AR Interfaces for Mid-Air 6-DoF Alignment: Ergonomics-Aware Design and Evaluation

Daniel Andersen* Voicu Popescu†

Purdue University

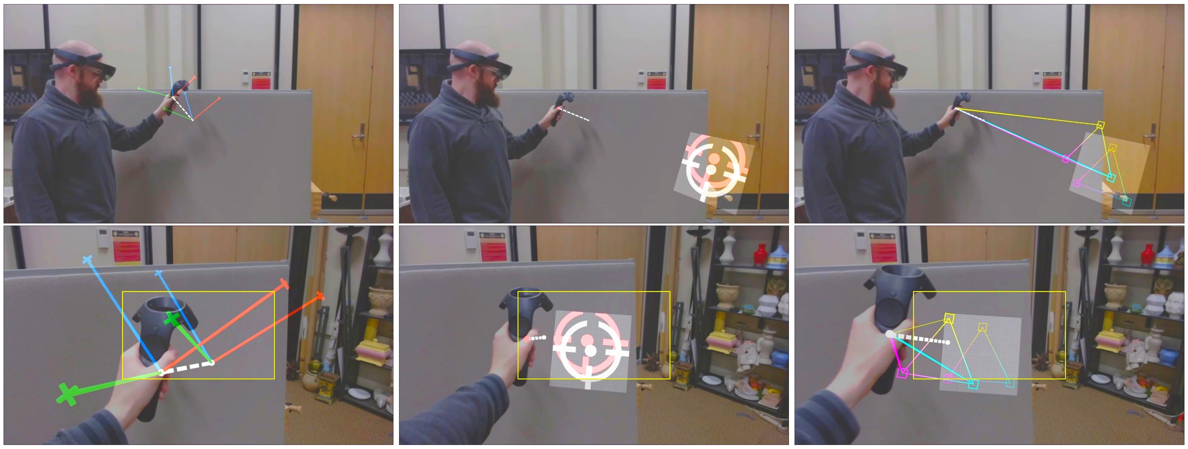

Figure 1: Third-person (top) and first-person (bottom) illustration of AR interfaces for six degrees-of-freedom mid-air alignment of a handheld object: AXES (left), PROJECTOR (middle), and PRONGS (right). The highlighted region represents the field of view of the augmented reality headset.

ABSTRACT

Aligning hand-held objects into mid-air positions and orientations is important for many applications. Task performance depends on speed and accuracy, and also on minimizing the user's physical exertion. Augmented reality head-mounted displays (AR HMDs) can guide users during mid-air alignments by tracking an object's pose and delivering visual instruction directly into the user's field of view (FoV). However, it is unclear which AR HMD interfaces are most effective for mid-air alignment guidance, and how the form factor of current AR HMD hardware (such as heaviness and low FoV) affects how users put themselves into tiring body poses during mid-air alignment. We defined a set of design requirements for mid-air alignment interfaces that target reduction of high-exertion body poses during alignment. We then designed, implemented, and tested several interfaces in a user study in which novice participants performed a sequence of mid-air alignments using each interface.

Results show that interfaces that rely on visual guidance located near the hand-held object reduce acquisition times and translation errors, while interfaces that involve aiming at a faraway virtual object reduce rotation errors. Users tend to avoid focus shifts and to position the head and arms to maximize how much AR visualization is contained within a single FoV without moving the head. We found that changing the size of visual elements affected how far out the user extends the arm, which affects torque forces. We also found that dynamically adjusting where visual guidance is placed relative to the mid-air pose can help keep the head level during alignment, which is important for distributing the weight of the AR HMD.

*e-mail: andersed@purdue.edu

†e-mail:popescu@purdue.edu

Index Terms: Human-centered computing—User studies; Computing methodologies—Mixed / augmented reality

1 INTRODUCTION

Quick and accurate placement of hand-held objects into mid-air poses is a key component of many real-world tasks. For example, positioning hand-held cameras into specific positions and rotations is important for visual comparison against a historical photograph, for workspace inspection, or for high-quality 3D reconstruction of real-world scenes. Other hand-held tools that function at a physical distance from a target (such as lights being positioned on a film set, spray cleaners for disinfecting surfaces, fire extinguishers, or firearms) require the user to efficiently hold the object in a certain pose in mid-air for the tool's purpose to be achieved. Even tools whose function involves physical contact with a nearby surface (such as surgical instruments interacting with a patient's body) often require considering the tool's alignment prior to making any contact with the surface [12]. The success of these tasks depends on the user's ability to align hand-held objects quickly, accurately, and with minimal physical exertion. Task performance could benefit from real-time guidance about how to align objects in mid-air, either during training or when performing the task itself.

Augmented reality head-mounted displays (AR HMDs), when combined with real-time tracking of hand-held objects, are a promising method of delivering alignment guidance directly into a user's field of view. Modern AR HMDs use on-board sensors to track and analyze the user's surroundings, and the headset's display lets the user see the real world, including any hand-held objects. The headset supports rendering virtual content in 3D at the appropriate depth due to its stereo display. As opposed to hand-held AR on a phone or tablet, the imagery rendered on an AR HMD is always available to the user, which frees up the user's hands to manipulate a hand-held object into mid-air alignment.

What is needed is an AR HMD interface that empowers a user to perform mid-air alignments of a hand-held object with speed, accuracy, and low amounts of physical exertion. However, while prior research has investigated methods of visualizing user guidance via AR, it has largely focused on alignment of tools that contact physical surfaces, and it has not explored the specific ergonomic factors that are intrinsic to mid-air alignment and that can be affected by the current hardware of AR HMDs.

Mid-air alignment represents a greater challenge than alignment against a physical surface. Visual landmarks on a surface assist a user in localizing a pose, haptic feedback from the surface lets a user hold certain degrees of freedom fixed while manipulating other degrees of freedom, and the surface's support of the weight of the object and the user's arm can enable superior motor skills. In contrast, mid-air alignment offers no direct visual landmarks for reference, the user must complete the alignment in a single sustained arm motion, and the user is responsible for supporting the weight of the object and the arm.

The form factor of current AR HMDs may influence physical exertion during mid-air alignment tasks. The headset's weight is tiring when tilting the head for sustained periods of time. The low field of view (FoV) means that the user may only see a small part of the visuals, or may only be able to assume a restricted set of body poses to adequately see the visualization. Holding a hand-held object in a mid-air pose for a sustained period of time increases exertion, especially if the arms are extended far from the body (which increases torque forces). However, it is an open question how different AR HMD alignment interfaces will affect user motion, and whether or not certain approaches will cause users to assume high-exertion body poses.

In this work, we present our investigation into the research question of AR HMD guidance for hand-held mid-air alignment. We first identify a set of design requirements that enumerate the criteria for a suitable interface for mid-air alignment, and describe a design space from which we sample several candidate interfaces. Finally, we present the results of a user study where participants wore an AR HMD and used each interface to complete a set of mid-air alignment tasks with a tracked hand-held object.

Our study revealed the preferences of users during mid-air alignment tasks with AR HMDs. Interfaces that place all guidance visualization near the hand-held object lead to shorter alignment times and reduced translation errors, while interfaces that involve aiming at distantly-placed virtual objects lead to reduced rotation errors. Users will tend to position themselves to maximize the amount of visual guidance that is visible at once on the FoV of the AR HMD. We discovered that adjustments in the size and placement of visual elements can significantly alter how the user positions their head and arms into tiring poses during the alignment process. Our findings also demonstrate the importance of finding a balance between providing enough visual information to perform alignment accurately while also avoiding overloading the user's visual field.

2 Prior Work

Here, we discuss prior research about 6-DoF alignment in general, and we examine different approaches of AR alignment guidance.

2.1 6-DoF Alignment

The task we seek to address in our work can be categorized as isomorphic 6-DoF docking, where an object is placed into a particular position and orientation using guidance with 1-to-1 correspondence [33, 56]. Krause et al. noted that interfaces that overlay a physical hand-held object with virtual/augmented imagery can lead to faster docking completion [29].

Prior work found that a user's proprioception is an important factor in performing movement and docking tasks in a virtual environment, and that aligning virtual objects that appear in the user's hand is faster than aligning virtual objects at some offset from the user's hands [28,38]. All interfaces we examine in our work use this knowledge and represent target poses as mid-air points with which the user should align their grip center. Users also tend to implicitly separate 6-DoF docking tasks into a 3-DoF positioning sub-task and a 3-DoF orientation sub-task, and to alternate between these sub-tasks during refinement [36]. As mid-air 6-DoF alignment does not afford the user a physical surface to hold fixed certain degrees of freedom while refining other degrees of freedom, this further motivates our work to find visual interfaces that help maintain accuracy in positioning while refining orientation, and vice versa. Several of our alignment approaches are designed to help reveal if, and to what extent, the visualization can affect the user's separation of the alignment task into these sub-tasks. Prior work also found that the aligned object's silhouette is important during alignment, and that visualizations that emphasize the virtual object's silhouette may lead to improved alignment performance [35]. We draw upon these lessons in our interface design, and extend it by investigating the interaction not just between an object's visible contour and how it is shaped or rendered, but also how far away the object is held and how much of the object is visible at a time on a low-FoV AR display.

