Hands-free Selection in Scroll Lists for AR Devices

Heiko Drewes
LMU Munich
Munich, Germany
heiko.drewes@ifi.lmu.de

Yara Fanger
LMU Munich
Munich, Germany
yarafanger@gmail.com

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

ABSTRACT

While desktops and smartphones have established user interface standards, they are still lacking for virtual and augmented reality devices. Hands-free interaction for these devices is desirable. This paper explores utilizing eye and head tracking for interaction beyond buttons, in particular, selection in scroll lists. We conducted a user study with three different interaction methods based on eye and head movements, gaze-based dwell-time, gaze-head offset, and gaze-based head gestures and compared them with the state-of-the-art hand-based interaction. The study evaluation of quantitative and qualitative measurement provides insights into the trade-off between physical and mental demands for augmented reality interfaces.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design**.

KEYWORDS

hands-free interaction, eye tracking, head tracking, head gestures, selection in list boxes

ACM Reference Format:
Heiko Drewes, Yara Fanger, and Sven Mayer. 2024. Hands-free Selection in Scroll Lists for AR Devices. In *Proceedings of Mensch und Computer 2024 (MuC '24), September 01–04, 2024, Karlsruhe, Germany.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3670653.3670671

1 INTRODUCTION

The higher-level goal of our research is to understand how user interfaces could look if we want to transfer large portions of the smartphone functionalities into AR glasses to get something we call smart glasses. Typical smartphone tasks are making telephone calls, reading and writing emails, surfing the internet, and using a navigation application. Most of these tasks need selections in scroll lists, for example, choosing an entry in a phone list. Desirably, such a user interface should keep the hands free for other tasks. Keeping the hands free would be a good argument to switch from a smartphone to smart glasses, but there is also a demand for industrial use cases where the hands need protective gloves or have to handle tools.

MuC '24, September 01–04, 2024, Karlsruhe, Germany

ACM ISBN 979-8-4007-0998-2/24/09
https://doi.org/10.1145/3670653.3670671

Smartphone user interfaces use standard interaction objects like buttons, menus, list boxes, etc., which are more or less the same as those used for interaction with desktop devices. However, desktop systems use mouse and keyboard input, while smartphones get the input from a touch-sensitive display. Keeping the standard interaction objects for AR glasses as familiar to the users means finding interaction methods for these interaction objects. AR devices typically can track head movements, and many AR devices come along with a built-in eye tracker. So, the idea of using head and eye movements to interact with AR devices lies near. Additionally, there is considerable research on selection and pressing buttons with gaze, head movements, or both. However, there is not much research on more complex interaction tasks.

In this paper, we investigate the selection in scroll lists. We designed three different methods to interact with scroll lists using eye and head movements and conducted a user study ($N = 25$) where we compared these methods against each other and additionally compared them to the *Hands* interaction, which is the default interaction method provided by the device manufacturer.

Our observations and insights are that lifting the arms for mid-air gestures is physically demanding, while interaction with the eyes demands cognitive effort. Users prefer head movements against arm movements as this is less physically demanding and wish only very few gaze interactions as this creates a cognitive load. Cognitive load influences human motor performance [30], but humans are well-trained to perform controlled arm, hand, and finger movements in the presence of cognitive load. However, most individuals are not well-trained in eye movements in the presence of cognitive load, especially in eye movements for interaction. The big question for the future of gaze interaction is whether training reduces cognitive load and will make controlled eye movements less dependent on other cognitive loads and stress.

2 RELATED WORK

The first eye-tracking devices for interaction date back to the early 1980ies. These systems provided eye-typing applications for disabled people. In 1981, Bolt gave a vision of using gaze for interaction [2]. Jacob did the first systematic research on how to use eye trackers for interacting with graphical user interfaces in 1990 [12].

Despite four decades of research, there was no other eye-tracking application in the wild other than eye typing. However, there is new hope that eye tracking will become an interaction technology for the masses with the introduction of AR and VR glasses. In contrast to public gaze-aware displays, a one-time calibration is no obstacle for a personal device. Building eye trackers into glasses also alleviates problems with outdoor usage caused by changing environmental light conditions. Eye tracker devices have become better and cheaper in the last few years and many hardware manufacturers equipped their AR and VR glasses with eye trackers, such as the HoloLens2[1] and the HTC Vive Pro Eye[2]. However, the standard interaction with these devices works with controllers or hand gestures.

The Apple Vision Pro[3] is controllable with eyes, hands, and voice. The gaze addresses the interaction object, and a finger pinch gesture makes the selection [26]. The hands can be down at the side or in the lap and do not have to be in mid-air to avoid the gorilla arm. We are curious whether Apple's product will make eye tracking a standard interaction technique similar to the touch gestures introduced with the iPhone. However, Apple's Vision Pro interface is not hands-free and does not use the gaze for advanced interaction, such as selecting from scroll lists.

There has been research on eye tracking in VR since at least the beginning of this millennium [7]. Since then, many publications on gaze interaction with AR and VR glasses appeared, e.g., [9, 13, 18, 25, 27]. Some research propagates positive expectations with statements "that eye tracking will soon become an integral part of many, perhaps most, HMD systems" [1], while other research on heads-up computing [29] mention gaze interaction only in the related work. In our study with scroll lists, gaze for interaction is not the only solution. Besides standard interaction with scroll lists via pointing devices or directly with the hand, it is also possible to utilize novel devices such as a wristband [8]. The question of whether gaze interaction will be established as a standard interaction method and which second input modality will be used in combination - hand and finger gestures, head movements, or controllers like finger rings or wristbands - is still open. If gaze interaction turns out to be problematic, "fallback modalities could be leveraged to ensure stable interaction" [32] as Sidemark et al. proposed.

