positions were shown in one sequence stage, previous position was shown in grey. Symbol placed at the bottom left of each gesture, if present, indicated that the specific gesture either had a dichotomous mirrored counterpart in the vocabulary (yellow symbol), dichotomous counterpart (orange symbol), activity appeared in both parts of the full study but gesture appeared in part two only (green symbol), same gesture was performed for more than one activity (light blue symbol) or the same gesture was also performed using the non-dominant hand to “hold” the object (dark blue symbol).

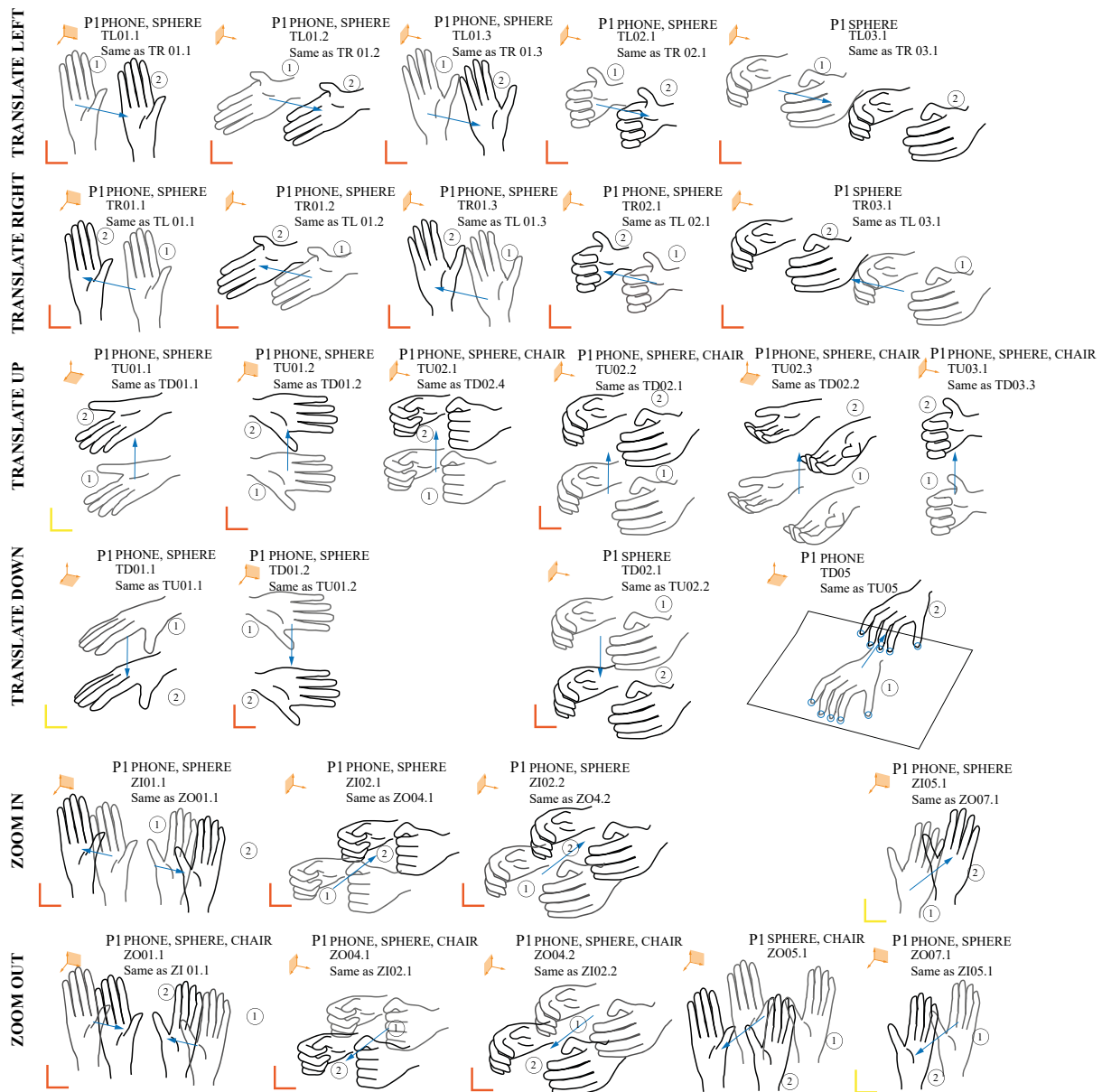


Figure 14 Consensus set (first half)

It is noticeable that some gestures were used only for interaction with the sphere, for example (in Figure 14) three out of four translate activities used both hands to translate the sphere e.g. TL and TR 03.1, and TD 02.1. However, for other activities similar hand shape and both hands were used for all other shapes as well, although less frequently. Majority of other activities were performed for at least two objects of distinctly different shapes. This was considered to be beneficial for the vocabulary being built with the intention of becoming a base vocabulary set for design, as it indicates they might be applicable to objects of varied other shapes. Dichotomous gestures show a high degree of consistency, use reversible gestures, and in some cases, for example TU 01.1 and TD 01.1 in Figure 14, there was a degree of symmetry e.g. mirroring the hand pose for the similar activity but following a path in a different direction.

While participants have not reported major issues with gesture performance, it should be noted that four participants found select/deselect gesture activity hard to understand. Eight participants noted that hexagon creation was more difficult than the rest of the objects, due to its complex shape.

