Comparing hand-centric designs for a menu selection task in augmented reality

Willem van Kemenade
Utrecht University
Utrecht, The Netherlands
w.vankemenade@students.uu.nl
S.N.: 8716471

Thesis supervisors:
Lynda Hardman & Wolfgang Hürst

13/07/2022

Abstract

Selection in Augmented Reality (AR) is a common task that can be used to, for example, select an item from a menu. There are many approaches to creating a selection interaction such as controller, gesture or gaze input, each having their pros and cons. Independent of any particular implementation, the interaction should be fast and usable since it is a common task. Furthermore, the cognitive load should be kept as low as possible for the best user experience.

A typical menu in augmented and virtual reality is a window with buttons fixed in virtual space, which forces the user to keep in mind where the menu is located or forced to relocate to use it. One type of design that solves this is hand-centric design. This design anchors objects in the virtual space around the user's hand which allows the interaction to always be in the same relative location to the user.

We created a finger based (SwipePie) and a palm based design (The Widget Pallet). These designs were implemented in a prototype AR system called DatAR and evaluated using a human-centred design approach. DatAR is part of a research endeavour to explore interacting with large numbers of neuroscience publications [28]. DatAR provided an environment to implement the designs and provide suitable tasks for evaluating them. Our goal is both to improve the selection interaction in the DatAR system, and to create a general solution that can be used for similar systems.

The results from a pilot study indicated that participants preferred the Widget Pallet designs over the SwipePie designs. A second evaluation examined how the Widget Pallet designs compared based on our three criteria: speed, cognitive load and usability. The results indicate no statistical difference on all criteria.

1 Introduction

Selecting an item from a menu is a common task that is seen in different places such as: the computer, mobile phone, virtual reality

Figure 1: Visualisation of depression co-occurence | image taken from [32][1]

(VR) and in augmented reality (AR). There are three main ways of interacting with a menu in VR and AR. The first is a controller where the most common method is the ray cast [21]. The second is a gesture-based method which uses a specific type of head or hand movement to select an option [25]. The third is a gaze-based method that uses head or eye movement to gaze at a menu button to select it [23]. An example of a system that uses menu selection often is the DatAR project, which is an AR project that aims to help neuroscience researchers with the exploration of their domain literature [28]. Offering this system in AR provides a visual way to explore a large amount of literature and discover how various topics, in particular brain diseases, relate to brain regions and each other. DatAR is built using widgets, which are objects containing one functionality that can be connected with other widgets. For example, an inlet widget provides data, Figure 2, and an outlet widget uses data to create visualisations, Figure 3.

Interacting with widgets and their inputs consists of two main actions: selecting which objects to create in virtual space and connecting them. The user selects which widgets and inputs to use and then connects these together to receive results from the system. An example of what can be done is to use these widgets and inputs to create a visualisation that shows which regions of the brain are found to be affected by the disease depression, see Figure 1. The creation of a widget is done through factory widgets, Figure 4, where the user pinches their thumb and index finger on one of the objects and drags it away. This essentially works as the menu for the user to select which option they would like to use. The input can be created in the same way and connection is done by dragging it into the input location that can be seen in Figure 2. Our goal is to both improve the selection interaction in DatAR and to create a sufficiently general solution for similar systems. To be clear, the aim is not to change the functionality but how the user interacts with it.

The process of manually grabbing

[1]From the video DatAR: An Immersive Literature Exploration Environment for Neuroscientists | AIVR conference 2020 `https://www.youtube.com/watch?v=PnOPECRNc_w`

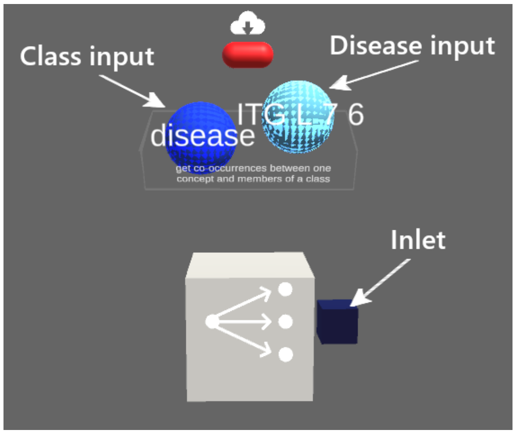


Figure 2: Inlet widget example

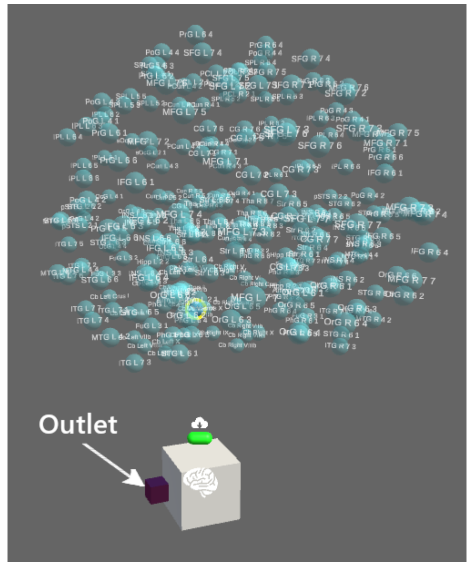


Figure 3: Outlet widget example

inputs and widgets takes considerable time and effort to complete **(speed)**. The pinching interaction does not always work flawlessly and can become frustrating **(user experience)** when it takes multiple tries before the interaction functions **(cognitive load)**. Trying out the system and experiencing how much effort it took to complete tasks was a source of inspiration for this project. A different form of interaction could improve the selection task for the creation of widgets, leading to the following research question:

Can we create a selection interaction that works fast, with low cognitive load and a pleasant user experience?

We address each of the three aspects separately:

- **Can we create a fast selection interaction?**
  - Our hypothesis is that a fast method will lead to a good experience. A slower method that is usable with low cognitive load could still be a good solution.
- **Can we create an interaction that requires low cognitive load?**
  - Our hypothesis is that having a low cognitive load during interaction will lead to a good user experience.
- **Can we create an interaction with a pleasant user experience?**
  - Our hypothesis is that a fast interaction and an interaction that requires low cognitive load will impact the user experience in a positive way. In addition to these aspects it is important to examine if the designed interaction is usable, intuitive and enjoyable.

To evaluate the proposed designs we use a human centered design approach, which means that at the end of a designing phase we use participants to examine how they use it and to receive feedback on what works well and what could be improved. We split the evaluation into two studies. In the first study, the described metrics of speed, cognitive load and usability will not be used, since answering surveys for all four designs will take too much time to complete. The first study

Figure 4: Factory widgets

is to see which two designs are most prefered by the users and the second study will compare the chosen designs based on the three described metrics

The remainder of this document shows the related work which explains different forms of selection interaction, our chosen design philosophy and how cognitive load and user experience relate to the user interface. This is followed by our design requirements which we used to create our finger and palm-based designs. We describe the pilot study with its results, most notable feedback and selected designs. The selected designs were compared in the last study using our chosen metrics to answer the main and sub-research questions in the discussion and conclusion. Finally, we explain the limitations, describe some possibilities for future work and reflect on some aspects of the project.