The basic publication for the eye-dwell method researched in our study is from Jacob [12]. Majaranta et al. [20] researched feedback for dwell time-based eye typing. Isomoto et al. [11] focus on dwell selection in AR and VR. Other research deals with dynamic and adjustable dwell times [19, 22]. There is research on scrolling in gaze-based interfaces by Kumar and Winograd [16], and for auto-scrolling when reading text by Sharmin et al. [31]. Sharmin et al. [31] also point to three US patents on the topic in their references. An important paper on the head-gesture method is "Eye-Based Head Gestures" by Mardanbegi et al. [21] from 2012, who used the vestibulo-ocular reflex for separating natural head movements from intended head movements for interaction. The idea of using the vestibulo-ocular reflex was presented already in 2003 by Nonaka [23]. Also, Špakov and Majaranta [34] and Nukarinen et al. [24] presented interaction methods based on gaze and head movements but without mentioning the vestibulo-ocular reflex. The head-gaze offset method for scrolling is our idea, but the selection method with head-gaze offset was also inspired by Sidenmark et al. [33]. As smart glasses should provide a pedestrian navigation system and our future research also aims for interaction with maps, it is worth mentioning the research on interacting with maps on optical head-mounted displays of Rudi et al. [28] and Liao et al. [17].

[1] https://www.microsoft.com/en-us/hololens/hardware
[2] https://www.vive.com/us/support/vive-pro-eye/
[3] https://www.apple.com/apple-vision-pro/

3 INTERACTION DESIGN FOR SCROLL LISTS

AR and VR devices allow for free movement of the user, and the first decision is the placement of the list box with which to interact. As the interface should work anywhere, it should not be world-stable. As we need the head orientation relative to the interaction object, a head-stable display would not work. For adequate interaction, we need a body-stable projection of the interface. In our implementation, we realized this by taking the head position and ignoring the head orientation.

According to the principles of VR interaction by Bowman et al. [3], our user study task consists of manipulation of the list to get the desired list item in the view, in Figure 1 to 4 on the left side, and a subsequent selection, depicted in the Figures on the right side. Several interaction methods can be used for the two sub-tasks, and our design decision was to use interaction methods of the same type of interaction for both sub-tasks to get a consistent interface. We used interaction methods from the literature, dwell-time [12], head gestures [21], and as a novel method, the offset between head and gaze vector where the selection sub-task is similar to Gaze-Activated Head-Crossing [33].

Implementing the interaction methods demanded decisions on parameter values for sizes, times, speeds, and angles. The choice of these parameter values influences the results, which should be considered when comparing the methods. We conducted a pilot study to estimate these values. The study goal, however, was not to find optimal parameter values but to find the optimal method.

3.1 Hands

Hand tracking is one of the standard interaction methods provided by the device manufacturers. The system shows the tracked hand as a virtual object; see Figure 1. The reason for implementing the *Hands* interaction in our study was to have a baseline for comparison with the eye-based interaction techniques. For this reason, the scroll list looked the same in all the tasks, except the eye-dwell method, which had two additional buttons above and below the list. One constraint

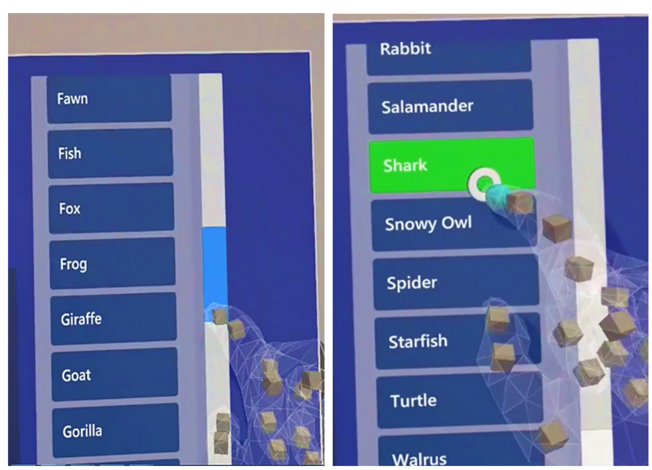

Figure 1: The *Hands* interaction method. The left side shows the interaction with the scroll bar with a hand, and the right side shows the selection of a list item.

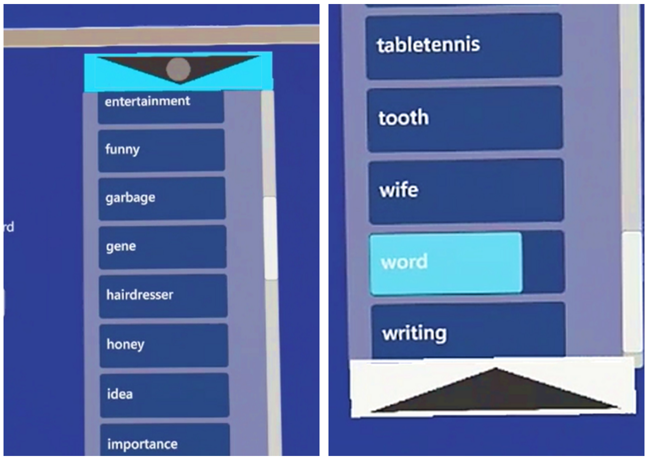

Figure 2: The *Eye-Dwell* interaction method. The light blue color indicates gaze feedback. The left side shows the dwell time button for scrolling (highlighted in light blue). The right side shows an item selection with a progress bar indicating the dwell time (light blue bar on the covered item "word").