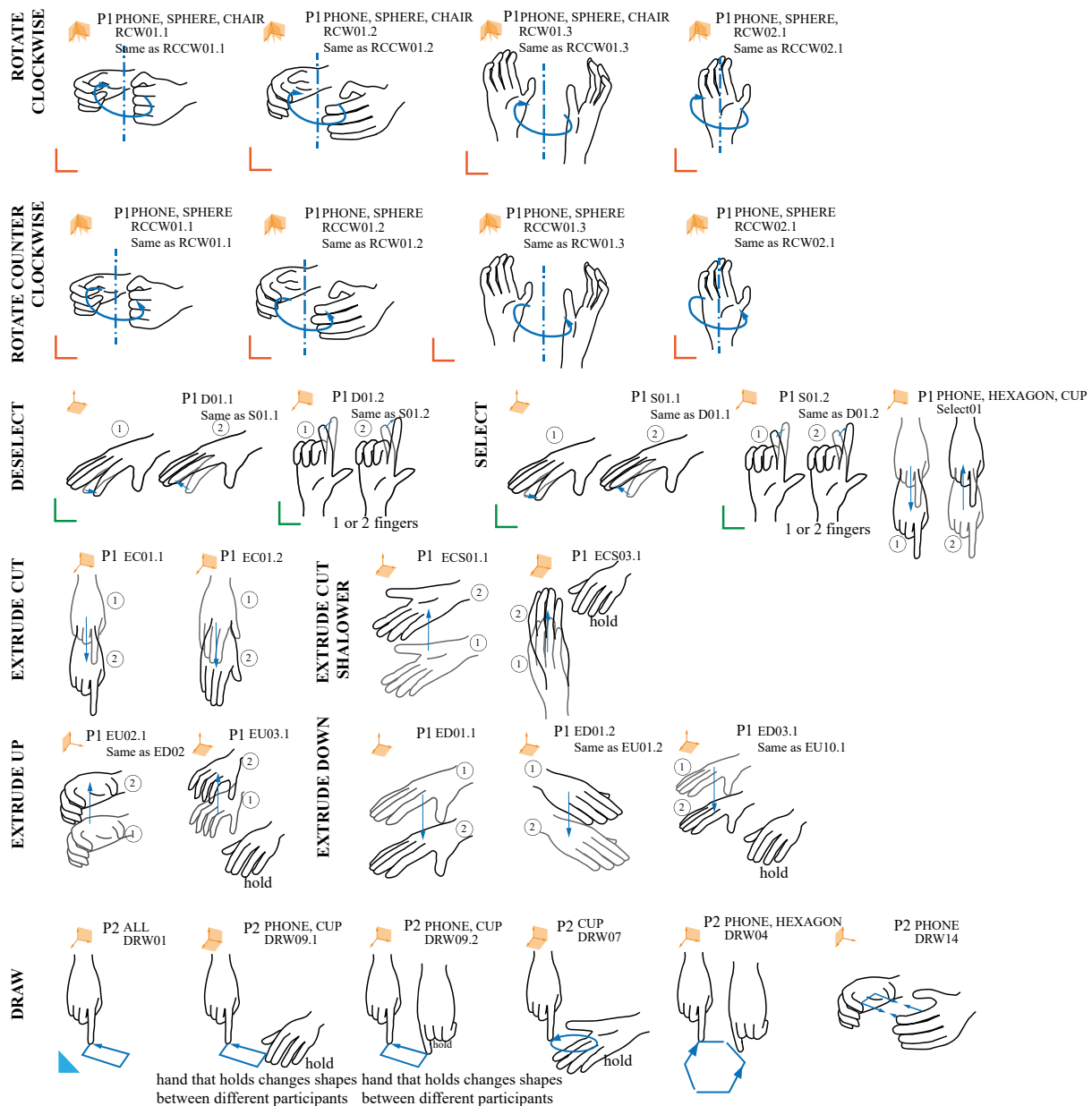


Figure 15 Consensus set (second half)

6. Evaluation of the consensus set

In order to further generalise the vocabulary and make it easily adoptable by those without specific design training, ten other professionals with varied backgrounds evaluated both the consensus set and the excluded set. The goal was to refine the vocabulary set so that it is natural, intuitive, and does not require extensive training to use. Refinement also aims to reduce bias design students may have introduced during gesture elicitation.

Evaluation consisted of abstract evaluation and VR evaluation. During the abstract evaluation the participants were asked to perform the gestures from both consensus sets using their hands. They were shown a photograph of the shape they were imagining they were interacting with, and the researcher running the study demonstrated the gestures. During the VR evaluation four variants of gestures for four key manipulative gestures were evaluated using a VR application. The gestures were four variants of *translate left* and *right*, *translate up* and *down*, *rotate clockwise* and *counterclockwise*, and *zoom in* and *zoom out*. The gestures were the highly repeated gestures from the consensus set. The system employed a Vive headset with a LEAP sensor mounted on it, Unity, and Steam platforms and a high specification desktop computer. System in use is shown in Figure 16. This was done primarily to compare the outcomes of abstract and VR aided evaluation, in order to

explore if the evaluation approach would introduce changes to the results. During both evaluations the participants were asked to rate the performed gestures, and then select their preferred gesture among the tested gestures, for each activity where more than one gesture was being evaluated. Another variant of the consensus set was created as an outcome, illustrated in Figure 20.

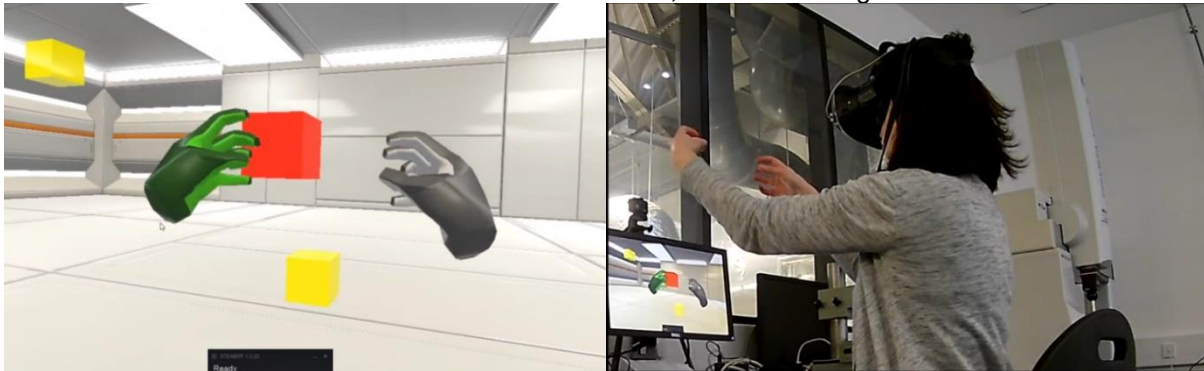


Figure 16 Participant performing the VR evaluation

During both evaluations the participants answered a number of questions following each activity. The questions were adapted from Wobbrock et al. (2009) and Piumsomboon et al. (2013) to better fit the needs of the evaluation study, and two questions were added to the VR evaluation. For the abstract evaluation, the questions were designed to evaluate appropriateness of the gesture for the activity, and ease of performance:

1. Was the gesture you just imitated was a good match for the current activity (i.e. would that gesture be a good way to execute that activity)?
2. Was the gesture you just tried was easy to perform (i.e. rate the difficulty of carrying out the gesture's physical action)?

In VR evaluation questionnaire, the questions about the gestures and application capabilities were separated. The questions were designed to evaluate appropriateness of the gesture for the activity, and ease of performance regardless of implementation qualities within the system, satisfaction with the result of the activity, and difficulty or ease of the use of the application:

1. Was the gesture you just imitated was a good match for the current command (i.e. would that gesture be a good way to execute that command).
2. Was the gesture you just tried was easy to perform (i.e. rate the difficulty of carrying out the physical action)?
3. Did the gesture resulted in the action you expected?
4. How difficult is it to perform the gesture (considering the VR technology use)?

Seven point Likert scale was used for the responses: Strongly Disagree, Disagree, Mildly Disagree, Neither Agree nor Disagree, Mildly Agree, Agree, Strongly Agree. Participants were also asked to provide any further comments of any nature they had at any point of the process.

The participants were two researchers specialising in organisational management, two mechanical engineers, two electrical engineers, two architects, one marketing manager, and one teaching associate specialising in cost forecasting. Their average age was 33 years old. They had between one and 15 years of professional experience in their profession (4.4 on average), and with the exception of one of the electrical engineers, did not have extensive experience using AR or VR environments. Some have used CAD in the past (for between two and ten years), but use of CAD was not a core element of their daily work in their current role, or in the case of architects' CAD is primarily used for layout of rooms in a building rather than 3D shape creation, manipulation or modification. Architecture and mechanical design do have some crossover with the field of design. While all types of conceptual design include a process of creating forms, functions and behaviours to a certain extent (Benami and Jin, 2002), in different fields focus is not put equally on all of them. This study focused primarily on the form and visualisation of design concepts. Product design students that participated in the gesture elicitation have had extended training in form generation, that did not necessarily have to have a link to function or behaviour in all stages of development. In product design conceptual design is often referred to as a collection of activities which "determine the form of an engineering product" (French et al., 1985) or the initial stage of the design process, during which fundamental, but approximate, outlines and form of a product are created (Keinonen and Takala (2010), pg 17; Ulrich

and Eppinger (2011), pg 18). Mechanical engineers primarily focus on the function as a solution to the problem generated via conceptual design (Chakrabarti and Bligh, 1994), or at least explore function and behaviour during conceptual design before they move on to form (Welch and Dixon, 1994). Architects may consider both form and function. However when they focus on form, particularly since computer generated forms have become a standard in the architectural design, form takes precedence over its optimization or manufacturing (Grobman et al., 2009). However, depending on the approach taken function can be linked to form, or form can be a separate entity superimposed on the function (Reveron, 2009). The latter would be similar to the form exploration observed in this study. When architects focus on the function, they focus on spatial configuration which is either defined by function or helps define the function (Reveron, 2009), and in these cases it is more akin to the approach taken in mechanical engineering (although the functions observed can have a very different nature). Hence, both mechanical engineers and architects were considered to be far enough removed from product designers to be considered appropriate participants in the evaluation, as they would introduce diversity both in training and mind-set.