Unlike prior work about virtual 6-DoF docking tasks, we focus on alignment of a physical object in a specific physical pose rather than manipulation of a virtual object in a virtual environment. Because a virtual object has no weight, the user can use a "clutch" feature to release the virtual object and keep it floating before proceeding with fine-tuning. This differs from aligning a physical object in mid-air, which must be moved in one smooth, rapid action to prevent arm fatigue from sustained manipulation [51]. Purely virtual manipulation also allows for special "fine-tuning" modes where large physical movements on an input device correspond to much finer virtual movements [35, 48, 49], or input amplification where small physical movements correspond to large virtual motions [53]. Purely virtual interfaces allow for constraining certain degrees of freedom or automatically snapping approximate alignments into place [41,42]. None of these approaches are feasible for our ultimate task of alignment of a physical hand-held object in mid-air.

2.2 Augmented Reality for Alignment Guidance

There is a long history of using AR for guidance for motion of held objects. For example, an early work by White et al. overlaid visual hints onto a tracked handheld object, though these were large-scale gestural movements rather than precision alignment of position/orientation [54]. Sukan et al. investigated interfaces for 3-DoF orientation alignment for monoscopic AR displays [46, 47], and Hartl et al. presented an approach for 3-DoF AR-guided capture of a planar surface from multiple viewing angles [18].

In this work, we focus on the use of AR for alignment guidance, rather than VR. We do this because we see more use cases for isomorphic physical alignment in AR rather than in VR (where alignment of purely virtual objects is more compelling). Prior work suggests an improvement in task completion time for alignment tasks in AR versus VR [30]. We also focus on headset-based AR rather than AR on a hand-held screen [52]. Using a smartphone or tablet located at the hand-held object may be suitable for the alignment task, but would require different visualization choices: imagery would be limited to the screen only, the type of hand-held object would be limited to those that support attachment of a physical screen, and the range of alignment poses would be limited to those where the user could directly view the screen.

We distinguish our work from the related but separate issue of *information localization* in AR HMDs, where the goal is to help a user discover virtual content placed outside the user's FoV. Various cues have been used to inform a user of out-of-view content and prompt them to look toward the content's location [10,16,34]. While any mid-air alignment task requires the user to first discover the pose in the scene, we instead focus on the interaction between when a user first sees a mid-air pose, and when the user has achieved alignment with the pose. Consequently, our alignment interfaces are all implemented with additional contextual guides to help the user orient themselves toward the location of the guidance.

One use case for mid-air alignment guidance in AR is for rephotography, where a photograph is captured from the same pose as a previous photo, for the purposes of inspection or historical comparison. Bae et al. presented a non-self-contained approach for interactive camera guidance and alignment by showing arrow icons on a nearby monitor [4]. Shingu et al. demonstrated using hand-held AR to align a camera into a virtual cone rendered relative to a tracked surface; however, this approach only offered 5-degree-of-freedom alignment [45]. A similar work by Sukan [46] sought to guide a user wearing an AR headset to assume a particular head pose; however, this is the reverse of our task where the goal is to place a hand-held object in a certain pose, while allowing the user's head and body to assume whatever pose is comfortable.

Another application of mid-air AR guided alignment is for 3D acquisition of real-world scenes. Andersen et al. presented an approach in which a tracked hand-held camera was superimposed with AR HMD guidance to help a user align the camera with a set of mid-air views to achieve complete scene coverage [2]. The authors' user study suggested that their specific AR guidance unintentionally encouraged holding the arm further from the body than was necessary or ergonomic. This motivates our own work on finding interfaces that encourage users to keep hand-held objects in ergonomic poses.

AR guidance is also useful in the area of improving interaction with occluded objects. Lilija et al. investigated different methods for improving tactile interaction with occluded objects while wearing an AR HMD [32]. Some of our interfaces also allow for alignment without directly looking at one's hand; however, our context is not in overcoming physical occlusion but in avoiding unnecessary forces on the head from viewing a mid-air pose above or below the user's head while wearing a heavy AR HMD.

Some AR alignment applications involve physical contact between a hand-held tool and a surface (e.g., for surgical needle injection) [22]. However, these tasks allow the user to cleanly split the 6-DoF task into a placement phase and an orientation phase. Purely mid-air 6-DoF alignment does not allow for pure separation of these two phases and so novel forms of visualization are needed. Another aspect of alignment involving physical surfaces is that they allow for the use of projector-based AR [19–21]. We use a similar concept of projection in some of our interfaces by projecting a virtual light from the hand-held object onto a virtual plane floating in space.

One prior interface is the "virtual mirror," which helps prevent tools from colliding with nearby structures [7, 40]. However, such interfaces rely on rendering virtual reflections of objects that are near the instrument, which is less applicable to mid-air alignment in open space. Another approach uses animations to show depth [11, 21]. However, this requires the user to wait while the animation plays to receive guidance, which can increase physical exertion.

Some prior alignment interfaces communicate a threshold of "good enough" alignment to the user, either explicitly into interface design with elements such as color-coding, or implicitly in a user study design that requires participants to reach a certain level of alignment before proceeding [14, 15, 19, 20, 26]. Because such thresholds are use-case-dependent and require a prior definition of a suitable threshold value, we exclude such approaches in our work which investigates the generic mid-air alignment problem in AR, independent of specific use-case-dependent criteria.

3 AR Alignment Interface Design

In this section, we formally define the 6-DoF alignment task (Sect. 3.1), we enumerate a list of design requirements to prioritize to satisfy the alignment task (Sect. 3.2), and we describe a possible design space from which to sample a set of interfaces to be evaluated in our user study (Sect. 3.3).

3.1 Definition of Task

The task we focus on is for a user to place a physical hand-held object into a specified mid-air position and orientation. We assume that the user may need to perform a sequence of dozens or even hundreds of these alignments. We also assume that the user is standing in an open space without nearby physical surfaces against which to rest.

To assist the user in the alignment task, we assume that the user wears an AR HMD that can superimpose stereo graphics onto the user's view. In our interface design, we take into account the FoV of modern AR HMDs, which tends to be lower than that of VR HMDs. We also assume that the poses of the AR HMD and of the hand-held object are tracked in real-time.

3.2 Design Requirements

Based on the alignment task we have defined, and the equipment and tools assumed to be available to the user, we next define a set of design requirements against which a potential AR alignment interface can be evaluated. We first note the following general objectives for an alignment interface:

(R1) *Low translation and rotation error*: In our work we measure both translation and rotation error; the threshold for "good enough" alignment accuracy is use-case-dependent. The minimum amount of error is also bounded by tracking accuracy and by how well a user can steady the movements of an outstretched arm.

(R2) *Short alignment time*: Whether a specific alignment time is "fast enough" is use-case-dependent. Alignment time is especially important for mid-air alignment because sustained time with an outstretched arm may lead to fatigue and may adversely impact accuracy in future alignments.

(R3) *Minimal visual clutter*: AR HMDs have a low FoV, and an alignment interface that takes up more screen real-estate will leave little room for any other visuals in an AR application.

(R4) *Low physical strain*: Fatigue in various parts of the body adversely affects user performance. The ability of an interface to increase or decrease physical strain is limited, because the user must in all cases move their hand into the same mid-air pose. However, the interface may affect how the user positions their arms, hands, wrist, and neck. also, eye fatigue may be affected by the complexity of the interface as well as the distance from the user at which the visuals are rendered (due to the fixed focal length of many AR headsets).

Because a user's physical strain is difficult to quantify directly, we also determine specific high-exertion body poses that the user should be discouraged from assuming, and encourage the following types of body poses during alignment:

(R4.1) *Keep the arm close to the body*, to avoid the fatiguing situation of *"gorilla arm"* [1, 9, 44]. Specifically, the horizontal distance between the hand and the head/body should be kept small, so that torque forces on the arm are reduced.

(R4.2) *Have low levels of head pitch*, so that the weight distribution of a heavy AR HMD does not cause neck fatigue by a user needing to look up and down for sustained amounts of time.

(R4.3) *Avoid lining up the head relative to the hand-held object*. If an interface requires the user to view the object from a small range of viewing angles during alignment, then that means that the user's body motions will need to be constrained and encumbered in order to achieve those limited viewing angles.

3.3 Design Space of Alignment Interfaces

An AR alignment interface visualizes the current pose, the desired pose, and (optionally) the spatial relation between the two. There are many options for each of these visualizations, and so the design space is extremely large. From initial observations, including referencing prior literature about 6-DoF docking tasks, we defined a design space in terms of the following dimensions: *redundancy*, *alignment strategy*, and *sensitivity*.

Redundancy We were inspired by the observation that alignment visualizations differ in the number of redundant elements that are presented to the user, and that even simple representations of coordinate systems in many 3D applications usually contain some redundant elements (such as the arrow tips on XYZ axes in 3D modeling software). *Redundancy* describes how many visual elements are available to inform the user of the alignment or misalignment of the hand-held object in each degree of freedom. On one hand, having additional visual elements to convey the translation/rotation alignment may help the user to align more accurately or more ergonomically, especially when some visual elements might be outside the FoV of the AR HMD. On the other hand, redundancy may introduce visual clutter and impose cognitive load, especially in the case of self-occlusion from visual elements overlapping other visual elements.

