

# Translational and Rotational Arrow Cues (TRAC) Navigation Method for Manual Alignment Tasks

DAVID E. USEVITCH, ADAM J. SPERRY, and JAKE J. ABBOTT, University of Utah

Many tasks in image-guided surgery require a clinician to manually position an instrument in space, with respect to a patient, with five or six degrees of freedom (DOF). Displaying the current and desired pose of the object on a 2D display such as a computer monitor is straightforward. However, providing guidance to accurately and rapidly navigate the object in 5-DOF or 6-DOF is challenging. Guidance is typically accomplished by showing distinct orthogonal viewpoints of the workspace, requiring simultaneous alignment in all views. Although such methods are commonly used, they can be quite unintuitive, and it can take a long time to perform an accurate 5-DOF or 6-DOF alignment task. In this article, we describe a method of visually communicating navigation instructions using translational and rotational arrow cues (TRAC) defined in an object-centric frame, while displaying a single principal view that approximates the human's egocentric view of the physical object. The target pose of the object is provided but typically is used only for the initial gross alignment. During the accurate-alignment stage, the user follows the unambiguous arrow commands. In a series of human-subject studies, we show that the TRAC method outperforms two common orthogonal-view methods—the triplanar display, and a sight-alignment method that closely approximates the Acrobot Navigation System—in terms of time to complete 5-DOF and 6-DOF navigation tasks. We also find that subjects can achieve 1 mm and 1° accuracy using the TRAC method with a median completion time of less than 20 seconds.

CCS Concepts: • **Human-centered computing** → **Graphical user interfaces**; **Virtual reality**; • **Computing methodologies** → *Motion capture*;

Additional Key Words and Phrases: Image-guided surgery, visual guidance, pose matching

## ACM Reference format:

David E. Usevitch, Adam J. Sperry, and Jake J. Abbott. 2020. Translational and Rotational Arrow Cues (TRAC) Navigation Method for Manual Alignment Tasks. *ACM Trans. Appl. Percept.* 17, 1, Article 1 (February 2020), 19 pages.  
<https://doi.org/10.1145/3375001>

## 1 INTRODUCTION

Many image-guided surgery (IGS) tasks require a clinician to manually position an instrument in space, with respect to a patient, with five or six degrees of freedom (DOF), comprising a 3-DOF position and a 2-DOF or 3-DOF orientation. For example, in our group, we are interested in positioning robotic [4] and magnetic [14] devices in the proper location with respect to a patient's head, based on preoperative CT scans, during robot-assisted

Research reported in this publication was supported in part by the National Institute of Deafness and Other Communication Disorders of the National Institutes of Health under Award Number R01DC013168.

Authors' addresses: D. E. Usevitch, A. J. Sperry, and J. J. Abbott, Department of Mechanical Engineering and the Robotics Center, University of Utah, 1495 East 100 S, Salt Lake City, UT 84112; emails: {david.usevitch, adam.sperry, jake.abbott}@utah.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

1544-3558/2020/02-ART1 \$15.00

<https://doi.org/10.1145/3375001>

ACM Transactions on Applied Perception, Vol. 17, No. 1, Article 1. Publication date: February 2020.

cochlear implant surgery. Displaying the current and desired pose of the object on a 2D display (e.g., a computer monitor) is straightforward. However, providing guidance to accurately and rapidly navigate the object in 5-DOF or 6-DOF is challenging. Guidance is commonly accomplished using a triplanar display, which simultaneously shows three distinct orthogonal viewpoints of the workspace, and requires simultaneous alignment in all views [1, 4, 6, 8, 13, 19]. Minor variations of this implementation are sometimes used [22]. When working with the human body, the sagittal, coronal, and axial planes are typically used as references for view placement [6]. The arrangement of the three orthogonal views on the monitor differs widely, with no universally accepted or consistent definition. Although the triplanar display is commonly used, it is quite unintuitive, and it can take a relatively long time to perform an accurate 5-DOF or 6-DOF alignment.

Previous studies have examined alternate methods to the triplanar display to improve speed and user intuition in navigation tasks. DiGirola et al. [7] recognized that little effort had been made to provide an accurate tool guidance system or strategies for surgeons to improve implant alignment and bone preparation. They developed the HipNav system, which, after preoperative planning and registration, displays virtual crosshairs to be matched on a virtual view of the patient to align the implant. Joskowicz et al. [11] developed the FRACAS system to align bone fragments using a virtual reality view of the bone fragment combined with CT images. Their goals included reducing the surgeon's cumulative exposure to radiation during fluoroscopic imaging, improving the surgeon's hand-eye coordination, reducing the mental burden on the surgeon, reducing the overall intraoperative time, and reducing the skill acquisition curve. Traub et al. [19] compared three methods for use in a surgical drilling task: the monitor-based triplanar display, a see-through augmented reality visualization method, and a hybrid method that combined both the triplanar display and augmented reality methods. After testing one novice user and two experienced surgeons, they found that using an augmented reality display (for both alternate methods) did not improve accuracy but did reduce overall navigation time. The hybrid display was preferred over both the triplanar display and the other augmented reality display. The navigation method used with the Acrobot Navigation System (ANS) uses a principal view aligned with the long axis of the object's target pose [2]. A ring on the proximal end of the virtual object, and crosshairs on its distal end, are matched with similar features on a target object. A virtual disc inside of the ring changes size to give information about translation along the long axis of the object. A second, orthogonal view is also provided. Sun et al. [18] built a real-time visual guidance system for a 6-DOF ultrasound probe navigation task that involves pose matching of virtual pyramids. They found that it was easiest to first rotate the probe to the correct orientation, then translate it to the proper placement on the skin. Labadie and Fitzpatrick [13] described several navigation methods that use not only visual feedback but also auditory and tactile feedback, and they ultimately concluded that current IGS systems "provide little help in guiding a surgeon toward a specific target."

The problem of manual navigation (i.e., orientation) may seem related to the model rotation that commonly occurs when using modern 3D-modeling/CAD software. Model rotation enables the user to better view and understand the model, and the user may rapidly rotate the model many times to examine it from many different angles. However, the desired poses are generated under the user's own volition, and the resulting viewpoints are 2D projections. This is fundamentally different from the type of manual navigation tasks of interest here, in which the target pose is provided to the user, and the alignment task is 5-DOF or 6-DOF in nature.

As we consider how to better facilitate intuitive 5-DOF and 6-DOF navigation, which require both translation and rotation to achieve a desired pose, we can turn to the literature on how humans conceive such spatial motions. Ware and Arsenault [21] found that a large orientation mismatch between visual and haptic frames (e.g., the up-to-90° mismatch in a given orthogonal view) are detrimental to humans' ability to perform an alignment task. They also found that orientation mismatches of less than approximately 45° have a negligible effect. Jacob et al. [10] explored the concept of *integrality and separability* in the cognitive function of tasks, and found that rotation and translation tasks are perceived as mentally distinct (separable) operations. As a result, when performing a precise 6-DOF manual alignment task, humans actually switch back and forth between performing pure translations and pure rotations. Although translation tasks have been found to be straightforward, general

