3D Selection in Mixed Reality: Designing a Two-Phase Technique To Reduce Fatigue

Adrien Chaffangeon Caillet, Alix Goguey, Laurence Nigay

3D Selection in Mixed Reality: Designing a Two-Phase Technique To Reduce Fatigue

Adrien Chaffangeon Caillet* Alix Goguey† Laurence Nigay‡

Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG

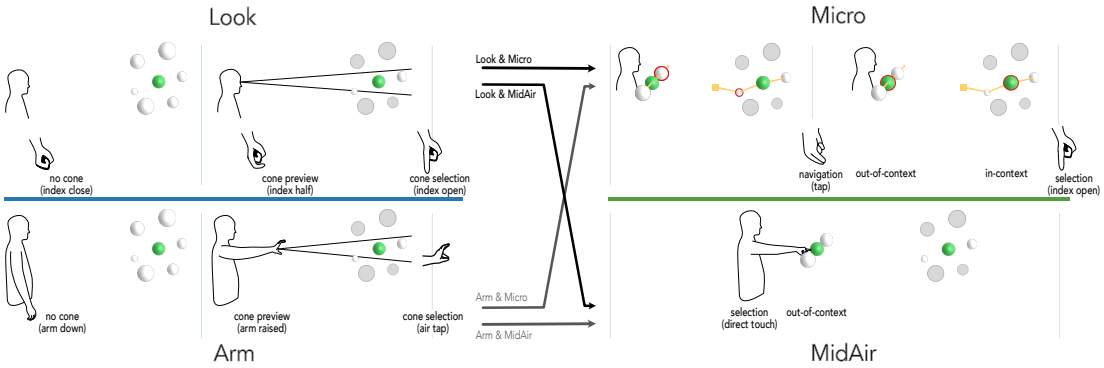


Figure 1: Using eye-gaze and microgestures (top) to reduce fatigue in AR in a two-phase technique. The first phase (Look) consists of pre-selecting objects using a cone directed along the eye-gaze and controlled by microgestures. The second phase (Micro) consists of selecting the target by navigating the list of pre-selected objects in the context of the 3D scene, or if needed outside of its context. Baseline modalities: arm-based raycasting and direct touch (bottom). The first phase (Arm) consists of pre-selecting objects using a cone directed along the arm and validated by an air-tap. The second phase (MidAir) consists of selecting the target from an out-of-context list using direct touch. We derive four techniques from these modalities: Look & Micro, Look & MidAir, Arm & Micro, Arm & MidAir.

Abstract

Mid-air pointing is widely used for 3D selection in Mixed Reality but leads to arm fatigue. In a first exploratory experiment we study a two-phase design and compare modalities for each phase: mid-air gestures, eye-gaze and microgestures. Results suggest that eye-gaze and microgestures are good candidates to reduce fatigue and improve interaction speed. We therefore propose two 3D selection techniques: Look&MidAir and Look&Micro. Both techniques include a first phase during which users control a cone directed along their eye-gaze. Using the flexion of their non-dominant hand index finger, users pre-select the objects intersecting this cone. If several objects are pre-selected, a disambiguation phase is performed using direct mid-air touch for Look&MidAir or thumb to finger microgestures for Look&Micro. In a second study, we compare both techniques to the standard raycasting technique. Results show that Look&MidAir and Look&Micro perform similarly. However they are 55% faster, perceived easier to use and are less tiring than the baseline. We discuss how the two techniques could be combined for greater flexibility and for object manipulation after selection.

Index Terms: 3D selection—eye-gaze—microgesture;

1 Introduction

Pointing is a basic task universally present in 2D and 3D graphical user interfaces. Facilitating pointing is thus an important and active research topic in the field of Human-Computer Interaction (HCI), and researchers have proposed numerous techniques to make pointing faster and more accurate. However, after more than twenty years of research into 3D selection techniques for mixed reality (MR), commercial headsets for AR, e.g. HoloLens, or VR, e.g. Oculus or HTC Vive, still use an arm-based ray-casting technique that is not suitable for distant and/or small objects, or for dense environments with partially or fully occluded objects. Moreover, using the arm to point at targets leads to arm fatigue and discomfort, an effect known as ”gorilla arm”. [20].

In this paper, we explore two modalities to reduce fatigue: eye-gaze and microgestures, i.e fast and subtle finger movements [6, 7]. The combination of gaze and microgestures has already been implemented for 2D selection of sparse targets displayed on a cockpit screen, and induced less fatigue than the use of cockpit physical controllers [52]. This paper examines whether this combination can be used for 3D selection in mixed reality.

Eye-gaze suffers from the inaccuracy of eye trackers and the constant natural movements of the human eye. To overcome these problems, we are investigating the use of a cone directed along our gaze. To solve the problem of midas touch for eye-gaze, we use microgestures to activate, visualize and validate the cone selection. If several objects are selected within the cone, we use microgestures to navigate the list of selected objects that is displayed both in the context of the 3D scene (by directly linking objects in the 3D scene) and linearly in front of the user, outside the context of the 3D scene.

After a first experiment comparing eye-gaze and microgestures with baseline modalities (see Fig. 1), we derive two 2-phase selection techniques for MR: Look&MidAir and Look&Micro. They combine eye-gaze and respectively direct mid-air touch or microgestures (see Fig. 1). Both techniques include a first phase during which users control a cone directed along their eye-gaze. Using the flexion of their non-dominant hand index finger, users pre-select the objects intersecting this cone. If several objects are selected, a disambiguation phase is performed using direct mid-air touch for

*e-mail: adrien.chaffangeon@univ-grenoble-alpes.fr
†e-mail:alix.goguey@univ-grenoble-alpes.fr
‡e-mail:laurence.nigay@univ-grenoble-alpes.fr

Look&MidAir or thumb-to-finger microgestures for Look&Micro.

The contributions of this work are twofold: (1) we experimentally explore several modalities both in terms of fatigue and performance (2) and based on the results, we propose and evaluate two 2-phase 3D selection techniques for MR that minimize fatigue, Look&MidAir and Look&Micro. The results show the drastic gain in speed for the two techniques, 55% faster than the baseline arm-based raycasting technique that is commonly used today in MR. Finally, we highlight the possibility of combining the two techniques for greater flexibility and for the manipulation of objects after selection.

2 Related Work

Our related work first reviews different designs of a 3D selection technique, then explores how to design a two-step 3D selection technique and the microgesture modality.

2.1 3D selection

This section is structured following Delamare et al. phase division of a selection task [9]: a pointing phase (i.e., how to point at targets) and a disambiguation phase (i.e, how to identify one target amongst several pointed at).

Pointing modality – In order to point at something, one needs a pointer that the system can recognize. There are various kind of pointers, from physical pointers to user's body parts. Commercial VR headsets, such as the Oculus [31] or the HTC Vive [50], use a controller to select and interact with 3D objects. Controllers are devices held by users. Controllers integrate buttons, trackpads, and/or joysticks. Other types of controllers exist, including pens [51]. Myopoint [18] uses the forearm, equiped with a Myo band to measure electromyographic and 3D orientation (using IMUs), to point on a large distant screen. The HoloLens 2 [34] and the Motion Leap 2 [28], two of the most recent MR headsets, detect the direction of the user's forearm, with cameras on the headset, to point at objects and a hand gesture to validate the selection. These selection techniques require mid-air interaction that may causes arm fatigue and discomfort, a side-effect referred as *gorilla arm* [20]. Thus, even if mid-air interaction does not necessarily induce physical fatigue, especially when the arm is not fully extended [58], no technique relying solely on mid-air interaction meets our requirement of low physical fatigue to avoid any potential impact. Therefore, researchers have tried other means such as head- and eye-gaze as an effortless pointer.