2 Related work

We review different aspects of design and interaction to provide background information and inspiration for creating a new interaction design. body- and hand-centric design could help with ease of use and reducing cognitive load. Selection in VR/AR has many forms of interaction and there are possibilities of using gesture or gaze for the new interaction. Cognitive load and user experience is important to keep in mind when designing interactions for users.

2.1 Body- and hand-centric design

Mobile phones are limited in their interaction space, it only goes as far as the screen size. Chen et al. [5] proposed body-centric design with the idea that the mobile device could be used to navigate through content with the body as the anchor point. This means that no matter where the user is located the interaction will always be around the body. The problem of limited interaction space is not an issue in VR/AR since it is much larger than a mobile device, but the idea of a body-centric design is still useful.

Objects in VR/AR usually have a fixed location in 3D space. If it is something that the user needs to interact with often the user needs to remember where the location of it is or make sure to keep it close. With a body-centric design this object will be anchored to the location of the user, which makes it so that the user does not need cognitive effort to remember where the object is located.

Xu et al. [34] proposes a hand-centric design that takes the same idea as body-centric but using the hand as the centre point of the interaction. An interface goes from the wrist to the tip of the finger and is used for pointing tasks. The research concludes with a set of design considerations when using a hand-centric design, which are the following:

- Do not rotate the interface when the user's hand rotates.
- Keep the position of the interface inside the comfort level.
- Interface in the slightly left bottom direction works best and use flexion

movement.

Interacting with the widgets is done with the hands. Therefore, we expect that it will be faster, reduce cognitive load and easier to use the hand as the centre point of the interaction instead of the body as a whole. Similar to body-centric this adds the benefit that the interaction can always easily be within reach.

2.2 Selection interaction in VR/AR

The selection interaction in VR/AR is the type of interaction that we want to improve upon. Selection in this context just means to select an object in the system not a type of aiming or gaming related interaction. There are three types of selection: controller, gesture, and gaze interaction. The aim of this section is to look into the possibilities of using gesture or gaze interaction for improving the interaction.

2.2.1 Gesture interaction

Gestures convey information through body movements [19]. In VR and AR the Head Mounted Display (HMD) takes movements from the hands or head and interprets those movements as a type of action. It can be used for different tasks such as controlling a menu [7], browsing a collection [16], or switching modes. Modes are distinct user interface inputs where different actions result from the same input [24]. An example of a mode switch in VR/AR is during a drawing task where the user can draw with a pencil and switches modes with a gesture to change the input to act as an eraser. There are papers examining different methods for switching modes such as by head gestures [25] or hand gestures [26].

In the DatAR system there are two distinct modes the user can be in: in the default **widget interacting/manipulation mode** the user can grab widgets to move them in 3D space. If the user wants to create new widgets they need to switch to the **widget creation mode**. Using a hand-centric design the interaction could be in front of the user's hand and if the user tries to grab a widget the system will get confused whether a button is pressed or a widget is grabbed. This is why there needs to be distinct modes to avoid conflicting interactions.

Head and hand gestures both work well, although the head based gestures are most beneficial in a task where both hands are used simultaneously. In DatAR it is possible to interact with widgets using both hands though it is not required or necessary. The research on hand gestures shows that using your dominant hand for interaction and the non-dominant hand for mode switching works the best and is most prefered by the user. This would be a good way to interact with the system where the dominant hand can interact with the widgets while the non-dominant hand could use a fist or palm gesture to switch between the modes.

Gestures can have many different meanings that can change depending on the context or culture. This makes them complex but is able to simplify a task because we as humans are already familiar with using gestures. The issue is that there are no explicit standards for which gesture fits which context. This makes it a difficult task to decide what fits in a system and makes sense to a user [35].

2.2.2 Gaze interaction

Gaze interaction uses the head- or eye-tracking of the headset to know where the user is looking. The easiest to use and implement interaction is the raycast method where rays are cast from the eyes to highlight and select items [29].

A comparison study between gaze

and controller aiming showed that gaze aiming has less cognitive load but lower accuracy than a controller [20]. Furthermore, with a pointing task no significant difference was found between the controller and gaze using the ray cast method [30]. The difference between controller and gaze selection has also been studied in different contexts such as a drawing task [23]. The results from the study showed that the Dunk brush performed the best and was prefered by most users while the Dwell gaze was the worst and was least prefered. Although, it did show that the gaze interaction had the least cognitive load compared to other methods. The Dunk brush and Dwell gaze may have different effects in another context such as using the pallet to choose widgets. Choosing widgets will not be as much interaction as choosing a colour and drawing, meaning that the Dwell gaze could still be useful in a different context even if it ranked the lowest. When using gaze interaction, specifically eye-tracking there is an issue with involuntary eye movements and the Midas touch problem making it more inaccurate to use [33].

2.3 Cognitive load in user interfaces

Cognitive load is the working memory that processes information that we as humans receive through our senses [1]. Our short-term memory is the same as RAM, it has a limited capacity, and it stores information that only needs to be remembered for a short time. If the information is sufficiently important, it can be put through to the hard drive or long-term memory [10]. This working memory can be compared to using an interface where the user is using their senses to navigate and use the interface. When there is too much information to process in the working memory, it becomes too much to handle and this leads to a cognitive overload. Cognitive overload when using an interface can lead to a worse user experience.

In cognitive psychology, there are three types of cognitive load: intrinsic, extraneous, and germane. Intrinsic cognitive load is “the inherent level of difficulty associated with a specific instructional topic” [4]. Extraneous cognitive load is “generated by the manner in which information is presented to learners and is under the control of instructional designers” [4] and germane cognitive load is “the processing, construction, and automation of schemas” [27]. Meaning that germane cognitive load handles the creation of long-term storage of information. The type of cognitive load that can be handled by a good UX design is the extraneous cognitive load, because it deals with anything that takes up mental resources that are not related to the main task at hand [15]. For example, click on the correct menu item for the page you want. If the menu does not work correctly or not the way the user expected then this adds to the extraneous cognitive load. Small frustrations can add up to a bad experience for the user.

There are multiple ways to measure the perceived mental effort. One is a construct that provides a means of comparing instructional conditions, combining it with mental effort ratings and performance scores [22]. Another way of measuring is by using instruments such as a heart rate and/or blood pressure monitor to measure the cognitive and physical workload [8]. The pupillary response is also a way that has been used to measure cognitive load. Greater pupil dilation is found to be associated with high cognitive load [9]. The last type of measurement is a subjective survey such as the NASA-TLX which is a widely used tool [11] that rates perceived workload to assess a task, system, or team’s effectiveness [12]. The NASA-TLX uses pairwise comparisons to add weights to the scales. To keep the number of questions low we removed this and will use the raw version of the questionnaire. Examining the difference between normal and raw NASA-TLX shows no conclusive evidence which is better [11].

2.4 User experience

User experience examines if a created interaction is intuitive, usable and enjoyable in combination with reducing the cognitive load of a user [13]. A well designed interaction should work with the limits of the cognitive load and therefore reduce to improve the experience [31]. To make sure this is the case, all elements should be measured when testing the interaction. There are numerous surveys that evaluate the user experience, such as SUS [3], UEQ [17], PUEU [6], and CSUQ [18]. These all have slightly different ways of asking how well the interaction is perceived by the user.