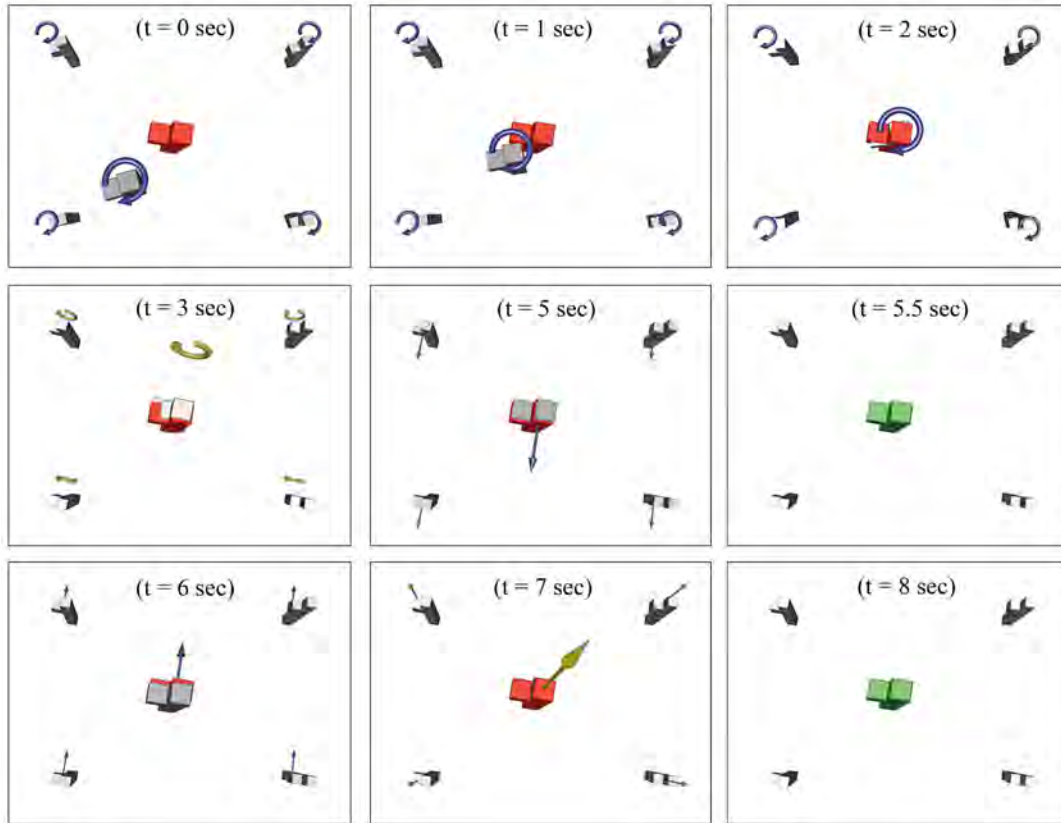


Fig. 1. A typical navigation task with the TRAC method. The red stationary object is the target pose. The gray objects move with the handheld object; the main object in the center both translates and rotates, but the four alternate views in the corners only rotate with fixed positions. In this example, the user quickly performs a gross alignment within the first 2 seconds without using the arrow cues. Then, using only a single arrow cue at a time, the user converges on the target within 8 seconds. Note that the subject briefly passes through the specified accuracy tolerance at  $t = 5.5$  seconds but did not maintain a sufficient accuracy for the required duration.

rotation matching has been found to be quite unintuitive. Parsons [17] found that humans have difficulty in visualizing and performing rotations of objects about arbitrary axes, but performance improves significantly when object rotations are represented about obvious object-centric axes (as opposed to world-centric or egocentric axes).

In this article, we describe a method of visually communicating navigation instructions using translational and rotational arrow cues (TRAC) defined in an object-centric frame while displaying a single principal view that approximates the human's egocentric view of the actual object (Figure 1). The target pose of the object is provided, which is useful for the initial gross alignment. However, with only a single view of the target pose, it is difficult, if not impossible, to perform an accurate 5-DOF or 6-DOF alignment by trying to match the object's pose with the target. Rather, during this accurate-alignment stage, the user must follow the arrow commands provided. The user first performs rotations to align the orientation (in 2-DOF or 3-DOF, as specified by the task) to be within some specified tolerance. Then the user performs translations to align the 3-DOF position to be within some specified tolerance, and this process is iterated until both the orientation and position are simultaneously

within the specified tolerances. During the respective orientation or translation alignments, the user is shown only a single arrow at a time, chosen to correspond to the DOF with the largest error and with that arrow maintained until that DOF is within the specified tolerance. Additional views of the object are provided, which are only slightly rotated from the central object (i.e., less than  $45^\circ$ ), to prevent the arrow cues from becoming occluded or ambiguous. We, somewhat arbitrarily, choose (1) to use four views, one for each corner, although as few as one additional view may be sufficient; (2) to rotate these corner views by  $25^\circ$  from the central object, in their respective offset directions; and (3) to scale these corner views to be 50% of the size of the central object.

This article comprises four human-subject experiments to evaluate the performance of the TRAC method. In general, when using the TRAC method, we observe that users quickly (typically less than 2 seconds) perform the gross alignment without using the arrow cues. During the subsequent accurate-alignment stage, users move cautiously, making slower and more deliberate motions, ready to stop moving when the arrow disappears to avoid overshoot. Consequently, we are primarily interested in the accurate-alignment stage, and our experiments are designed to characterize this stage. Throughout this article, we quantify performance using the completion time to perform a 5-DOF or 6-DOF alignment task to within some specified accuracy.

In Experiment 1, we compare two different methods of presenting the arrow cues to the user, considering a 3-DOF position alignment and 3-DOF orientation alignment separately. In both cases, arrow cues are shown in object-centric axes. However, in one case, we show all three arrows simultaneously (3-A method), and in the other case, we show only the arrow chosen to correspond to the largest error and keep showing that arrow until that DOF is within the specified tolerance (1-A method). We find that showing only a single arrow at a time results in significantly faster alignment times than showing all three arrows at a time, for both position and orientation. The results of this experiment helped us define the TRAC method as described earlier. It should also be noted that, in pilot testing, we considered a third option in which only the arrow corresponding to the largest error at any given instant in time is shown, but we found the resulting rapid switching between arrow cues to be confusing. Neither the 3-A nor 1-A method had problems with flickering like this alternate method.

In Experiment 2, we compare the TRAC method versus a triplanar display and a sight-alignment (SA) method inspired by and adapted from the ANS, for 6-DOF navigation tasks. The name for the SA method comes from aligning a cross and ring similar to the alignment of sights on a firearm. The ANS is primarily used for 5-DOF alignment tasks, and therefore an adaptation for 6-DOF was necessary for this experiment. Visual examples of both the triplanar and SA methods are included in Experiments 2 and 3. We choose to test these two alternate navigation methods against the TRAC method due to their common clinical use based on literature searches and expert opinions. Although various other guidance approaches have been suggested, we find them to be variants on the two basic approaches that we consider here. We find that the TRAC method results in significantly faster alignment times than both the triplanar display and the SA methods. In addition, we find evidence that the worst-case outliers seen when using the triplanar display and the SA methods are substantially slower than the worst-case outliers seen when using the TRAC method. This experiment utilizes subjects without any prior training with any of the methods, indicating that the TRAC method is inherently more intuitive than the others. Our results comparing the TRAC method to a triplanar display were first presented in Usevitch and Abbott [20].

Since many tasks require only 5-DOF alignment, Experiment 3 compares the TRAC method versus the triplanar display and SA methods again, but for 5-DOF navigation tasks. The SA method in this experiment is also an adaptation from the ANS, designed for 5-DOF tasks. The TRAC method again results in significantly faster alignment times and similar results with respect to outliers as found in Experiment 2.

Finally, in Experiment 4, we characterize the time-accuracy trade-off when using the TRAC method in 6-DOF alignment. We find that subjects are able to achieve 1 mm and  $1^\circ$  accuracy with a median completion time of less than 20 seconds. This result serves as a conservative estimate for 5-DOF alignment. Additionally, we find that subjects perform the alignments significantly faster in the second session of the experiment compared to the first session, which were separated by 2 days, indicating that training with the TRAC method results in significant improvement over its already intuitive first use.

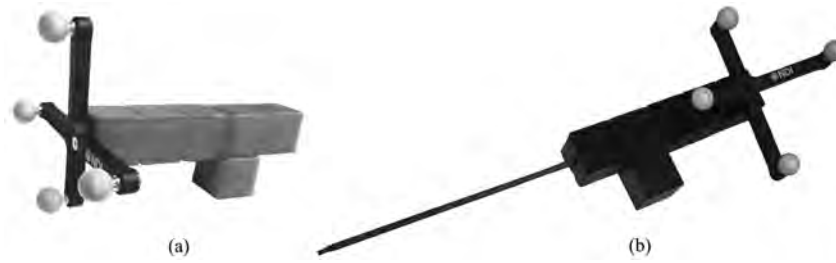


Fig. 2. (a) Object with attached motion-tracking markers used in Session A of Experiment 1, as well as in Experiments 2 and 4. (b) Modified version of the object designed to be used with 5-DOF tasks in Experiment 3. A needle-like attachment is included to better represent a 5-DOF medical instrument.

This work does not make any new contribution to the problem of registration; the TRAC method can be used with any previously developed registration method. The TRAC method is also independent of the type of tracking system used. In all of our experiments we utilize optical tracking, but other methods such as electromagnetic or linkage-based (e.g., [2]) tracking could be used as well.

## 2 EXPERIMENTAL METHODOLOGY

The following describes the typical methodology used throughout all experiments. Details describing variances for specific experiments are provided in their respective experiment sections.

### 2.1 Methods

**2.1.1 Subjects.** A human-subject study was conducted for Experiments 1, 2, and 4 comprised of four adult males and four adult females, and for Experiment 3 comprised of six adult males and six adult females, from the university student population. The subjects had normal (corrected) vision and normal motor functions by self-report. The study was conducted with the approval of the University of Utah Institutional Review Board. Many of the subjects participated in more than one experiment.

