Simulating the Impact of Navigation Waypoint Accuracy on Soldier Performance

Dr. John Graybeal
U. S. Army DEVCOM C5ISR Center
UNITED STATES OF AMERICA

William Heath Sharp
Planned Systems International Inc.
UNITED STATES OF AMERICA

Dr. Emily Lasko
CACI International Inc.
UNITED STATES OF AMERICA

Kiran Bagalkotkar
U. S. Army DEVCOM C5ISR Center
UNITED STATES OF AMERICA

Dr. Jasmine Dang
U. S. Army DEVCOM C5ISR Center
UNITED STATES OF AMERIC

ABSTRACT

The maturation of multiple technologies has made it possible to display various types of augmented reality (AR) information in Soldier night-vision devices (NVD). However, augmenting NVD optics with additional information will only improve Soldier performance if the information is sufficiently accurate. Further research is needed to better understand how Soldiers will perform when provided with realistically imperfect information. The U.S. Army DEVCOM C5ISR Center researched how Soldiers perform when NVD systems contain realistically imperfect navigation assistance. In such cases, Soldiers will need to leverage both their own navigational skills and system guidance to maximize their efficiency. In this experiment, 40 U.S. Army Soldiers navigated through a simulated urban environment using two desktop monitors. The primary monitor displayed a simulated NVD sensor view. Soldiers were instructed to follow an assigned route through the urban environment, which was displayed to Soldiers via a simulated paper map on the secondary computer monitor. Soldiers were instructed to navigate through the simulated environment as quickly as possible without deviating from the assigned route. They completed 48 scenarios with four different types of AR assistance. In the first experimental condition, Soldiers navigated without any AR assistance. In the second, AR navigation waypoints were perfectly placed along the route and were visible in the NVD sensor feed. In the final two conditions, Soldiers received AR navigation waypoints, but the waypoints were displayed with either minor or more severe geospatial registration accuracy errors. This enabled evaluation of the potential effects of providing Soldiers with realistically imperfect information, relative to unassisted Soldier performance and to Soldier performance with perfect AR guidance. The experiment's primary metrics of Soldier performance included the time to complete each route and deviations from the assigned route. All AR navigation assistance was beneficial to Soldiers, but the benefits were progressively reduced by greater errors in geospatial registration.

Keywords: augmented reality, navigation, human performance, sensors, displays, perception

1.0 INTRODUCTION

The U.S. Army DEVCOM C5ISR Center is exploring how to improve future night vision devices (NVD). This includes exploring advanced capabilities that could be incorporated into future NVDs through virtual prototyping. Virtual prototyping involves using simulation to create an instance of technology for presentation and/or evaluation. This paper describes efforts to leverage virtual prototyping to simulate the effects of adding augmented reality (AR) navigation information to a NVD sensor in order to evaluate the extent to which that information would be beneficial.

This quantitative evaluation used a virtual navigation simulation and a simulated AR navigation overlay. This type of AR overlay would allow Soldiers to see relevant information inside the NVD sensor's field of view (FOV) when completing navigation tasks. If successful, this type of AR overlay could improve Soldier navigation performance without disrupting other Soldier tasks.

However, many factors impact the quality of AR information. For AR information to be beneficial to Soldiers, it must be timely, relevant, and accurate. If the quality of the AR information is sufficiently poor in any of these (or other) dimensions, Soldier performance will not improve and may even decline. Consequently, engineers hoping to utilize AR information need to consider the accuracy of the information they provide to Soldiers. The accuracy of navigation waypoints in an operator's display may change over time, and could be limited by any number of hardware, software, human, or environmental factors. Consequently, this experiment examines the effects of spatial inaccuracies in the registration of navigation waypoints in a simulated NVD display.

We performed a system test of the virtually prototyped NVD AR overlay. The experiment gathered performance data from 40 Soldiers who interacted with the simulated AR overlay and used it to navigate a series of virtual routes. The goal of this experiment was to examine the potential improvements to Soldier performance by providing navigation assistance from within the NVD. Furthermore, the experiment aimed to characterize the impact of realistic system errors on Soldier performance. In order to establish baseline performance metrics, Soldiers completed some of the routes without AR assistance. In other trials, Soldiers completed the routes with perfect AR guidance to characterize an upper limit for potential system benefits. In the majority of trials, Soldiers completed a prescribed route with imperfect navigation waypoints that appeared nearby, but not directly at, their intended location. Researchers subsequently analyzed the speed and accuracy of each Soldier's attempt to navigate the prescribed routes.