Alignment strategy We also noted prior literature on 6-DoF docking, which found that users tend to implicitly decompose the 6-DoF into two 3-DoF (translation and rotation) tasks, between which the user alternates during refinement [36]. This raised the question of whether an alignment interface would be more suitable if it explicitly presented the 6-DoF task to the user in its decomposed form. *Alignment strategy* describes how the interface guides or imposes a particular method of alignment onto the user. One option is for an alignment interface to decouple the degrees of freedom during alignment by guiding the user to perform translation alignment first and then to proceed with rotation alignment. Another option is for the interface to present both rotation and translation guidance together and allow the user to focus on them simultaneously.

Sensitivity Finally, the literature on 6-DoF docking found that users tended to be less accurate in the dimension parallel to the user's viewing direction: positioning a virtual object closer or further away was more challenging than adjusting it horizontally or vertically relative to the user's viewpoint. This highlighted the importance of considering an alignment interface in context of the user's perceived visual change during alignment in each degree of freedom. *Sensitivity* describes how much visual change results from a change in translation or rotation alignment. To easily fine-tune alignment in different degrees of freedom, the user should receive salient visual feedback as a result of small adjustments in those degrees of freedom. Sensitivity depends not just on what visual elements make up an interface, but also on how the visual elements are oriented relative to the user's viewpoint, and on how much of the visual elements are currently within the user's FoV.

4 AR ALIGNMENT INTERFACES INVESTIGATED

Here, we describe each interface we tested: AXES (which we tested as two variations named SHORTAXES and LONGAXES), PROJECTOR, and PRONGS (which we tested as two variations named STATICPRONGS and DYNAMICPRONGS). For each of these five main interfaces, we also tested a version with an optional RADIUSBUBBLE visualization added to each interface, yielding a total of ten interfaces. We detail how each interface addresses our design requirements and where each interface falls within our design space.

We refer the reader to the accompanying video, which further illustrates each interface in use. The illustrations shown in this paper and in the video were generated via a mixed-reality overlay of a video camera that was enhanced with a 6-DoF VR tracker.

4.1 AXES Interfaces

Here we describe the two AXES interfaces we tested (labeled SHORTAXES and LONGAXES). This interface type acts as a familiar base-case for 6-DoF alignment (Fig. 1, left column). It consists of orthogonal red/green/blue-colored cylindrical axes affixed to the

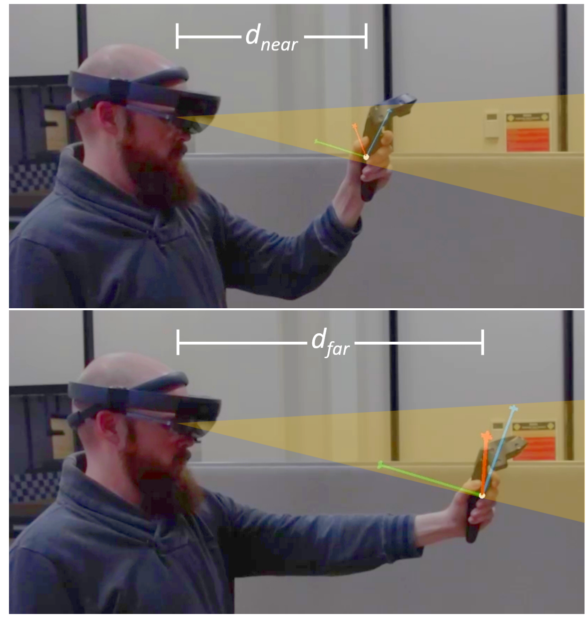

Figure 2: Comparison of our two variations of the AXES interface. Top: SHORTAXES is sized to fill the FoV (yellow cone) when at a distance d_{near} from the user. Bottom: LONGAXES is sized to fill the FoV when at a distance d_{far} from the user.

alignment point on the hand-held object. To indicate the tip of each axis is an additional cross-shape pattern. Even with this basic design, there are many possible variations in how the axes are displayed that might affect performance in an alignment task. One variation we explore is the length of the axes and its possible effect on user ergonomics during alignment.

Our design requirements establish that an effective interface should let the user keep their arm close to the body rather than unnecessarily stretching it out **(R4.1)**. If the axes are too large to be fully seen with the low FoV of the AR HMD, then the user may compensate by stretching out the arm further from the body, or by rotating the head to see the entire visualization.

We define two variations of the AXES interface (labeled LONGAXES and SHORTAXES). At runtime, we determine two distances d_{far} and d_{near}, where d_{far} is the head-to-hand distance (projected onto the floor plane) when the user's arm is fully stretched out in front, and where d_{near} is the head-to-hand distance when the user's arm is at a comfortable hinged angle in front of the user (Fig. 2). We set the axes' lengths so that LONGAXES fills the FoV of the AR HMD when held d_{far} from the head, and so that SHORTAXES fills the FoV of the AR HMD when held d_{near} from the head.

The AXES interfaces represent a moderate amount of visualization redundancy. In principle, all that is needed to disambiguate a 6-DoF pose would be two color-coded line segments (representing three points in space). Translation and rotation redundancy is achieved by each of the three axes and the axes end points. The orthogonal axes offer a large visual change for at least two of the three axes, while the axis that is most parallel to the user's view direction will offer less guidance. In terms of alignment strategy, the AXES interfaces place all visuals in the same area, which encourages simultaneous alignment of translation and rotation without focusing on different visuals in the scene. The sensitivity of these interfaces depends on how the three orthogonal axes are oriented relative to the user: two axes could appear tangent to the viewing direction while the third is parallel to the viewing direction, or the length of all three axes could be simultaneously visible.

4.2 Projector Interface

Projector emphasizes rotational alignment with a virtual flashlight metaphor (Fig. 1, middle column). The position of the mid-air pose is represented with a floating sphere, and a virtual plane is rendered at a fixed offset from the position of the mid-air pose.

The virtual plane is positioned so that, when the user is standing on the opposite side of the plane at a distance d_{near} from the mid-air pose, the plane is approximately 2m away, which matches the focal distance of our AR HMD, with the goal of reduced eye fatigue **(R4)**. The size of the virtual plane is adjusted so that, from this viewing distance of 2m, the plane is entirely contained within the vertical FoV of the AR HMD. On the plane is a gray, rotationally asymmetric target pattern. A matching red target pattern is emitted from a fixed orientation relative to the hand-held object using projective texture mapping, such that when the object is in the correct 6-DoF pose, the red target pattern from the controller lines up with the gray target pattern on the virtual plane.

This interface represents a low level of redundancy, as well as an alignment strategy that strongly decouples the translation and alignment tasks. Only the floating sphere at the mid-air pose location clearly indicates translational alignment, and only the patterned virtual plane indicates rotational alignment. If the user observes only one of these interface components at a time, then the user only has partial knowledge of the current 6-DoF alignment state. Projector shows strong visual cues when translating horizontally/vertically relative to the view direction, but weak cues when translating parallel to the view direction. It shows strong visual cues for rotational alignment, but only when the rotation is close enough to have the virtual flashlight projecting onto the plane. The visual sensitivity of this interface is highest when the user's viewpoint simultaneously includes both the sphere representing the target position and the target plane representing the target orientation.

4.3 Prongs Interfaces

Here we describe the two Prongs interfaces we tested (labeled StaticProngs and DynamicProngs). This interface type represents a combination of the principles behind the Axes interfaces and the Projector interface (Fig. 1, right column). A virtual plane is rendered relative to the mid-air pose, and is marked with a colored triangle pattern. An elongated wireframe tetrahedron is rendered relative to the user's hand-held object, so when the hand-held object is aligned with the mid-air pose, the base of the pyramid aligns with the triangle pattern on the plane.

One of our design requirements is that an alignment interface should allow for low levels of head pitch **(R4.2)**. Sustained periods of looking up or down with a heavy AR HMD can place additional physical strain on the user's neck. The user's head could be kept level regardless of a mid-air pose's position or orientation if all necessary visual instruction was rendered at head height.

To investigate this, we define two variations of Prongs, which we label StaticProngs and DynamicProngs (Fig. 3). These two interfaces vary in how the target plane is positioned relative to the mid-air pose, and correspondingly how the tetrahedron on the hand-held object is oriented.

In StaticProngs, the target plane is placed at a fixed position and orientation relative to the mid-air pose, just as it is in Projector. Similarly, the tetrahedron that emerges from the hand-held object always has the same orientation relative to the hand-held object. This interface therefore requires that, depending on the orientation of the mid-air pose, the user may have to tilt their head up or down to view the target plane.

In DynamicProngs, the target plane's position relative to the mid-air pose is altered such that the target plane appears at the user's head height. Given a user height h from the headset to the floor, we rotate the rendering of the mid-air pose so the target plane's center is at height h. We correspondingly rotate the rendering of the tetrahedron on the user's hand-held object by the same angle, so that when the triangle patterns are lined up, the physical orientation of the mid-air pose matches the original target mid-air pose orientation. The DynamicProngs interface ensures that the user does not need to tilt the head up or down to view the target plane **(R4.2)**.