Head- [8, 36] and eye- [22] gaze have long been studied for 3D interaction. Studies [4, 23] have compared them and found that head-gaze is more stable but is slower than eye-gaze. The two studies have opposite results on users' preferences. On the contrary, Qian et al. [40] found eye-gaze to be slower than head-gaze in VR. Nevertheless, in most of these studies [23, 40], the authors discuss the potential impact of the eye tracker's low accuracy on the performance of eye-gaze. In addition to hardware limitations, human eye limitations such as visual acuity, involuntary flick, drift or saccade-like eye movements [55] reduce the accuracy of eye-gaze to select objects. Since selecting a target requires us to first look at it [56], eye-gaze is in theory the fastest pointer. Moreover, during text entry in VR [17], eye-gaze was reportedly less physically demanding than head-gaze. Therefore, we use eye-gaze as the base of our pointer to minimize fatigue while providing the highest speed.

Both eye and head gazes suffer from the Midas touch problem [22], i.e. while moving their head or eyes, the system will recognize an interaction input, even if it is not the user's intention. Therefore, when using head- or eye-gaze we need two actions: one to activate the selection mode, and one to validate the selection. In the literature, activating and validating an eye-gaze selection is commonly performed using dwell time [22]. However, optimal dwell time is dependent on the tasks and is different for each user [21]. Moreover, in a study using ISO 9241 - part 9 [10] on desktop [57], using dwell was slower than pressing the space bar on a keyboard. Pursuits interaction correlates the movements of proxies around targets to the eye-movements [13, 25]. To select a target, the eyes simply follow the associated proxy. However, studies [13, 25], tested only up to 8 visible targets. Moreover, when testing with 8 targets error rates varied between 20% and 35% [13]. Thus, these interaction techniques are not suited for dense environments. The HoloLens 1, an AR headset, comes with a clicker [33] to validate the head-gaze selection. However it requires additional hardware. Gaze+Pinch [39] uses one hand gesture to validate the selection, a pinch gesture. However, a single hand gesture cannot be used for the two actions of activating the mode and validating the selection. Therefore, we explore microgestures to increase the gesture set used with eye-gaze.

As discussed above, eye-gaze suffers from accuracy issues In order to cope with the human eye limitations and the low accuracy of the eye tracker, the eye-gaze will not control a ray, which requires precision to select an object, but a cone, as used in previous works [14, 24, 29]. Objects intersecting the cone directed along the eye-gaze are potential candidates, requiring a disambiguation mechanism to select one of them.

Disambiguation mechanism – Several objects can be intersected with a cone selection, thus requiring a disambiguation mechanism to select only one object. Beyond cone selection, a disambiguation mechanism is required for techniques using another type of volume for selection or using ray selection. Delamare [9] distinguishes three mechanisms: no disambiguation, automatic disambiguation, and interactive disambiguation.

Techniques with no disambiguation phase select the first object intersected by the ray. This is the standard raycasting, like the laser gun [29]. Even though it works well in most scenarios, it is hard to select small and/or distant objects, especially if they are partially hidden by other objects.

Automatic disambiguation techniques select an object based on heuristics [16, 41]. For instance, with the Smart Ray [15], the object in contact with the ray for the longest time is selected, and with the spotlight selection [29], the selected object is the closest object inside a cone using a metric that yields ellipsoidal isodistance surfaces. Given the automatic nature of such techniques, users will partly lose the explicit control over the interaction, which might induce frustration.

Interactive disambiguation techniques allow users to modify, refine or go through a list of objects pre-selected by a volume selection or by a ray. The disambiguation phase can either occur during the pointing movement, or afterwards, as a separate phase. For instance, parallel disambiguation was chosen for the Depth Ray [15] or the RayCursor [2]: users move a cursor along the ray while pointing. In this case, many techniques use a bubble mechanism to select the target closest to the ray, or a cursor on the ray [2, 30]. SQUAD [26] uses a separate disambiguation phase: after pre-selecting an initial set of objects, users recursively refine the set until only one object is left.

Our goal is to take advantage of eye-gaze for its speed and low physical demand properties. To disambiguate objects within the eye-gaze directed cone, we discard parallel disambiguation due to eye-gaze instability. Therefore, we consider a separated disambiguation phase, using another modality to navigate the list of objects inside the cone.

2.2 Designing a distinct disambiguation phase

With the aim of designing a two-stage technique, we are exploring different designs for a separate disambiguation phase using microgestures.

Separated disambiguation phase – Delamare et al. presented a design space for disambiguation mechanisms in the case of physical target selection [9]. The design space organizes a set

of design options into two groups: the *interaction* group and the *disambiguation system* group.

The *interaction* group includes design features related to the display and control space. **Display space.** The pre-selected objects can be displayed either 1) within the scene, maintaining context awareness and avoiding visual focus changes, as in the lock ray technique [15], or 2) within a remote space, which can be useful in specific cases as exposed in SQUAD [26]. **Control space.** In addition, pre-selected objects can be selected either by interacting directly within the scene or in the remote space. In our study, we allow both disambiguation within the scene and within a remote space for flexibility.

The *disambiguation system* group includes the information used for disambiguation. Pre-selected objects can be disambiguated using information about their appearance (e.g. shape, color, or size), functionalities (e.g. associated commands), relative position to each other, or other properties (e.g. current state or identifier). Techniques based on a remote display and control space, such as SQUAD [26], involve information about the preselected objects that allows them to be differentiated. The information about the pre-selected objects displayed in the remote space depends on the context of use. In this paper, we focus on the performance of the modalities independently of the context of use. Therefore, the study does not consider information about pre-selected objects that are displayed in the remote space. To do so, we use a stripped-down scene with abstract spherical objects and a clearly identified target, as has been done in other studies [2, 16, 30].

Finally, as pointed out in [26], in terms of performance, using a separate disambiguation phase implies a trade-off between the selection phase and the disambiguation phase. Indeed, the faster and more imprecise the selection phase is, the greater the number of pre-selected objects and therefore the longer the disambiguation phase is. Several modalities can be used for the disambiguation phase, such as voice [37, 49] or touch on a handheld device [45]. We study the reuse of microgestures for a fast disambiguation phase and a straightforward transition between the selection phase and the disambiguation phase.

MICROGESTURES: GOING BEYOND A PINCH – Microgestures are defined as fast and subtle finger movements [6, 7]. In Mixed Reality (MR), some microgestures are already used. In the HoloLens 2, after pointing with the forearm, an air tap is used to confirm the selection. Similarly, in Gaze+Pinch [39], after looking at an object, a pinch gesture is used to confirm the selection. However, these techniques do not take full advantage of the expressiveness of the microgesture-based modality. Both techniques are direct pointing techniques, i.e. the first target intersected by the ray is selected, yet, by using more microgestures it should be possible to do an interactive disambiguation system. Indeed, Soliman et al. proposed a design space for thumb-and-finger microgestures from which they instanciated a set of 50+ microgestures [44]. The microgesture-based modality is gaining much attention from its ability to be used eyes-free while performing a primary task [54], such as driving [12, 19] or biking [46], and more generally while grasping an object [43]. Wambecke et al. proposed M[eye]cro, a 2D selection technique to interact in a cockpit which combines eye-gaze and microgestures. The relative inaccuracy of the gaze was compensated for by target expansion, a target plane close to the user and the spreading of targets in 2D space. Compared to interacting with the cockpit physical controllers, the combination of eye-gaze and microgestures was faster when performed in multitasking configuration, and induced less fatigue. This motivates our study to go beyond traditional air tap and pinch microgestures for 3D selection.

3 DESIGNED TECHNIQUES

In the following, we refer to *modality* as an interaction technique used during one phase, e.g. eye-gaze or microgesture, *technique* as a compound of two modalities, e.g. Look&Micro or Look&MidAir.

From the literature, we identified two promising modalities for a 3D selection technique minimizing physical fatigue: eye-gaze and microgesture. Due to eye-gaze low accuracy, the techniques require two interaction phases, a cone selection phase and a disambiguation phase. For both phases, we identified a baseline modality and a modality minimizing physical fatigue.

For the cone selection phase, during which users pre-select candidate objects using a cone, the baseline modality is mid-air interaction. As in the recent MR headset HoloLens 2, our baseline modality uses the forearm and a hand gesture (ARM) to roughly select a set of objects. Our modality minimizing fatigue uses eye-gaze and finger flexion (LOOK).

For the disambiguation phase, during which users select one of the pre-selected objects, the baseline modality uses direct mid-air touch (MIDAIR) in the list of pre-selected objects. Our modality minimizing fatigue uses thumb-to-finger tap (MICRO) to navigate the list of pre-selected objects.