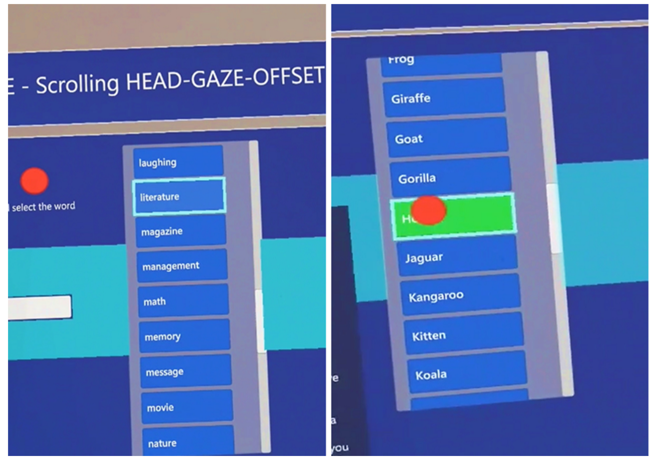

Figure 3: The *Head-Gaze Offset* interaction method. The left side shows the scrolling with the head direction above the middle of the list and the gaze below the head direction. The right side shows the selection of a list item by placing the red dot and the gaze onto the list item. The red dot shows the head direction with a 9° offset to the left.

of interaction with the hand, similar to the interaction with real objects, is that the interaction objects must be within arm's reach. Consequently, we displayed the list box at a virtual distance of 45 cm while using 1 m for the gaze interaction techniques.

Figure 1 shows the scroll list implementation, which has the familiar design of a vertical list with a scroll bar slider on the right side. Putting the finger into the slider allows for scrolling the list as opposed to the finger movement. Placing the finger into a list entry selects the entry when pulling out the finger.

3.2 Eye-Dwell

The standard gaze-only interaction method, typically an accessibility option, is dwell time. The user has to look for a certain time, the dwell time, at the interaction element. Other options for gaze-only scrolling, such as auto-scrolling as presented by Sharmin et al. [31], would also be worth studying. However, we wanted to provide a simple and easy method for those participants who might struggle with the other methods.

For the *Eye-Dwell* interface, we placed a button at the top of the list for scrolling down and another button at the bottom of the list for scrolling up. While looking at this button, the list scrolls down or up. To select a list entry, the user has to look longer at the list item. The list item provides feedback with a growing bar indicating the time already elapsed, see right side of Figure 2.

The height of both the dwell button and the list entry was exactly 2°, which is sufficient to avoid problems with the eye tracker accuracy. The optimal dwell time depends on the user's experience and can be as low as some hundred milliseconds. However, as we did not expect experienced users for our study, we set the dwell time to 2 seconds. Due to the problems with dynamic scrolling discussed later in Section 3.5 we used a static scroll speed of 6.4°/s.

3.3 Head-Gaze Offset

For scrolling up with the *Head-Gaze-Offset* method, the head direction has to be above the middle of the list, at least 7°, while the gaze has to be below the head direction. The scrolling speed depends on the angle between the head and gaze direction. The bigger the angle, the quicker the scrolling, which ranges from 4.8°/s to 12.8°/s. In contrast to the dwell-time method, the eyes are on the list content and can recognize when the item to select comes into the field of view. Once the gaze is on the item, the eyes can follow the moving item, which decreases the angle between the gaze and head and reduces the scroll speed. Eventually, the list stops to scroll. When the scrolling stops, the gaze is already on the item to make the selection.

The red dot represents the head direction with a 9° offset to the left (see Figure 3). For the selection, the red dot has to be brought onto the selected item while the gaze also stays on the item. This means that the head has to turn 9° to the right. We expect that it is not necessary to display the red dot after some practice with this method.

3.4 Head Gestures

With the head-gesture method, the list scrolls down when the head direction is at least 7° above the middle of the list and scrolls up when it is below. The eyes are not involved in scrolling.

While keeping the eyes on the list item, a head gesture using the vestibulo-occular reflex, triggers the selection, see Figure 4. We chose a roll movement for the head gesture with a minimal roll angle of 9°. Both gesture type and angle were the results of our pilot study. The head gesture detection works the same way as for the gaze gestures introduced by Drewes and Schmidt [6], however, based on angles.

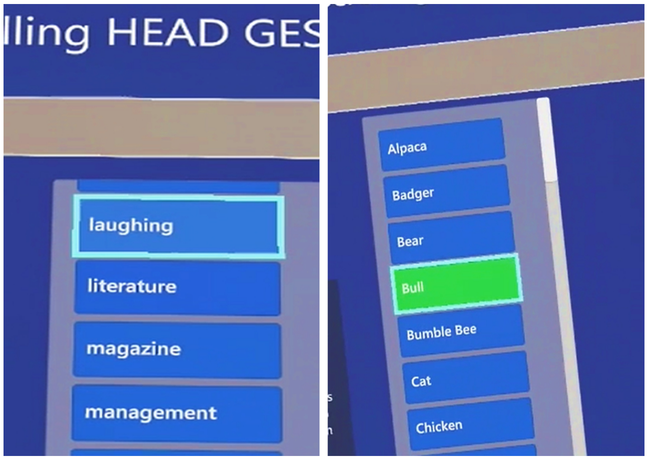

Figure 4: The *Head Gesture* interaction method. The left side shows the scrolling with just the head direction above the middle of the list. The right side shows the selection of a list item by looking at the list item and performing a head gesture (roll).

3.5 Interface Design Decisions

Implementing the interaction methods means making decisions on many details, such as the width and height of the list. We did a pilot study with three individuals to estimate reasonable values for some of the parameters, such as scroll speed, the distance of the interface, and preferred head gesture. We chose other parameters, such as the number of list items, arbitrarily but based on plausible assumptions.