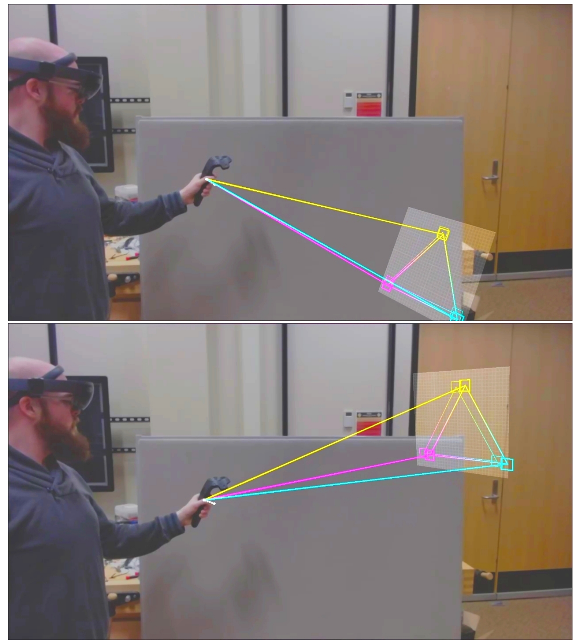

Figure 3: Comparison of StaticProngs and DynamicProngs variations, given the same indicated physical 6-DoF pose for the user's controller. Top: in StaticProngs, the target plane is always positioned at a consistent orientation relative to the target mid-air pose. Bottom: in DynamicProngs, the target plane is adjusted to always be at the user's standing head height.

Within our design space, the Prongs interfaces have high visual redundancy. The entire perimeter of the triangle pattern represents alignment information to the user, and the gradual color change along each line gives finer detail of alignment than the solid colors of the Axes interface. The Z-buffering of the wireframe tetrahedron against the virtual plane also illustrates, for each point on the triangle pattern, whether it is in front of or behind the virtual plane. Because the triangle patterns to be aligned are 3D objects rather than just a projection, there is a stronger visual cue when the hand-held object is moved towards the user and away from the user. However, because the target plane is far from the user, the visual cues are weaker than in the Axes interfaces. This interface also separates the rotation and translation alignment tasks, like Projector does. As with Projector, this interface's visual sensitivity is highest when all visual elements appear simultaneously within the user's FoV.

4.4 RadiusBubble Visualization Addition

An additional visualization, RadiusBubble, can be added to each previously described interface (the Axes interfaces, the Projector interface, and the Prongs interfaces). A dynamically resizing semi-transparent sphere is added to any alignment points in the interface. The sphere's radius equals the distance between the current position of the alignment point on the user's hand-held object and the position of the alignment point on the indicated mid-air pose (Fig. 4). For the Prongs interfaces, we also added a color-coded RadiusBubble to each point of the alignment triangle.

The presence of the RadiusBubble adds redundancy and sensitivity by providing an additional visual cue for translation alignment and by offering a strong visual change as the user's translational

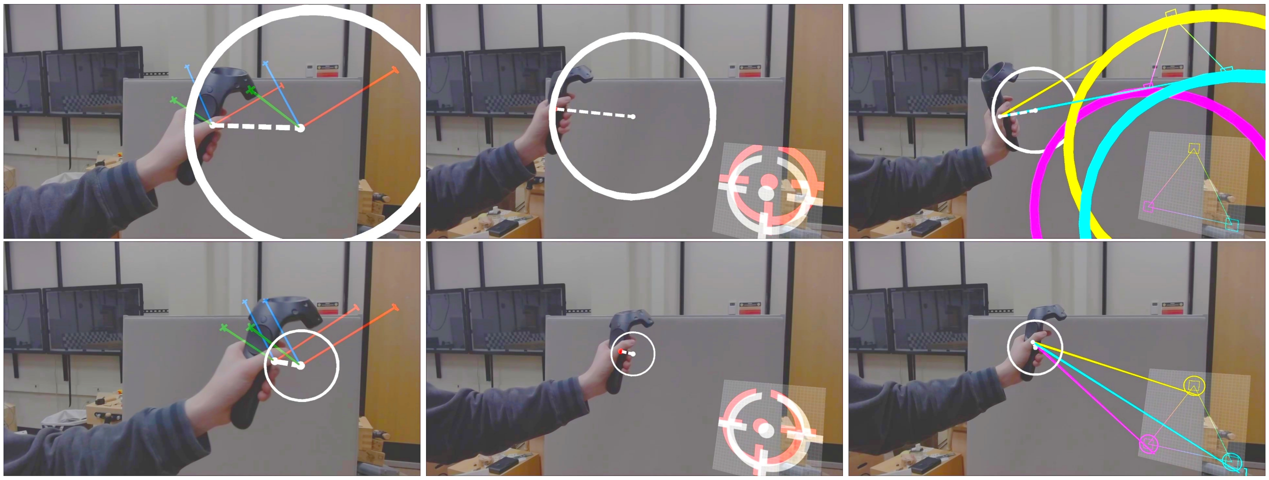

Figure 4: Addition of the RADIUSBUBBLE visualization to AXES (left column), PROJECTOR (middle column), and PRONGS (right column). The top row illustrates when the alignment distance is large, and the bottom row illustrates when the alignment distance is small.

alignment changes, especially in the user's view direction which would otherwise have weak visual cues. However, this addition comes at the cost of additional visual clutter **(R3)**.

5 INTERFACE EVALUATION USER STUDY

To determine which specific interface factors impacted user performance and ergonomics, we conducted a within-subjects user study in which 23 participants who were first-time users of our interfaces tested each of our 10 interfaces to perform a set of mid-air 6-DoF alignments. In this section, we describe the study task, the variables evaluated during the study, and the experimental procedure.

5.1 Participants

We recruited 23 unpaid participants from our university (17 male, 6 female; age: 28.3 ± 6.3)[1]. 21 participants were right-handed and 2 were left-handed. On a 5-point Likert scale, participants self-reported much experience with video games (4.2 ± 1.0), and moderate experience with VR (3.1 ± 0.9) and AR (2.5 ± 0.9).

5.2 Task

Participants wore an AR HMD and held a tracked 6-DoF controller. For each interface, participants performed a sequence of 12 mid-air alignments while standing. Given a mid-air pose to be aligned, participants were asked to place the controller so the indicated alignment point on the controller appeared aligned with the indicated mid-air pose, and then to press the controller trigger. The participants were asked to perform the task both quickly and accurately according to their own judgment. After each trigger press, the next mid-air pose was presented, and the participant proceeded with the next alignment until all 12 poses in the sequence had been aligned. Participants were encouraged to move their arms and head and to walk around and modify their body poses in whatever way felt appropriate.

The set of 12 poses to be aligned were identical for all interfaces. To prevent memorization, the order in which the poses were presented was randomized each time. The pose positions were randomly generated within a 3D volume centered in front of the user and sized approximately $1\text{m} \times 0.5\text{m} \times 0.05\text{m}$. To account for variations in participants' body-types, the position and size of this 3D volume was scaled based on participants' height and armspan. The pose rotations were generated by rotating between 15° and 45° about a random axis from a neutral up-oriented rotation of the controller.

[1]Due to concerns about the COVID-19 pandemic near the end of the data-gathering period, we halted testing of the last planned participant.

5.3 Metrics

As is common for virtual environment interaction studies, we collect performance metrics of time and accuracy [43]. We also collect metrics that encode the ergonomics of participants' body poses.

Alignment time We measure the time between when the mid-air pose first appears within the FoV of the participant's AR HMD and when the participant has confirmed the alignment.

Translation and rotation error We measure the distance and angle between the controller's pose and the indicated pose.

Horizontal headset-to-controller distance At the moment of alignment, the vector from the participant's HMD to the controller is projected onto the floor plane and its magnitude (in meters) is recorded. We use this as a proxy for the torque forces on the participant's outstretched arm during alignment.

Headset pitch At the moment of alignment, we measure the magnitude of the pitch angle of the participant's headset.

Headset lineup concentration We also measure how much the user lines up their head in a specific direction relative to the controller during alignment. At each alignment, we project the headset's position onto a unit sphere in the controller's local coordinate frame. Over the course of the 12 alignments in a single trial, these projected points form a cluster of viewing directions. For each trial, we fit the 12 viewing directions to a von Mises-Fisher distribution to get a mean direction μ and a concentration parameter κ [5]. (The von Mises-Fisher distribution can be thought of as a spherical analogue to the normal distribution, with κ analogous to $\frac{1}{\sigma^2}$). We take κ to be the headset lineup concentration; the higher the value, the more tightly-clustered the viewing direction is for that trial.

Questionnaire We also measure the participants' perceived workload and self-reported fatigue in different parts of the body, as recorded by post-session questionnaires consisting of a NASA-TLX [17] and a Borg CR-10 scale [8]. We provide additional detail about the questionnaires in Sect. 5.5.

5.4 Conditions

We set up our study as a test of two factors, defined as *Interface* and *RadiusBubble*. The *Interface* variable had five levels: SHORTAXES, LONGAXES, PROJECTOR, STATICPRONGS, and DYNAMICPRONGS. The *RadiusBubble* variable had two levels, referring to either the absence of or the addition of the RADIUSBUBBLE visualization to the different interfaces (see Sect. 4.4).

5.5 Experimental Procedure

Each participant's study session lasted about 60 minutes. After filling out a demographics questionnaire, the participant put on the AR HMD. Our application then recorded the standing height h of the participant, and the horizontal headset-to-controller distances d_{far} and d_{near} when the arm was first fully outstretched and then when the arm was in a comfortable hinged position. These distances were used to adjust the sizing of various elements of the interfaces, as described throughout Sect. 4.