The three microgestures used by the techniques are described using the µGlyph notation [6]: PINCH (), TAP (), and STRETCH (). In the following, their names are used to describe the interaction.

Our exploration of the design space by considering two modalities for each phase of the technique leads us to define four techniques.

ARM&MIDAIR– With the Arm&MidAir technique, as currently implemented in the HoloLens 2, the resting position (i.e., the idle interaction state) corresponds to when users' arms are alongside their body. Users lift up their dominant hand forearm to start the cone selection phase.

Along the axis of their forearm, we define an infinite cone with an aperture of 3° corresponding to Microsoft's recommendation for a comfortable experience [32], which at 6m has a diameter of 30cm. During the cone selection phase, all the objects of the scene are grayed out except the ones intersecting the cone. When users PINCH their thumb and index finger of the same hand, all the objects intersecting the cone are pre-selected, and the disambiguation phase begins. In the disambiguation phase, copies of the pre-selected objects are displayed as a linear list at 60 cm in front of the users (29° vertically under the field of view). This design is similar to the HoloLens 2 main menu. A visual feedback at the bottom of their field of view informs the users that the list is available. The order of the objects in the list is defined by their proximity to the user in the scene (e.g, the leftmost in the list is the closest to the user). To select an object, users simply touch it in the list with their index finger.

If users lower their arms, the entire selection process is canceled.

LOOK&MIDAIR– The Look&MidAir technique combines eye-gaze with simple microgestures to control the cone during the cone selection phase. The cone, whose properties are identical to the previously described one, is directed with the users' eye-gaze. To avoid the Midas problem [22], the cone selection phase has to start with users' left hand fully closed. The cone is triggered when the index finger starts stretching out. As previously, all objects are grayed out except the ones intersecting the cone. Users can adjust the cone simply by changing their eye-gaze direction. When the index is completely stretched out, the objects intersecting the cone are pre-selected and the disambiguation phase, which is the same as in Arm&MidAir, begins.

If users close their hand, the entire selection process is canceled.

LOOK&MICRO– The Look&Micro technique implements the Look&MidAir cone selection phase but uses microgestures for the next phase. In the disambiguation phase, we take advantage of the index finger, already being stretched out from the first phase, to navigate pre-selected objects. We use a red cursor outlining both representations (in the 3D scene and in the list representation in front of the users) of a pre-selected object. This synchronized cursor allows users to navigate pre-selected objects with or without contextual knowledge. To reinforce views' synchronization, pre-

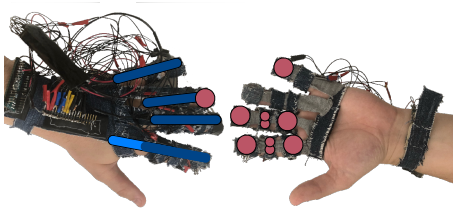


Figure 2: Glove to capture thumb-to-finger and flexion microgestures. Red circles represent pressure sensors. Blue lines represent flex sensors. The brighter blue line is the flex sensor added for the second experiment.

selected objects of the scene are connected as in the list (i.e., with orange lines). To move the red cursor, users TAP the top and middle phalanx of the index finger with their thumb to move the cursor away / rightwards or closer / leftwards. The cursor is initially placed on the nearest / leftmost object. Participants could freely choose to disambiguate in- or out-of-context, by simply looking at the scene or the at the list. To select the object outlined by the cursor, users simply TAP their ring finger nail.

As previously, if users close their hand, the entire selection process is canceled.

ARM&MICRO– To be able to compare the modalities of the second phase without the bias of the modalities of the first phase, we also test the fourth combination of modalities. The Arm&Micro techniques implements the Arm&MidAir cone selection phase and the Look&Micro disambiguation phase.

4 EXPLORATORY EXPERIMENT

In this first study, we explore the modality design space for both phases, the selection phase and the disambiguation phase. Our goal is to compare the performance of all the modality combinations presented above.

HARDWARE – With a focus on interaction rather than recognition, we built a simple glove to recognize thumb to finger and stretch microgestures, inspired from Wambecke et al [52] The glove contains fourteen sensors, four flex sensors and ten pressure sensors (see Fig. 2 for the layout of the sensors). Pressure sensors are 17 mm wide, except on middle phalanxes where two smaller 7.62 mm sensors have been placed. This was done to obtain a large input area on middle phalanxes, without hindering closing of the fingers. Flex sensors are 55.37mm long. Every sensor is linked to an arduino Micro. To get smoother analog values from the sensors, we use a $10k\Omega$ resistor between the ground and the sensor output. The position of the pressure sensors can be adjusted to be perfectly positioned on each phalanx, regardless of the finger or hand sizes. We use an HoloLens 2 for Mixed Reality. The connection between both is made through a nodejs server hosted on a Windows 10 computer. The software running on the HoloLens 2 is using the Mixed Reality Toolkit v2.3.0 [35].

TASK – In a given trial, participants have to select a green target in a cloud of white points. The cloud of points remains the same throughout the block. The cloud of points density and point size vary between each block. In each block, the cloud of points layout is pseudo-randomly generated and is the same for each participant. The cloud of points describes a 0.5m radius sphere which center is placed at 6m in front of the participants, similar to the sets of points across a room used by Lu et al. [30]. In comparison, at the same distance, the cross-section of the cone passing through the sphere center is a circle of 0.16m. The positions of the target vary throughout a block of trials. The position of a target is randomly chosen in a list of predefined positions (given in spherical coordinate around the center of the cloud of points): $(r = 0m, \theta = 0, \phi = 0)$ and $(r = 0.5m, \theta \in [0°, 90°, 180°], \phi \in [0°, 90°, 180°, 270°])$. Participants can rest as long as they wish between blocks. Between each trial of a block, a message on screen asks participants to deactivate the technique, i.e. close their hand for Look&Micro and Look&MidAir or lower their arm for the Arm&MidAir or Arm&Micro, and to press a button to continue. In case of error, the trial is rescheduled at the end of the block, thus ensuring the trial is completed once with no error. We distinguish two types of errors: selection errors, i.e. selecting a set of objects that does not contain the target, and disambiguation errors, i.e. the selected object during the disambiguation phase is not the target.

PARTICIPANTS – We recruited 16 participants (mean age = 24.9yo, $\sigma = 9$, gender: 10 males, 6 females). None of them had used an MR headset before. Four of them had already participated in an experiment three months earlier involving thumb-to-finger taps. The microgestures were used in a different context than MR. Since our glove is designed to be worn on the left hand only, we chose right-handed participants only to avoid bias from using the dominant or non-dominant hand.

DESIGN – We used a within-subject design with factors: TECHNIQUE, TARGET NUMBER and TARGET SIZE. We decided to test two densities of target inside the cloud of points: 20% and 60%. We broke down the density into two parameters: the number of targets in a volume and the size of the targets. First, we fixed the TARGET NUMBERS following a voxel approximation algorithm of a sphere [38] to divide the volume containing the objects to select from. As a parameter, this algorithm takes a desired number of voxels used to divide a sphere diameter. In order to ensure a voxel matching the sphere center, we used odd numbers as parameters: a diameter of 3, yielding a sphere of 19 voxels, and a diameter of 5, yielding a sphere of 81 voxels. We then chose each voxel to contain exactly one target, which thus yields two TARGET NUMBER conditions: LOW (19 targets) and HIGH (81 targets). Second, we computed the TARGET SIZES based on the number of targets to obtain the two aimed densities, 20% and 60%. For a density α, a target has a diameter equal to $\alpha \times$ VOXEL SIDE. TARGET SIZE conditions are thus 3.3cm and 10cm in diameter for LOW, and 2cm and 6cm in diameter for HIGH. Finally, each target was positioned with pseudo-randomly generated jitters around its voxel center, respecting a rule, that the target must be entirely contained within its voxel. Prior to each block TECHNIQUE x TARGET NUMBER x TARGET SIZE, participants performed 4 training selection tasks. This data was discarded for the analysis.