**2.1.2 Apparatus.** Data in this study were primarily gathered using a Polaris Spectra (Northern Digital Inc.) optical tracker, which detects the position and orientation of an object with retroreflective spheres rigidly attached. An object inspired by Ware and Arsenault [21] was chosen because it has three clearly discernible orthogonal axes (Figure 2(a)); this object was used in Experiments 1, 2, and 4. The object was modified as shown in Figure 2(b) for the 5-DOF tasks of Experiment 3.

For all sessions, a monitor was positioned 1.83 m (6 ft) in front of a subject sitting at a table. For Experiments 1 and 4, a 483-mm (19-in.) monitor was used. For Experiments 2 and 3, a 1.02-m (40-in.) monitor was used. The Polaris system was placed either above or below the monitor. Chair height was adjusted for each subject to comfortably rest the elbows on the table or chair armrests.

**2.1.3 Design.** Each experiment used a counterbalanced approach for both male and female subjects in a repeated-measures design, examining completion time  $T$  to match a virtual object, whose movement was controlled by a handheld physical object, to a static virtual target pose in either 5-DOF or 6-DOF. Each session tested completion time using different navigation methods and/or accuracy thresholds, depending on the experiment. A single session lasted from 9 to 62 minutes, with large variability observed between subjects and sessions. Sessions were separated by at least 2 days to mitigate effects of learning and fatigue.

For each experiment, a randomly generated set of poses was used in all sessions of that experiment by all subjects. The order of the poses was randomized for each session and subject, to mitigate the effects of presentation order (e.g., learning, fatigue). For each pose in Experiments 1 and 2, translational offsets were assigned



along each of the three principal axes drawn randomly from uniform distributions in the range  $\pm 40$  mm from the home position. For each pose in Experiments 3 and 4, translational offsets were assigned as fixed displacements of 15 mm in random directions sampled uniformly on the sphere. This change in methodology between Experiments 2 and 3 was based on the observation that alignment results were insensitive to small differences in the magnitude of the initial translation error. For all four experiments, a rotational offset of  $15^\circ$  was assigned, using the angle-axis formulation, about an axis drawn randomly from a uniform distribution on the sphere.

Before every pose in each session, object homing was required to ensure that the object would not go out of range of the tracking system due to rotational or translational drift. To ease the burden of homing on the subject, the required accuracy thresholds during homing were relaxed, which should have a negligible effect on the results, due to the speed of the initial gross-alignment stage.

The error in translation is computed simply as the vector from the current position of the object to the target position. This error vector is then projected onto the three axes of the object's frame. The error in orientation is computed as follows (see Lynch and Park [15]). Let  ${}^wR_o$  and  ${}^wR_t$  be the rotation matrices describing the orientations of the object and target, respectively, in the world frame. The rotation matrix describing the orientation error between the object and the target is then  ${}^oR_t = {}^wR_o^T {}^wR_t$ . This rotation can be expressed as a vector  $\theta$ , parameterized by an amount  $\|\theta\|$  (in radians) and an axis of rotation  $\hat{\theta} = \theta/\|\theta\|$ , using the angle-axis formulation. These two representations of orientation error are related through the matrix exponential and the matrix logarithm as

$${}^oR_t = e^{\mathbb{S}\{\theta\}} \iff \log {}^oR_t = \mathbb{S}\{\theta\}, \quad (1)$$

where  $\mathbb{S}\{\theta\}$  is the skew-symmetric matrix packing of the vector  $\theta$ :

$$\mathbb{S}\{\theta\} = \begin{bmatrix} 0 & -\theta_3 & \theta_2 \\ \theta_3 & 0 & -\theta_1 \\ -\theta_2 & \theta_1 & 0 \end{bmatrix}. \quad (2)$$

The values  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  are the angular errors projected onto the three axes of the object's frame.

A generalized linear mixed model (GLMM) was chosen to test whether the response variable completion time  $T$  was influenced by other treatment factors, due to the non-normal distribution of  $T$ , and to account for repeated testing of subjects. A gamma distribution was found to best fit the output distribution of  $T$  for each experiment. Fixed-effect and random-effect variables are outlined in each experiment based on the analysis performed. Methods for statistical analysis were similar to Moscatelli et al. [16]. The conventional significance for each analysis was determined at  $\alpha = 0.05$ .

**2.1.4 Procedure.** In each session, subjects were instructed to first home the virtual object before each pose. Once homed, they were told to align their virtual object with the target pose as quickly as possible. This process repeated until all target poses were completed. Specific instructions relating to each navigation method were also given. For the TRAC method, subjects were instructed to perform gross alignment first, then use the given arrow cues to perform the final accurate alignment. For the triplanar display method, the three orthogonal views of the manipulated object in the triplanar display were explained to the subject. For the SA method, a description of the orthogonal views and additional visual aids were given to the subject. Additional instructions related to avoiding camera workspace boundaries were given in all sessions. Each session was preceded by a short training stage in which the subject became familiar with the method before data collection began. Each session was separated into 2 to 4 subsessions, and subjects were required to take a break between subsessions for at least 30 seconds (but longer if desired) to avoid effects of fatigue. At the start of each subsession, the subject completed two warm-up trials that were not recorded and did not count toward the completion of the set of poses assigned to that subsession.

During each trial, when the object was within the specified position- and orientation-accuracy thresholds, the object changed colors to signal a match, and the subject was required to maintain the object within this

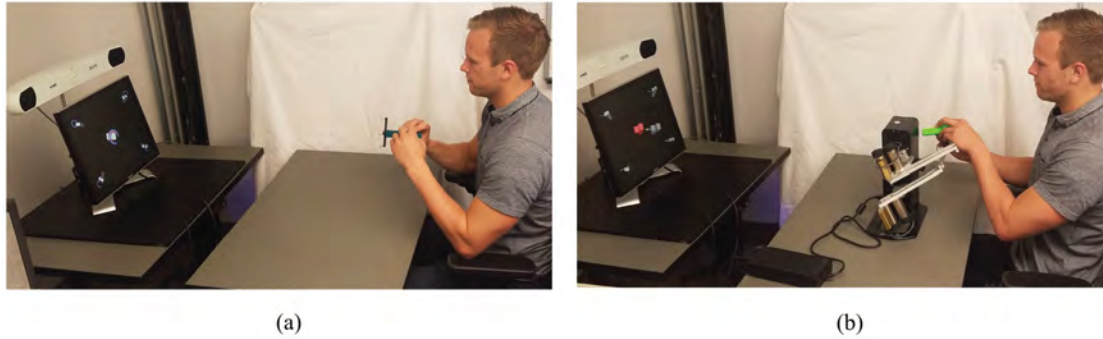


Fig. 3. (a) Setup for Session A of Experiment 1, and Experiment 4. (b) Setup for Session B of Experiment 1. The Polaris Spectra is still in place but not actually used in Session B. The subject in this figure is positioned closer to the monitor than in actual testing to make the photo more compact.

threshold for 1.5 seconds, at which point  $T$  was recorded and the 1.5-second hold time was subtracted. In other words,  $T$  measures the time from the moment of successful homing until the moment of successful alignment. The assigned thresholds for pose matching are given in the respective experiment sections and are similar to those achieved by various other studies involving navigation tasks [3, 12].

## 2.2 Results

All experimental results are shown using notched box-whisker plots displaying a logarithmic transformation of completion time  $T$  (i.e.,  $\log_{10}(T)$ ). In all such plots, the red line represents the median datum; the notch represents the confidence interval of the median; the lower and upper edges of the box represent the first and third quartiles of the data, respectively; the lower/upper whiskers represent the smallest/largest datum within 1.5 IQR (interquartile range, or middle 50% of the data); and the red crosses represent outliers.

All screen captures are shown with a white background to improve the appearance of the images (particularly in print). However, a black background was used in the actual experiments.

## 3 EXPERIMENT 1: REFINING THE TRAC METHOD

### 3.1 Methods

**3.1.1 Subjects.** All eight subjects were inexperienced.

**3.1.2 Apparatus.** The experiment was divided into two sessions. Session A was designed to evaluate pure rotational guidance and utilized the optical tracker as previously described (Figure 3(a)). Session B was designed to evaluate pure translational guidance. In Session B, an object of identical shape and size to the object in Session A was attached to an Entact W6D 6-DOF haptic device to be manipulated by the subject. The haptic device was positioned on the table, as shown in Figure 3(b). The haptic device was programmed to restrict object rotation using a stiff 3-DOF rotational spring implemented as a proportional-derivative controller. This enabled us to perform pure translation experiments without the confounding factor of rotation, which was not possible with the unconstrained object.