For each interface being tested (LONGAXES, SHORTAXES, PROJECTOR, STATICPRONGS, and DYNAMICPRONGS), participants completed two trials: one where the RADIUSBUBBLE visualization was included with the interface, and one where there was no RADIUSBUBBLE. Each trial was composed of 12 mid-air poses, for a total of $5 \times 2 \times 12 = 120$ alignments for each participant. To avoid order effects, each participant tested each of the $5 \times 2 = 10$ interface permutations in counterbalanced order.

Before each trial, the participant was taught via instruction and practice how to use the interface for that trial. The participant was instructed to perform the alignments both quickly and accurately, according to the participant's own judgment; the participant was also informed that prolonged time with an extended arm may lead to fatigue and reduced accuracy. During each trial, the poses of the headset and controller were logged every 50ms for later analysis.

After the participant completed the trials for one of the five interfaces, they removed the headset and completed a questionnaire which consisted of a NASA-TLX [17] and questions on a Borg CR-10 scale (with text anchor labels) about the participant's fatigue in different parts of the body: head, eyes, neck, arms, wrist, and back [8]. At the end of the questionnaire was an optional section where participants could write general comments and feedback about the interface that was tested. The participant then took a short rest before continuing by putting on the headset again and completing the next two trials for the next interface. This was repeated until all trials were finished.

5.6 Method

To evaluate the alignment interfaces, we implemented each in a Unity application [50] [2]. We deployed the application on a testbed prototype system which wirelessly networked an AR HMD with a VR system for 6-DoF tracking of a hand-held controller. For the AR HMD, we used a Microsoft HoloLens 1 (FoV = 30° x 17.5°; focal distance = 2m) [37]. One possible approach for tracking a 6-DoF hand-held controller relative to the AR HMD would be to use the HoloLens' RGB camera to track a fiducial marker [2, 3]. However, in practice, we found tracking from this camera was insufficiently robust, especially when the user was not directly looking at the controller. Instead, we tracked a VR controller via an HTC Vive connected to a desktop computer [24]. The controller's 6-DoF pose was transmitted wirelessly over a local network to the HoloLens [39]. The coordinate systems of the HoloLens and of the Vive were merged by a one-time calibration step in which a HoloLens world anchor was placed at the origin of the Vive coordinate system. The full latency of our testbed system was tested with a high-speed camera pointed through the HoloLens' display; the latency between physical motion of the hand-held controller and an update of the controller's pose in the HoloLens was 42ms. For all tested interfaces, the hardware and tracking latency were consistent, and all visualizations ran at the HoloLens' maximum framerate of 60fps.

For all interfaces, we place the alignment target point at the grip center of the user's controller. In informal pilot trials, we found this helped the user more easily manipulate controller orientation while keeping controller position relatively fixed. Prior work also found improved results from aligning a virtual object attached to the hand rather than attached to an offset relative to the hand [38].

As mentioned in Sect. 2, this work focuses not on discovery of out-of-FoV AR content, but on performing alignment given already-discovered AR content. For this reason, we include for all interfaces some additional guidance to help the user locate the mid-air pose. First, we render a dashed line from the target point on the controller to the position of the mid-air pose (Fig. 4). Second, in the PROJECTOR and PRONGS interfaces, when the target plane is outside the user's FoV, we render a screen-space line pointing from the center of the user's view toward the target plane's location.

[2] Source code for our application can be found at `https://github.com/DanAndersen/ARMidAirAlignment`.

6 RESULTS AND DISCUSSION

In this section, we describe the statistical analysis of our user study, we report our results, and we discuss how these results shed light on our design space in the context of our design requirements.

6.1 Statistical Analysis

For each per-alignment metric, we took the median value across all alignments per trial and tested for within-subjects effects with a two-way repeated measures ANOVA, where the two factors were *Interface* (5 levels) and *RadiusBubble* (2 levels). We used the Shapiro-Wilk test to check for normality; where normality was violated, we used a Box-Cox transform to transform the data, setting λ to a value for each metric that would yield normally-distributed data; where applicable, we report our statistics using the transformed values. For each within-subjects effect we conducted Mauchly's Test of Sphericity; where sphericity was violated, we used Greenhouse-Geisser correction to adjust the degrees of freedom. We report on main effects of *Interface* and *RadiusBubble* and on interaction effects between *Interface* and *RadiusBubble*. We performed post-hoc tests on *Interface* for pairwise comparisons using Bonferroni correction. In the case of the metric of headset lineup concentration, we found the data to be highly non-normally distributed, and so instead used an aligned rank transform for nonparametric analysis of variance, followed up by post-hoc tests comparing estimated marginal means with Bonferroni correction [27, 55]. For the questionnaires described in Sect. 5.5, we performed a Friedman test; statistically significant results were followed up with post-hoc analysis with Wilcoxon signed-rank tests with Bonferroni correction, setting $\alpha = 0.05/10 = 0.005$.

6.2 Results

Here, we list the results of the metrics described in Sect. 5.3.

Alignment time SHORTAXES had the shortest alignment times, with the PROJECTOR and PRONGS interfaces requiring longer alignment times (Fig. 5, top left). The presence of the RADIUSBUBBLE increased mean alignment time, as participants had an additional visual cue to refine translation error. The increased alignment time was most prominent for STATICPRONGS and DYNAMICPRONGS, likely due to the increased cognitive load of manipulating multiple RADIUSBUBBLE visualizations at the same time.

To ensure normality, we used a Box-Cox transformation with $\lambda = 0$. We found a main effect of *Interface* (Greenhouse-Geisser corrected: $F(2.851, 62.726) = 11.125, p < .001, \eta_p^2 = .336, 1-\beta = .998$) of *RadiusBubble* ($F(1,22) = 44.578, p < .001, \eta_p^2 = .670, 1-\beta > 0.999$), and an interaction effect ($F(4,88) = 4.587, p = .002, \eta_p^2 = .173, 1-\beta = .935$). Post-hoc testing revealed significant differences between SHORTAXES and PROJECTOR ($p = .003$), STATICPRONGS ($p = .003$), and DYNAMICPRONGS ($p < .001$), as well as between LONGAXES and DYNAMICPRONGS ($p = .004$).

Translation and rotation error For both translation error and rotation error, we see a significant difference between the interfaces (Fig. 5, top right and middle left). Mean translation error is low for both AXES interfaces, and higher for PROJECTOR and PRONGS interfaces. The presence of the RADIUSBUBBLE helps reduce translation

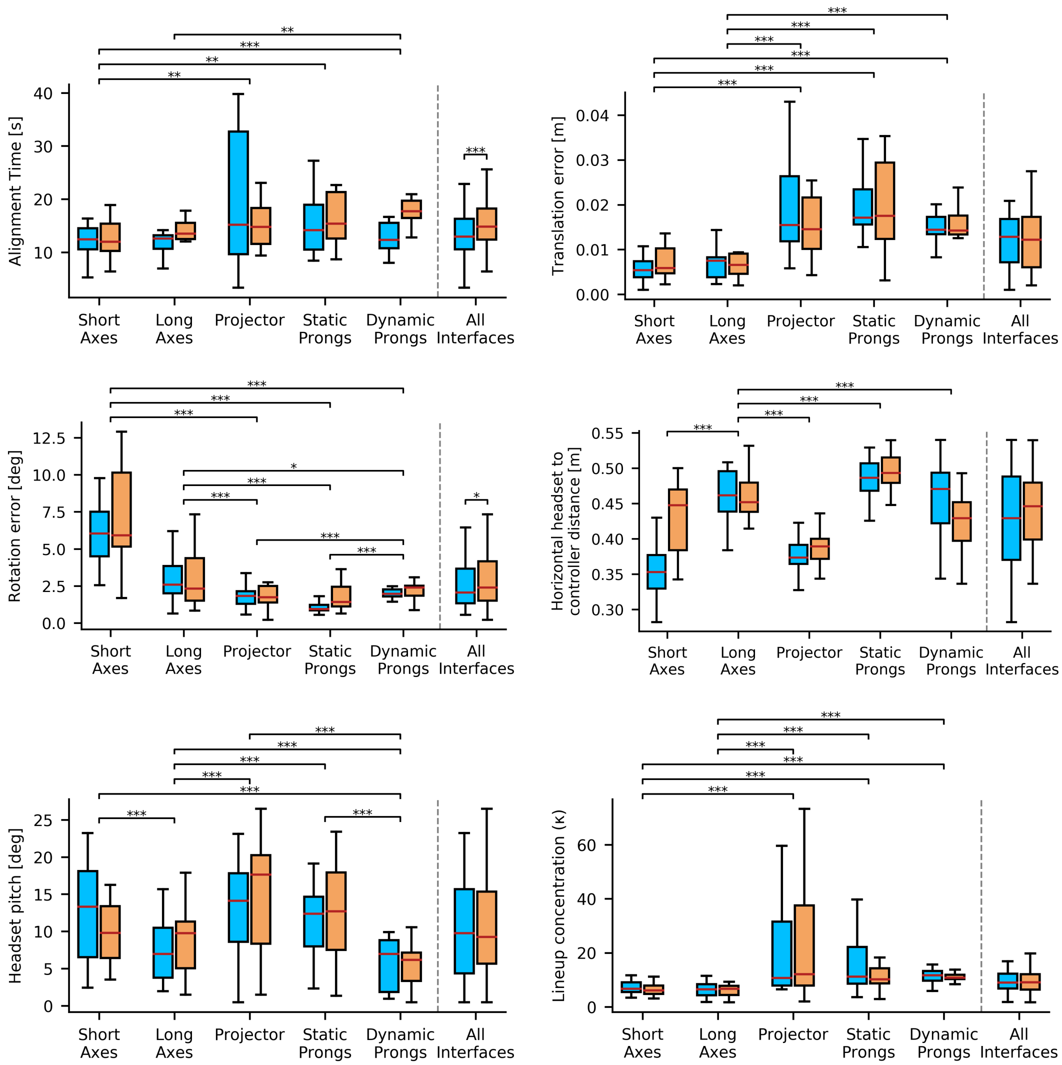

Figure 5: Participant results for alignment time (top left), translation error (top right), rotation error (middle left), horizontal headset-to-controller distance (middle right), headset pitch (bottom left), and headset lineup concentration (bottom right). Blue: without RADIUSBUBBLE. Orange: with RADIUSBUBBLE. Statistical significance indicated for $p < .05$ (*), $p < .01$ (**), and $p < .001$ (***).