A total of 1792 selections were collected, from 16 participants x 7 selections x 4 TECHNIQUE x 2 TARGET NUMBER x 2 TARGET SIZE. After finishing the experiment for a technique, participants were asked to fill out a raw Nasa TLX with 7 Likert scale. We used a raw Nasa TLX instead of single scale measures (e.g., Borg scale [5])to measure both mental and physical fatigue separately. Moreover, to capture the origin of the fatigue, we asked users to comment their answers every time fatigue was reported. After having performed the experiment with each technique, participants were asked to rank the techniques according to their perceived speed, accuracy, and both mental and physical fatigue as well as their preference. The experiment lasted around 1h.

Results

TIME – For the time analysis, our data was positively skewed and thus not following a normal distribution (Shapiro test $p < 0.001$). Following Tip 12 and 15 of Dragicevic [11], we applied a log transform then used n-way repeated ANOVA for our main effect analysis and pairwised t-test, with Bonferroni p-value adjustment, for posthoc analysis. Confidence intervals were computed using bootstrap.

As shown in Fig. 3A, the fastest technique is Look&MidAir (mean: 3.64s, confidence interval: [3.30, 3.99]), followed by Look&Micro (m: 5.06s, ci: [4.70, 5.47]), Arm&MidAir (m: 6.22s, ci: [5.55, 6.91]), and Arm&Micro (m: 6.9s, ci: [6.21, 7.71]).

TECHNIQUE has a significant main effect on SELECTION TIME

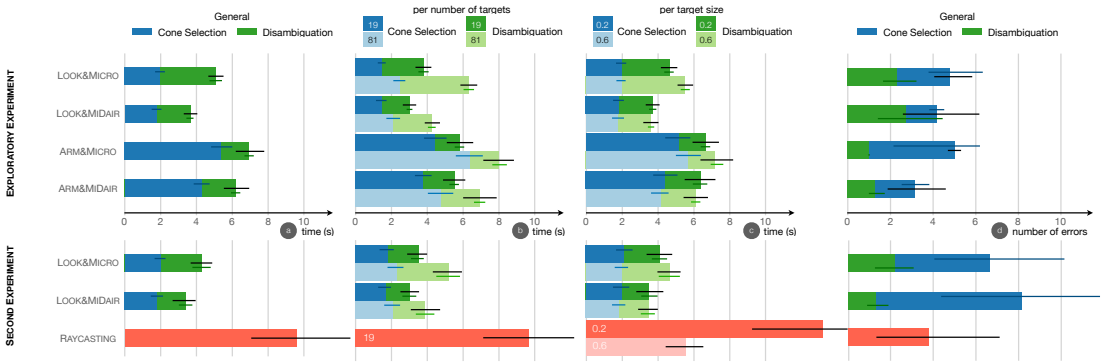


Figure 3: Results of of both experiments. A) Mean completion time and 95% CIs for each technique and phase. B) Mean completion time and 95% CIs for each technique, phase and number of targets. C) Mean completion time and 95% CIs for each technique, phase and target size. D) Average number of errors and 95% CIs per participant for each technique and phase.

($F_{(3,45)} = 42.92, p < 0.001, \eta^2 = 0.62$). Posthoc tests showed no significant difference between Arm&Micro and Arm&MidAir ($p = 0.79$), a significant difference between Arm&MidAir and Look&Micro ($p = 0.009$), and strong significant differences between every other pairs ($p < 0.001$). As shown in Fig. 3B, TECHNIQUE × TARGET NUMBER has a significant effect ($F_{(3,45)} = 9.22, p << 0.001, \eta^2 = 0.05$). As shown in Fig. 3C, TECHNIQUE × TARGET SIZE has a significant effect ($F_{(3,45)} = 5.46, p = 0.003, \eta^2 = 0.03$).

We also analyzed time performance per phase and studied the impact of the different modalities. For the cone selection phase, techniques using eye-gaze (m: 1.85s, ci: [1.66, 2.05]) were significantly faster than techniques using forearm (m: 4.81s, ci: [4.32, 5.32]) ($F_{(1,15)} = 213.07, p << 0.001, \eta^2 = 0.85$). For the disambiguation phase, techniques using direct touch (m: 1.91s, ci [1.84, 1.99]) were significantly faster than techniques using microgestures (m: 2.3s, ci: [2.19, 2.42]) ($F_{(1,15)} = 5.10, p = 0.04, \eta^2 = 0.10$). However, if we consider the four techniques separately during the disambiguation phase, microgestures perform poorer (Look&Micro 3.12s ± 0.33) and similar to direct touch (Arm&Micro 1.52s ± 0.21). For this phase, TECHNIQUE has a significant main effect on SELECTION TIME ($F_{(1,15)} = 5.10, p = 0.04, \eta^2 = 0.10$). Posthoc tests showed a significant difference between Look&Micro and all the others (all $p << 0.001$) but no difference between Arm&Micro and Arm&MidAir, as well as between Arm&Micro and Look&MidAir (both $p > 0.08$).

Further analysis focusing on the number of objects pre-selected in the list showed that: if the list includes less than 10 objects, the difference between microgestures (0.54s) and direct touch (0.45s) is not significant ($t = 2.02, df = 15, p = 0.06$); if the list has 10 or more objects, the difference between microgestures (0.97s) and direct touch (0.82s) is significant ($t = 2.36, df = 15, p = 0.03$). It is worth noting that 78% of selection trials had less than 10 objects in the list during disambiguation, and a list of more than 10 objects occurred only with dense environments of small targets (i.e., TARGET SIZE of 0.6 and TARGET NUMBER of 81).

ERROR – We consider two types of errors: cone error (i.e., the pre-selected objects do not contain the target object), and selection error (i.e., the wrong object was selected). Fig. 3D shows the average number of errors per participant for each technique. TECHNIQUE has a significant effect on cone errors ($F_{(3,42)} = 4.92, p = 0.005, \eta^2 = 0.19$). Posthoc tests shows significant differences only between Look&Micro and Arm&Micro ($p = 0.037$), and between Look&MidAir and Arm&Micro ($p = 0.027$). All others pairs have $p > 0.2$. For selection errors, there is no significant effect of TECHNIQUE ($F_{(3,42)} = 0.86, p = 0.47, \eta^2 = 0.04$).

RAW NASA TLX – We used a simplified raw Nasa TLX with a 7 likert scale. Since we did not put pressure on participants to go faster during the trials, we have replaced the question on how rushed they felt with a question on how fast they felt. Raw data and means are presented in Fig. 4a. We use Align-and-rank data for a non-parametric ANOVA [53] and found significant effect of TECHNIQUE on Mental ($F_{(3,45)} = 5.05, p = 0.004$), Physical ($F_{(3,45)} = 7.20, p = 0.0005$) and Stress ($F_{(3,45)} = 3.02, p = 0.04$). Going deeper to test our fatigue hypothesis, posthoc tests revealed significant differences between Look&Micro and both Arm&MidAir and Arm&Micro (all $p < 0.01$). It suggests that eye-gaze combined with microgestures might indeed reduce fatigue.

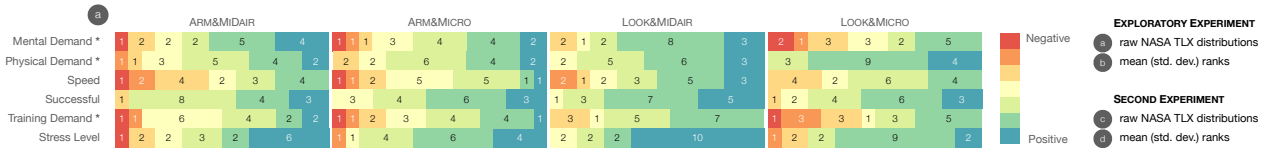
RANKING – Fig. 4b shows mean ranks and standard deviations.

Takeaways from the exploratory study

LOOK – Using eye-gaze for the cone selection phase was the best option in the experiment. Both techniques using eye-gaze during the first phase are the fastest ones and rank best for physical demand and participants' preferences. Two participants stated that controlling the cone state using the index flexion was more natural than the HoloLens 2 selection gesture, because of the smooth transition between the two phases of the technique. However, even though using eye-gaze and index flexion has the lowest physical demand, it is worth noting that two other participants reported pain from the repeated index flexions, and one participant would have preferred flexing of the index in the opposite direction (i.e., from fully opened to fully closed), thus using the same metaphor as clicking on a mouse.