UEQ is the most appropriate for our task because it asks questions about whether the interaction is usable on elements such as efficiency, perspicuity, dependability but also user experience aspects such as stimulation and novelty. Other than answering the question if the solution is usable we would also like to know if participants are motivated to use it or think if it is a novel solution.

Furthermore, the authors of the survey offer a handbook with helpful information about the survey and spreadsheets which make the analysis easier, because it provides multiple graphs and statistical tests.

3 Design requirements

When creating a new interaction concept, it is important to consider design requirements because they describe what the design should focus on and achieve. Ideally, all requirements would be met by a suitable design although this is not always possible. Every requirement is colour coded with green as high priority, orange as medium priority and red as low priority. Our aim is to minimally satisfy all high priority requirements. Requirements that include no citations were created from personal intuition based on using the DatAR system and programming/design experience.

- **Speed**
  - **Fast**: The new interaction should work as quickly as possible. Having the user spend a lot of time on simple interactions will lead to bad usability [34].
- **Cognitive load**
  - **Low cognitive load**: The interaction should require low cognitive load from the user. It should feel as a smooth experience with as few frustrations as possible.
- **User experience**
  - **Within reach**: Our focus is to create a **hand-centric interaction** so the interaction is always within reach of the user regardless of their location [34].
  - **Comfortable**: the experience of using the new interaction should be comfortable and not have issues such as fatigue or unnatural positions [34].
  - **Limited options**: The designs should not contain too much visual clutter. The more options given to the user the longer the decision time will take [14].
  - **Intuitive**: Users should be able to understand how the new interaction works with as little explanation as possible.
  - **Customizable**: Users have different preferences and ways of interacting. It should be customizable for the user so that it

is possible to adjust elements.

4 Initial interaction designs

We propose two designs, called *SwipePie* and *The Widget Pallet*, that aim to be fast, low cognitive load and a pleasant user experience. This section explains how they function and which papers inspired us to create these designs. Explanations which show that a design requirement is met are labeled by the name of the requirement in parentheses. The requirements fast, low cognitive load, comfortable and intuitive need to be evaluated before we can state if the requirement is satisfied.

4.1 SwipePie

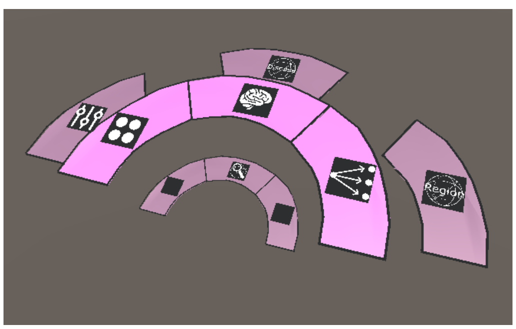


Figure 5: SwipePie design

SwipePie is a hand-centric design **(within reach)** that follows the dominant hand's palm location, Figure 5. This design was inspired by Banovic et al. [2] that created a pie style menu for a touch interaction. The middle row of buttons is located below the fingers and can be interacted with by tapping. If a button is tapped, it will change colour to indicate that a widget has been created. The top and bottom row are not interactable, but are there to show which widgets will be available when changing the button rows **(limited options)**. All buttons are at a slight angle to make sure that the icons are clearly visible. The interaction can be turned on and off by holding a fist with the non-dominant hand for 0.5 seconds. Turning it on switches the state to **widget creation mode** and turning it off changes the state to **widget interacting mode**. Switching these modes is to make sure that no accidental tapping of the buttons happens when trying to grab and move widgets. Changing the rows can be done by making a peace sign with the non-dominant hand and moving the hand up and down like a swipe gesture. We first tried to make the gesture with only the index finger to initiate a swipe, but this proved to be difficult for the Hololens 2 since it would often think it is a fist gesture. The inspiration for the swiping gesture comes from how smartphones are used to scroll, but from a paper by Koutsabasis et al. [16] as well. In the conclusion they state that swiping up and down gestures are more usable than a wheel gesture because that leads to more fatigue.

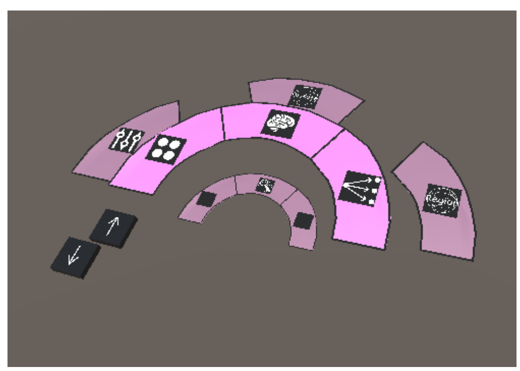


Figure 6: SwipePie with thumb buttons

In addition to the gesture confusion, it was also noticeable that switching between creating the widgets with one hand and swiping rows with the other hand felt like a lot of cognitive work. Therefore, we created an alternative to this design where the thumb is used for tapping buttons that change the rows, Figure 6. This way all the main interacting with this design is with the dominant hand and only the non-dominant hand would be used for turning the interaction on or off.

4.2 The Widget Pallet

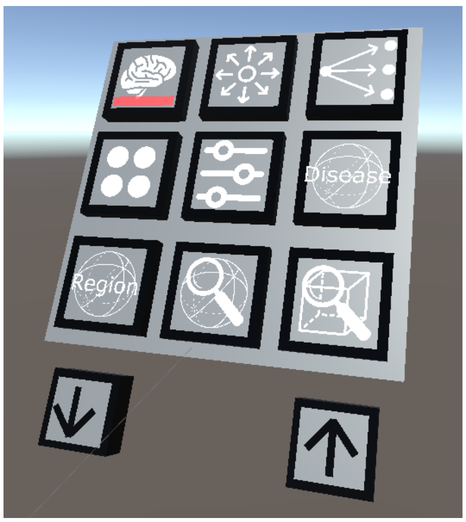


Figure 7: The Widget Pallet design with gaze feedback in the top left button

The widget pallet is a hand-centric design **(within reach)** where the user uses their non-dominant hand to make the pallet appear and their dominant hand *(Dunk brush)* or head gaze *(Dwell gaze)* to choose an option, see Figure 7. Instead of colour options, the pallet consists of the widget icons that the user can choose from. The pallet appears by twisting the wrist to show the palm as if the user is holding a physical paint pallet. This makes it an intuitive gesture because it represents the real world hand positioning. The pallet is placed next to the hand, since placing the pallet on top of the hand makes the Hololens confused which hand is being used for interaction or accidentally open the Hololens palm menu. Selecting an option from the pallet can be done by tapping the button with the index finger or using head gaze by pointing the gaze cursor on an option for one second. When interacting by tap the colour of the button changes and a sound is played to indicate that a tap is performed while the gaze interaction shows a red line that extends from the middle to the width of the button to show the second long dwell time. There are arrow buttons placed below the pallet to change rows when more than nine widgets are available in the system. This keeps the design flexible and not limited to only nine total widgets **(limited options)**.

5 Pilot study

There are four designs that should be tested with participants. The two goals of the pilot study is to get initial feedback on the designs and to determine which two designs are considered most preferable. The two best will be improved based on the feedback and compared using our chosen criteria in the following study.