**3.1.3 Design.** There are many possible variations of presenting arrow cues as guidance. Two methods that were found by the authors to be intuitive in pilot studies were tested. Both involved displaying arrows along obvious object-centric axes of the manipulated object. The first method displays three arrows (on all three orthogonal axes) simultaneously, and individual arrows disappear when they fall within the given convergence bounds, as shown in Figure 4(a). The second method displays a single arrow at a time to direct user movement,



(a) 3-A Method: Session A, pure orientation (left); Session B, pure translation (right).



(b) 1-A Method: Session A, pure orientation (left); Session B, pure translation (right).

Fig. 4. Examples of 3-A (a) and 1-A (b) navigation methods.

as shown in Figure 4(b); we select the arrow corresponding to the single DOF with the largest error, and show that arrow until convergence in that DOF, at which point we switch to the new arrow corresponding to the largest error, iterating until all three DOF have converged. We refer to these methods as the 3-arrows (3-A) method and 1-arrow (1-A) method, respectively. In both cases, we choose arrow colors arbitrarily assigned to directions but distinct so they may be recognizable by individuals affected by deuteranomaly—the most common form of color blindness [9].

The experiment was divided into two sessions designed to separate translational from rotational guidance: Session A displays only rotational (circular) arrows; Session B displays only translational (straight) arrows. In Session A, the position of the virtual object was locked in the center of the monitor, and translations of the physical object had no effect on the virtual object. Session B used the Entact haptic device to physically restrict rotation as described earlier.

A set of 30 target poses was randomly generated for both the 1-A and 3-A methods, all of which were included in each session. This amounted to a total of 60 trials per session from both methods. The same poses with randomized orders were used in both the pure orientation and pure translation sessions for all subjects. Target thresholds were set to  $1^\circ$  for pure orientation tests and 1 mm for pure translation tests.

For both the pure-orientation and pure-translation tests, a GLMM was used to evaluate the effect of treatment factor *arrow method* and the interaction of treatment factors *arrow method* and *gender* (both treated as fixed-effect factors) on response variable *completion time T* with *subject* and *target pose* treated as random-effect factors.

**3.1.4 Procedure.** In Session B, the haptic device automatically homed the object before each trial and then automatically configured itself to that trial's orientation. Subjects were informed of this and instructed to allow the device to pull them to the home position, then wait a few seconds for the next trial to begin.



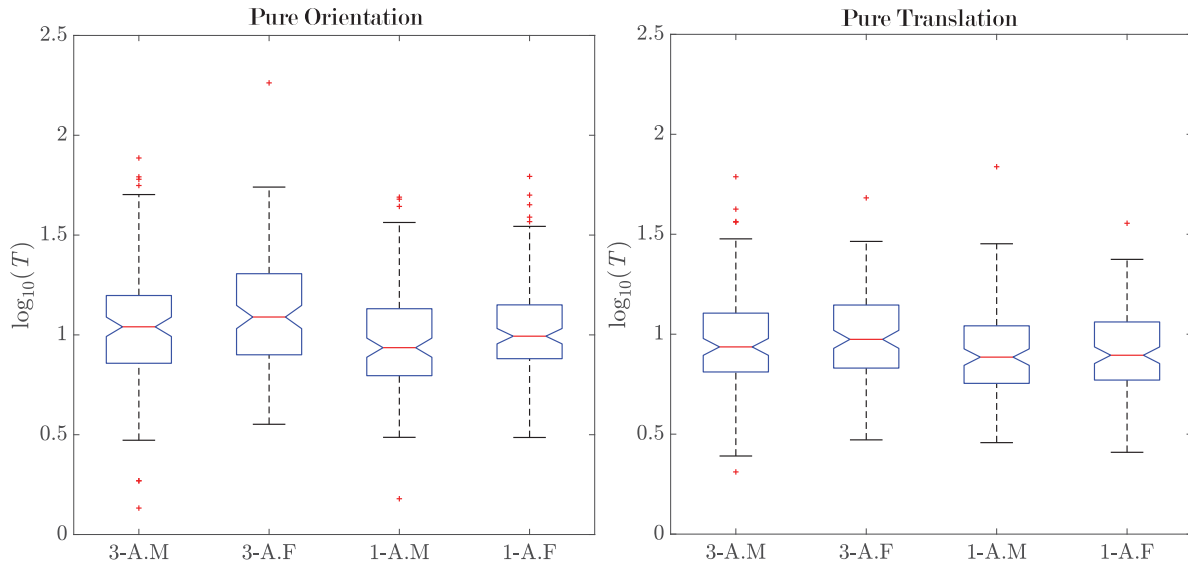


Fig. 5. Results for Experiment 1, for all male (M) and female (F) subjects, comparing  $\log_{10}(T)$  for the 3-A and 1-A methods, for pure orientation using an accuracy of  $1^\circ$  (left) and pure translation using an accuracy of 1 mm (right).

### 3.2 Results

The results of Experiment 1 are presented in Figure 5. For both the pure-orientation and pure-translation tests, there was a main effect of *arrow method* and no interaction between *arrow method* and *gender*. The 1-A method is significantly faster than the 3-A method for both pure orientations and pure translations for both males and females. For pure orientations (from Session A), the median  $T$  of the 3-A method was 29% longer than the median  $T$  of the 1-A method. For pure translations (from Session B), the median  $T$  of the 3-A method was 17% longer than the median  $T$  of the 1-A method.

Based on these results, we will exclusively utilize the 1-A method in all subsequent experiments, and in the TRAC method generally. It is our conjecture that the 1-A method is superior to the 3-A method because it reduces the cognitive burden of the task by always presenting a clear unambiguous guidance command to the user, whereas the 3-A method ultimately requires the user to make a choice about which arrow to follow first.

## 4 EXPERIMENT 2: TRAC VERSUS TRIPLANAR DISPLAY AND SA IN 6-DOF

### 4.1 Methods

**4.1.1 Subjects.** All eight subjects were inexperienced. None had participated in Experiment 1.

**4.1.2 Apparatus.** All sessions of this experiment used the setup shown in Figure 6.

**4.1.3 Design.** This experiment was divided into three sessions designed to compare the performance of different navigation methods in 6-DOF. Session A was designed to evaluate guidance using the TRAC method, as shown in Figure 7(a). The size of the objects in the TRAC method were constrained such that the principal object view was the same size as the objects in the triplanar display (which dictated the maximum size that could be rendered). This choice was made to avoid a potential confounding factor of object size in this experiment.

Session B was designed to evaluate guidance using a triplanar display, with three orthogonal views in an ego-centric frame. There is no standard arrangement used in prior work, so we chose an arrangement that correlates closely with Bruns and Webster [4], in which we position a front view at the bottom left corner, a side view at



Fig. 6. Setup for Experiment 2, in which the monitor shows the triplanar display (Session B). The subject in this figure is positioned closer to the monitor than in actual testing to make the photo more compact.

the lower right, and a top view at the top left, as shown in Figure 7(b). This arrangement is also consistent with common mechanical-drawing techniques. The goal in the design of the triplanar display was to make it both as intuitive as possible for the subjects and consistent with modern applications of this method.

Session C was designed to evaluate guidance using the SA method, inspired by the ANS method, as shown in Figure 7(c). This method uses two orthogonal views: one view displays an axial view of the long axis of the target object, and the second view is orthogonal to the first. These views correspond with a front and top view with respect to the egocentric frame of the subject holding the object. A ring appears on the proximal end of the virtual object, and crosshairs appear at the distal end of the virtual object; these are matched with a similar ring and crosshairs on the target object. A transparent disc inside of the ring increases and decreases in size to represent translation error along the SA axis. This disc size matches the ring and turns green when the position in the axial direction is correct. The rest of the objects turn green when all thresholds for position and orientation are matched. Note that we maintained the axial-view object sizes consistent with the triplanar display, and the orthogonal view sizes appear slightly smaller, consistent with Barrett et al. [2]. The design description of the ANS is somewhat ambiguous, and many design decisions seem somewhat arbitrary. Despite this, the fundamental navigation method of the ANS is interesting and warrants comparison. Thus, the SA method was designed to include the key elements of the ANS and similar methods, and includes several design adaptations to fit individual experiments. For example, in this experiment, the crosshairs rotate with the object to better describe the 6-DOF pose, better facilitating 6-DOF alignment.