MIDAIR – In the disambiguation phase, directly touching the objects was overall faster than using microgestures. However, when comparing techniques with similar first phase, the ones using mid-air touch for the second phase have a higher physical demand. With a long list of large objects (i.e., 81 targets of size 0.6%), participants sometimes had to physically move to reach the targets that were at the end of the list. Direct touch might therefore be cumbersome to use in dense environments.

MICRO – After analyzing participants' feedback, we have two explanations for the difference between Look&Micro and Arm&Micro during the disambiguation phase. 1/ For some participants, the glove was too sensitive to the index flexion. With Look&Micro, after opening their index to select the cone, participants had to focus on how they perform the subsequent thumb to finger taps to navigate the list. If they curled their index too much, the system would recognize a closed index canceling the list (Fig. 5). This behavior was due to the flex sensor being too short and not covering the joint between the index and the palm. After few errors, participants would be more careful increasing both time of the disambiguation phase and mental demand. 2/ Some participants had trouble validating on the ring finger while keeping their hand perfectly closed (except for the index finger and the thumb). As for 1/, sensors on the glove would recognize an open hand leading to starting all over again. So



b

Technique	Mental demand *	Physical demand *	Preference	Fastest	Best performance
LOOK&MIDAIR	**1.56 (0.73)**	2.19 (0.83)	**1.56 (0.83)**	**1.50 (0.73)**	**1.56 (0.73)**
LOOK&MICRO	3.38 (1.09)	**1.62 (0.89)**	2.56 (1.26)	2.56 (1.15)	2.50 (1.26)
ARM&MIDAIR	2.06 (0.77)	3.25 (1.13)	2.69 (0.79)	2.75 (1.13)	3.06 (1.06)
ARM&MICRO	3.00 (0.89)	2.93 (0.93)	3.19 (0.98)	3.19 (0.75)	2.88 (0.81)

d

Technique	Mental demand *	Physical demand *	Preference	Fastest	Best performance
LOOK&MIDAIR	**1.33 (0.65)**	1.67 (0.65)	**1.50 (0.67)**	**1.25 (0.45)**	**1.33 (0.65)**
LOOK&MICRO	2.25 (0.75)	**1.50 (0.67)**	1.75 (0.75)	1.92 (0.67)	1.83 (0.58)
RAYCASTING	2.42 (0.67	2.83 (0.39)	2.75 (0.45)	2.83 (0.39)	2.83 (0.39)

* : scale were inverted compared to what was asked to the participants, to harmonise the presentation of the results. / for the ranks results, lower values are better

Figure 4: Raw Nasa TLX results and Mean (std. dev.) ranks results for all techniques of both experiments.

Figure 5: Depending on how participants perform thumb to finger tap to navigate the list during the disambiguation phase, the system would recognize the index as A) open or B) closed.

participants would be more careful to avoid such error, increasing time and mental demand.

Additionally, from participants' feedback, the participants' mental model for both Look&Micro and Arm&Micro includes two distinct steps: move the cursor and then validate. When the target object was the first element of the list, participants hesitated before validating or moved one object forward in the list and then back before validating.

5 Refined Design

The exploration experiment shows that eye-gaze for the cone selection phase was the best option. Both techniques using eye-gaze during the first phase are the fastest ones and rank best for physical demand. Look&MidAir is the best technique in terms of selection time, mental demand and participants' preference. Look&MidAir is therefore our strongest candidate technique. From the results and participants' feedback, we also think that Look&Micro could be a strong candidate technique given few improvements to reduce its mental demand and selection time.

In the refined version of Look&MidAir, we changed the aperture of the cone from 3° to 2°, which according to Microsoft's recommendation still provides a comfortable experience. With a smaller aperture we aim at reducing the size of the list of objects in the disambiguation phase, avoiding users having to move to reach the end of the list. In addition, if users only pre-select one object in the cone, the disambiguation phase is skipped and the object is automatically selected.

In the refined version of Look&Micro, we also changed the aperture to 2°. To avoid the additional thinking when the target is the first object of the list, we added a non-selectable cube at the start of the list, forcing users to tap at least once to navigate the list. This design solution seems to match the two-step mental model of the disambiguation phase.

To further reduce the mental demand and gesture difficulties during the disambiguation phase, we abandoned the ring finger validation to be consistent with the gestures of the cone selection phase. Since the index finger is half closed for navigation taps, straightening it back validates the currently selected object in the list, as shown in Fig. 1B. Finally, we added another flex sensor on the index to cover the base joint, as shown in Fig. 2, which solves the detection problem described above.

6 Second Experiment

In this second experiment, we evaluate and compare the refined version of Look&Micro and Look&MidAir.

Baseline – The closest multimodal technique from the literature, namely Gaze&Pinch with disambiguation as implemented in [27], is a 2D selection technique. Gaze&Pinch combines eye-gaze with a secondary modality based on mid-air gestures for the disambiguation phase. The users point with their eyes and a pinch gesture positions a cursor on the 2D plane of the targets at the position of the detected gaze. After the initial selection with the eye, horizontal and vertical movements of the hand control the cursor on the target plane. A release gesture selects the target. To adapt this technique to 3D with an eye-gaze cone selection, a cursor must be positioned in the cone after the pinch gesture. The cursor position is then controlled by vertical, horizontal, and depth movements of the hand. But this adaptation of Gaze&Pinch to 3D is not straightforward, and requires further study to determine the best initial cursor position in the cone and the most appropriate control-display gain to enable the cursor to move quickly and accurately inside the 6-meter-long cone. In the absence of a reference multimodal technique combining gaze with mid-air gestures or microgestures, we used raycasting [1] as a reference in the experiment. While raycasting used to be done using a controller in the first VR headsets, it is now overshadowed by arm-based raycasting in AR headsets, i.e. Hololens 2, and is provided "out of the box" with all mainstream AR/MR devices. Moreover, arm-based raycasting is already used in the literature [18, 39, 49, 59]. Therfore, we used arm-based raycasting as implemented in Hololens 2, i.e. using the forearm to point and a tap of the index finger on the thumb to select, to facilitate replication. In the following, we use the term *raycasting* to refer to *arm-based raycasting*.

Hypothesis – From the literature and our previous experiment, we expect Look&MidAir and Look&Micro to be significantly faster and with less impact on physical fatigue than raycasting (H_1). Since Look&MidAir provides a direct access to the objects of the list, whereas Look&Micro a sequential access linearly increasing access time, we also expect Look&MidAir to be faster than Look&Micro in dense environments, i.e. when the disambiguation list will be long, but equivalent in scarce environments, i.e. when the disambiguation list will be short (H_2). Following the results of our previous experiment, we expect Look&MidAir to have lower mental demand (H_M) but higher physical demand (H_P) than Look&Micro.

Participants – We recruited 12 participants (mean age = 26.77yo, $\sigma = 5.90$, gender: 9 males, 3 females). None of them had used an MR headset before, nor participated in an experiment using thumb-to-finger tap. Since our glove is designed to be worn on the left hand only, we chose right-handed participants for non-dominant hand microgestures only.

Hardware, Task and Design – We reused the equipment of the exploratory experiment, with a slight modification of the glove. We added an additional flex on the index finger to capture the angle

between the index finger and the palm, see Fig. 2 for the exact layout. We use the same task and experimental design as in the exploratory experiment. The only difference is that we tested 3 TECHNIQUE instead of 4: Look&Micro, Look&MidAir and Hololens 2 raycasting as the baseline. For the raycasting, we only tested it for TARGET NUMBER = 19 as after pilot testing we concluded that the TARGET NUMBER = 81 was way too hard.

Prior to each block TECHNIQUE x TARGET NUMBER x TARGET SIZE, participants performed 7 training selection tasks. This data was discarded for the analysis. A total of 840 selections were collected, from 12 participants x 7 selections x 2 TECHNIQUE x 2 TARGET NUMBER x 2 TARGET SIZE + 12 participants x 7 selections x 1 TECHNIQUE x 1 TARGET NUMBER x 2 TARGET SIZE. The experiment lasted around 1h.