One of these details, which is worth discussing in depth, is the scroll speed. For the *Hands* method, the scroll speed is the speed of the hand and depends on the list length and how quickly the user moves the hand. For the other methods, the scroll speed is a value that is coded in the source code. Eventually, it should be adjustable in the "preferences." The choice of these values, scroll speed, dwell time, and number of list items, influence the task completion times and make comparisons between the four methods questionable. Nevertheless, we will make comparisons, but interpret them carefully.

We intended to offer a dynamic scroll behavior as this eases selection in long lists, and we expect such behavior for a real product. There are two possibilities for realizing the dynamic scroll for the *Head-Gaze Offset* and the *Head Gesture*. The first option is to increase the scroll speed over time. The other option is to use the angle between the head and gaze direction (*Head-Gaze Offset*) or between the head direction and the horizon (*Head Gesture*) as a control parameter for the scroll speed.

For the *Eye-Dwell* method, the only option is to increase the scroll speed over time. However, we encountered a problem. With the *Head-Gaze Offset* and the *Head Gesture*, the head direction may be above or below the list, but the gaze stays on the list to recognize whether the desired item comes into view. In contrast, the eye-dwell method requires that the gaze is on the button above or below the list, and for this reason, the eye can not see whether the desired item appeared already. In consequence, the user has to look at the list from time to time, and this would reset the dynamic scroll behavior. We decided not to implement dynamic scroll for the eye-dwell method. Additionally, we decided to use a slow scrolling speed of $6.4°/s$ and a long dwell time of two seconds.

4 USER STUDY

We used a HoloLens2 for the study. The study design followed the common standards with a training phase, a questionnaire for demographic data, tasks for each interaction method with randomized order according to Latin square, a questionnaire after every task, and a final questionnaire with questions on how the four interaction methods compare to each other.

4.1 Procedure

After we informed the participants about the study, they signed a consent agreement and filled out the demographic questionnaire. Next, participants familiarized themself with the device and went through the eye tracking calibration, and then they entered a training phase for all four interaction methods. When the participants ensured that they understood the interaction methods, the training phase ended, and the main part of the study started. Here, we ask participants to perform five selections for each condition.

For each selection, we presented a scroll list with 50 alphabetically sorted entries from which eight were visible at a time. We used a list of items different from the one in the training to avoid learning effects. The first two list items to select were randomly chosen. The next three list items were from the list positions top (within the first 10 entries), middle (10 entries around the 25th entry), or bottom (within the last 10 entries) in random order. The start position of the list for the first task was the first item at the top. The start position for the subsequent tasks was the position from the end of the previous task.

After each condition, we asked users to fill in the raw NASA Task Load Index (NASA-TLX) [10] and the System Usability Scale (SUS) [4].

After the experiment, we asked participants to rate their overall experience with the experiment to ensure result validity. For this, we asked the first three items from the Simulator Sickness Questionnaire (SSQ) [14]: *General Discomfort*, *Fatigue*, and *Headache* on a 4-point scale.

4.2 Participants

We conducted a within-group user study with 25 people. The age ranged from 23 to 60 years, with an average age of 37 years (SD = 12.0). The gender distribution was 60% male and 40% female.

5 OBSERVATIONS AND INSIGHTS

The results of the study depend on the design decisions for parameter values. This makes it questionable whether it is legitimate to compare the results. For example, a shorter dwell time value would shorten the gaze-dwell method's completion time. Consequently, it would be possible to tweak the results to the desired outcome. For this reason, we prefer to speak about observations. Nevertheless, we did significance tests as this is common scientific practice. Despite the training, we excluded the first two trials per condition to

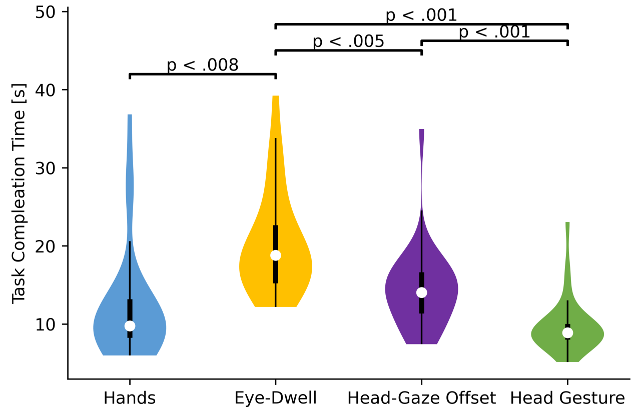

Figure 5: Average task completion time for the four interaction methods with the post hoc p-values indicated.

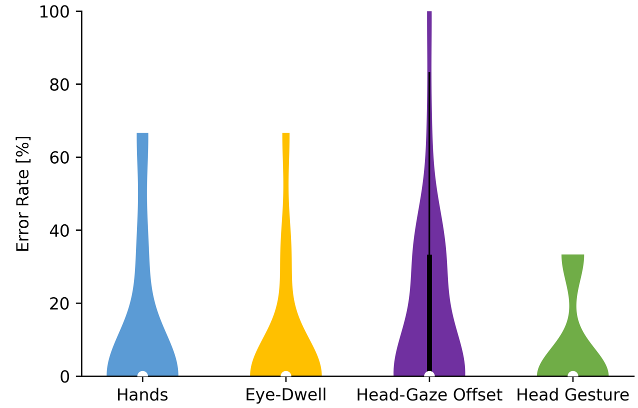

Figure 6: Error rate for the four methods.

ensure they understood the interaction techniques, and no on-set training effects were present.