5.1 Evaluation method

They were asked to sign a consent form, then given an explanation that DatAR is a system with the aim to help neuroscientists explore their domain literature where one of the core ideas is to use a widget system to explore the data. The five widgets used in the evaluation were explained in a video to show the participant how to use inputs, connect widgets and the functionality of each widget. Each participant used all four designs (SwipePie Swipe, SwipePie Thumb, Widget Pallet Touch and Widget Pallet Gaze) for the same tasks. The order of the designs was randomised using a Latin Square (appendix A) to ensure that the learning effect would not heavily influence the opinion on a design. Keeping the

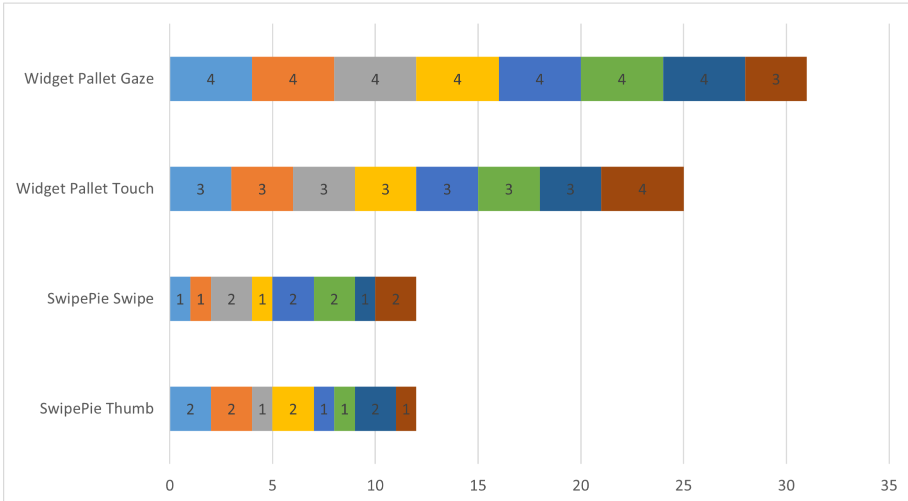


Figure 8: Pilot study design ranking results

order the same for each participant could lead to a more positive opinion for the third and fourth design, since the participants become more used to the system and the tasks near the end of the evaluation. Explanations of how to use the different designs were provided through text and GIFs. When participants understood the explanation, they put on the headset and performed the tasks shown in the virtual space (appendix B.1). The tasks that were created for this evaluation are actions that a user would perform in a normal setting of the system. This consists of creating widgets, input and connecting them together. We chose this over specific selection tasks so that we can examine how it performs in the system under realistic conditions. After completing all three tasks, the participant provided feedback on their experience on the topics of aesthetics, usability and improvements that could be made.

5.2 Results and feedback

Eight master students from the GMT (6), HCI (1) and AI (1) programmes participated in return for a €10 movie voucher for the hour-long evaluation. After performing the tasks with each design, the participants were asked to rank the designs from best to worst. 4 points mean the design is most prefered and 1 point is least prefered. Based on the responses from the participants the designs rankings are shown in Figure 8.

The SwipePie design had some technical and design issues that ended up frustrating participants. The following list is the most notable feedback (F) and our observations (O) during the evaluation.

- F/O Inaccurate gesture performance
  - **Observation:** All participants had issues using the gestures. They performed inaccurately and would not always respond to what the participant was doing. This inaccurate performance led to frustration for the participants.
  - **Analysis:** The Hololens would often be confused. Resting your hand on the table was seen as the fist gesture and enabled the inter-

action by accident. It would also confuse the swipe gesture for the fist gesture, which made the interaction disable while the participant was trying to switch the rows. Additionally, when it came to the swiping, it would only work sometimes and at other moments it would change multiple rows while the gesture was only performed once.

- F/O Accidental button tapping
  - **Observation:** Every participant accidentally tapped the buttons and created widgets that were not intended.
  - **Analysis:** To avoid accidentally creating widgets, the design needed to be switched to widget manipulation mode before interaction with widgets. This was something that participants would forget which led to accidental button tapping.
- F/O Gorilla arm effect
  - **Observation:** The interaction requires participants to hold their arm up to use the interaction. All participants told us during the evaluation that they felt fatigue.
  - **Analysis:** The arm position created a lot of tension in the muscles. Even for short periods of time (30 - 60 seconds) it showed that fatigue was felt. This is not an issue with the Widget Pallet because it uses a more relaxed position.
- O Different hand sizes
  - **Observation:** The design was fit to a small hand, participants with larger hands needed to perform uncomfortable movements to interact with the buttons.
  - **Analysis:** When implementing we made sure to have a right and left handed version, but did not consider different hand sizes in the initial design. This showed that when creating a design based on the fingers it is important to make it work and test for different hand sizes.
- F One hand for all the main interactions
  - **Observation:** Two participants mentioned that they did not like using both hands actively for the interaction.
  - **Analysis:** The comments show that switching between hands for performing the swipe gesture made it feel like a lot of work cognitively. It would be better to keep all active interactions based around the dominant hand. Interactions such as enabling the design could still be used with the non-dominant hand since it doesn't need to be used often and thus not an active interaction.

The two Widget Pallet designs were clear favourites of the participants. The following list is the most notable feedback (F) and our observations (O) during the evaluation.

- F/O Intuitive and easy to use
  - **Observation:** Four participants found the touch and gaze methods to be easy to use for completing the tasks.
  - **Analysis:** The Widget Pallet design is simple compared to the more complex SwipePie design. This design is easy to understand, since the participants only need to think about opening the pallet and choosing a button option.
- F Colour contrast
  - **Observation:** One participant noted that it could use other colours to provide more contrast.

  - **Analysis:** The colour or aesthetics were not considered during the initial design. The light grey was chosen as a default option but this colour made it difficult to distinguish the buttons and icons at first glance.
- F/O Accidental button tapping
  - **Observation:** Five participants told us that they would sometimes accidentally tap the wrong button or thought that the wrong button responded to the tap.
  - **Analysis:** Examining the recordings it became clear that, because the hand is difficult to see when using the headset, the participants misjudged the location of their finger and moved it to a different button than the one intended. Participants suggested making the pallet bigger to avoid this issue.
- F Gaze deemed better than touch
  - **Observation:** Three participants mentioned that, even though the touch variant felt most natural to them, the gaze variant was perceived to be more accurate and thus gave more confidence to use.
  - **Analysis:** Compared to the original paper [23], the one second dwell time was not experienced as a long time by six participants. One participant mentioned that it felt too long and another said it felt too short. The reason we think that most participants did not mind the dwell time, is because this interaction does not require them to immediately perform grabbing actions when creating widgets. Most participants would first get all the elements needed to perform the task and then place the input and widgets together. This is different from Pfeuffer et al.'s task, where the dominant hand needs to be used for painting immediately after choosing a colour
- F/O Gaze selection indication
  - **Observation:** One participant gave feedback that the red bar showing the gaze timer was not clear enough to show that a button was selected. In the recordings it can also be seen that participants in the beginning were confused if a widget was created or not.
  - **Analysis:** When the red bar disappears it can mean that the gaze selection is completed, but it can also mean that the participant looked away and ended the gaze timer. This led to confusion because it gave no other visual or audio feedback.
- F Static pallet
  - **Observation:** One participant told us that to use the pallet you need to sit straight and not move to any side because the pallet does not move with you. This made it difficult to use the pallet in other positions.
  - **Analysis:** The pallet was created to be static to make it easier to tap the buttons. This decision was based on the suggestion from Xu et al. [34] which stated to not rotate the interface with the user's wrist. The comment made us realise that this would restrict the user to only a single position instead of being able to use all the virtual space.