Results

TIME – For the time analysis, our data was positively skewed and thus was not following a normal distribution (Shapiro test $p < 0.001$). As previously explained, we used a log transform followed by pairwised t-test, with Bonferroni p-value adjustment.

As shown in Fig. 3A, the fastest technique is Look&MidAir (mean: 3.41s, confidence interval: [2.97, 3.92]), followed by Look&Micro (m: 4.31s, ci: [3.80, 4.82]), and raycasting (m: 9.65s, ci: [7.28, 12.35]).

TECHNIQUE has a significant main effect on SELECTION TIME ($F_{(2,22)} = 53.05, p < 0.001, \eta^2 = 0.63$). Using pairwised t-test, we found no significant difference between Look&Micro and Look&MidAir ($p = 0.20$), and strong significant differences between both techniques and the raycasting ($p < 0.001$). As shown in Fig. 3B, TECHNIQUE × TARGET NUMBER has no significant effect $F_{(1,11)} = 6.83, p = 0.24, \eta^2 = 0.03$. As shown in Fig. 3C, TECHNIQUE × TARGET SIZE has a significant effect $F_{(2,22)} = 52.70, p << 0.001, \eta^2 = 0.31$.

Considering only the disambiguation phase, direct touch (m: 1.61s, ci [1.35, 1.92]) is significantly faster than microgestures to navigate the list (m: 2.34s, ci: [1.99, 2.70]) ($F_{(1,11)} = 15.81, p < 0.002, \eta^2 = 0.28$.

ERROR – Given the nature of the raycasting interaction we considered all errors in the raycasting condition as selection errors. A t-test between Look&MidAir and Look&Micro shows no significant effect of TECHNIQUE on cone errors ($t = 0.99, df = 11, p = 0.35$). For selection errors, there is no significant effect of TECHNIQUE ($F_{(2,22)} = 0.1.84, p = 0.18, \eta^2 = 0.10$).

RAW NASA TLX – As done for the previous experiment, we use Align-and-rank data for a non-parametric ANOVA [53] followed by posthoc tests using estimated marginal means on the ART model. When considering only the second phase, we removed the raycasting from our data. Raw data and means are presented in Fig. 4c. We found no significant effect of TECHNIQUE on Mental ($F_{(2,22)} = 1.12, p = 0.34$), but a significant effect on Physical ($F_{(2,22)} = 10.89, p << 0.001$), Speed($F_{(2,22)} = 15.52, p << 0.001$), Success ($F_{(2,22)} = 10.45, p << 0.001$) and Stress ($F_{(2,22)} = 16.47, p << 0.001$). Posthoc tests revealed significant difference between raycasting and both Look&MidAir and Look&Micro for physical demand, speed, success and stress (all $p < 0.002$).

RANKING – Fig. 4d shows mean ranks and standard deviations.

7 DISCUSSION

Both Look&MidAir and Look&Micro are respectively 60% and 55% faster, are preferred and have a lower workload than raycasting (validating H_1). The main issue with raycasting was the accuracy. P0 (Participant 0) and P6 explained that it was more intuitive to extend their arm to grab an object, but also declared, as other participants (P2, P3, P9), that it was too frustrating to use. The frustration came from both 1) the difficulty to be precise with the ray moving around the target and 2) the ray moving out of the target when performing the HoloLens 2 selection gesture, resulting in a missed selection. In the following, we only discuss the results for Look&MidAir and Look&Micro, since the purpose of the experiment to compare touch and microgesture modalities for the disambiguation phase.

TIME – For Look&Micro, microgestures are used for linear access to objects in the list. Depending on the context of use, the disambiguation phase can be performed within the scene (e.g., to preserve the spatial layout of objects) or in a remote list of copied objects displayed in front of the user (e.g., to access targets partially hidden in the scene). Since Look&MidAir provides a direct access to the objects, we expected Look&MidAir to be faster than Look&Micro. However, although participants (P0, P1, P7, P8, P10) felt that Look&Micro was slower, our posthoc analysis did not reveal a significant difference between Look&MidAir and Look&Micro. Looking at Fig. 3, a difference in the disambiguation phase still seems lurking. As in our exploratory experiment, we found a significant difference, using t-test, between the two techniques for selection involving 3 or more thumb-to-finger taps. However, this represents 23% of the cases in scarce environments, and 65% of the cases in dense environments. Therefore, partly validating H_2: Look&MidAir is faster than Look&Micro only for the case of a long list of pre-selected objects.

ERRORS – Comparing errors between the two conducted studies, it seems that reducing the size of the cone in the second study leads to more cone errors, i.e. the target not being inside the cone. A smaller cone induces a smaller list of pre-selected objects and thus a shorter disambiguation phase. However, with a smaller cone, participants have to be more precise in the selection phase, which has an impact on the error rate. Therefore, if one wants to decrease the size of the cone to facilitate the disambiguation phase, they need to take into account the impact on the first phase.

FATIGUE – We expected Look&MidAir to suffer from the Gorilla arm effect thus inducing physical demand. However, based on the raw Nasa TLX results and the rankings, there is no clear difference between Look&Micro and Look&MidAir on physical demand. Thus, rejecting H_P. We believe that the arm movement in Look&MidAir is fast enough to not suffer from the gorilla arm, as opposed to raycasting pointing with the forearm. P10 explicitly said that Look&Micro induced fatigue in his index because both phases of the technique imply index gestures. Even though only one participant reported it, one needs to be careful when designing a technique with microgestures to not induce fatigue with repetitive movements. P6 reported eye fatigue due to prolonged use of the Hololens 2 and wished to take breaks between techniques.

MENTAL DEMAND AND LEARNING CURVE – There is no significant difference in the raw Nasa TLX results. However, Look&Micro is ranked higher than Look&MidAir for mental demand. We explain this higher mental demand by the fact that participants have to perform several microgestures that must be mentally planned. This higher mental demand for Look&Micro might have an impact on its time performance. Three of our participants (P0, P1, P5) explained that Look&MidAir is the easiest technique to learn because it is more intuitive to directly select a ball than navigate the list of balls. Therefore, it seems that microgestures require more training than direct touch. Thus, our hypothesis (H_M) is valid.

USERS' PREFERENCE – From the rank analysis, Look&MidAir seems to be slightly preferred to Look&Micro. Look&MidAir ranked first 7 times out of 12 and second 4 times. Look&Micro ranked first 5 times out of 12 and second 5 times. Therefore, even though there is no clearly preferred technique, both techniques are ranked better than raycasting, which is encouraging for future work on eye-gaze. An encouraging playful feeling was reported. For instance P5 and P9 preferred Look&Micro because it was "more fun".

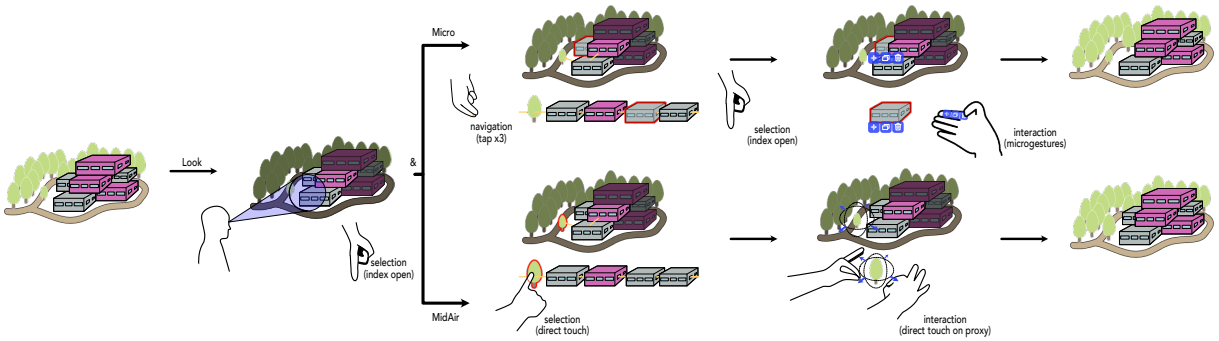


Figure 6: Aspen is an architect student refining her final project. She wants to remove a module from her building. Her gaze preselection cone intersects a tree and several modules, so she uses Look&Micro to select the desired module in the context of the 3D scene already created with several similar modules. She then uses the contextual menu to remove it. Around her building, Aspen placed a park and realized that she had placed a tree that was too small. She uses Look&MidAir to select this tree because she can easily spot it in the list in front of her. She then interacts with the tree proxy inside the list to resize it.