Training Time. We first analyzed the training time to understand the first-time investment in learning the interaction. A Shapiro–Wilk test showed that the training time was not normally distributed ($W = .875$, $p < .001$). Consequently, we performed a Friedman test showing significant training time differences ($\chi^2 = 9.048$, $p < .027$). Moreover, post-hoc Wilcoxon signed-rank tests with Bonferroni correction applied showed that *Head-Graze Offset* ($M = 137.4s$, $SD = 83.3$) was significantly slower learn than all other methods; vs *Hands* $p < .033$ ($M = 85.0s$, $SD = 49.1$); vs *Eye-Dwell* $p < .019$ ($M = 81.3s$, $SD = 42.7$); vs *Head-Gaze Gesture* $p < .007$ ($M = 82.4s$, $SD = 36.6$). All other comparisons are $p > .05$.

Experimental Validity. Overall, participants rated the 3-items of the SSQ [14] very low. For the items *General Discomfort* the avg response was 0.4 ($SD = .8$), *Fatigue* was raged with 0.4 ($SD = .5$), and *Headache* got a mean response of 0.2 ($SD = .5$). Combined with the individual raw NASA-TLX [10] results, we argue that the overall implementation had a low impact on the experiment.

5.1 Task Completion Time

The time to successful selection is the task completion time (TCT) measured in seconds. First, we confirmed that the TCT was not normally distributed using a Shapiro–Wilk test ($W = .856$, $p < .001$). Consequently, we performed a Friedman test comparing the four methods. The results showed significant differences ($\chi^2 = 34.295$, $p < .001$). All significant post hoc comparisons using Wilcoxon signed-rank tests with Bonferroni correction applied are indicated in Figure 5. The fastest interaction method was the *Head Gesture* method, which was even quicker than the *Hands* selection. However, it is without significance. The head-gesture method was also the one with the least incorrect selections. The slowest interaction was *Eye-Dwell* method. The reason was the slow and non-dynamic scroll speed and the long dwell time for the selection, which was an intended design decision to provide an easy interface method for participants who might be overstrained by the other interaction methods. A higher scroll speed for gaze-dwell would reduce the execution time to a value similar to the other methods.

5.2 Error Rate

We counted an error if at least one wrong selection was made within one trial and depicted the results in Figure 6. Participants continued after a wrong selection until they selected the correct object. While this could have led to consecutive wrong selections, this prolonged interaction has already been penalized by the increased TCT. A Shapiro–Wilk test ($W = .584$, $p < .001$) confirmed that the error rate was not normally distributed. Consequently, we performed a Friedman test comparing the four methods, which showed no significant differences ($\chi^2 = 1.637$, $p = .713$). The higher number of incorrect selections with the *Head-Gaze Offset* was unexpected. Participants reported that they found this method particularly challenging and complex to use. The wrong selections with the hand method also surprised us. Participants told us that the scroll bar was not wide enough and too close to the list of items. Maybe the occlusion with the hand and parallax effects are further reasons for the wrong selections.

5.3 Scroll Speed

Figure 7 shows the scrolled distance over time for the four interaction methods measured in the study. Theoretically, the data points should lie on a curve but with some dispersion of the values as human performance varies.

There are predictive models for classical scroll methods [5] how much time it takes to acquire a list item. However, for the *Hands* method, the interaction process is complex and not fully understood. In our data for the *Hands* method, see Figure 7, the dispersion is high, and a functional relation is not recognizable. Although the *Hands* method was not the fastest interaction, it achieved the highest scroll speeds. Only for the *Hands* method does the scroll speed depend on the number of items in the list [5].

The data (see Figure 7) reflect the constant scroll speed for the gaze-dwell method nearly perfectly. The data points are mostly in a straight line. The scroll speed, which is the slope of the regression line, does not depend on human abilities but was a design decision.

For the dynamic scroll methods, e.g., *Gaze-Head Offset* and gaze-gesture method, with a constant increase in speed, the data should lie on a parabola. However, because of the high dispersion, this is not recognizable.

5.4 User Ratings

Figure 8 shows the average SUS score [4] for the four interaction methods. We confirmed that the data is not normally distributed ($W = 947$, $p < .001$). Next, a Friedman test showed that the conditions are significantly different ($\chi^2 = 17.766$, $p. < .001$). All significant post hoc comparisons using Wilcoxon signed-rank tests with Bonferroni correction applied are indicated in Figure 8. Again, the head gesture method got the best rating.

Figure 9 shows the average ratings in six categories from the raw NASA-TLX questionnaire [10]. The head gesture method has the lowest frustration and effort and the best performance.

6 DISCUSSION

According to the SUS score, none of the interaction methods is completely inoperative. The favorite interaction method, however, was the head-gesture method. Interestingly, the *Head Gesture* method has the least portion of gaze interaction of all the interaction methods using gaze. *Eye-Dwell* interaction is mentally demanding, while body movements are physically demanding.

6.1 Fatigue Effects

There is a clear hierarchy of how physically demanding body movements are. Lifting the arms for mid-air gestures is physically demanding, a well-known effect called gorilla arm syndrome. Besides the wish to have the hands free for other tasks, the gorilla arm syndrome is a reason why HCI researchers look for other ways of interaction.

Moving the head is much less demanding, so people bend the head down to look at their smartphone display and do not lift the arm holding the smartphone. Eye movements cause nearly no physical demand as the eyes constantly move, even while we sleep. In contrast, the mental demand for intentional eye movements seems to be high. Maybe the reason for the cognitive demand is the novel way to interact, which needs high concentration, and the cognitive demand will get lower after some practice. However, less eye movements will always be less demanding.