error, but only for the non-AXES interfaces. We see the opposite trend for rotational error, with relatively high rotational error (about $5°$) for the AXES interfaces and lower rotational error (about $2-3°$) for the PROJECTOR and PRONGS interfaces.

We ensured normality with a Box-Cox transformation with $\lambda = -1$ for translation error, and with $\lambda = 0$ for rotation error. Translation error showed a main effect of *Interface* (Greenhouse-Geisser corrected: $F(2.260, 49.716) = 105.882, p < .001, \eta_p^2 = .828, 1-\beta > .999$), did not show a main effect of *Bubble*, but did show a crossover interaction effect between *Interface* and *Bubble* (Greenhouse-Geisser corrected: $F(2.492, 54.830) = 2.970, p = .049, \eta_p^2 = .119, 1-\beta = .617$). Post-hoc testing showed significant differences between the set of AXES interfaces and the set of non-AXES interfaces ($p < .001$). Rotation error showed a main effect of *Interface* ($F(4, 88) = 70.302, p < .001, \eta_p^2 = .762, 1-\beta > 0.999$), and of *RadiusBubble* ($F(1, 22) = 7.959, p = .010, \eta_p^2 = .266, 1-\beta = .769$). Post-hoc testing showed significant differences between SHORTAXES and all other non-AXES interfaces ($p < .001$), between DYNAMICPRONGS and all interfaces besides LONGAXES ($p < .001$), and between LONGAXES and PROJECTOR ($p < .001$).

Horizontal headset-to-controller distance Fig. 5, middle right, shows for each interface how far from the body the controller was held. Data was found by Shapiro-Wilk to be normally distributed without need for transformation. We found a main effect of *Interface* ($F = 17.933$, $df = 4, p < .001, \eta_p^2 = .449, 1-\beta > 0.999$) but not of *RadiusBubble*; we also found an interaction effect between *Interface* and *RadiusBubble* (Greenhouse-Geisser corrected: $F = 3.275$, $df = 2.659, p = .032, \eta_p^2 = .130, 1-\beta = .685$). Post-hoc testing showed significant difference between LONGAXES and all other interfaces ($p < .001$).

Headset pitch To ensure normality, we used a Box-Cox transformation with $\lambda = 0$. We found a main effect of *Interface* (Greenhouse-Geisser corrected: $F(1.403, 30.869) = 68.094, p =< .001, \eta_p^2 = .756, 1-\beta > 0.999$), but not of *RadiusBubble*. Post-hoc testing showed a significant difference between DYNAMICPRONGS and all other interfaces ($p < .001$). DYNAMICPRONGS had the lowest headset pitch (Fig. 5, bottom left). Post-hoc testing also showed a significant difference between LONGAXES and all other interfaces

($p < .001$), which we interpret as a consequence that longer axes lead to holding the controller further from the body.

Headset lineup concentration Data was analyzed nonparametrically using an aligned rank transform. We found a main effect of *Interface* ($\chi^2 = 143.754, df = 4, p < .001$), but not of *RadiusBubble*. Post-hoc testing found a significant ($p < .001$) difference between each AXES interface and each non-AXES interface.

Fig. 5, bottom right, shows a contrast between PROJECTOR/STATICPRONGS and other interfaces. Participants lined themselves up behind the controller much more consistently, in order to view both the mid-air alignment point and the target plane. We also see a much wider variance for these interfaces; participants were varied in just how committed they were to lining up behind the controller at the cost of extreme body poses. When using these interfaces, about one-third of participants were observed to crouch or even kneel on the floor during the alignment process. In contrast, DYNAMICPRONGS, by placing all guidance at head level, relaxed the need for a specific viewing direction relative to the mid-air pose.

Questionnaire Here we report the statistically significant findings of the questionnaire provided to participants after completing each of the five interfaces. The NASA-TLX showed a significant ($\chi^2(4) = 16.750, p = .002$) effect of *Interface* on Mental Demand (NASA-TLX-1), with post-hoc testing showing SHORTAXES to have lower Mental Demand than each of the PRONGS interfaces ($p < .003$ with Bonferroni-corrected $\alpha = .005$). We hypothesize that the added visual complexity of the PRONGS interfaces, especially with the multiple RADIUSBUBBLEs added, led to increased cognitive load. Our analysis also showed a significant ($\chi^2(4) = 10.935, p = .027$) effect of *Interface* on Effort (NASA-TLX-5), but no significant post-hoc pairwise comparisons were found after Bonferroni correction.

Responses for self-reported fatigue of different body parts revealed no statistically significant effect of *Interface*. It is challenging for participants to isolate which stimulus affected cumulative fatigue during a long user study session. However, the mean fatigue levels for each body part show which kinds of fatigue were most strongly felt. Fatigue in the arms was most highly self-reported (3.2/10), followed by the wrist (1.5/10), neck (1.0/10), and eyes (1.0/10). Fatigue on the head (0.8/10) and back (0.7/10) was relatively low.

6.3 Discussion

Here we interpret the results of our user study in the context of our design requirements (Sect. 3.2) and our design space (Sect. 3.3).

6.3.1 Conformance to Design Requirements

Overall, the AXES interfaces lead to the fastest alignment times **(R2)** and the lowest translation errors **(R1)**. This is likely because all visuals are located near the user's hand, and because the interface is familiar to novice users. The high redundancy of the axes allows for multiple visual cues to the user during alignment.

However, the AXES interfaces show lower rotational accuracy than the PROJECTOR and PRONGS interfaces **(R1)**; the tips of the XYZ axes are a weak visual cue, and because they are near the edges of the FoV, they may not be visible to the user. In contrast, interfaces that involve aiming at a far-away "target" better lend themselves to rotational accuracy. PROJECTOR has high visual sensitivity to rotational error, which has both pros and cons: visually magnifying the misalignment allows for greater fine-tuning but also reveals the natural shakiness of the arm and hand, leading some participants to say they had low confidence in their alignment performance.

The RADIUSBUBBLE showed a crossover effect where its presence slightly increased the translation error **(R1)** for the AXES interfaces, and slightly decreased the translation error for the PROJECTOR and PRONGS interfaces. For the non-AXES interfaces, the RADIUSBUBBLE was the primary indicator of 3D position, whereas it was a redundant indicator of position for the AXES interfaces. The RADIUSBUBBLE also increased alignment time **(R2)**, showing the speed-accuracy tradeoff common to many tasks. Participants were generally pleased with the presence of the RADIUSBUBBLE when fine-tuning the alignment near the correct position, but many reported that the RADIUSBUBBLE was distracting when the controller was far from the target point, and especially when there were multiple RADIUSBUBBLEs at the same time (i.e., in the PRONGS interfaces) **(R3)**. When the spheres are so large that the curvature is not easily visible within the AR HMD's FoV, it is hard to interpret the directional cue. Several participants mentioned that when the RADIUSBUBBLE was present, it became their primary focus during alignment and they would ignore the other visual elements until the sphere was reduced in size. This may be due to the relative brightness/thickness of the RADIUSBUBBLE compared with other visual elements, or the fact that it dynamically changes its size.

Comparing the results from SHORTAXES and LONGAXES reveals how just changing the size of AR visuals alters the user's body pose. Longer axes results in users holding the controller further from the body, which increases fatiguing torque forces on the arm **(R4.1)**.

Our study results answer the research question of how a user wearing a low-FoV AR HMD will deal with hand-held AR content that extends outside the field of view. The dominant behavior we saw was that users will choose against moving the head back and forth to see different portions of the alignment interface. Instead, the user will attempt wherever possible to adjust their body pose so that all parts of the interface are simultaneously visible within the same field of view. In the case of LONGAXES, this means stretching out the arm far from the body, even though participants reported more fatigue in the arms than in other parts of the body. In the case of PROJECTOR and STATICPRONGS, where the target plane was located far from the mid-air pose, many participants assumed exaggerated and tiring poses such as leaning back, bending over, and even kneeling on the floor just to keep all visual elements in the field of view at the same time **(R4.3)**. While we had placed the target plane far from the pose so as to match the AR HMD's focal distance (to follow conventional recommendations about placing AR content far from the user), a better approach would be to instead keep the target plane relatively close to the mid-air pose.