5.3 Widget Pallet version two

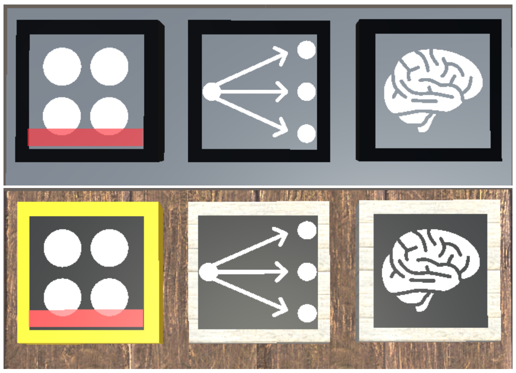


Figure 9: Contrast and gaze feedback comparison between version 1 (top) and version 2 (bottom)

Increasing the size of the pallet was considered due to comments by participants. This change was not made because increasing the size would make it more difficult to have the pallet fit inside the limited FOV of the Hololens. Nevertheless, three aspects of the pallet were improved for the next evaluation. The pallet changed from a grey colour to a light and dark wood texture. This was done to make it more aesthetically pleasing and provide more contrast to make the icons more recognizable. The selection feedback for the gaze variant was also improved. A yellow colour is displayed to indicate that the user is gazing at one of the buttons and a sound is played when the gaze is completed. This provides better feedback to the user that a widget has been created. See Figure 9 to compare the difference in contrast and gaze feedback between the first and second version. The final improvement is that the pallet rotates with the user's wrist movement to ensure that the pallet is always facing the user. This way no matter where the user is positioned the pallet can always be used.

6 Widget Pallet comparison study

The results from the pilot study showed that participants had the highest preference for the widget pallet designs. The designs have been improved and will be compared in this study. The comparison study has two goals. First, to determine which of the two designs performs best based on our chosen criteria. Second, is to use the results to answer the main research and sub research questions.

6.1 Evaluation method

Similar to the previous study, they were asked to sign a consent form, received information about DatAR, the widgets and explanations of how to use the designs. There were three tasks to perform per design. The first task was to get familiar with the interaction and the following two tasks were timed (appendix B.2). The tasks were the same for both designs. Compared to the pilot study there are two changes made to the tasks. One task has been removed and replaced with a short task that contains all basic interactions to test whether the participant understood how to use the design. The last task description has been altered, because the min-max filters can be inaccurate when releasing so a larger range was given for the participant to use. This change was made because in the previous study participants were too focused on getting the exact number given and the point is to perform the steps in the task not test a specific widget. After completing the tasks, the participant answered the NASA-TLX questionnaire, the UEQ questionnaire and an open question that asked if there were any improvements to be made on the design. To examine whether there is a performance difference between participants familiar with the designs and participants

	Task 1 (16 participants)	Task 2 (16 participants)
Touch	00:32,57 +/- 00:14,94	01:48,30 +/- 00:54,14
Gaze	00:32,95 +/- 00:19,29	01:51,48 +/- 00:34,51

Table 1: Average task time divided per design. Time result after the +/- is the standard deviation

	Task 1 (16 participants)		Task 2 (16 participants)	
	Pilot	New	Pilot	New
Touch	00:28,21 +/- 00:11,58	00:36,93 +/- 00:17,34	01:35,22 +/- 00:41,73	02:01,38 +/- 01:04,40
Gaze	00:23,28 +/- 00:05,50	00:42,62 +/- 00:23,53	01:45,87 +/- 00:35,48	01:57,09 +/- 00:34,95

Table 2: Average task time divided per study group and design. Time result after the +/- is the standard deviation

that use it for the first time, the participants were split into two groups. One group containing the pilot study participants and the other group containing the new participants. To balance the learning effect for the timed tasks, the two groups were split in half once more where 4 participants of the group started with the touch design and the remaining half started with the gaze design.

6.2 Results

15 master students and 1 computer science professor participated in the second evaluation study. All 8 participants from the pilot study and 8 new participants took part in the study.

6.2.1 Time performance

The time performance results give an indication of how well the participants are able to complete a task using the designs. While the tasks were created to make them as realistic as possible, human factors can influence the performance, such as: proficiency in grabbing objects, proficiency in connecting widgets and understanding the task. Some participants were more proficient in using the Hololens than others and this needs to be considered when looking at the times.

The touch design has a slightly faster average than the gaze design. To find differences in the designs we chose the Mann Whitney U test, because the Shapiro Wilk test showed that the data is not normally distributed and the Levene test shows that the data is of equal variance. The Mann Whitney U test shows no significant difference (α = 0.05) was found between touch vs gaze in task 1 (p = 0.78) and touch vs gaze in task 2 (p = 0.42). The average time results can be seen in table 1.

The pilot group was faster than the new participants group in all tasks. The Mann Whitney U test shows there is a significant difference (α = 0.05) in the pilot group being faster than the new group in gaze task 1 (p = 0.01). However, no sig-

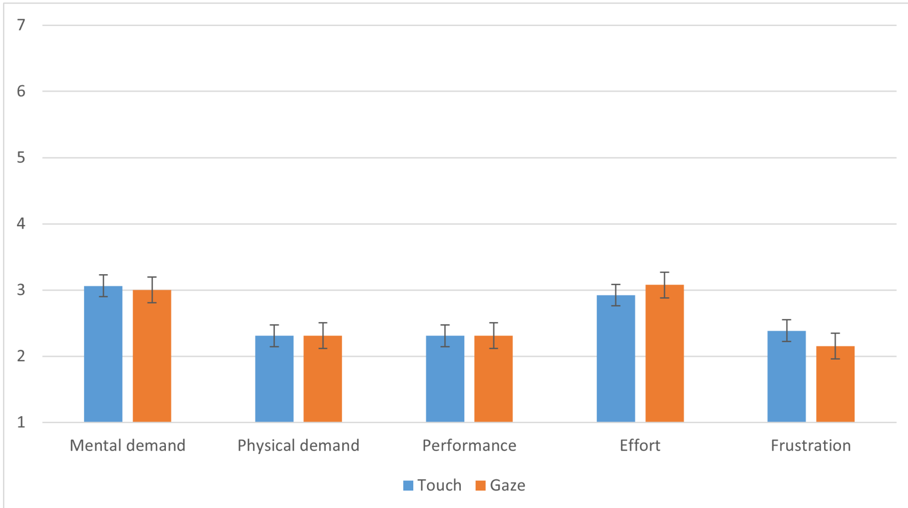


Figure 10: NASA-TLX average score result per scale

nificant difference was found between the groups in touch task 1 (p = 0.38), touch task 2 (p = 0.32) and gaze task 2 (p = 0.50). The average times for the task per group can be seen in table 2.