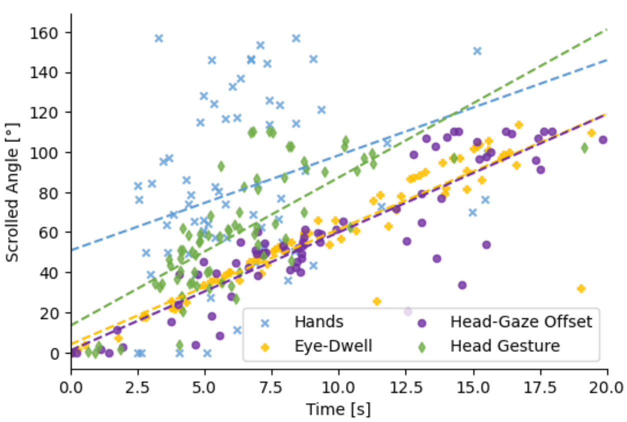

Figure 7: Scrolled distance over time with the four methods. The dashed line shows the trend line for the four methods.

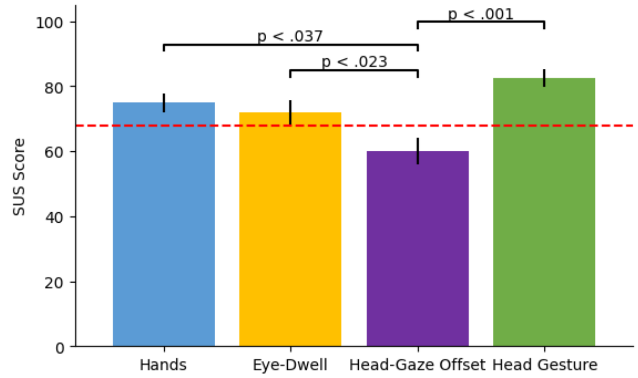

Figure 8: Results from the SUS questionnaire with the post hoc p-values indicated. The red line indicates the threshold of 68 to make systems that are considered to be below average. Error bars represent standard error.

6.2 Multi-Modal Interaction Approach

There is a trade-off for a multi-modal interaction method with body and eye movements between mental and physical demands. The head-gesture method, which involves scrolling by head movement only and selection with a small head movement while looking at the intended item, minimizes both the physical and the mental demands. At the same time, future research must examine the methods' robustness against accidental selection while doing other interactions.

It seems that using a roll movement for the head gesture was a good choice. In previous studies, we used nodding (tilt) and shaking (pan), which was a source of problems. Participants tended to nod or shake their heads too vigorously, and the heavy head-mounted

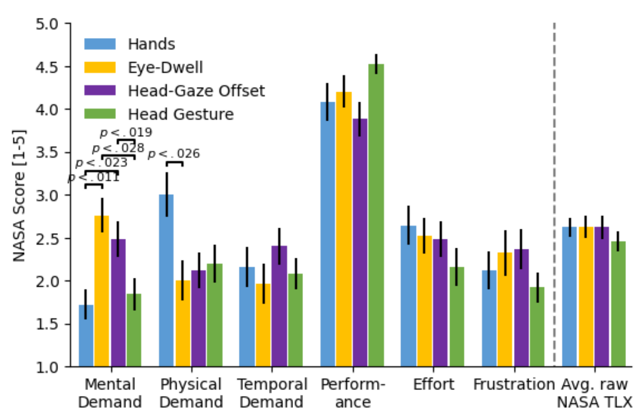

Figure 9: Results from the raw NASA-TLX questionnaire [10].

device got out of place and eventually spoiled the eye-tracker calibration. In the study, the participants performed the roll movement much softer. For answering yes-no questions with a head gesture, nodding and shaking the head is more intuitive. However, rolling the head may be the better option for selecting an item.

Selecting an item from a scroll list consists of two basic interactions: scrolling the list and selecting the item. Both basic interactions can be done with the dwell-time approach, the offset between head and gaze vector, or head gestures utilizing the vestibulo-occular reflex. We designed our study with both interactions being from the same type for consistency. However, this is not mandatory, and combining interaction methods of different types, for example, head-gaze offset for scrolling and dwell-time for selection, needs to be explored in future research.

6.3 Effect of List Entries

Another interesting question is the influence of the number of list entries on the quantitative and qualitative results. A limitation of our study was that we used only a fixed number of list entries. Also, dynamic scrolling behavior needs further investigation. Other questions are about selection in unsorted lists or how helpful a page-scroll mechanism is.

6.4 Relation To Cognitive Load

Observation from our previous studies on gaze interaction suggests that gaze interaction works best when the users are in a relaxed mood, while stressed users typically perform poorly. Stress seems to downgrade manual fine motor skills and also influences eye movements. Hands gesticulate unconsciously, but we are used to performing willful movements even under stress. In contrast, the eyes are much less used to performing willful movements, and stress degenerates this ability [15]. The big question for gaze interaction is whether willful eye movements under stress are doable on a similar level as hands or the head can do or whether gaze interaction is only feasible with relaxed users.

6.5 Limitations

As with every study, the presented study has many limitations. One of the limitations is that the study was a laboratory study with controlled conditions. AR devices are partially transparent, which means the study results may depend on what the users see in the environment. Within an office environment, the participants chose a view direction with a smooth background and nothing moving in the field of view.

Another limitation lies in the AR device used for the study. The HoloLens is quite heavy, which may affect the performance of head gestures. Future AR devices will probably have less weight. The study only shows how users react to novel interaction techniques but does not tell how users will perform after some days of practice.

Finally, the study design needed many decisions on parameter values for list entry sizes, angle thresholds for head movements, scroll speeds, scroll speed acceleration, dwell times, etc. Other parameter values will lead to different task completion times, error rates, and user judgments.