Informal follow-up conversations with participants revealed three main reasons for the trend against head motion. First, they were concerned about the visuals becoming lost outside of the FoV once they had been initially sighted. Second, the AR HMD's weight made head motion unpleasant. Third, distracting "color separation" artifacts (a consequence of the AR HMD's hardware) would occur when rapidly rotating the head.

The results from DYNAMICPRONGS show that adjusting the placement of visual guidance can help redirect the user's body into more ergonomic poses. Placing the target plane at head level reduces the user's head pitch to avoid neck strain **(R4.2)**. And while participants still attempt to line themselves up behind the visual guidance, they do not bend over or kneel to do so, because all visuals are kept at head level. However, DYNAMICPRONGS has a disadvantage in that, for each pose, the user must relearn the relationship between the controller's orientation and the augmented visuals, which increases rotational error versus STATICPRONGS **(R1)**. We also noticed a difference in how several participants held the controller during DYNAMICPRONGS: instead of holding it firmly by the grip, participants would hold it by the fingertips, so that they could more easily manipulate the orientation without placing strain on the wrist.

6.3.2 Implications for Interface Design Space

Our user study results inform conclusions about these interfaces from the perspective of the design space described in Sect. 3.3.

In the context of *redundancy*, interfaces that present multiple indicators of translation/rotation alignment can improve task performance. However, too many redundant visual elements cause cognitive overload, such as when multiple RADIUSBUBBLEs are placed on the PRONGS interfaces. Rather than attempting alignment with only a partial view of a redundant visualization, users attempt to view all parts simultaneously. Therefore, interfaces should be designed to carefully balance these trade-offs.

In the context of *alignment strategy*, the AXES interfaces that allow simultaneous guidance and refinement of both position and orientation lead to faster alignment and less translation error than interfaces that explicitly separate translation and rotation alignment guidance. It is hard for users to keep a mid-air position fixed while looking away to receive guidance on orientation alignment. Even though users alternate between fine-tuning position and fine-tuning orientation, users should have all visual information available at once so they can choose which degrees of freedom to refine.

In the context of *sensitivity*, we observe that users tend to position themselves relative to the indicated pose such that sensitivity is maximized during the alignment process. That is, users seek to magnify how much visual change results from adjusting the position or rotation of the hand-held object. In the case of the PROJECTOR and PRONGS interfaces, this results in strained body poses as users try to place both translation and rotation guidance into the same FoV. Therefore, we recommend designing interfaces so that the user pose that maximizes visual sensitivity is also an ergonomic pose.

6.3.3 Variability of Poses

In our analysis, we aggregated the 12 alignments per trial by taking the median, which allows robustness to outliers. Here, we explore the variability of the measurements across the poses. For each per-alignment measurement, we measured the standard deviation across each trial. The standard deviation differs depending on the participant and on the alignment interface being used, but examining the mean standard deviation across all trials can give some insight.

For alignment time, the standard deviation across trials was 3.87 seconds, suggesting a broad spread where some poses took more time or less time for the user to complete alignment. It is possible that some poses initially appear out of the user's FOV, making them more challenging to discover even with the guidance provided by the alignment interface. On the other hand, translation error showed an average per-trial standard deviation of 1.30cm, which suggests that users were consistent in their judgment of whether the controller was sufficiently aligned with the indicated pose. Similarly, rotation error showed an average per-trial standard deviation of 1.88 degrees. Because participants were in control of judging when an alignment was "good enough" (by choosing when to pull the trigger on the VR controller), this self-judgment remained consistent across poses. In contrast, the metrics for headset-to-controller distance and headset pitch had larger variations. The average per-trial standard deviation for headset-to-controller distance was 5.30cm, and the average per-trial standard deviation for headset pitch was 6.04 degrees. The distance that a user outstretches their arm, or the angle at which the user tilts their head, depends not just on the individual participant or on the interface but also on the mid-air pose's specific location.

6.3.4 Limitations and Future Work

While our alignment interfaces reasonably span across the possible design space and reveal important findings, no single user study can test all possibilities for user interfaces. For example, the thickness of the Fresnel effect in our RADIUSBUBBLE visualization was constant across all interfaces, rather than investigating whether that detail influenced participants' performance and preference. Our work informs future interfaces to be validated with additional studies.

Similarly, our study only used one type of AR HMD. Although our proposed alignment interfaces are parameterized based on the AR HMD's FoV and focal distance, and thus can be easily transferred to other headset models, additional research should be done to determine how AR HMDs with different weights and FoVs affect user pose during alignment. Such studies could be done using a high-FoV VR headset with a synthetically rendered AR overlay, but special care is needed to fully replicate properties of real-world AR headsets (e.g., color separation artifacts).

Our approach in this work was to examine generic mid-air alignment without specifying the hand-held object being aligned; the VR controller we selected for the user study was designed to be general-purpose for different VR applications. However, future work could investigate how the shape and weight of the hand-held object affects the performance of different AR alignment guidance. As mentioned in Sect. 6.3.1, we observed that several participants gripped the controller differently when using the DYNAMICPRONGS visualization, holding it by the fingertips rather than gripping with the palm. Prior work has found that a user's grip style can influence pointing tasks in virtual environments [6, 31]. A promising future avenue of research would be to investigate in depth the full interplay between AR alignment visualization, the user's preferred grip style, and the user's alignment performance.

The consumer-level hardware in our user study offered reasonably robust tracking; however, some minor tracking noise and controller jitter can occur intermittently. While the level of tracking noise would be consistent across all our alignment interfaces, the interfaces that visually magnify the user's hand movements (e.g. LONGAXES vs SHORTAXES) might appear more shaky and less stable to the user. An interesting area of future work would be to investigate how different alignment interfaces affect the user's perception of the level of tracking noise, and how the user's motions change to compensate for the perceived tracking instability.

One limitation of our study design was that each participant completed a total of five questionnaires (one for each *Interface*), rather than one for each permutation of *Interface* and *RadiusBubble*. While this was chosen to avoid cognitive fatigue on the part of the participants, the questionnaire responses did not capture differences in self-reported metrics between the presence and absence of the RADIUSBUBBLE for a single interface.

To evaluate the ergonomics of our different interfaces, we use proxy measurements such as how far from the body the arm is outstretched. A more detailed endurance model from full tracking of all limbs and joints could provide additional insights [13, 23, 25].

Recent research into tracking limb poses from head-mounted cameras could enable future alignment guidance interfaces that detect high-exertion poses in real time and prompt the user to assume more ergonomic poses. Such tracking could also enable a third-person "out-of-body" AR interface, enabling the user to perform mid-air alignment without ever needing to look at one's hands.

7 CONCLUSION

Aligning hand-held objects in mid-air is important for many real-world tasks. Combining AR HMDs with object tracking can enable fast and precise mid-air alignment. In this work, we have explored the design space of AR HMD mid-air alignment interfaces and have tested several potential interfaces in a user study against a set of proposed design requirements. The study results reveal how the specific visualization affects task performance and also the way in which users position and orient their own bodies during the task, which has important implications for ergonomics. This helps motivate the direction of future AR HMD hardware and interface development.

ACKNOWLEDGMENTS

We thank the members of the CGVLab and the ART group at Purdue University for their feedback during the development of our interface design. This work was supported in part by the National Science Foundation under Grant DGE-1333468. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the NSF.

REFERENCES

[1] B. Ahlström, S. Lenman, and T. Marmolin. Overcoming touchscreen user fatigue by workplace design. In *Posters and Short Talks of the 1992 SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, p. 101–102. ACM, New York, NY, USA, 1992.

[2] D. Andersen, P. Villano, and V. Popescu. AR HMD guidance for controlled hand-held 3D acquisition. *IEEE Transactions on Visualization and Computer Graphics*, 25(11):3073–3082, 2019.

[3] E. Azimi, L. Qian, N. Navab, and P. Kazanzides. Alignment of the virtual scene to the tracking space of a mixed reality head-mounted display. *arXiv preprint arXiv:1703.05834*, 2017.

[4] S. Bae, A. Agarwala, and F. Durand. Computational rephotography. *ACM Transactions on Graphics*, 29(3), July 2010.

[5] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra. Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research*, 6:1345–1382, Dec. 2005.

[6] A. U. Batmaz, A. K. Mutasim, and W. Stuerzlinger. Precision vs. power grip: A comparison of pen grip styles for selection in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 23–28. IEEE, New York, NY, USA, 2020.

[7] C. Bichlmeier, S. M. Heining, M. Feuerstein, and N. Navab. The virtual mirror: A new interaction paradigm for augmented reality environments. *IEEE Transactions on Medical Imaging*, 28(9):1498–1510, 2009.

[8] G. A. Borg. Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, pp. 377–381, 1982.

[9] S. Boring, M. Jurmu, and A. Butz. Scroll, tilt or move it: Using mobile phones to continuously control pointers on large public displays. In *21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, OZCHI '09, p. 161–168. ACM, New York, NY, USA, 2009.