6.1. Results of the abstract evaluation

Averaged ratings for the abstract evaluation, across all ten participants, are given in Figure 17 for the gestures with the highest agreement from part one of the study and Figure 18 for the part two of the study. Where more than one gesture was evaluated for an activity, the preferred gesture across the ten participants is encased in a box.

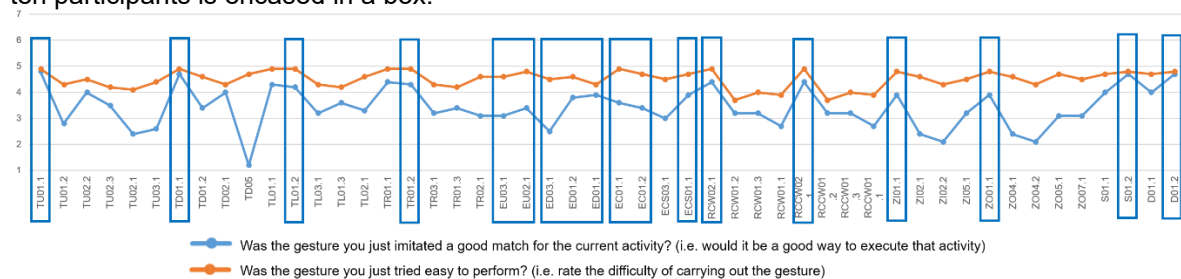


Figure 17 Averaged ratings across the ten participants for the gestures with highest agreement from the Part 1 of the study

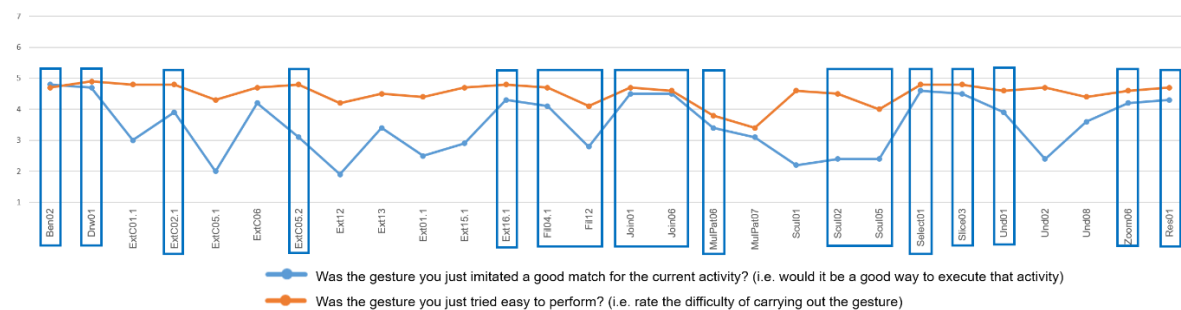


Figure 18 Averaged ratings across the ten participants for the gestures with highest agreement from the Part 2 of the study

For both parts, participants found majority of the gestures easy to perform, and average ratings were on the agree side of the scale. For the most frequently chosen preferred gestures, it was clear that the higher agreement ratings consistently corresponded well with the instances where participants on average preferred one gesture for a specific activity. Examples of these were TU01.1, TD01.1, TR01.2, ECS01.1, RCW02.1, RCCW02.1, ZI01.1, ZO01.1, S01.2, D01.2, Ext 16.1. In instances where there were disagreements between the participants on the preferred gestures, and multiple gestures were chosen, the average appropriateness ratings were on average still on the agree side of the scale. The examples of this are preferred gestures for *translate left* (TL01.1 and TL 01.2) and *extrude down* (ED01.2 and ED 01.1), which were chosen with similar frequencies. Participants were required to choose one of the gestures they were evaluating without introducing changes to them or suggesting different gestures they would prefer instead, and this meant that in some cases there were disagreements with the proposed gestures. Average ratings for appropriateness of gestures for TD05 indicated strong disagreement. All sculpting gestures were rated on the disagree side of the scale. In the instances where a single gesture was tested for the activity the average appropriateness ratings were on the agree side, often on the higher end (mildly agree), like for Ben02, Drw01, Res 01. This

indicated that the gestures identified from the study analysis were appropriate and acceptable to the evaluators.

6.2. Results of the VR evaluation

Averaged ratings for the VR evaluation, across all ten participants, are given in Figure 19. In terms of the consistency between the choice of a preferred gesture for an activity and average appropriateness rating for it, the observations were similar to the abstract evaluation. Satisfaction with the achieved results was consistent with the appropriateness and ease of performance ratings, and generally positive for the majority of gestures. However, the ratings for ease or difficulty of performance in the VR system did not always follow the same trend as the remaining three parameters. Technology used was not able to detect grasping gestures as well as an open hand or a fist. Due to occlusion, detection of open hand had proven more difficult when the side of the hand was the only surface detectable by the LEAP sensor. In some instances, these issues did not seem to have a great effect on the choice of preferred gestures. For example, the discrepancy between rates for appropriateness of the gesture, which was rated as high, and ease of performance, which was rated as low, was highest for RCW/RCCW 02.1 and Select01. Often comment was made that the gesture was hard to perform in the system, but that participants would prefer it if it had worked well. Where all four rated parameters were low (e.g. TL/TR03.1, TU02.2/TD02.1 or RCW/RCCW 01.2), comments indicated that decision against using the particular gesture was not due to ratings but because they preferred not to use both hands.

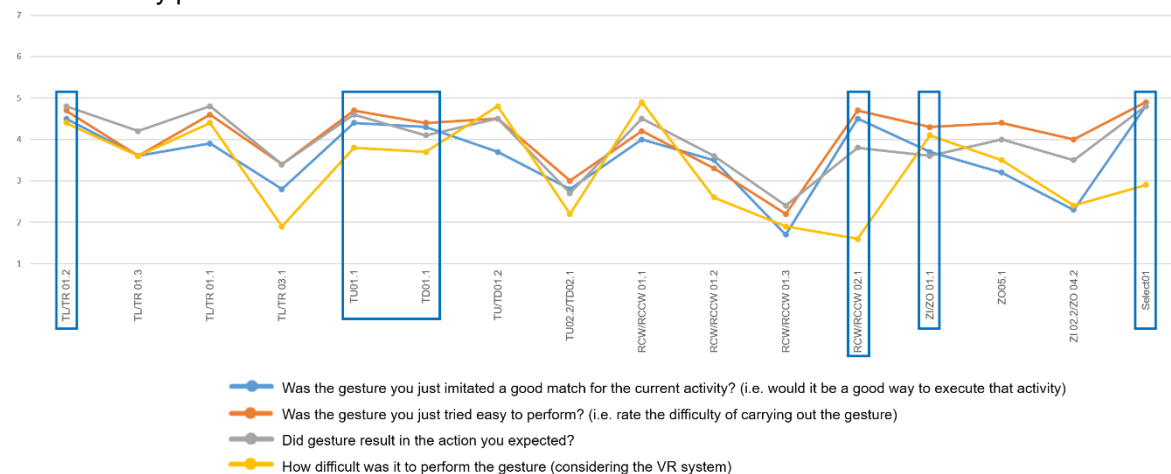


Figure 19 Averaged ratings across the ten participants for the VR application

It was noticeable that six out of ten individual participants explicitly stated a preference towards using one or both hands, unprompted. Where this was the case, majority of the participants (five of them) preferred to use one hand only as it was less tiring and more comfortable. One participant stated that using both hands was preferred as it gave them more control over the object. One participant stated that they would use both hands only if very precise motion was required, but would otherwise prefer one-handed gestures.