7 CONCLUSIONS AND FUTURE WORK

Despite the many limitations, the study brought valuable insights. First, the study shows that it is possible to interact with scroll lists only using gaze and head movements. TCTs and error rates are in an acceptable range for all tested interaction methods. The eye-dwell method had the longest TCT, but reducing the long dwell time of 2 seconds to a dwell time below one second and implementing dynamic scroll behavior would bring the TCT down to the value of the other methods. Compared to the hands-only interaction, the gaze and head movement interaction allows for bigger distances between the user and the interaction objects, which feels subjectively nicer.

The study's main takeaway is that the users preferred the interaction method with the least gaze interaction. As a general design rule for developing hands-free interaction in AR and VR we recommend using mostly head movements with only a little support by eye movements.

We intend to investigate the effects of long-term usage. Effects like the gorilla arm syndrome become only obvious in longer studies. The same is true for training effects. After some training with a dwell-time approach, it is possible to reduce the dwell-time period, and the interaction will be more efficient. Experienced users may not need the red dot for the *Gaze-Head Offset* method, resulting in less distraction. Working several days with one of the tested interaction methods may change the picture.

Scrolling is a frequent interaction in graphical user interfaces, not only in lists but also in documents. Existing interfaces do not only provide scrolling by one list entry or text line but also scrolling by page. For long lists or documents, an option to scroll to the start or end of the list could be beneficial. Implementing these additional options and testing them in a study is another task for future work.

The main research question, however, for future research is to find out why gaze interaction is so out of favor for the users. The big question is whether training will lead to ease of use and increase the acceptance of gaze interaction or whether gaze interaction has a general conflict with unconscious eye movement.

8 OPEN SCIENCE

We encourage readers to reproduce and extend our results. Therefore, we made the data collected in our study and our analysis scripts available on the Open Science Framework https://osf.io/q4e7k/.

ACKNOWLEDGMENTS

Sven Mayer was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 251654672-TRR 161. Yara Fanger was a student trainee at Siemens AG where her work and research were supervised and supported by T SDT IPV-DE Team. Her supervisor during the time was Nischita Sudharsan-Gröber (Siemens AG).

REFERENCES

[1] Isayas Berhe Adhanom, Paul MacNeilage, and Eelke Folmer. 2023. Correction to: Eye Tracking in Virtual Reality: A Broad Review of Applications and Challenges. *Virtual Real.* 27, 2 (2023), 1569–1570. https://doi.org/10.1007/s10055-023-00781-4
[2] Richard A. Bolt. 1981. Gaze-Orchestrated Dynamic Windows. *SIGGRAPH Comput. Graph.* 15, 3 (1981), 109–119. https://doi.org/10.1145/965161.806796
[3] Doug A. Bowman, Donald B. Johnson, and Larry F. Hodges. 1999. Testbed Evaluation of Virtual Environment Interaction Techniques. In *Proceedings of the*