6.2.2 NASA-TLX

The touch and gaze design both have a low scoring in all scales and it shows only small differences between mental demand, effort and frustration with no difference between physical demand and performance. A Mann Whitney U test was used, because the data contains ordinal values, and found no significant difference ($\alpha = 0.05$) in the scales in mental demand (p = 0.92), physical demand (p = 0.83), performance (p = 0.42), effort (p = 0.78) and frustration (p = 0.95). Figure 10 shows the average scores per scale.

6.2.3 UEQ

There is no baseline available to compare the two designs against. Only comparing the designs against each other will not show if the design on its own is considered usable. For this reason, we first examine the scores against the UEQ benchmark which contains the results of 468 product evaluations. It scores on the scales of bad, below average, above average, good and excellent. Following the benchmark is the comparison of the two designs to see if there are any significant differences found between the scales.

6.2.3.1 Touch benchmark The benchmark in Figure 11 shows that the touch version of the widget pallet scores above average (only slightly below good) in the efficiency and novelty scales and good in the attractiveness, perspicuity, dependability and stimulation scales.

6.2.3.2 Gaze benchmark The benchmark in Figure 12 shows that the gaze version of the widget pallet scores above average in efficiency, good in attractiveness, perspicuity, dependability and novelty and scores excellent in stimulation.

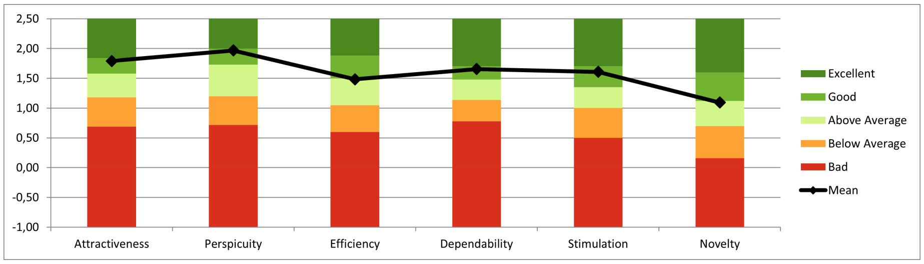


Figure 11: Touch design UEQ benchmark result

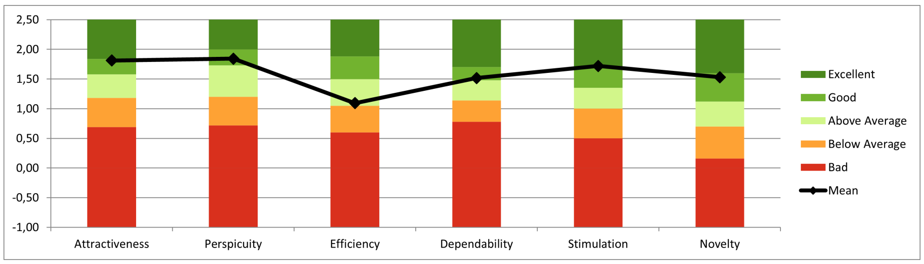


Figure 12: Gaze design UEQ benchmark result

6.2.3.3 Comparing designs The results in Figure 13 show slight differences between the attractiveness, perspicuity, dependability and stimulation scales. There are also some bigger gaps where efficiency is scored higher for touch and gaze being scored higher for the novelty scale. Performing a two sample T-Test with unequal variance ($\alpha = 0.05$) shows that there is no significant difference found in all scales of attractiveness (p = 0.94), perspicuity (p = 0.73), efficiency (p = 0.31), dependability (p = 0.69), stimulation (p = 0.74), novelty (p = 0.27).

6.2.4 Design requirements

Based on the results shown, table 3 shows a summary of the design requirements and if the design has met the requirement.

7 Discussion

7.1 Can we create a fast selection interaction?

No significant difference in speed of task completion was found between the touch and gaze design, but there was a difference between the pilot group and new group for gaze task 1 (section 6.2.1)

7.1.1 Touch vs Gaze

We expected the gaze design to work slower than touch because of the one second dwell time. This does not seem to have an impact when performing tasks. A reason for this could be the use of the dominant hand. In the touch design the dominant hand is close to the pallet to tap the button and then moves to the widgets to grab and connect them. This is not the case with the gaze design because the hand is free and does need to move back and forth to the pallet to use it. This difference could

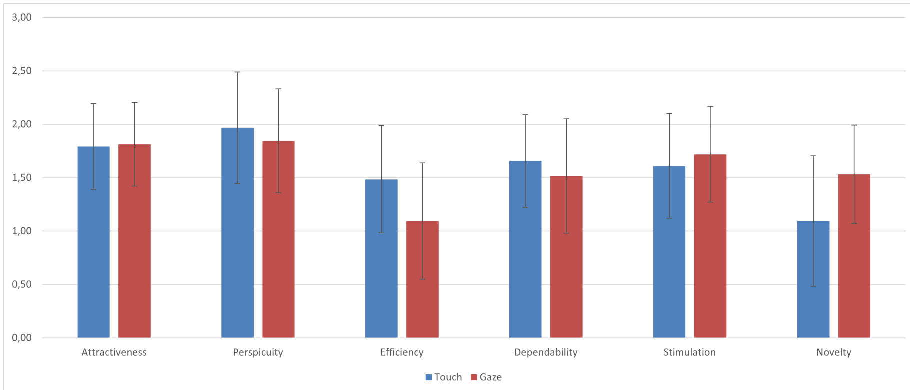


Figure 13: UEQ product comparison graph

Requirement	Widget Pallet Touch	Widget Pallet Gaze
High priority		
- Fast	✓	✓
- Low cognitive load	✓	✓
- Within reach	✓	✓
- Comfortable	✓	✓
- Limited options	✓	✓
Medium priority		
- Intuitive	✓	✓
Low priority		
- Customizable		

Table 3: Results of the design requirements

be a factor that the times are more similar to each other, besides the aspect that some participants were more proficient in using the hololens than others.

7.1.2 Pilot group vs New group

The difference between the pilot group (8 participants) and the new group (8 participants) for the first gaze task indicates that with practice users may become more familiar with the interaction and thus work faster. The pilot group was already familiar with how the designs worked and were able to use it proficiently whereas new participants were trying to understand how it worked. This statistical difference is not seen in the second task because the task contained many steps and every participant had moments where they forgot the next step and had to read the task again.

7.2 Can we create an interaction that requires low cognitive load?

Participants did not experience any noticeable difference between the touch and gaze designs and both scored low in cognitive load. The most notable results are the slight differences in average between mental demand, frustration and effort (section 6.2.2). Mental demand and frustration is a combined section because the higher mental demand influenced frustration.

7.2.1 Mental demand and Frustration

The touch design has a slightly higher score in mental demand and frustration because participants would sometimes press the wrong button. This is because their hand is occluded by the pallet which makes it difficult to estimate the location of the hand. Suggestions were made, similar to those in the pilot study, to make the pallet larger or to space the buttons out more. The feedback of participant 10 gave a new insight saying *"Maybe if the button gets highlighted where the finger is hovering over it would be easier to compare the position of your finger to the widget menu and you can adjust easier and therefore click easier."* Highlighting the button where the finger is hovering over could be a solution to the issue of accidental button tapping.