Sessions A and B were counterbalanced equally between male and female subjects. A study that compared only the TRAC method and the triplanar display (Sessions A and B) was presented in Usevitch and Abbott [20]. Session C was added subsequently, approximately 6 months after Sessions A and B, and as a result all subjects performed this session last. One of the original subjects was not available for Session C, so a new subject was tested in all three sessions over the course of a week; as a result, the data for Sessions A and B are slightly different from the results in Usevitch and Abbott [20]. A set of 30 poses was randomly generated for use in all sessions and by all subjects. For all sessions, the alignment thresholds were set to  $5^\circ$  and 3 mm. These thresholds, which are fairly large and somewhat arbitrary, were selected based on pilot tests so as to be generally attainable so that no contingencies would be required in the event that a subject was unable to complete a pose-matching trial.

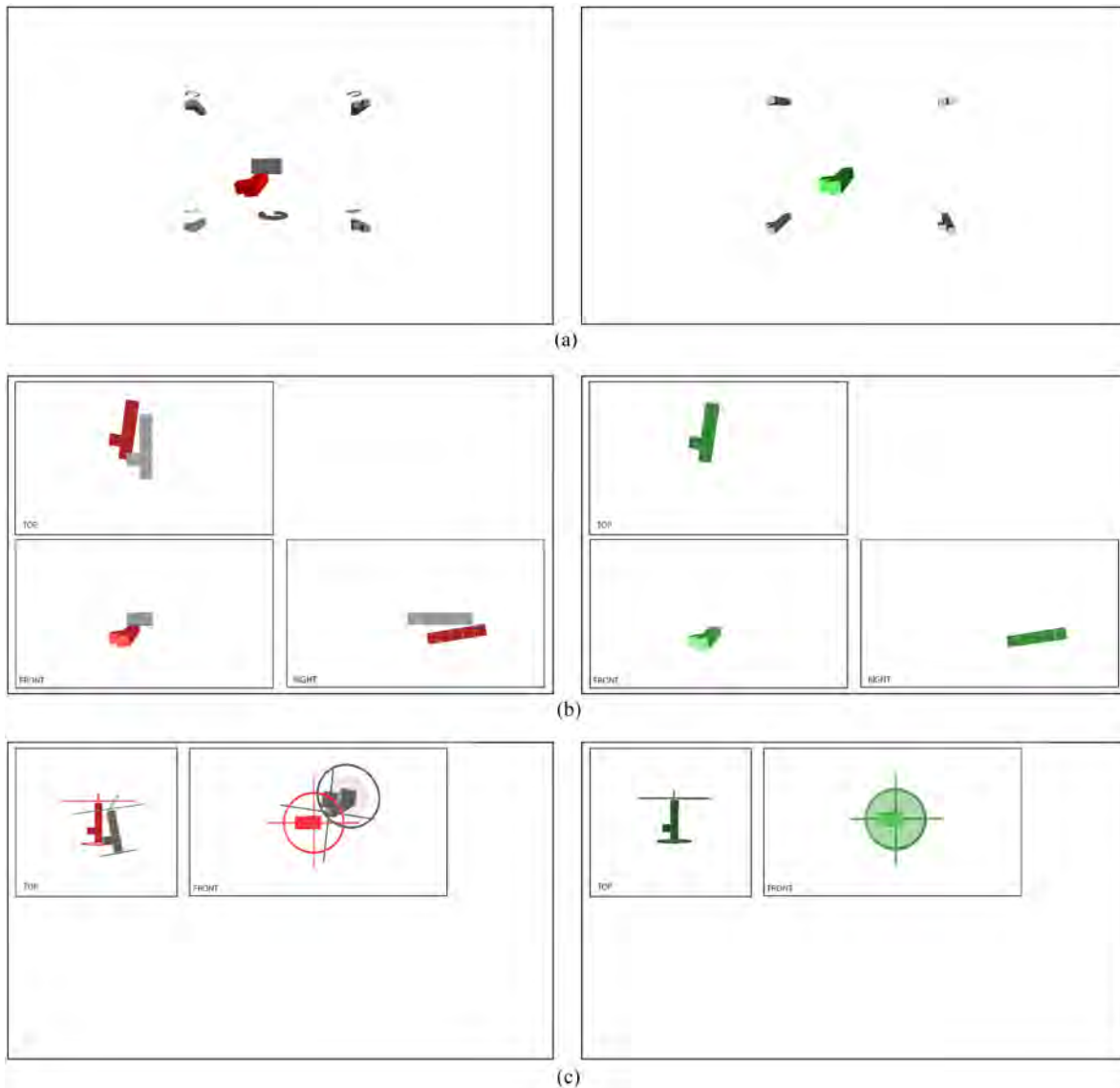


Fig. 7. Displays from Experiment 2. The left images depict the same target pose (red) and current pose of the handheld object (gray) for all three methods, and the right images depict a successful alignment. (a) TRAC method. The user must follow a series of single arrow cues. (b) Triplanar display showing three orthogonal views in a world-centric frame. The user must align the object in all three views simultaneously. (c) SA method showing two orthogonal views in a target-pose-centric frame. Each object contains a ring at the proximal end and crosshairs at the distal end, which must both be aligned. A transparent disc inside of the ring increases and decreases in size to represent translation error along the SA axis.

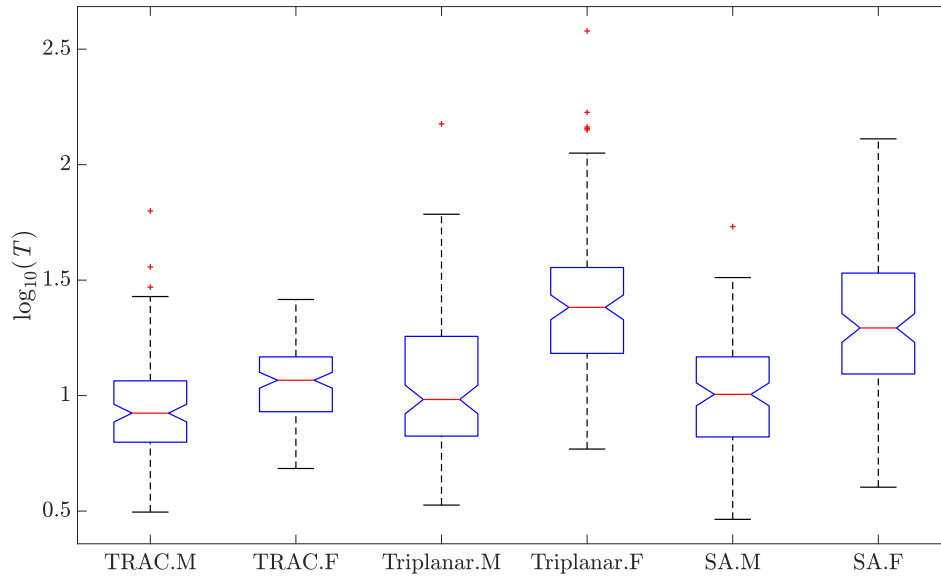


Fig. 8. Results for Experiment 2 (6-DOF alignment), for all male (M) and female (F) subjects, comparing  $\log_{10}(T)$  for the TRAC method, the triplanar display, and the SA method, using an accuracy of  $5^\circ$  and 3 mm.

A GLMM was used to evaluate the effect of treatment factor *navigation method* and the interaction of treatment factors *navigation method* and *gender* (both treated as fixed-effect factors) on response variable *completion time*  $T$  with *subject* and *target pose* treated as random-effect factors.

**4.1.4 Procedure.** In this and subsequent experiments, subjects were instructed to home the device in a position that felt most comfortable to them. Once the 3-DOF orientation was homed, the initial position was reset to the current object location, provided it was within the acceptable workspace boundaries of the experiment.

## 4.2 Results

The results of Experiment 2 are presented in Figure 8. There was a main effect of *navigation method* and interaction between *navigation method* and *gender*. TRAC navigation is significantly faster than the triplanar display for both males and females. TRAC is significantly faster than the SA method for females; TRAC's improvement over SA for males was not statistically significant. Overall, when using the triplanar display, the median  $T$  was 63% longer than the median  $T$  of the TRAC method. The median  $T$  using the SA method was 39% longer than the median  $T$  of the TRAC method. The worst-case  $T$  when using the triplanar display was more than six times longer than the worst-case  $T$  using the TRAC method. The worst-case  $T$  when using the SA method was over twice as long as the worst-case  $T$  using the TRAC method.

## 5 EXPERIMENT 3: TRAC VERSUS TRIPLANAR DISPLAY AND SA IN 5-DOF

### 5.1 Methods

**5.1.1 Subjects.** Ten of the 12 subjects had participated in a previous experiment.

**5.1.2 Apparatus.** The setup of Experiment 3 is the same as for Experiment 2, shown in Figure 6, with the exception of using the object of Figure 2(b), which includes a needle-like element to better represent a 5-DOF task. Instruments used in 5-DOF tasks typically include features (e.g., protruding geometry, labels, optical-tracking markers) that enable visual identification of the full 6-DOF pose, so we believe the object of Figure 2(b) is an

adequate representation for such instruments. The position of the retroreflective spheres and initial homing positions were adjusted for proper view by the optical tracker and to provide room for grasping. In the homing positions, the subjects held the object with the needle of the object pointing down toward the table.