6.3. Variation of the consensus set including evaluation outcomes

Taking both full study analysis and evaluation of gestures by other professionals into account, a variation of the consensus set was created, shown in Figure 20. Gestures evaluated in abstract and VR evaluation are marked with a green and/or black circular shape to the top left of each gesture sketch, respectively. Gestures most frequently chosen as preferred for the abstract evaluation are circled in grey dotted line, are green dotted line for the VR evaluation. Eight out of eleven choices made following the VR evaluation matched the choices from the abstract evaluation. Where the choices did not match, both were retained. While the numbers of participants in the evaluation process and gestures explored across both evaluations were relatively low compared to the gesture elicitation study reported in Section 3, the overlap between chosen gestures from both evaluations may indicate resulting variations of gesture vocabularies between the evaluations are not highly divergent. While a VR evaluation of the complete vocabulary would likely introduce some changes compared to the results of the abstract evaluation, and should be further tested in future research, based on the outcomes of this evaluation the differences are not expected to be extensive.

The gestures from both consensus set and the discarded set (remainder of the consensus set compiled prior to elimination of activities where AR or Fleiss' kappa were not high enough) were tested, to ensure gestures were not eliminated prematurely, as the conditions during the two parts of the study varied. For the consensus set it was noticeable that most of the bi-manual gestures were eliminated, with the exception of gestures for *zoom in* and *out*, and one of the *extrude up* gestures. Additionally, gestures that were more frequently performed for only one shape were not chosen as preferred by the evaluators. It was also noticeable, that majority of the chosen gestures included an element of dimensionality e.g. for the translation gestures, hand tends to use all three planes rather than just remain in the vertical plane the objects were initially showed in to the original participants. For dichotomous activities, where a specific gesture was chosen for one activity and not for its pair, both were retained e.g. ZI05.1 was chosen and its paired gesture for *Zoom out* and ZO07.1 was not, but both were retained in the set. Where gesture was in the consensus set being tested, but originated in Part 2 of the study and was the only gesture of a specific category to be included in the set, it was retained in the final consensus set if the rate for appropriateness of the gesture for the activity and the ease of performance were high e.g. gesture for *selection* activity "Select01".

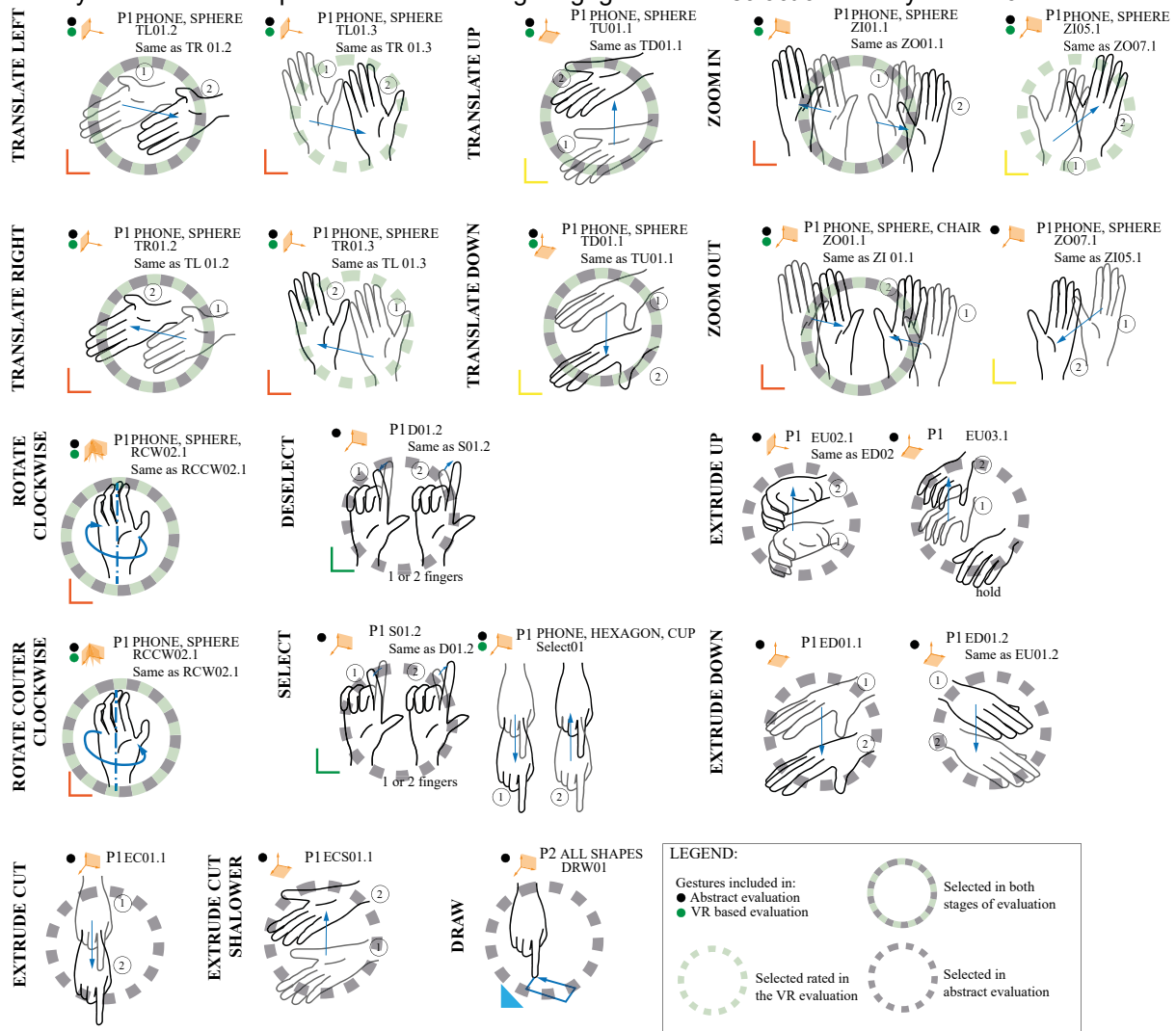


Figure 20 Variation of the consensus set including evaluation outcomes

Variation of the discarded set including evaluation outcome is given in the supplementary material. *Bend*, *Undo*, *Sculpt*, *Multiply/Pattern* and *Loft* were gestures that were initially not included in the consensus set due to low agreement during statistical analysis. Additionally, it seemed like the shape of the object being modified has influenced gestures suggested for those activities. *Loft* had a very low number of repetitions, and was excluded from the consensus set, but the remaining three gestures were reconsidered for inclusion following the evaluation. Ben 02 and Und01 have had high ratings for appropriateness of gestures for the activity, MulPat06 and Mulpat07 were on the lower end of the appropriate scale, while Scul02, Scul05 and Und02 were assessed as either not appropriate for

the activity, or between appropriate and inappropriate. Two participants have provided unprompted comments that for sculpt activity none of the gestures were ideal, and one participant thought the same about multiply/pattern activity. This combined with the generally low estimates of appropriateness for these gestures, and combined with the fact that statistical analysis has not shown significant results has led to the decision to retain gestures for these activities in the discarded set.