8 Limitations

Hardware limitations – We are using a glove with sensors to detect the microgestures when using eye-gaze, but we are using a camera to detect the HoloLens 2 select gesture (native option). Therefore, participants had to adapt to the hardware in order to get their gestures recognized, e.g. learn how to place their hand so that the HoloLens 2 can detect the gestures. Analysing participants' feedback, we observed that the hardware difference had an impact on preferences, since some participants were annoyed with the recognition errors from the camera. However, we decided to use HoloLens 2 raycasting as a baseline reference for future comparisons with our techniques, providing details about the scenes implemented in our experiments. Similarly, in our study, we used a glove for fast, cheap and reliable recognition. While wearing gloves seems unrealistic for everyday use, there is no less intrusive recognition hardware with similar recognition rates yet. Finally, the size of the cone is based on the error cone of the HoloLens 2 eye-tracker. Improved eye-tracking will further reduce the size of the cone. This will significantly reduce the size of the list of pre-selected objects during the disambiguation phase, thus improving the overall performance of our techniques.

Loss of spatial layout – When using the list in front of them, participants loose spatial and contextual information. With Look&MidAir, they can hover a target to find it in the scene. With Look&Micro, the list is also displayed in the scene, allowing disambiguation with spatial and contextual information. However, since our goal was to compare the modalities for 3D selection, we did not compare the performance of Look&Micro for disambiguation in the scene versus in the list. From our participants' comments, we can expect that if the environment is sparse, users will disambiguate in the scene, but in a dense environment, users will disambiguate in the list (P1, P3). P3 added that he preferred to disambiguate in the scene because looking at the list required head movements. A future study testing different scene configurations and information types [9] would provide deeper insight into what drives a user to look at the list rather than the scene.

Participants – Our participants were unfamiliar with mixed reality headsets and microgestures, which made it easier to compare user performance between standard and new techniques, since they were novices in both cases. But this may also have an impact on the generalizability of our work to experienced MR users. However, we believe that if experienced users can achieve better results with raycasting of the Hololens 2, they will get similar results for Look&MidAir and Look&Micro. Finally, our population is slightly skewed towards young men. Future studies based on our techniques could take into account a more gender-balanced and experienced population.

Fitts's Law – Fitts's Law is a commonly used model for 2D selection using the ISO-9241. However, there is no directly applicable extension from 2D (nor 2.5D) to 3D, and most importantly there is no consensus on a formula to use given the 6-DOFs as well as changes of apparent target size due to depth [47, 48]. Moreover, to the best of our knowledge, Fitts's models do not take into account occlusion or have 3D extension for a task with distractors / dense environments [3]. Since no Fitts's model would reflect our context (i.e. dense 3D arrangement), we used a setup similar to previous studies [2, 30] on dense 3D arrangement of targets with occlusion.

9 Conclusion and Future Work

In this paper we presented the design of Look&MidAir and Look&Micro, two 2-phase 3D selection techniques for Mixed Reality, using eye-gaze and index finger flexion during a first selection phase, and respectively direct touch or thumb-to-finger tap during a second disambiguation phase. During the first phase, users control a cone directed along the eye-gaze. By extending their index finger, they pre-select all the objects intersecting with the cone. In the second phase, using Look&MidAir, they directly touch the desired object. Using Look&Micro, they navigate the list of selected objects and select one of them by performing thumb-to-finger taps. A first exploratory study showed that eye-gaze combined with microgesture is twice faster and has a lower impact on physical fatigue than using the forearm to point. We also found that both direct mid-air touch and microgesture modalities to be promising for the second phase. In a second study, we compared Look&MidAir and Look&Micro to the Hololens 2 raycasting, using the forearm to point and a pinch hand gesture to validate. We found Look&MidAir and Look&Micro to be at least twice faster, with lower physical fatigue and mental demand than arm-based raycasting. While Look&MidAir was found faster than Look&Micro, they were equivalent in terms of physical fatigue and users' preference.

Future work – Finally, we discuss the possibility of combining the two techniques Look&MidAir and Look&Micro. In cases where disambiguation requires spatial context, Look&MidAir can be expected to be slower than Look&Micro because it requires back and forth between the list, to hover an object, and the scene, to see where the object is in the 3D scene. Moreover, previous studies [42, 43] have shown that microgestures can be used while grasping physical objects whereas this can be tedious with mid-air interaction. Therefore, since both techniques can work together, offering both as a combined technique would let users decide which technique to use depending on their current context. This combination foreshadows great flexibility for 3D selection according to the context of use (e.g. types of target objects, density of the environment, holding or not a physical object) and for object manipulation after selection, as illustrated in Fig. 6.

Acknowledgments

This work was partly supported by the French National Research Agency (ANR) project MIC (ANR-22-CE33-0017).

References

[1] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013. doi: 10.1016/j.cag.2012.12.003

[2] M. Baloup, T. Pietrzak, and G. Casiez. RayCursor: A 3D Pointing Facilitation Technique based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pp. 1–12. ACM Press, Glasgow, Scotland Uk, 2019. doi: 10.1145/3290605.3300331

[3] R. Blanch and M. Ortega. Benchmarking pointing techniques with distractors: Adding a density factor to fitts' pointing paradigm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 1629–1638. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1979180

[4] J. Blattgerste, P. Renner, and T. Pfeiffer. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3206343.3206349

[5] G. A. Borg. Psychophysical bases of perceived exertion. *Medicine and science in sports and exercise*, 14(5):377–381, 1982.

[6] A. Chaffangeon Caillet, A. Goguey, and L. Nigay. µglyph: a microgesture notation. In *In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. New York, NY, USA, To Be Published 2023. doi: 10.1145/3544548.3580693

[7] L. Chan, R.-H. Liang, M.-C. Tsai, K.-Y. Cheng, C.-H. Su, M. Y. Chen, W.-H. Cheng, and B.-Y. Chen. FingerPad: Private and subtle interaction using fingertips. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology - UIST '13*, pp. 255–260. ACM Press, St. Andrews, Scotland, United Kingdom, 2013. doi: 10.1145/2501988.2502016

[8] R. M. S. Clifford, N. M. B. Tuanquin, and R. W. Lindeman. Jedi forceextension: Telekinesis as a virtual reality interaction metaphor. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 239–240. IEEE, Los Angeles, CA, USA, 2017.

[9] W. Delamare, C. Coutrix, and L. Nigay. Designing disambiguation techniques for pointing in the physical world. In *Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '13, p. 197–206. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2494603.2480309

[10] S. A. Douglas, A. E. Kirkpatrick, and I. S. MacKenzie. Testing pointing device performance and user assessment with the iso 9241, part 9 standard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, p. 215–222. Association for Computing Machinery, New York, NY, USA, 1999. doi: 10.1145/302979.303042

[11] P. Dragicevic. HCI Statistics without p-values. Research Report RR-8738, Inria, June 2015.