7.2.2 Effort

The gaze design shows a slightly higher score in effort because using the head movement to create a widget is perceived to be only slightly more effort than tapping the button. Since the gaze cursor comes from the forehead, not between the eyes, it forces the participant to either look further down than it is comfortable or to move their hand up to select an item from the bottom row of buttons. Participant 9 suggested *"It would be easier if the panel was not tilted and instead straight up, so you can look directly at it."*. Changing the angle from tilted to straight up could lead to less head movement being required to select a widget.

7.3 Can we create an interaction with a pleasant user experience?

No statistical difference was found between the touch and gaze design for the user experience. However, there are some slight differences in average scores between perspicuity, efficiency, dependability, stimulation and novelty (section 6.2.3). Dependability and stimulation are combined into one section because they have similar reasons for the differences in results.

7.3.1 Perspicuity

The slightly lower score for gaze in perspicuity is because it took a little more effort to get accustomed to. People in general are more used to tapping a button than needing to gaze at it. This made it less intuitive for the participants and even a small group of participants tried to tap the buttons when it was explained to use the gaze method.

7.3.2 Novelty

The slightly higher score in novelty for the gaze design is because it is an interaction method that participants were not used to. This made it feel more innovative and creative to them.

7.3.3 Efficiency

The time performance result shows no clear difference between the designs but participants did score the gaze design lower in efficiency. This shows us that participants do experience the dwell time to be slower than touch. Participants have mentioned during the evaluation that they like the instant response of the touch pallet compared to waiting for the dwell time. Participant 5 mentioned in the feedback *"I* *would prefer the physical touch of the pallet buttons, as they feel more natural and don't have a forced-delay in my head movement."*. Participants 16 and 8 did note that they had trouble understanding the force needed to press the button saying: *"The feedback for when a button was pressed was not always as clear to me."* and *"I sometimes had to push harder to confirm a 'click' than I anticipated"*. To avoid the button being pressed by just moving the finger over it, there is some force required to push in the button.

7.3.4 Dependability and stimulation

The touch design scores higher in dependability because the instant response of pressing the button makes participants feel more in control. However, a small group of participants mentioned during the evaluation that gaze felt more accurate than touch because of the issue with pressing the wrong button. This explains why the stimulation score for gaze is slightly higher because the accuracy makes participants more motivated to use it. Sometimes there was an issue with the gaze during evaluation where a slight head movement would move the cursor away from the button and stop the dwell time. Participant 16 wrote in the feedback *"I had a bit of trouble keeping my gaze stable on the menu on my arm"*.

8 Limitations

This research contains two limitations in the method and one limitation that is a learning point when performing research in the future. The comparison study contained a small sample size and almost all of the results were non conclusive. However, the averages in the metrics show differences, for example, the average of efficiency being much lower for gaze. We think that evaluating the designs with a larger sample size could show possible statistical differences for the NASA-TLX and UEQ results. The differences in time performance results are small. Therefore, we do not think a statistical difference will be found there. Furthermore, we made the decision to use realistic tasks to best test the designs with DatAR. The goal states that the solution should work in DatAR but also be general enough to work for similar systems. We did not conduct an experiment to specifically test the selection performance of the designs. We believe that the designs are usable in other systems, but this should be tested to get a more general idea of how it performs.

The learning point we discovered, while writing the pilot and comparison study, is that we should have taken notes during evaluation. In the text there are places where "some participants" is mentioned, or a variation of it, because these comments were said during evaluation but were not written down. Keeping better track of how many participants made a particular comment will make the writing stronger in the future.

9 Conclusion and Future work

To conclude this thesis and to answer the main research question: ***Can we create a selection interaction that works fast, with low cognitive load and a pleasant user experience?*** The results show that both designs meet the criteria of working fast, with a low cognitive load and a pleasant user experience. Because there are no significant differences found between the designs, we can say that the best choice of design will come down to the preference of the user. Table 4 shows the pros and cons for each design to help with this choice.

There are three possible avenues for future work. We were unable to use controllers, but it could be interesting to com-

	Touch	Gaze
Pros	• Creating the widget works instantly • Perceived as more natural • Easy to use	• Perceived as more accurate • Easy to use
Cons	• Difficult to see hand which leads to pressing the wrong button • Difficulty understanding force needed to press the button	• Long dwell time • Slight head movements can reset the dwell timer • Using head movement is perceived to be more effort

Table 4: Pros and cons list of the touch and gaze designs

pare the designs with controllers against the hand tracking of the hololens to examine if there are any differences there. It is also possible to look further into the mentioned limitations to evaluate with a larger sample size to see if there are any statistical differences to determine if one design performs better than the other and to study the performance of the designs using a general selection task to evaluate the selection time and number of errors.

10 Reflection

Reflecting on the thesis we are content with the choices made regarding the method and environment. We mentioned that not having a general selection performance task is a limitation and it is something that could be further researched. However, it does not make the thesis weaker for not having it. The designs perform well and could be used in other systems, but to better generalise this should be further tested. Making the tasks as realistic as possible was a conscious decision because it was the best option for the environment and it showed that the tasks were not biased to make the designs look better than they are.

DatAR was a great environment to use for the project because it is still in development which meant there were enough possibilities for creating a thesis project. Another advantage was that the system could be used and adjusted instead of needing to spend time building a system from scratch. The biggest issue was the data not working, but since the thesis was design focused this had little impact.

The biggest hurdle during the thesis was trying to find participants for the studies, especially the pilot study. It was a mistake to try and hold evaluations during an exam period, which was something we did not check before sending out the message. It also became clear that holding evaluations on campus was more effort than other thesis, where the participant only needs to open a link and answer a survey. This meant only a small group of people were interested in participating. By sending messages directly to other students it eventually got to the number of people we wanted to. It was easier for the comparison study because there were more responses and all the participants from the previous study agreed to come back for the next one.

References

[1] A Baddeley and Graham Hitch. *The psychology of learning and motivation: Advances in research and theory, chapter working memory (pp. 47–89)*. 1974.

[2] Nikola Banovic et al. "Design of unimanual multi-finger pie menu interaction". In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. 2011, pp. 120–129.

[3] John Brooke. "Sus: a "quick and dirty'usability". In: *Usability evaluation in industry* 189.3 (1996).

[4] Paul Chandler and John Sweller. "Cognitive load theory and the format of instruction". In: *Cognition and instruction* 8.4 (1991), pp. 293–332.

[5] Xiang'Anthony' Chen et al. "Extending a mobile device's interaction space through body-centric interaction". In: *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. 2012, pp. 151–160.

[6] Fred D Davis. "Perceived usefulness, perceived ease of use, and user acceptance of information technology". In: *MIS quarterly* (1989), pp. 319–340.

[7] Matthew M Davis et al. "Depth-based 3D gesture multi-level radial menu for virtual object manipulation". In: *2016 IEEE Virtual Reality (VR)*. IEEE. 2016, pp. 169–170.

[8] Tycho K Fredericks et al. "An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads". In: *International Journal of Industrial Ergonomics* 35.12 (2005), pp. 1097–1107.

[9] Eric Granholm et al. "Pupillary responses index cognitive resource limitations". In: *Psychophysiology* 33.4 (1996), pp. 457–461.