[10] F. Bork, U. Eck, and N. Navab. Birds vs. fish: Visualizing out-of-view objects in augmented reality using 3d minimaps. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct*, ISMAR '19, pp. 285–286. IEEE, New York, NY, USA, 2019.

[11] F. Bork, B. Fuers, A. Schneider, F. Pinto, C. Graumann, and N. Navab. Auditory and visio-temporal distance coding for 3-dimensional perception in medical augmented reality. In *2015 IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '15, pp. 7–12. IEEE, New York, NY, USA, 2015.

[12] W. Chan and P. Heng. Visualization of needle access pathway and a five-dof evaluation. *IEEE Journal of Biomedical and Health Informatics*, 18(2):643–653, 2014.

[13] N. Cheema, L. A. Frey-Law, K. Naderi, J. Lehtinen, P. Slusallek, and P. Hämäläinen. Predicting mid-air interaction movements and fatigue using deep reinforcement learning. In *2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. ACM, New York, NY, USA, 2020.

[14] Z. Chen, C. G. Healey, and R. S. Amant. Performance characteristics of a camera-based tangible input device for manipulation of 3D information. In *43rd Graphics Interface Conference*, GI '17, p. 74–81. Canadian Human-Computer Communications Society, Waterloo, CAN, 2017.

[15] S. Drouin, D. A. D. Giovanni, M. Kersten-Oertel, and D. L. Collins. Interaction driven enhancement of depth perception in angiographic volumes. *IEEE Transactions on Visualization and Computer Graphics*, 26(6):2247–2257, 2020.

[16] U. Gruenefeld, L. Prädel, and W. Heuten. Locating nearby physical objects in augmented reality. In *18th International Conference on Mobile and Ubiquitous Multimedia*, MUM '19, p. 1–10. ACM, New York, NY, USA, 2019.

[17] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, eds., *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139 – 183. North-Holland, 1988.

[18] A. D. Hartl, C. Arth, J. Grubert, and D. Schmalstieg. Efficient verification of holograms using mobile augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1843–1851, 2016.

[19] F. Heinrich, F. Joeres, K. Lawonn, and C. Hansen. Comparison of projective augmented reality concepts to support medical needle insertion. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2157–2167, 2019.

[20] F. Heinrich, F. Joeres, K. Lawonn, and C. Hansen. Effects of accuracy-to-colour mapping scales on needle navigation aids visualised by projective augmented reality. In *Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery*, CURAC '19, pp. 1–6, 2019.

[21] F. Heinrich, G. Schmidt, K. Bornemann, A. L. Roethe, W. I. Essayed, and C. Hansen. Visualization concepts to improve spatial perception for instrument navigation in image-guided surgery. In *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, Medical Imaging '19, pp. 559–572. SPIE, Bellingham, WA, USA, 2019.

[22] F. Heinrich, L. Schwenderling, M. Becker, M. Skalej, and C. Hansen. HoloInjection: augmented reality support for CT-guided spinal needle injections. *Healthcare Technology Letters*, 6(6):165–171, 2019.

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

[24] HTC. VIVE, 2017.

[25] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani. Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion. In *2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 3328–3339. ACM, New York, NY, USA, 2017.

[26] M. Kalia, N. Navab, and T. Salcudean. A real-time interactive augmented reality depth estimation technique for surgical robotics. In *2019 International Conference on Robotics and Automation*, ICRA '19, pp. 8291–8297. IEEE, New York, NY, USA, 2019.

[27] M. Kay and J. O. Wobbrock. *ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs*, 2020. R package version 0.10.7. doi: 10.5281/zenodo.594511

[28] T. Kim and J. Park. 3D object manipulation using virtual handles with a grabbing metaphor. *IEEE Computer Graphics and Applications*, 34(3):30–38, 2014.

[29] F. Krause, J. H. Israel, J. Neumann, and T. Feldmann-Wustefeld. Usability of hybrid, physical and virtual objects for basic manipulation tasks in virtual environments. In *2007 IEEE Symposium on 3D User Interfaces*, pp. 87–94. IEEE, New York, NY, USA, 2007.

[30] M. Krichenbauer, G. Yamamoto, T. Taketom, C. Sandor, and H. Kato. Augmented reality versus virtual reality for 3D object manipulation. *IEEE Transactions on Visualization and Computer Graphics*, 24(2):1038–1048, 2018.

[31] N. Li, T. Han, F. Tian, J. Huang, M. Sun, P. Irani, and J. Alexander. Get a grip: Evaluating grip gestures for vr input using a lightweight pen. In *2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. ACM, New York, NY, USA, 2020.

[32] K. Lilija, H. Pohl, S. Boring, and K. Hornbæk. Augmented reality views for occluded interaction. In *2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–12. ACM, New York, NY, USA, 2019.

[33] T. Louis and F. Berard. Superiority of a handheld perspective-coupled display in isomorphic docking performances. In *2017 ACM International Conference on Interactive Surfaces and Spaces*, ISS '17, p. 72–81. ACM, New York, NY, USA, 2017.

[34] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, and E. Kruijff. Non-visual cues for view management in narrow field of view augmented reality displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '19, pp. 190–201. IEEE, New York, NY, USA, 2019.

[35] A. Martin-Gomez, U. Eck, and N. Navab. Visualization techniques for precise alignment in VR: A comparative study. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces*, VR '19, pp. 735–741. IEEE, New York, NY, USA, 2019.

[36] M. R. Masliah and P. Milgram. Measuring the allocation of control in a 6 degree-of-freedom docking experiment. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, p. 25–32. ACM, New York, NY, USA, 2000.

[37] Microsoft. Microsoft HoloLens, 2017.
[38] M. R. Mine, F. P. Brooks, and C. H. Sequin. Moving objects in space: Exploiting proprioception in virtual-environment interaction. In *24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, p. 19–26. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1997.
[39] Mirror. Mirror Networking - Open Source Networking for Unity, 2020.
[40] N. Navab, M. Feuerstein, and C. Bichlmeier. Laparoscopic virtual mirror new interaction paradigm for monitor based augmented reality. In *2007 IEEE Virtual Reality Conference*, pp. 43–50. IEEE, New York, NY, USA, 2007.
[41] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *28th Annual ACM Symposium on User Interface Software and Technology*, UIST '15, p. 405–415. ACM, New York, NY, USA, 2015.
[42] P. Olsson, F. Nysjö, J. Hirsch, and I. Carlbom. Snap-to-fit, a haptic 6 DOF alignment tool for virtual assembly. In *2013 World Haptics Conference (WHC)*, pp. 205–210. IEEE, New York, NY, USA, 2013.
[43] A. Samini and K. L. Palmerius. Popular performance metrics for evaluation of interaction in virtual and augmented reality. In *2017 International Conference on Cyberworlds*, CW '17, pp. 206–209. IEEE, New York, NY, USA, 2017.
[44] A. Sears. Improving touchscreen keyboards: design issues and a comparison with other devices. *Interacting with Computers*, 3(3):253–269, 12 1991.
[45] J. Shingu, E. Rieffel, D. Kimber, J. Vaughan, P. Qvarfordt, and K. Tuite. Camera pose navigation using augmented reality. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '10, pp. 271–272. IEEE, New York, NY, USA, 2010.
[46] M. Sukan. *Augmented Reality Interfaces for Enabling Fast and Accurate Task Localization*. PhD thesis, Columbia University, 2017.
[47] M. Sukan, C. Elvezio, S. Feiner, and B. Tversky. Providing assistance for orienting 3D objects using monocular eyewear. In *2016 Symposium on Spatial User Interaction*, SUI '16, p. 89–98. ACM, New York, NY, USA, 2016.
[48] J. Sun, W. Stuerzlinger, and B. E. Riecke. Comparing input methods and cursors for 3D positioning with head-mounted displays. In *15th ACM Symposium on Applied Perception*, SAP '18, pp. 1–8. ACM, New York, NY, USA, 2018.
[49] R. J. Teather and W. Stuerzlinger. Guidelines for 3D positioning techniques. In *2007 Conference on Future Play*, Future Play '07, p. 61–68. ACM, New York, NY, USA, 2007.
[50] Unity Technologies. Unity Real-Time Development Platform, 2019.
[51] V. Vuibert, W. Stuerzlinger, and J. R. Cooperstock. Evaluation of docking task performance using mid-air interaction techniques. In *3rd ACM Symposium on Spatial User Interaction*, SUI '15, p. 44–52. ACM, New York, NY, USA, 2015.
[52] P. Wacker, A. Wagner, S. Voelker, and J. Borchers. Heatmaps, shadows, bubbles, rays: Comparing mid-air pen position visualizations in handheld AR. In *2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–11. ACM, New York, NY, USA, 2020.
[53] J. Wentzel, G. d'Eon, and D. Vogel. Improving virtual reality ergonomics through reach-bounded non-linear input amplification. In *2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–12. ACM, New York, NY, USA, 2020.
[54] S. White, L. Lister, and S. Feiner. Visual hints for tangible gestures in augmented reality. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ISMAR '07, pp. 47–50. IEEE, Washington, DC, USA, 2007.
[55] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *ACM Conference on Human Factors in Computing Systems*, CHI '11, pp. 143–146. ACM, New York, NY, USA, 2011.
[56] S. Zhai and P. Milgram. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, p. 320–327. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1998.