**5.1.3 Design.** This experiment was designed to be similar to Experiment 2, with the exception that all trials required only 5-DOF alignment. This was done by removing any consideration of rotation around the axis of the needle when determining alignment accuracy. The virtual object was modified to match the geometry of the modified physical object and therefore comprised both an object handle and a needle. The target object, however, was only the needle without the handle. Thus, the 6-DOF object appeared correctly aligned on the screen if the needle portion appeared correctly aligned in 5-DOF (Figure 9). A circular disk was also added proximate to the needle to remove ambiguity regarding which direction the needle was pointing. For the SA method, the crosshairs on the target object were automatically adjusted in real time to align with the crosshairs on the virtual object such that no rotation about the needle axis was ever required.

The experiment was divided into the same three sessions as described in Experiment 2. The session order was randomized for each subject resulting in six different order permutations. Each permutation was assigned randomly to one male and one female subject.

A GLMM was used to evaluate the effect of treatment factor *navigation method* and the interaction of treatment factors *navigation method* and *gender* (both treated as fixed-effect factors) on response variable *completion time T* with *subject* and *target pose* treated as random-effect factors.

**5.1.4 Procedure.** The procedure for this experiment is the same as Experiment 2. Subjects were made aware of the 5-DOF nature of the task and correct alignment using the added circular disk on the target needle.

## 5.2 Results

The results of Experiment 3 are presented in Figure 10. There was a main effect of *navigation method* and interaction between *navigation method* and *gender*. TRAC is significantly faster than both the triplanar-display and the SA methods for both males and females in 5-DOF. Overall, using the triplanar display, the median *T* was 112% longer than the median *T* of the TRAC method. The median *T* using SA navigation was 84% longer than the median *T* of the TRAC method. The worst-case *T* when using both the triplanar-display and the SA methods was more than nine times longer than the worst-case *T* using the TRAC method.

## 6 EXPERIMENT 4: TIME-ACCURACY TRADE-OFF AND LEARNING WITH THE TRAC METHOD

### 6.1 Methods

**6.1.1 Subjects.** Six of the eight subjects had participated in a previous experiment.

**6.1.2 Apparatus.** The setup of Experiment 4 is the same as used in Session A of Experiment 1, shown in Figure 3(a).

**6.1.3 Design.** This experiment was designed to evaluate the time-accuracy trade-off of the TRAC method in 6-DOF, which will serve as a conservative estimate for 5-DOF. We tested position accuracy thresholds of 1 mm and 2 mm, and orientation accuracy thresholds of 1° and 2°, for a total of four position-orientation accuracy combinations. The thresholds were also chosen to display feasibility of subjects to simultaneously achieve 1 mm and 1° accuracy in orientation and position.

A set of 30 target poses was randomly generated for each of the four position-orientation accuracy thresholds. This made up a total of 120 trials for use by all subjects. The order of 120 trials was randomized for each subject and split into two sessions of 60 trials separated by at least 2 days.

In addition to characterizing the time-accuracy trade-off with the TRAC method, we are interested in determining if subjects' performance improves with practice when using the TRAC method. A GLMM was used to



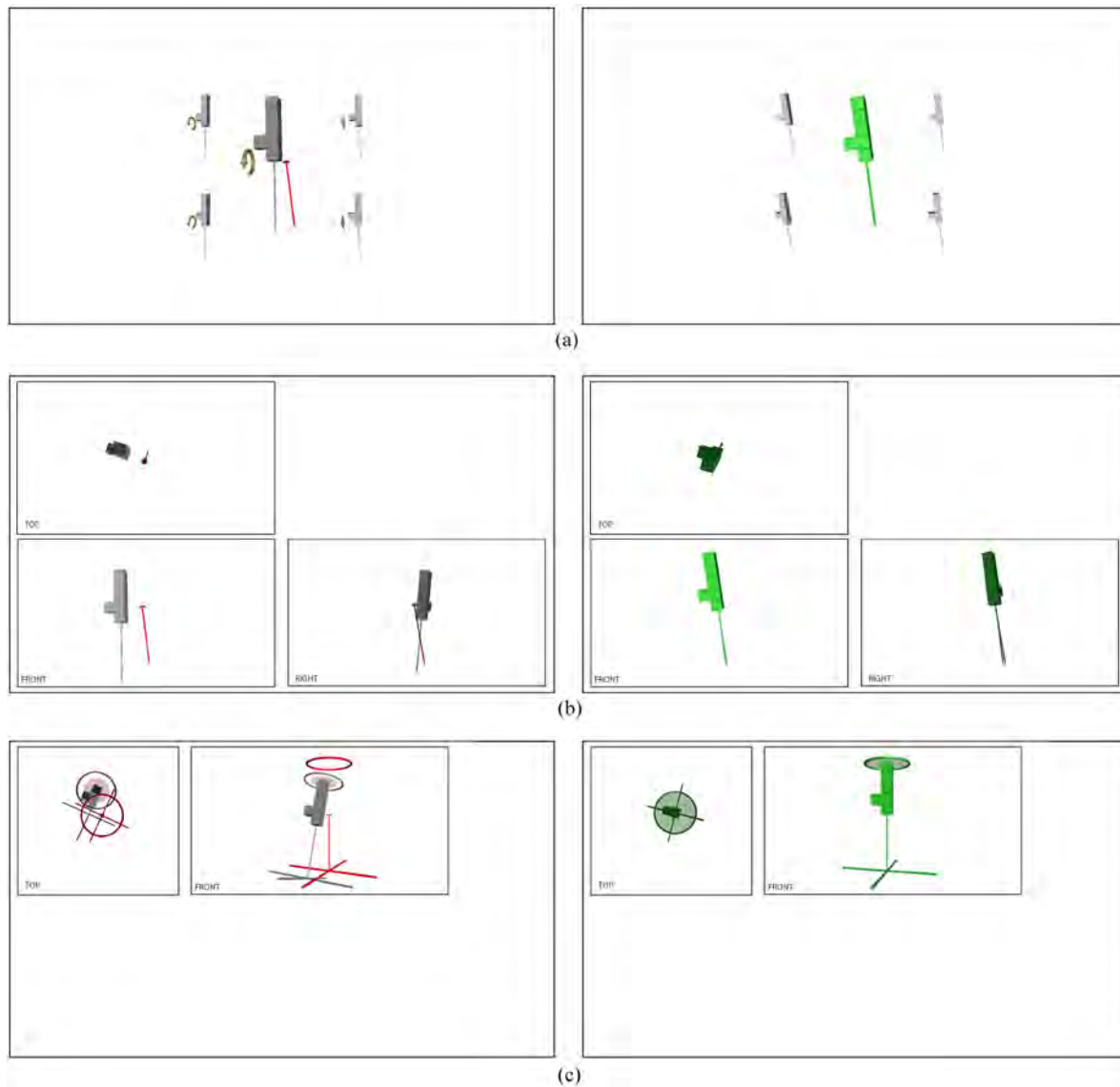


Fig. 9. Displays from Experiment 3. The left images depict the target pose of the needle (red) and current pose of the handheld object with attached needle (gray) for all three methods, and the right images depict a successful alignment. (a) TRAC method. The user must follow a series of single arrow cues. (b) Triplanar display showing three orthogonal views in a world-centric frame. The user must align the needle in all three views simultaneously. (c) SA method showing two orthogonal views in a target-pose-centric frame. Each object contains a ring at the proximal end and crosshairs at the distal end, which must both be aligned. The crosshairs on the target rotate automatically about the needle's axis of symmetry to match the handheld object such that only the origins of the crosshairs must be aligned. A transparent disc inside of the ring increases and decreases in size to represent translation error along the SA axis.