7. Discussion

Research reported in this paper resulted in three contributions which can be summarised as:

1. Knowledge about the natural gestures for conceptual design in 3D space, as the key contribution, which provides a basis for future development of CAD like systems for the support of conceptual design and could inform the future technology development requirements. This contribution will be discussed in Section 7.1.
2. Knowledge about the implementation of the most repeated gestures using currently readily available sensing technology, including ease of use, fatigue, limitations of existing technology and the effect those might have on gesture implementation, highlighting areas for future research and technology development. This contribution will be discussed in Section 7.2.
3. A mixed methods approach for studying natural hand gestures in 3D space, which has several strengths over existing approaches, along with some weaknesses, that could be improved through future research. This contribution will be discussed in Section 7.3.

7.1. Natural gestures for conceptual design and their envisaged application

The gesture elicitation study reported in this paper was designed to only provide instruction in terms of what participants can expect and what they will react to using their hands. No detailed instruction was given on how to perform the gestures. The intention behind this was to discover if the gestures identified were genuinely those participants would naturally suggest, without verbal influence from the authors. Instead, the setting in which the study was performed “pushed” them towards gestures performed in 3D space, by providing large amount of space they could use and 3D visualisations they responded to. Since the pool of gestures they could have performed was theoretically infinite, while the authors hoped that the repetition of gestures would take place, there was a risk of the outcomes being at a level where the repetition rates were not meaningful. Therefore, the finding that for 22 gestures correlated with specific activities the repetition was not likely to have happened by chance and that it was at statistically significant levels was a high level contribution on its own. The resulting vocabulary of gestures for conceptual design is considered to be the main contribution of this paper. This vocabulary is not purporting to become a definitive prescribed vocabulary of gestures for design, fit for implementation in a system in its current form. It is instead defining easy to perform, intuitive and natural gestures that can be a first step towards a more natural interaction with a CAD like system for conceptual design. It is envisaged that, in the future, the identified gestures may serve as a starting point in a system that adapts itself to the user, and learns from them over time, so that each user commands a unique system that is natural and intuitive to them.

Over the past three decades, the development of sensing technology has slowly become more and more applicable in interaction interfaces. Hundreds of gesture based interactive systems have been developed for different purposes, but other than in gaming very few have reached large scale implementation and adoption. This is particularly the case in the field of engineering design. Gestures may be a more natural and intuitive way of interaction with a 3D space and the virtual objects within it, but there are still many constraints present. The constraints are different from those present in WIMP (Windows Icon Pinter Mouse) systems. However, they still mean that users either have to learn the gestures predefined by system developers and perform them in a very specific way in order to achieve seamless recognition (Huang et al., 2018), or use the gestures for activities where free-form gestures are fit for use (Vinayak et al., 2013, Vinayak and Ramani, 2015). This leads to an increase in the cognitive load, detracts from the design goals and puts the focus on the process of interaction with the CAD system instead. The authors of the paper believe that these constraints are the key obstacle to more widespread adoption of gesture interfaces and provide the vocabulary of gestures for conceptual design as a set of initial default gestures that could be used in development of a novel

CAD like system that requirements for the technology to support it can be extrapolated from. This system could in the future potentially adapt to each individual user, by learning from them and modifying the gesture language used to control it accordingly and continuously, akin to predictive text feature on mobile phones.

The gestures comprising the vocabulary of conceptual design identified in this paper are largely manipulative gestures (translate, rotate, zoom in/out), gestures used to sketch the profiles (which are then extruded) and extrusion gestures (whether increasing the height of a part, decreasing it or cutting shapes out of it). The authors envisage the initial application would start with manipulative gestures and free sketches, as the manner those are currently performed using WIMP systems is not the most intuitive and natural. It is time consuming and often tedious to manipulate an object in a 3D environment using a mouse, and the sequence of activities that need to be performed differs from how objects would be interacted with in a 3D world. Utilisation of sketching features in current CAD systems similarly includes activities unlike those performed when sketching using a pen and a paper. There have been developments focused on improving the sketching abilities particularly in the automotive industry, but they have not propagated to the wider field of industrial design. Following the successful implementation of these types of gestures in a gesture controlled system, further research into both procedural structure of CAD based activities and ways it could potentially change depending on the shape being created could be explored. These combined could then inform the development of the system that provides the key capabilities “out of the box” and then adapts to the users and their specific workflows as it is being used.

7.2. Implementation challenges

The primary goal of the research reported was to find the most appropriate gestures for the activities, regardless of implementation technology limitations. However, the identified gestures were implemented and evaluated using a purpose built VR interaction system, in order to explore which gestures may be more appropriate for more general use without the need for extensive training designers typically require to become proficient in use of CAD systems. Observations made during this process that may be informative for future research in the field form the second contribution of this paper.

The limited VR implementation reported in Section 6.2 was aiming to reproduce the 16 most repeated gestures selected for implementation. However, due to limitations of the LEAP system used for gesture detection, some gestures had limited usability. Detection of grasping gestures was intermittent and only functioned well for a narrow window of angles between the fingers and the palm. Similarly, detection of open hand when the side of the hand was visible to the LEAP was intermittent. Participants needed multiple attempts to successfully use them and stated they had difficulties performing these gestures in the system. They have nevertheless chosen a grasping gesture for rotation (RCW02.1/RCCW02.1) and a gesture including a side view of the hand visible for the translations (TL01.2/TR01.2 and TL01.3/TR01.3). This was considered acceptable as the VR system was used primarily to evaluate the usability of gestures themselves and not to achieve the ultimate system implementation or evaluate gesture detection capability. Had the full implementation taken place immediately, the gestures would have likely needed to be modified to increase the detection rates and thus usability of the system. The authors believe that LEAP in particular is not able to support the full implementation of gestures for conceptual design in its current form. However, technology is developing rapidly and affordable systems with larger field of vision and higher detection capabilities are expected to be available in the near future. It should be considered that other limitations may materialise then, although they may be different from those we can foresee now. If that was to happen the vocabulary identified in this paper may need to be reassessed for those limitations, and potentially refined to comply with both designers’ preferred motions and technological capabilities of systems supporting the implementation.

During the implementation, the question of fatigue and ease of gesture performance became important. During the study gestures were initially elicited in, none of the participants mentioned the number of hands used or being tired, and the gestures performed were roughly evenly split between unimanual (47.8%) and bimanual gestures (52.2%). The consensus set had a slightly lower proportion of bi-manual gestures, but still comparable, 22 out of 54 gestures were bi-manual (41%). During the evaluation five participants stated, unprompted, that they preferred one handed gestures, as it was

less tiring and more comfortable. Only two out of 22 gestures (9%) included in the variation of the consensus set following the evaluation outcomes (see Figure 20) were bi-manual gestures. This further contributes to the finding that the gesture vocabulary elicited from participants without consideration for technology may need to be re-evaluated once the suitable technology is chosen for its practical implementation. The value of elicited gestures considered most natural may be compared to the value of similar gestures slightly adapted to account for the capabilities of the technology.

7.3. Mixed methods approach for studying natural hand gestures in 3D space

The methodology followed in the study is a combination of established approaches, adapted for 3D interaction, expanded with more robust statistical analysis. These adaptations could also be considered to be minor methodological contributions of the research.