[12] C. Endres, T. Schwartz, and C. A. Müller. Geremin": 2D microgestures for drivers based on electric field sensing. In *Proceedings of the 15th International Conference on Intelligent User Interfaces - IUI '11*, p. 327. ACM Press, Palo Alto, CA, USA, 2011. doi: 10.1145/1943403.1943457

[13] A. Esteves, D. Verweij, L. Suraiya, R. Islam, Y. Lee, and I. Oakley. Smoothmoves: Smooth pursuits head movements for augmented reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, p. 167–178. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3126594.3126616

[14] A. Forsberg, K. Herndon, and R. Zeleznik. Aperture based selection for immersive virtual environments. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology - UIST '96*, pp. 95–96. ACM Press, Seattle, Washington, United States, 1996. doi: 10.1145/237091.237105

[15] T. Grossman and R. Balakrishnan. The design and evaluation of selection techniques for 3d volumetric displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, p. 3–12. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1166253.1166257

[16] G. d. Haan, M. Koutek, and F. H. Post. IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. In E. Kjems and R. Blach, eds., *Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2005. doi: 10.2312/EGVE/IPT_EGVE2005/201-209

[17] J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard. A fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3206343.3206344

[18] F. Haque, M. Nancel, and D. Vogel. Myopoint: Pointing and clicking using forearm mounted electromyography and inertial motion sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, p. 3653–3656. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2702123.2702133

[19] R. Hauslschmid, B. Menrad, and A. Butz. Freehand vs. micro gestures in the car: Driving performance and user experience. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 159–160. IEEE, Arles, France, Mar. 2015. doi: 10.1109/3DUI.2015.7131749

[20] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 1063–1072. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2557130

[21] A. Huckauf and M. H. Urbina. Object selection in gaze controlled systems: What you don't look at is what you get. *ACM Trans. Appl. Percept.*, 8(2), Feb. 2011. doi: 10.1145/1870076.1870081

[22] R. J. K. Jacob. 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 Empowering People - CHI '90*, pp. 11–18. ACM Press, Seattle, Washington, United States, 1990. doi: 10.1145/97243.97246

[23] S. Jalaliniya, D. Mardanbeigi, T. Pederson, and D. W. Hansen. Head and Eye Movement as Pointing Modalities for Eyewear Computers. In *2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops*, pp. 50–53. IEEE, Zurich, Switzerland, June 2014. doi: 10.1109/BSN.Workshops.2014.14

[24] A. Jogeshwar, G. Diaz, S. Farnand, and J. Pelz. The cone model: Recognizing gaze uncertainty in virtual environments. vol. 32, 01 2020. doi: 10.2352/ISSN.2470-1173.2020.9.IQSP-288

[25] M. Khamis, C. Oechsner, F. Alt, and A. Bulling. Vrpursuits: Interaction in virtual reality using smooth pursuit eye movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, AVI '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3206505.3206522

[26] R. Kopper, F. Bacim, and D. A. Bowman. Rapid and accurate 3d selection by progressive refinement. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 67–74. IEEE, Singapore, Singapore, 2011.

[27] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. *Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality*, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2018.

[28] M. Leap. Magic leap 2 website, https://www.magicleap.com/magic-leap-2, 2023.

[29] J. Liang and M. Green. JDCAD: A highly interactive 3D modeling system. *Computers & Graphics*, 18(4):499–506, 1994. doi: 10.1016/0097-8493(94)90062-0

[30] Y. Lu, C. Yu, and Y. Shi. Investigating bubble mechanism for ray-casting to improve 3d target acquisition in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 35–43, 2020. doi: 10.1109/VR46266.2020.00021

[31] Meta. Oculus website, https://www.oculus.com/, 2023.

[32] Microsoft. Eye tracking on hololens 2, https://docs.microsoft.com/en-us/windows/mixed-reality/eye-tracking, 2023.

[33] Microsoft. Hololens 1 clicker, https://docs.microsoft.com/en-us/hololens/hololens1-clicker, 2023.

[34] Microsoft. Hololens website, https://www.microsoft.com/en-us/hololens/, 2023.

[35] Microsoft. Mrtk v2.3.0, https://github.com/microsoft/mixedrealitytoolkit-unity/releases/tag/v2.3.0, 2023.

[36] M. R. Mine. Virtual environment interaction techniques. Technical report, 1995.

[37] D. Miniotas, O. Špakov, I. Tugoy, and I. S. MacKenzie. Speech-augmented eye gaze interaction with small closely spaced targets. In *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, ETRA '06, p. 67–72. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1117309.1117345

[38] Oranj. Voxel algorithm, https://github.com/oranj/voxel, 2023.

[39] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 99–108. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132180

[40] Y. Y. Qian and R. J. Teather. The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17*, pp. 91–98. ACM Press, Brighton, United Kingdom, 2017. doi: 10.1145/3131277.3132182

[41] G. Schmidt, Y. Baillot, D. Brown, E. Tomlin, and J. Swan. Toward disambiguating multiple selections for frustum-based pointing. In *3D User Interfaces (3DUI'06)*, pp. 87–94, 2006. doi: 10.1109/VR.2006.133

[42] A. Sharma, M. A. Hedderich, D. Bhardwaj, B. Fruchard, J. McIntosh, A. S. Nittala, D. Klakow, D. Ashbrook, and J. Steimle. Solofinger: Robust microgestures while grasping everyday objects. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445197

[43] A. Sharma, J. S. Roo, and J. Steimle. Grasping Microgestures: Eliciting Single-hand Microgestures for Handheld Objects. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pp. 1–13. ACM Press, Glasgow, Scotland Uk, 2019. doi: 10.1145/3290605.3300632

[44] M. Soliman, F. Mueller, L. Hegemann, J. S. Roo, C. Theobalt, and J. Steimle. FingerInput: Capturing Expressive Single-Hand Thumb-to-Finger Microgestures. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, pp. 177–187. ACM, Tokyo Japan, Nov. 2018. doi: 10.1145/3279778.3279799

[45] S. Stellmach and R. Dachselt. Look & touch: Gaze-supported target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 2981–2990. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208709

[46] Y. Tan, S. H. Yoon, and K. Ramani. BikeGesture: User Elicitation and Performance of Micro Hand Gesture as Input for Cycling. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, pp. 2147–2154. ACM Press, Denver, Colorado, USA, 2017. ZSCC: 0000008. doi: 10.1145/3027063.3053075

[47] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie. Effects of tracking technology, latency, and spatial jitter on object movement. In *2009 IEEE Symposium on 3D User Interfaces*, pp. 43–50, 2009. doi: 10.1109/3DUI.2009.4811204

[48] E. Triantafyllidis and Z. Li. The challenges in modeling human performance in 3d space with fitts' law. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411763.3443442

[49] E. Tse, M. Hancock, and S. Greenberg. Speech-filtered bubble ray: Improving target acquisition on display walls. In *Proceedings of the 9th International Conference on Multimodal Interfaces*, ICMI '07, p. 307–314. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1322192.1322245

[50] Vive. Vive website, https://www.vive.com/, 2023.

[51] P. Wacker, O. Nowak, S. Voelker, and J. Borchers. ARPen: Mid-Air Object Manipulation Techniques for a Bimanual AR System with Pen & Smartphone. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pp. 1–12. ACM Press, Glasgow, Scotland Uk, 2019. doi: 10.1145/3290605.3300849

[52] J. Wambecke, A. Goguey, L. Nigay, L. Dargent, D. Hauret, S. Lafon, and J.-S. L. de Visme. M[eye]cro: Eye-gaze+microgestures for multitasking and interruptions. *Proc. ACM Hum.-Comput. Interact.*, 5(EICS), May 2021. doi: 10.1145/3461732

[53] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963

[54] K. Wolf, A. Naumann, M. Rohs, and J. Müller. A taxonomy of microinteractions: Defining microgestures based on ergonomic and scenario-dependent requirements. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2011*, pp. 559–575. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[55] A. L. Yarbus. *Eye movements and vision*. Springer, Boston, MA, USA, 2013.

[56] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, p. 246–253. Association for Computing Machinery, New York, NY, USA, 1999. doi: 10.1145/302979.303053

[57] X. Zhang and I. S. MacKenzie. Evaluating eye tracking with iso 9241 - part 9. In *Proceedings of the 12th International Conference on Human-Computer Interaction: Intelligent Multimodal Interaction Environments*, HCI'07, p. 779–788. Springer-Verlag, Berlin, Heidelberg, 2007.

[58] D. Zielasko, M. Krüger, B. Weyers, and T. Kuhlen. Passive haptic menus for desk-based and hmd-projected virtual reality. 03 2019. doi: 10.1109/WEVR.2019.8809589

[59] K. Özacar, J. D. Hincapié-Ramos, K. Takashima, and Y. Kitamura. 3D Selection Techniques for Mobile Augmented Reality Head-Mounted Displays. *Interacting with Computers*, 29(4):579–591, 12 2016. doi: 10.1093/iwc/iww035