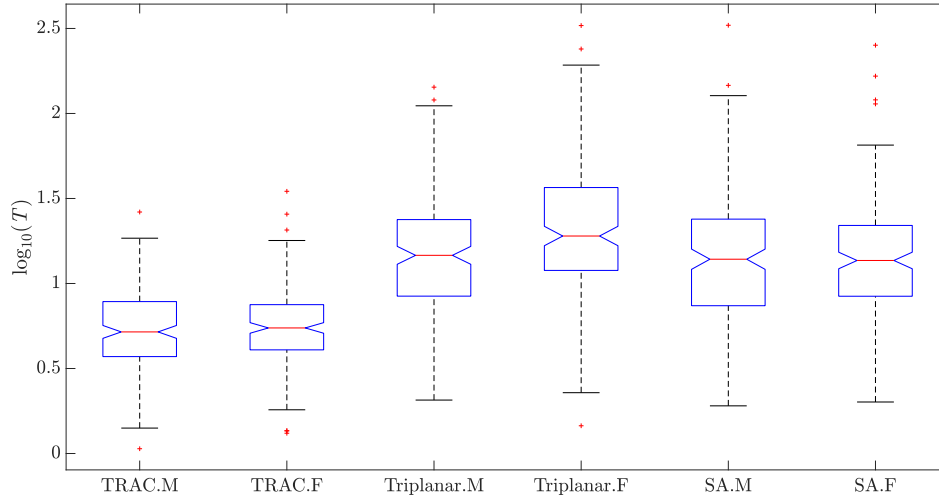


Fig. 10. Results for Experiment 3 (5-DOF alignment), for all male (M) and female (F) subjects, comparing  $\log_{10}(T)$  for the TRAC method, the triplanar display, and the SA method, using an accuracy of  $5^\circ$  and 3 mm.

evaluate the effect of treatment factor *session* and the interaction of treatment factors *session* and *gender* (both treated as fixed-effect factors) on response variable *completion time T* with *threshold values* treated as a fixed-effect factor and *subject* and *target pose* treated as random-effect factors.

## 6.2 Results

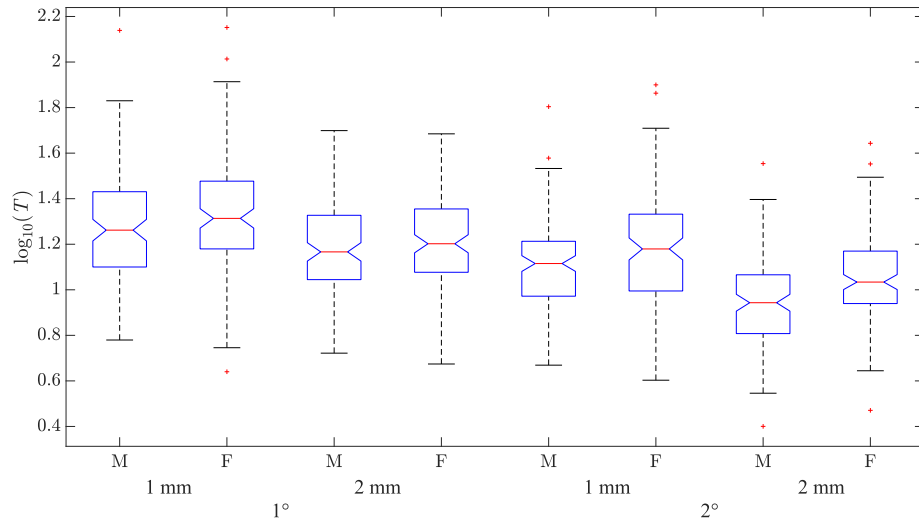
The results of Experiment 4 are presented in Figure 11(a). There was a main effect of *session* and no interactions between *session* and *gender*. The median *T* for a position-orientation accuracy of 1 mm and  $1^\circ$  was 19 seconds, and it was 9 seconds for a position-orientation accuracy of 2 mm and  $2^\circ$ .

The results of Experiment 4 showing the effect of learning are presented in Figure 11(b). Subjects were significantly faster in the second session (median *T* of 12 seconds) than in the first session (median *T* of 17 seconds). In other words, subjects' performance using the TRAC method improves with practice.

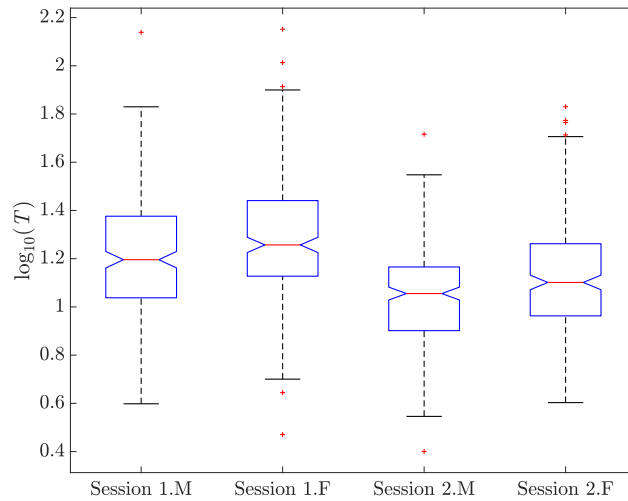
## 7 DISCUSSION

In Experiment 1, for both the pure orientation and pure translation tests, the 1-A method resulted in significantly faster pose alignment than the 3-A method. We went into this experiment without a strong conjecture about what the results would be, due to the lack of prior results from the literature, as well as our own lack of clear preference during pilot testing. Interestingly, subjects were equally divided in verbal declarations of preference for the two methods. However, the results of the superiority of the 1-A method over the 3-A method are clear. It is likely that the 1-A method's suggestion of the best direction to move (i.e., the direction with the largest error) enabled subjects to avoid making suboptimal choices in which they correct small errors before first correcting larger errors in orthogonal directions. Such an effect would be most observable in orientation alignment, as rotations do not commute, and we do in fact see that the gap between the 1-A and 3-A methods is largest in the pure orientation test.

In Experiments 2 and 3, we observed a significant difference in performance between genders. This result is consistent with similar results from the literature [5]. The fact that we did not see a significant difference between genders in Experiment 1 or Experiment 4, which considered only variations on the TRAC method, suggests that use of the TRAC method may be particularly beneficial for women.



(a) Comparison of  $\log_{10}(T)$  between each of the four combinations of position (1 mm, 2 mm) and orientation (1°, 2°) accuracy.



(b) Comparison of  $\log_{10}(T)$  between sessions.

Fig. 11. Results for Experiment 4, for all male (M) and female (F) subjects.

In Experiment 2, and subsequently in Experiment 3, we chose a relatively large convergence tolerance for our study because in pilot testing we found that many subjects struggled with completing alignments in a reasonable amount of time when using the triplanar display or the SA method if we used a tighter tolerance. Consequently, our results may under-represent the true benefit of the TRAC method relative to the triplanar display and the SA method.

There is also a potential for a bias in Experiment 2 due to the SA method (Session C) being tested after the completion of the study comparing the TRAC method to the triplanar display (i.e., Sessions A and B), but we believe that such a bias would be small (if it exists at all) and would tend to benefit the SA method. In each of

Sessions A and B, half of the subjects were complete novices to any navigation method, whereas in Session C, all of the subjects had completed both Sessions A and B, and were thus familiar with the use of navigation systems.

We observed a larger difference between the TRAC method and the SA method in Experiment 3 than in Experiment 2. This may be accounted for by the bias toward the SA method in Experiment 2 described earlier. However, we believe that another likely explanation is that the SA method is more sensitive to the relative orientations of the object, the user, and the monitor, leading to variability in performance from case to case when using the SA method. In Experiment 2, the SA axis happened to be aligned with the surface normal of the monitor (i.e., into and out of the monitor), which would not be practical in general during manual alignment tasks. The TRAC method was designed to be robust in this regard.

Subjects of Experiment 3 were asked to complete a written questionnaire. All 12 subjects indicated that the TRAC method was preferable and more intuitive than the others. Many reported that the single view of the TRAC method was desirable because it is difficult to divide attention between multiple views simultaneously. It is worth noting that one subject who had no prior experience with any method was unable to complete the tasks necessary for this experiment. This subject started with the triplanar display and spent approximately 1 hour to complete just six trials. When subsequently asked to attempt the TRAC method, the subject was able to complete the tasks with comparable proficiency to other subjects in the study.

In Experiment 4, we observed a factor-of-two decrease in the completion time when doubling (i.e., relaxing) both translational and rotational accuracy thresholds. The relative effect on completion time due to the accuracy thresholds for translation or rotation is unclear; therefore, it is difficult to make any statements regarding a time-accuracy trade-off. With median completion times under 20 seconds at 1 mm and 1° accuracy, there is reason to believe that submillimeter and subdegree accuracy are achievable. We also found that subjects' performance improved with increased exposure to the TRAC method. It is unclear how much practice would be required to plateau in learning, but it should not be assumed that the time-accuracy values reported in Figure 11(a) represent the best performance that could be expected from users with training.