They contributed to strengths of this study compared to similar studies in the field. This study includes a substantially higher number of participants. Participants were recruited until saturation of data was reached i.e. no new gestures were identified from a single participant. This led to 44 participants in this study compared to approximately 20 participants typically included in elicitation studies. The study setup was unique in that it focused on providing the participants with a theoretically infinite solution space in which they could suggest any gestures they found fit, but limiting the time they had to perform the activities. This limitation forced them to perform the gestures in real time without giving them the opportunity to create analogies with current ways of working and instead recording their intuitive responses. Statistical analysis of data was included to increase the robustness of the research. Results of the statistical analysis indicated that, for the included gestures, the agreement did not happen by chance. Any study where data is coded by human coders is inherently subjective. To counter this three coders were involved in order to increase the consensus in the coding process. After they reached an agreement on the parameters of the coding process, they coded 10% of the sample, and their outputs have shown that appropriate levels of inter-coder reliability were reached. Following the formation of the gesture vocabulary for conceptual design, the evaluation was performed by a different group of participants, comprised of professionals other than product designers. They confirmed the gestures were appropriate for the activities they were assigned to, and further refined the vocabulary by selecting their preferred gesture for each activity. This variation of the gesture vocabulary created following the evaluation could increase usability of a future system potentially making the design process accessible to other target groups.

The weaknesses of the research should also be acknowledged. While the number of participants was higher than in a typical study, majority of the participants (38 out of 44) were from the same institution. They would have received the same training and this may have contributed to the high levels of agreement. The participants were shown the visuals that did not include CAD commands or environment explicitly. However, due to their training, which included CAD systems, they could have conceivably linked the aesthetics of the visuals to a CAD system. This may have inadvertently affected the gestures performed in Part 1 of the study or the activities the participants chose to perform in the Part 2 of the study. Finally, as more extensively discussed in Section 7.2, the VR portion of the evaluation was limited by the capabilities of the technology used to implement it and it should be revisited in the future when technology is more developed and all of the gestures can be implemented seamlessly.

Observing the study set up and its effect on the outcomes, three areas were identified where introducing further methodological changes could lead to enhanced data collection in the future work. The conceptual design stage typically does not focus on the detail of dimensions and positioning. The study was set up to reflect this, and did not allow participants to focus on details. However, a number of participants noted that they might have acted differently if they were required to perform activities that were more specific or dimension some elements of the objects created. In the future it may be interesting to include dimensioning or positioning capability that requires more precision and compare the differences or changes in the gesture vocabulary this may lead to.

In Part 1 of the study all participants responded to the same activities viewed on the screen and in Part 2 they were free to devise their own activities in order to achieve the visualised outcome. Following the analysis it was noted that agreement rates were at significant levels for a larger number of activities in Part 1 of the study. The authors believe that this is largely due to the lower numbers of repetitions for each activity in Part 2, as more varied activities were chosen by the participants. In any

future study where the activity definition is at least partially in the hands of the participants, a larger number of participants should be included and the study performed until the repetition numbers for individual activities observed are at comparable levels.

During the coding, it was noticed that gestures for *Bend*, *Sculpt*, *Multiply/Pattern*, *Loft* and *Undo* depended highly on the shape of the object being manipulated. *Undo* was occasionally performed using a symbolic gesture that did not take into account the shape of the object, such as the hand emulating a gesture akin to wiping a surface. At times the activity being “undone” was the activity of object creation performed in reverse. For example, if a triangle was rotated around an axis during the *pattern* activity, to undo it the *pattern* activity is performed in reverse i.e. rotation is performed in the opposite direction. It may be necessary to modify the data collection practices for the shape dependent activities, and explore if shape dependence, in terms of gesture use, is desirable. To do this, inclusion of larger variety of different shapes in future studies may be needed.

8. Conclusions

A gesture elicitation study was performed including 44 product design engineering students, in order to create a novel user based vocabulary of gestures for conceptual design. The aim of the study was to discover which gestures designers would use during conceptual design if they were not influenced by what was achievable using existing technology. Performed gestures were parsed, sketched and coded based on the adopted taxonomy, following the user based gesture elicitation approach established in the field of HCI. Then they were categorised based on the hand form and the path travelled, in a number of categories and sub categories. Statistical analysis including combination of AR (agreement rate) and Fleiss κ were used to determine which gestures should be added to the consensus set for the gesture vocabulary for both parts of the study. Chi square analysis was then applied to the categories of gestures in order to determine the likelihood of number of repetitions for each category occurring by chance. Following Chi square analysis, a number of categories were disqualified from the consensus set, and then consensus sets for part one and part two of the study were merged into a unified consensus set of gestures for conceptual design. This consensus set represents the variants for 20 in-air gestures frequently used during conceptual design. A variant of the consensus set has been refined via evaluation activity performed using other professionals, in order to achieve higher adoptability of gestures by inexperienced users and reduce need for gesture learning. Gesture vocabulary was evaluated using abstract evaluation, employing participants’ imagination, and a subset of most frequently repeated gestures within the vocabulary was also evaluated using VR hardware. VR evaluation demonstrated that gesture vocabulary elicited from participants without consideration for technology may need to be re-evaluated once the suitable technology is chosen for its practical implementation, as it may require modifications. Effect of object shape was noticeable for more complex modification activities such as pattern or undo, where the gesture performed was often not generic but closely linked to the shape of the object being modified. These types of activities may require further exploration in future studies. However, it has been found that for majority of frequently performed conceptual design activities, gestures elicited from the user show significant agreement across the sample, and the resulting vocabularies of gestures for conceptual design have the potential to inform future human computer interaction development.

9. Acknowledgements

This research was supported by the United Kingdom’s Engineering and Physical Sciences Research Council (EPSRC) (grant number EP/M012123/1 – LH, AD, MG;), and two EPSRC/University of Strathclyde Research Studentships (EP/M508159/1 – GC; EP/M506643/1 – CM).

The funding for the development of the VR evaluation system was provided by AFRC (Advanced Forming Research Centre) CATAPULT Route to Impact funding awarded in 2018 - AFRC_CATP_1220_RTI.

10. References

AL-SHAMAYLEH, A. S., AHMAD, R., ABUSHARIAH, M. A., ALAM, K. A. & JOMHARI, N. 2018. A systematic literature review on vision based