ACM Symposium on Virtual Reality Software and Technology (London, United Kingdom) (VRST '99). Association for Computing Machinery, New York, NY, USA, 26–33. https://doi.org/10.1145/323663.323667
[4] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).
[5] Andy Cockburn and Carl Gutwin. 2009. A Predictive Model of Human Performance With Scrolling and Hierarchical Lists. *Human–Computer Interaction* 24, 3 (2009), 273–314. https://doi.org/10.1080/07370020902990402
[6] Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the Computer Using Gaze Gestures. In *Human-Computer Interaction – INTERACT 2007*. Springer Berlin Heidelberg, Berlin, Heidelberg, 475–488. https://doi.org/10.1007/978-3-540-74800-7_43
[7] Andrew T. Duchowski, Vinay Shivashankaraiah, Tim Rawls, Anand K. Gramopadhye, Brian J. Melloy, and Barbara Kanki. 2000. Binocular Eye Tracking in Virtual Reality for Inspection Training. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00)*. Association for Computing Machinery, New York, NY, USA, 89–96. https://doi.org/10.1145/355017.355031
[8] Jacqui Fashimpaur, Amy Karlson, Tanya R. Jonker, Hrvoje Benko, and Aakar Gupta. 2023. Investigating Wrist Deflection Scrolling Techniques for Extended Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 386, 16 pages. https://doi.org/10.1145/3544548.3580870
[9] Ajoy S. Fernandes, T. Scott Murdison, and Michael J. Proulx. 2023. Leveling the Playing Field: A Comparative Reevaluation of Unmodified Eye Tracking as an Input and Interaction Modality for VR. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2269–2279. https://doi.org/10.1109/TVCG.2023.3247058
[10] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909
[11] Toshiya Isomoto, Shota Yamanaka, and Buntarou Shizuki. 2022. Interaction Design of Dwell Selection Toward Gaze-Based AR/VR Interaction. In *2022 Symposium on Eye Tracking Research and Applications* (Seattle, WA, USA) *(ETRA '22)*. Association for Computing Machinery, New York, NY, USA, Article 39, 2 pages. https://doi.org/10.1145/3517031.3531628
[12] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. ACM, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246
[13] Pekka Kallioniemi, Tuuli Keskinen, Ville Mäkelä, Jussi Karhu, Kimmo Ronkainen, Arttu Nevalainen, Jaakko Hakulinen, and Markku Turunen. 2018. Hotspot Interaction in Omnidirectional Videos Using Head-Mounted Displays. In *Proceedings of the 22nd International Academic Mindtrek Conference* (Tampere, Finland) *(Mindtrek '18)*. Association for Computing Machinery, New York, NY, USA, 126–134. https://doi.org/10.1145/3275116.3275148
[14] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (01 Jul 1993), 203–220. https://doi.org/10.1207/s15327108ijap0303_3
[15] Thomas Kosch, Mariam Hassib, Paweł W. Woźniak, Daniel Buschek, and Florian Alt. 2018. Your Eyes Tell: Leveraging Smooth Pursuit for Assessing Cognitive Workload. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174010
[16] Manu Kumar and Terry Winograd. 2007. Gaze-Enhanced Scrolling Techniques. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology* (Newport, Rhode Island, USA) *(UIST '07)*. Association for Computing Machinery, New York, NY, USA, 213–216. https://doi.org/10.1145/1294211.1294249
[17] Hua Liao, Changbo Zhang, Zhao Wendi, and Weihua Dong. 2022. Toward Gaze-Based Map Interactions: Determining the Dwell Time and Buffer Size for the Gaze-Based Selection of Map Features. *ISPRS International Journal of Geo-Information* 11 (02 2022), 127. https://doi.org/10.3390/ijgi11020127
[18] Mathias N. Lystbæk, Peter Rosenberg, Ken Pfeuffer, Jens Emil Grønbæk, and Hans Gellersen. 2022. Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 145 (2022), 18 pages. https://doi.org/10.1145/3530886
[19] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast Gaze Typing with an Adjustable Dwell Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 357–360. https://doi.org/10.1145/1518701.1518758
[20] Päivi Majaranta, Anne Aula, and Kari-Jouko Räihä. 2004. Effects of Feedback on Eye Typing with a Short Dwell Time. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications* (San Antonio, Texas) *(ETRA '04)*. Association for Computing Machinery, New York, NY, USA, 139–146. https://doi.org/10.1145/968363.968390
[21] Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. 2012. Eye-Based Head Gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 139–146. https://doi.org/10.1145/2168556.2168578
[22] Martez E. Mott, Shane Williams, Jacob O. Wobbrock, and Meredith Ringel Morris. 2017. Improving Dwell-Based Gaze Typing with Dynamic, Cascading Dwell Times. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2558–2570. https://doi.org/10.1145/3025453.3025517
[23] Hidetoshi Nonaka. 2003. Communication Interface with Eye-Gaze and Head Gesture Using Successive DP Matching and Fuzzy Inference. *J. Intell. Inf. Syst.* 21, 2 (2003), 105–112. https://doi.org/10.1023/A:1024754314969
[24] Tomi Nukarinen, Jari Kangas, Oleg Špakov, Poika Isokoski, Deepak Akkil, Jussi Rantala, and Roope Raisamo. 2016. Evaluation of HeadTurn: An Interaction Technique Using the Gaze and Head Turns. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (Gothenburg, Sweden) *(NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, Article 43, 8 pages. https://doi.org/10.1145/2971485.2971490
[25] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180
[26] Ken Pfeuffer, Jan Obernolte, Felix Dietz, Ville Mäkelä, Ludwig Sidenmark, Pavel Manakhov, Minna Pakanen, and Florian Alt. 2023. PalmGazer: Unimanual Eye-hand Menus in Augmented Reality. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction* (, Sydney, NSW, Australia,) *(SUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 10, 12 pages. https://doi.org/10.1145/3607822.3614523
[27] Thammathip Piumsomboon, Gun Lee, Robert W. Lindeman, and Mark Billinghurst. 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 36–39. https://doi.org/10.1109/3DUI.2017.7893315
[28] David Rudi, Ioannis Giannopoulos, Peter Kiefer, Christian Peier, and Martin Raubal. 2016. Interacting with Maps on Optical Head-Mounted Displays. In *Proceedings of the 2016 Symposium on Spatial User Interaction* (Tokyo, Japan) *(SUI '16)*. Association for Computing Machinery, New York, NY, USA, 3–12. https://doi.org/10.1145/2983310.2985747
[29] Shardul Sapkota, Ashwin Ram, and Shengdong Zhao. 2021. Ubiquitous Interactions for Heads-Up Computing: Understanding Users' Preferences for Subtle Interaction Techniques in Everyday Settings. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 36, 15 pages. https://doi.org/10.1145/3447526.3472035
[30] Stoo Sepp, Steven J. Howard, Sharon Tindall-Ford, Shirley Agostinho, and Fred Paas. 2019. Cognitive Load Theory and Human Movement: Towards an Integrated Model of Working Memory. *Educational Psychology Review* 31, 2 (2019), 293–317. https://doi.org/10.1007/s10648-019-09461-9
[31] Selina Sharmin, Oleg Špakov, and Kari-Jouko Räihä. 2013. Reading On-Screen Text with Gaze-Based Auto-Scrolling. In *Proceedings of the 2013 Conference on Eye Tracking South Africa* (Cape Town, South Africa) *(ETSA '13)*. Association for Computing Machinery, New York, NY, USA, 24–31. https://doi.org/10.1145/2509315.2509319
[32] Ludwig Sidenmark, Mark Parent, Chi-Hao Wu, Joannes Chan, Michael Glueck, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. 2022. Weighted Pointer: Error-aware Gaze-based Interaction through Fallback Modalities. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3585–3595. https://doi.org/10.1109/TVCG.2022.3203096
[33] Ludwig Sidenmark, Dominic Potts, Bill Bapisch, and Hans Gellersen. 2021. Radi-Eye: Hands-Free Radial Interfaces for 3D Interaction Using Gaze-Activated Head-Crossing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 740, 11 pages. https://doi.org/10.1145/3411764.3445697
[34] Oleg Špakov and Päivi Majaranta. 2012. Enhanced Gaze Interaction Using Simple Head Gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 705–710. https://doi.org/10.1145/2370216.2370369