It should be noted that the objects we used throughout this study were small and light, and had three distinct visually discernible axes. It is possible that the time-accuracy trade-off, and even the achievable accuracy, will change with an increase in the object's size or mass. Gravity-compensation mechanisms may mitigate this effect. One consequence of providing arrows in an object-centric frame is that these arrows become ambiguous for objects without three distinct axes (e.g., a cube or a simple cylinder). If optical tracking is used, the ambiguity can be eliminated easily by including the rigid body comprising the optical-motion-tracking markers as part of the virtual object. It is unlikely that an instrument used in IGS will lack the geometry necessary to visually identify its full 6-DOF pose because even instruments used for 5-DOF tasks typically have some sort of handle, label, or protrusion. Despite challenges with arrow ambiguity for certain uncommon geometries, the 5-DOF experiment (Experiment 3) illustrates the robustness of the TRAC method to other nonideal geometries. The narrowness of the target needle in this experiment makes it difficult to visually discern model alignment on screen, a strategy upon which the triplanar display method heavily relies. The TRAC method, however, relies only on visual model alignment during the gross-alignment stage, after which the user can completely ignore the models and focus on the arrow cues.

An inherent benefit of the TRAC method over the other methods considered is that the visibility of the arrow guidance cues is invariant to the size of the pose error and the resolution of the display. As a result, it may be possible to utilize a relatively small display with the TRAC method without degrading performance, or even for the TRAC method to be integrated into (rather than replacing) existing navigation graphical user interfaces. It is also possible that the TRAC method's arrow cues could be overlaid on the actual physical object in an augmented-reality display. An indication to the user of error magnitude may be useful as a warning that the user is approaching the threshold, and could be implemented by varying the color or size of the arrows or by some other method; however, it is not clear as to which method would have a greater impact on performance. Color poses challenges for those with color blindness, whereas varying the size of the arrows undermines the visibility

benefits noted earlier. It is not obvious that these modifications will improve performance due to the slow guarded motions already employed by subjects during the accurate-alignment stage. We leave such investigations as open problems for future work. Finally, we note that in the future, the red-green color scheme to distinguish the target pose from successful alignment should be replaced with two colors that can be distinguished by people with various forms of color blindness.

## 8 CONCLUSION

We described a method of visually communicating navigation instructions using TRAC defined in an object-centric frame while displaying a single principal view that approximates the human's egocentric view of the physical object. In a series of human-subject studies, we showed that the TRAC method outperforms commonly used triplanar-display and SA methods in terms of time to complete 5-DOF and 6-DOF alignment tasks, and we found that subjects can simultaneously achieve 1 mm and 1° accuracy using the TRAC method with a median completion time of less than 20 seconds. The TRAC method can be useful in IGS, as well other tasks in which a human must manually position and orient an object in space with some predetermined accuracy. With faster navigation times, the TRAC method could decrease overall intraoperative time, reduce surgeons' cumulative exposure to radiation, and reduce the mental burden on surgeons during instrument alignment. The TRAC method can be combined with existing tracking and registration systems, and can be incorporated into existing graphical user interfaces for navigation.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Peter Kazanzides and Dr. Russell H. Taylor from Johns Hopkins University for their feedback and suggestions when [20] was presented at the 2018 Hamlyn Symposium on Medical Robotics. The authors would also like to thank Timothy S. Lau for his suggestions and help with the statistical analysis.

## REFERENCES

- [1] L. Adams, W. Krybus, D. Meyer-Ebrecht, R. Rueger, J. M. Gilsbach, R. Moesges, and G. Schloendorff. 1990. Computer-assisted surgery. *IEEE Comput. Graph.* 10, 3 (1990), 43–51.
- [2] A. R. W. Barrett, B. L. Davies, M. P. S. F. Gomes, S. J. Harris, J. Henckel, M. Jakopek, V. Kannan, F. M. Rodriguez y Baena, and J. P. Cobb. 2007. Computer-assisted hip resurfacing surgery using the Acrobot® Navigation System. *Proc. Inst. Mech. Eng. H* 221 (2007), 773–786.
- [3] E. M. Boctor, R. J. Webster, M. A. Choti, R. H. Taylor, and G. Fichtinger. 2004. Robotically assisted ablative treatment guided by freehand 3D ultrasound. In *International Congress Series*, Vol. 1268. Elsevier, 503–508.
- [4] T. L. Bruns and R. J. Webster III. 2017. An image guidance system for positioning robotic cochlear implant insertion tools. In *Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*. *Proc. SPIE* 10135, Article 101350O, 6 pages.
- [5] E. Coluccia and G. Louse. 2004. Gender differences in spatial orientation: A review. *J. Environ. Psychol.* 24, 3 (2004), 329–340.
- [6] G. Dagnino, I. Georgilas, P. Köhler, S. Morad, R. Atkins, and S. Dogramadzi. 2016. Navigation system for robot-assisted intra-articular lower-limb fracture surgery. *Int. J. Comput. Assist. Radiol. Surg.* 11, 10 (2016), 1831–1843.
- [7] A. M. DiGioia, B. Jaramaz, M. Blackwell, D. A. Simon, F. Morgan, J. E. Moody, C. Nikou, et al. 1998. Image guided navigation system to measure intraoperatively acetabular implant alignment. *Clin. Orthop. Relat. Res.* 355 (1998), 8–22.
- [8] R. L. Galloway, R. J. Maciunas, and C. A. Edwards. 1992. Interactive image-guided neurosurgery. *IEEE Trans. Biomed. Eng.* 39, 12 (1992), 1226–1231.
- [9] National Eye Institute. 2015. Facts About Color Blindness. Retrieved January 10, 2020 from [https://nei.nih.gov/health/color\\_blindness/facts\\_about](https://nei.nih.gov/health/color_blindness/facts_about).
- [10] R. J. K. Jacob, L. E. Sibert, D. C. McFarlane, and M. P. Mullen Jr. 1994. Integrality and separability of input devices. *ACM Trans. Comput. Hum. Interac.* 1, 1 (1994), 3–26.
- [11] L. Joskowicz, C. Milgrom, A. Simkin, L. Tockus, and Z. Yaniv. 1998. FRACAS: A system for computer-aided image-guided long bone fracture surgery. *Comput. Aided Surg.* 3, 6 (1998), 271–288.
- [12] G. A. Krombach, T. Schmitz-Rode, B. B. Wein, J. Meyer, J. E. Wildberger, K. Brabant, and R. W. Günther. 2000. Potential of a new laser target system for percutaneous CT-guided nerve blocks. *Neuroradiology* 42, 11 (2000), 838–841.
- [13] R. F. Labadie and J. M. Fitzpatrick. 2016. *Image-Guided Surgery: Fundamentals and Clinical Applications in Otolaryngology*. Plural Publishing, San Diego, CA.



- [14] L. Leon, F. M. Warren, and J. J. Abbott. 2018. Optimizing the magnetic dipole-field source for magnetically guided cochlear-implant electrode-array insertions. *J. Med. Robot. Res.* 3, 1 (2018), 1850004.
- [15] K. M. Lynch and F. C. Park. 2017. *Modern Robotics: Mechanics, Planning and Control*. Cambridge University Press.
- [16] A. Moscatelli, M. Mezzetti, and F. Lacquaniti. 2012. Modeling psychophysical data at the population-level: The generalized linear mixed model. *J. Vis.* 12, 11 (2012), 1–17.
- [17] L. M. Parsons. 1995. Inability to reason about an object's orientation using an axis and angle of rotation. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 6 (1995), 1259–1277.
- [18] S. Sun, M. Gilbertson, and B. W. Anthony. 2013. Computer-guided ultrasound probe realignment by optical tracking. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*. 21–24.
- [19] J. Traub, P. Stefan, S. M. Heining, C. Riquarts, T. Sielhorst, E. Euler, and N. Navab. 2006. Towards a hybrid navigation interface: Comparison of a slice based navigation system with in-situ visualization. In *Proceedings of the International Workshop on Medical Imaging and Virtual Reality*. 179–186.
- [20] D. E. Usevitch and J. J. Abbott. 2018. Translational and rotational arrow cues (TRAC) outperforms triplanar display for use in 6-DOF IGS navigation tasks. In *Proceedings of the Hamlyn Symposium on Medical Robotics*. 104–105.
- [21] C. Ware and R. Arsenault. 2004. Frames of reference in virtual object rotation. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*. 135–141.
- [22] Z. Yaniv, F. Banovac, D. Lindisch, P. Cheng, K. Cleary, V. Watson, E. Wilson, H. Abeledo, E. Campos-Nanez, and T. Popa. 2009. Needle-based interventions with the image-guided surgery toolkit (IGSTK): From phantoms to clinical trials. *IEEE Trans. Biomed. Eng.* 57, 4 (2009), 922–933.

Received August 2018; revised July 2019; accepted September 2019