[10] Danny Halarewich, About The Author Danny is a SaaS / e-commerce entrepreneur, and About The Author. *Reducing cognitive overload for a better user experience*. 2016. URL: https://www.smashingmagazine.com/2016/09/reducing-cognitive-overload-for-a-better-user-experience/.

[11] Sandra G Hart. "NASA-task load index (NASA-TLX); 20 years later". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 9. Sage publications Sage CA: Los Angeles, CA. 2006, pp. 904–908.

[12] Sandra G Hart and Lowell E Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research". In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 139–183.

[13] Setia Hermawati and Glyn Lawson. "Establishing usability heuristics for heuristics evaluation in a specific domain: Is there a consensus?" In: *Applied ergonomics* 56 (2016), pp. 34–51.

[14] William E Hick. "On the rate of gain of information". In: *Quarterly Journal of experimental psychology* 4.1 (1952), pp. 11–26.

[15] Aliia Kodzhoshalieva. "Reducing Cognitive Load for a Data-Intensive Web Application User". In: (2020).

[16] Panayiotis Koutsabasis and Chris K Domouzis. "Mid-air browsing and selection in image collections". In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 2016, pp. 21–27.

[17] Bettina Laugwitz, Theo Held, and Martin Schrepp. "Construction and evaluation of a user experience questionnaire". In: *Symposium of the Austrian HCI and usability engineering group*. Springer. 2008, pp. 63–76.

[18] James R Lewis. "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use". In: *International Journal of Human-Computer Interaction* 7.1 (1995), pp. 57–78.

[19] Xueshi Lu et al. "Exploration of hands-free text entry techniques for virtual reality". In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2020, pp. 344–349.

[20] Francisco Lopez Luro and Veronica Sundstedt. "A comparative study of eye tracking and hand controller for aiming tasks in virtual reality". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 2019, pp. 1–9.

[21] Mark R Mine. "Virtual environment interaction techniques". In: *UNC Chapel Hill CS Dept* (1995).

[22] Fred GWC Paas and Jeroen JG Van Merriënboer. "The efficiency of instructional conditions: An approach to combine mental effort and performance measures". In: *Human factors* 35.4 (1993), pp. 737–743.

[23] Ken Pfeuffer et al. "Empirical evaluation of gaze-enhanced menus in virtual reality". In: *26th ACM symposium on virtual reality software and technology*. 2020, pp. 1–11.

[24] Jef Raskin. *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional, 2000.

[25] Rongkai Shi et al. "Exploring Head-based Mode-Switching in Virtual Reality". In: *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2021, pp. 118–127.

[26] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. "Experimental analysis of barehand mid-air mode-switching techniques in virtual reality". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–14.

[27] John Sweller, Jeroen JG Van Merrienboer, and Fred GWC Paas. "Cognitive architecture and instructional design". In: *Educational psychology review* 10.3 (1998), pp. 251–296.

[28] Ghazaleh Tanhaei, Lynda Hardman, and Wolfgang Huerst. "DatAR: Your brain, your data, on your desk-a research proposal". In: *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE. 2019, pp. 138–1385.

[29] Vildan Tanriverdi and Robert JK Jacob. "Interacting with eye movements in virtual environments". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 2000, pp. 265–272.

[30] Vildan Tanriverdi and Robert JK Jacob. "Interacting with eye movements in virtual environments". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 2000, pp. 265–272.

[31] Janet Patton Tracy and Michael J Albers. "Measuring cognitive load to test the usability of web sites". In: *Annual Conference-society for technical communication*. Vol. 53. 2006, p. 256.

[32] Ivar Troost et al. “DatAR: An immersive literature exploration environment for neuroscientists”. In: *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE. 2020, pp. 55–56.

[33] Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. “Towards gaze-mediated interaction: Collecting solutions of the “Midas touch problem””. In: *Human-Computer Interaction INTERACT'97*. Springer. 1997, pp. 509–516.

[34] Xuhai Xu et al. “Hand range interface: Information always at hand with a body-centric mid-air input surface”. In: *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 2018, pp. 1–12.

[35] LI Yang et al. “Gesture interaction in virtual reality”. In: *Virtual Reality & Intelligent Hardware* 1.1 (2019), pp. 84–112.

Appendices

A Latin square pilot study

- SwipePie Swipe (A)
- SwipePie Thumb (B)
- Widget Pallet Touch (C)
- Widget Pallet Gaze (D)

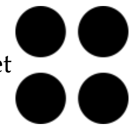
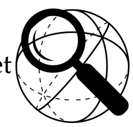
A	B	C	D
C	D	A	B
B	A	D	C
D	C	B	A

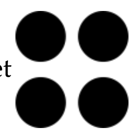
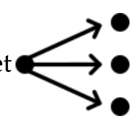
Table 5: Random order for pilot study

B Evaluation tasks

B.1 Pilot study tasks

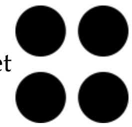
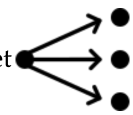
- Task 1: Inspect a resource
  - Create a disease class input 

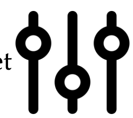
  - Create a concept of class widget 
  - Add disease input to widget
  - Create Resource inspector widget 
  - Add Anxiety as input
- Task 2: Brain region co-occurrence
  - Create a disease class input 

  - Create a concept of class widget 
  - Add disease input to widget
  - Create co-occurrence widget 
  - Create a region class input 

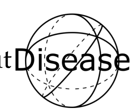
  - Add region and depression to co-occurrence widget
  - Create brain widget 
  - Connect widgets together
- Task 3: Brain region co-occurrence with filter

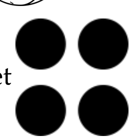
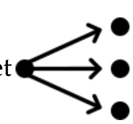
- Create a disease class input 

- Create a concept of class widget 
- Add disease input to widget
- Create co-occurrence widget 
- Create a region class input 

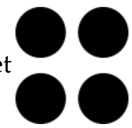
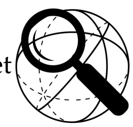
- Add region and depression to co-occurrence widget
- Create brain widget 
- Connect widgets together
- Create min max widget 
- Connect min max to co-occurrence and brain
- Filter the results for "number of times cerebral palsy co-occurs with a given region" to about a 100.

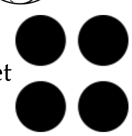
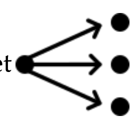
B.2 Comparison study tasks

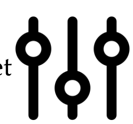
- Task 1: Interaction check task
  - Create a disease class input 

  - Create a concept of class widget 
  - Add disease input to widget
  - Create co-occurrence widget 

  - Create brain widget 
  - Connect widgets together
- Task 2: Inspect a resource
  - Create a disease class input 

  - Create a concept of class widget 
  - Add disease input to widget
  - Create Resource inspector widget 
  - Add Anxiety as input
- Task 3: Brain region co-occurrence with filter
  - Create a disease class input 

  - Create a concept of class widget 
  - Add disease input to widget
  - Create co-occurrence widget 
  - Create a region class input 

  - Add region and depression to co-occurrence widget
  - Create brain widget 
  - Connect widgets together
  - Create min max widget 

- Connect min max to co-occurrence and brain
- Filter the results for “number of times cerebral palsy co-occurs with a given region” between a range of 100 - 150