- gesture recognition techniques. *Multimedia Tools and Applications*, 77, 28121-28184.
- ALCAIDE-MARZAL, J., DIEGO-MÁS, J. A., ASENSIO-CUESTA, S. & PIQUERAS-FISZMAN, B. 2013. An exploratory study on the use of digital sculpting in conceptual product design. *Design Studies*, 34, 264-284.
- ANONYMOUS 2018. Details omitted for double-blind reviewing.
- ANONYMOUS 2019. Details omitted for double-blind reviewing.
- ARROYAVE-TOBÓN, S., OSORIO-GÓMEZ, G. & CARDONA-MCCORMICK, J. F. 2015. AIR-MODELLING: A tool for gesture-based solid modelling in context during early design stages in AR environments. *Computers in Industry*, 66, 73-81.
- BAILEY, S. K., JOHNSON, C. I. & SIMS, V. K. Using Natural Gesture Interactions Leads to Higher Usability and Presence in a Computer Lesson. Congress of the International Ergonomics Association, 2018. Springer, 663-671.
- BEATTIE, N., HORAN, B. & MCKENZIE, S. 2015. Taking the LEAP with the Oculus HMD and CAD-Plucking at thin Air? *Procedia Technology*, 20, 149-154.
- BENAMI, O. & JIN, Y. Creative stimulation in conceptual design. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2002. 251-263.
- BERINGER, N. Evoking gestures in SmartKom-Design of the graphical user interface. International Gesture Workshop, 2001. Springer, 228-240.
- BEYER, G. & MEIER, M. Music Interfaces for Novice Users: Composing Music on a Public Display with Hand Gestures. NIME, 2011. 507-510.
- BHUIYAN, M. & PICKING, R. 2011. A gesture controlled user interface for inclusive design and evaluative study of its usability. *Journal of software engineering and applications*, 4, 513.
- BOURDOT, P., CONVARD, T., PICON, F., AMMI, M., TOURAINÉ, D. & VÉZIEN, J.-M. 2010. VR-CAD integration: Multimodal immersive interaction and advanced haptic paradigms for implicit edition of CAD models. *Computer-Aided Design*, 42, 445-461.
- BUCHMANN, V., VIOLICH, S., BILLINGHURST, M. & COCKBURN, A. FingARtips: 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, 2004. ACM, 212-221.
- CHAKRABARTI, A. & BLIGH, T. P. 1994. An approach to functional synthesis of solutions in mechanical conceptual design. Part I: Introduction and knowledge representation. *Research in Engineering Design*, 6, 127-141.
- CHU, C.-C. P., DANI, T. H. & GADH, R. 1997. Multi-sensory user interface for a virtual-reality-based computeraided design system. *Computer-Aided Design*, 29, 709-725.

- DANI, T. H. & GADH, R. 1997. Creation of concept shape designs via a virtual reality interface. *Computer-Aided Design*, 29, 555-563.
- DAVE, D., CHOWRIAPPA, A. & KESAVADAS, T. 2013. Gesture interface for 3d cad modeling using kinect. *Computer-Aided Design and Applications*, 10, 663-669.
- DELLER, M., EBERT, A., BENDER, M. & HAGEN, H. Flexible gesture recognition for immersive virtual environments. Information Visualization, 2006. IV 2006. Tenth International Conference on, 2006. IEEE, 563-568.
- DIM, N. K., SILPASUWANCHAI, C., SARCAR, S. & REN, X. Designing Mid-air TV gestures for blind people using user-and Choice-Based Elicitation Approaches. Proceedings of the 2016 ACM Conference on Designing Interactive Systems, 2016. ACM, 204-214.
- DINH, D.-L., KIM, J. T. & KIM, T.-S. 2014. Hand gesture recognition and interface via a depth imaging sensor for smart home appliances. *Energy Procedia*, 62, 576-582.
- DONG, H., FIGUEROA, N. & EL SADDIK, A. An Elicitation Study on Gesture Attitudes and Preferences Towards an Interactive Hand-Gesture Vocabulary. Proceedings of the 23rd ACM international conference on Multimedia, 2015. ACM, 999-1002.
- ERICSSON, K. A. & SIMON, H. A. 1984. *Protocol analysis: Verbal reports as data*, the MIT Press.
- FINDLATER, L., LEE, B. & WOBBROCK, J. Beyond QWERTY: augmenting touch screen keyboards with multi-touch gestures for non-alphanumeric input. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012. ACM, 2679-2682.
- FOEHRENBACH, S., KÖNIG, W. A., GERKEN, J. & REITERER, H. 2009. Tactile feedback enhanced hand gesture interaction at large, high-resolution displays. *Journal of Visual Languages & Computing*, 20, 341-351.
- FRENCH, M. J., GRAVDAHL, J. & FRENCH, M. 1985. Conceptual design for engineers, Springer.
- GROBMAN, Y. J., YEZIORO, A. & CAPELUTO, I. G. 2009. Computer-Based Form Generation in Architectural Design — A Critical Review. *International Journal of Architectural Computing*, 7, 535-553.
- HAN, Y. C. & HAN, B.-J. 2014. Virtual pottery: a virtual 3D audiovisual interface using natural hand motions. *Multimedia Tools and Applications*, 73, 917-933.
- HARTMAN, N. W. 2009. Defining expertise in the use of constraint-based CAD tools by examining practicing professionals. *Engineering Design Graphics Journal*, 69.
- HAYES, A. F. & KRIPPENDORFF, K. 2007. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1, 77-89.

- HOLZ, C. & WILSON, A. Data miming: inferring spatial object descriptions from human gesture. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011. ACM, 811-820.
- HUANG, C.-C. S. 2007. Conceptual Modeling Environment (COMOEN). *Computer-Aided Architectural Design Futures (CAADFutures) 2007*. Springer.
- HUANG, J., JAISWAL, P. & RAI, R. 2018. Gesture-based system for next generation natural and intuitive interfaces. *AI EDAM*, 1-15.
- HUANG, J., JAISWAL, P. & RAI, R. 2019. Gesture-based system for next generation natural and intuitive interfaces. *Ai Edam*, 33, 54-68.
- ISBISTER, K. & MUELLER, F. F. 2015. Guidelines for the design of movement-based games and their relevance to HCI. *Human-Computer Interaction*, 30, 366-399.
- JÄÄSKELÄINEN, R. 2010. Think-aloud protocol. *Handbook of translation studies*, 1, 371-374.
- JAHANI, H. & KAVAKLI, M. 2018. Exploring a user-defined gesture vocabulary for descriptive mid-air interactions. *Cognition, Technology & Work*, 20, 11-22.
- KANG, J., ZHONG, K., QIN, S., WANG, H. & WRIGHT, D. 2013. Instant 3D design concept generation and visualization by real-time hand gesture recognition. *Computers in Industry*, 64, 785-797.
- KEINONEN, T. K. & TAKALA, R. 2010. *Product concept design: a review of the conceptual design of products in industry*, Springer Science & Business Media.
- KELA, J., KORPIPÄÄ, P., MÄNTYJÄRVI, J., KALLIO, S., SAVINO, G., JOZZO, L. & MARCA, D. 2006. Accelerometer-based gesture control for a design environment. *Personal and Ubiquitous Computing*, 10, 285-299.
- KHAN, S. & TUNÇER, B. 2019. Gesture and speech elicitation for 3D CAD modeling in conceptual design. *Automation in Construction*, 106, 102847.
- KIM, H., ALBUQUERQUE, G., HAVEMANN, S. & FELLNER, D. W. Tangible 3D: Hand Gesture Interaction for Immersive 3D Modeling. IPT/EGVE, 2005a. 191-199.
- KIM, H., ALBUQUERQUE, G., HAVEMANN, S. & FELLNER, D. W. 2005b. Tangible 3D: Hand Gesture Interaction for Immersive 3D Modeling. *IPT/EGVE*, 2005, 191-9.
- KIM, M. & LEE, J. Y. 2016. Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality. *Multimedia Tools and Applications*, 75, 16529-16550.
- KINECT, M. 2018. Available: <https://msdn.microsoft.com/en-gb/library/hh438998.aspx> [Accessed 13/02/2018 2018].
- KOUTSABASIS, P. & VOGIATZIDAKIS, P. 2019. Empirical research in mid-air interaction: A systematic review. *International Journal of Human-Computer Interaction*, 35, 1747-1768.

- KRIPPENDORFF, K. 2004. Content analysis: An introduction to its methodology Thousand Oaks. *Calif.: Sage*.
- LEAP MOTION INC. 2018. *LEAP MOTION* [Online]. Available: <https://www.leapmotion.com/> [Accessed 13/02/2018 2018].
- LEE, M. & BILLINGHURST, M. A Wizard of Oz study for an AR multimodal interface. Proceedings of the 10th international conference on Multimodal interfaces, 2008. ACM, 249-256.
- LIDDAMENT, T. 1999. The computationalist paradigm in design research. *Design studies*, 20, 41-56.
- LIKKANEN, L. A. & PERTTULA, M. 2009. Exploring problem decomposition in conceptual design among novice designers. *Design studies*, 30, 38-59.
- LOMBARD, M., SNYDER-DUCH, J. & BRACKEN, C. C. 2002. Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human communication research*, 28, 587-604.
- LOPES, D. S., DE FIGUEIREDO PARREIRA, P. D., PAULO, S. F., NUNES, V., REGO, P. A., NEVES, M. C., RODRIGUES, P. S. & JORGE, J. A. 2017. On the utility of 3D hand cursors to explore medical volume datasets with a touchless interface. *Journal of biomedical informatics*, 72, 140-149.
- MCAWEENEY, E., ZHANG, H. & NEBELING, M. User-driven design principles for gesture representations. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018. 1-13.
- MCNEILL, D. 1992. *Hand and mind: What gestures reveal about thought*, University of Chicago press.
- MEMO, A. & ZANUTTIGH, P. 2018. Head-mounted gesture controlled interface for human-computer interaction. *Multimedia Tools and Applications*, 77, 27-53.
- MILANI, F., DE MARCHI, A. C. B. & RIEDER, R. Usability Guidelines to Develop Gesture-Based Serious Games for Health: A Systematic Review. 2017 19th Symposium on Virtual and Augmented Reality (SVR), 2017. IEEE, 188-194.
- MORRIS, M. R., WOBROCK, J. O. & WILSON, A. D. Understanding users' preferences for surface gestures. Proceedings of graphics interface 2010, 2010. Canadian Information Processing Society, 261-268.
- NACENTA, M. A., KAMBER, Y., QIANG, Y. & KRISTENSSON, P. O. Memorability of pre-designed and user-defined gesture sets. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013. ACM, 1099-1108.
- NOOR, A. K. & ARAS, R. 2015. Potential of multimodal and multiuser interaction with virtual holography. *Advances in Engineering Software*, 81, 1-6.
- PISHARADY, P. K. & SAERBECK, M. 2015. Recent methods and databases in vision-based hand gesture recognition: A review. *Computer Vision and Image Understanding*, 141, 152-165.

- PIUMSOMBOON, T., CLARK, A., BILLINGHURST, M. & COCKBURN, A. User-defined gestures for augmented reality. IFIP Conference on Human-Computer Interaction, 2013. Springer, 282-299.
- QIN, S., WRIGHT, D. K., KANG, J. & PRIETO, P. 2006. Use of three-dimensional body motion to free-form surface design. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 220, 335-339.
- QUEK, F. 2004. The catchment feature model: A device for multimodal fusion and a bridge between signal and sense. *Eurasip Journal on Applied Signal Processing*, 2004, 1619-1636.
- RAUTARAY, S. S. & AGRAWAL, A. 2015. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial intelligence review*, 43, 1-54.
- REVERON, F. O. Developing spatial configuration abilities coupled with the space syntax theory for first year architectural studies. Proceedings of the 7th international space syntax symposium, 2009. 81-82.
- RIDUWAN, M., BASORI, A. H. & MOHAMED, F. 2013. Finger-based gestural interaction for exploration of 3D heart visualization. *Procedia-Social and Behavioral Sciences*, 97, 684-690.
- RIENER, A., FERSCHA, A., BACHMAIR, F., HAGMULLER, P., LEMME, A., MUTTENTHALER, D., PUHRINGER, D., ROGNER, H., TAPPE, A. & WEGER, F. Standardization of the in-car gesture interaction space. 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2013, October 27, 2013 - October 30, 2013, 2013 Eindhoven, Netherlands. Association for Computing Machinery, 14-21.
- ROBINSON, G., RITCHIE, J. M., DAY, P. N. & DEWAR, R. G. 2007. System design and user evaluation of Co-Star: An immersive stereoscopic system for cable harness design. *Computer-Aided Design*, 39, 245-257.
- RUIZ, J., LI, Y. & LANK, E. User-defined motion gestures for mobile interaction. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011. ACM, 197-206.
- SANTOS¹, C. A., SALEME¹, E. B. & DE ANDRADE¹, J. C. 2015. A systematic review of data exchange formats in advanced interaction environments. *International Journal of Multimedia and Ubiquitous Engineering*, 10, 123-140.
- SCHMIDT, A. 2015. Following or leading?: the HCI community and new interaction technologies. *interactions*, 22, 74-77.
- SCHULER, D. & NAMIOKA, A. 1993. *Participatory design: Principles and practices*, CRC Press.
- SHESH, A. & CHEN, B. Smartpaper: An interactive and user friendly sketching system. Computer Graphics Forum, 2004. Wiley Online Library, 301-310.
- SONG, J., CHO, S., BAEK, S.-Y., LEE, K. & BANG, H. 2014. GaFinC: Gaze and Finger Control interface for 3D model manipulation in CAD application. *Computer-Aided Design*, 46, 239-245.

- TSANDILAS, T. 2018. Fallacies of agreement: A critical review of consensus assessment methods for gesture elicitation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 25, 18.
- VARGA, E., DOROZHKIN, D. V. & VANCE, J. M. Analysis of the usability of hand motion language in shape conceptualization. DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 28.-31.07. 2007, 2007. 681-682 (exec. Summ.), full paper no. DS42_P_504.
- VATAVU, R.-D. & WOBBROCK, J. O. Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015. ACM, 1325-1334.
- VINAYAK, MURUGAPPAN, S., LIU, H. & RAMANI, K. 2013. Shape-It-Up: Hand gesture based creative expression of 3D shapes using intelligent generalized cylinders. *Computer-Aided Design*, 45, 277-287.
- VINAYAK & RAMANI, K. 2015. A gesture-free geometric approach for mid-air expression of design intent in 3D virtual pottery. *Computer-Aided Design*, 69, 11-24.
- VULETIC, T., DUFFY, A., HAY, L., MCTEAGUE, C., CAMPBELL, G. & GREALY, M. 2019. Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies*, 129, 74-94.
- WELCH, R. V. & DIXON, J. R. 1994. Guiding conceptual design through behavioral reasoning. *Research in Engineering Design*, 6, 169-188.
- WOBBROCK, J. O., AUNG, H. H., ROTHROCK, B. & MYERS, B. A. Maximizing the guessability of symbolic input. CHI'05 extended abstracts on Human Factors in Computing Systems, 2005. ACM, 1869-1872.
- WOBBROCK, J. O., MORRIS, M. R. & WILSON, A. D. User-defined gestures for surface computing. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2009. ACM, 1083-1092.
- XIAO, Y. & PENG, Q. A hand gesture-based interface for design review using leap motion controller. DS 87-8 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 8: Human Behaviour in Design, Vancouver, Canada, 21-25.08. 2017, 2017. 239-248.
- YUAN, C. Visual tracking for seamless 3d interactions in augmented reality. *International Symposium on Visual Computing*, 2005. Springer, 321-328.
- ZHONG, K., KANG, J., QIN, S. & WANG, H. Rapid 3d conceptual design based on hand gesture. *Advanced Computer Control (ICACC)*, 2011 3rd International Conference on, 2011. IEEE, 192-197.
- ZIMMERMAN, T. G., LANIER, J., BLANCHARD, C., BRYSON, S. & HARVILL, Y. A hand gesture interface device. *ACM SIGCHI Bulletin*, 1987. ACM, 189-192.