2.0 METHODOLOGY

2.1 General Concept

Each Soldier was tasked with navigating virtual routes in a simulated urban environment. Soldiers had a primary computer screen displaying the sensor feed of the simulated NVD. A secondary computer screen displayed a static map, simulating a "printed," paper map that Soldiers would carry with them. The static "printed" map depicted the prescribed route the Soldier needed to follow. Soldiers completed navigating routes both with and without the assistance of an AR overlay. This approach allowed Soldiers to gain a better understanding of how the information in the AR overlay influenced their navigation experience, and enabled researchers to collect quantitative metrics measuring how Soldier performance changed as an effect of providing AR information. In some experimental trials, the AR overlay system was realistically imperfect when displaying waypoints. Thus, in addition to evaluating how hypothetically perfect navigation support would influence task performance, researchers aimed to examine how potential system imperfections affected human performance. Specifically, the effects of imperfect AR navigation information were evaluated in relation to 1) baseline performance without AR assistance and 2) performance with perfectly accurate AR assistance. Soldiers received training on how to operate the system prior to experimental trials. Additionally, Soldiers were given adequate time to practice system controls. Soldiers completed the experimental trials over a two-day period.

2.2 Participants

Forty (40) Soldiers participated in the evaluation of the virtually prototyped AR navigation overlay across four groups of ten (10) Soldiers. Soldiers reported to our laboratory to participate in several quantitative and qualitative perception research experiments for a variety of systems under development. All quantitative and qualitative data were collected anonymously, and each Soldier was assigned an anonymous identification number, corresponding to all recorded data.

2.3 Procedure

2.3.1 Task Description

During each trial, Soldiers were given a pre-planned route through an urban environment that was drawn on a static "printed" map (see **Figure 1-1**). This map simulated a paper map that Soldiers would carry with them while navigating and was displayed on a secondary computer screen (see **Figure 2-1**). Soldiers were instructed to follow the route (depicted on the "printed" map) as accurately and efficiently as possible. Soldiers used a gaming controller to move through the environment and could only move along roads in the environment; they could not walk on the grass or go into buildings.

Along each route, Soldiers needed to complete a series of turns on their way to their destination, marked by a parked vehicle at the end of each route. Once Soldiers reached the vehicle, the trial ended. The next trial would not begin until the Soldier pressed a button on their controller. The pre-planned routes the Soldiers navigated were different on every trial.

Soldiers completed trials both with and without AR assistance. On baseline trials without AR assistance (No AR condition), Soldiers had to navigate routes with only the "printed" map as a guide. The printed map displayed "waypoints" using red dots on the map, which were all located at turns or the final destination (*i.e.*, the vehicle; see **Figure 1-1**). On trials with AR assistance, an AR overlay was displayed on their NVD sensor's FOV, including AR waypoints. Each waypoint was located at a corresponding turn on the "printed" map" and at the destination. On some trials, the location of these floating waypoints was *perfect* (Perfect Waypoint Accuracy condition), meaning the waypoint on the "printed" map and the waypoint in the simulated virtual environment were in the exact same location. In other trials, the location of these floating waypoints was *imperfect*, meaning the AR overlay waypoints in the simulated virtual environment were displaced by a realistic amount of system error compared to their correct positions depicted on the "printed" map. The imperfect AR waypoints could be off by varying amounts of translational error, in a random direction. However, the same amount and direction of translational error was applied to each waypoint within a given trial. To compare the effects of smaller versus larger translational errors, two AR conditions were defined. The High Waypoint Accuracy condition displayed AR waypoints with minor geospatial registration errors, while the Low AR Accuracy Condition displayed more severe geospatial registration accuracy errors. The exact magnitude of the translational errors varied within an AR condition in order to prevent Soldiers from learning patterns in the translational errors. However, both AR conditions varied the magnitude of translational errors, relative to the respective mean translational error, in the same way. In the Perfect Waypoint Accuracy condition, there was no translational error present in the AR overlay.

The AR overlay simultaneously displayed two sequentially numbered AR waypoints. When a trial began, the system displayed the first two waypoints along the route. When Soldiers approached the upcoming waypoint, the system automatically detected when the Soldier had reached it, which triggered several events. First, a "Waypoint Reached" message flashed across the screen. Subsequently, the system cycled waypoints, meaning that unnecessary information about waypoints was removed, existing information was updated, and new information was added. Specifically, the reached waypoint was deleted from the overlay, the second-nearest waypoint became the nearest waypoint, and information from the next upcoming waypoint was added to the display.

Soldiers also had the ability to "skip" a waypoint without reaching it. This was a necessary system feature that made it possible for Soldiers to interact with imperfect waypoints. This feature made it possible for Soldiers to receive progressive navigation information without walking close to each waypoint. This was useful if reaching an inaccurate waypoint would slow a Soldier down or require the

Soldier to deviate from the intended route. In some cases, the imperfect waypoints were off the road and unreachable. In such instances, Soldiers were instructed to skip the waypoint in order to move onto the next one. Lastly, if Soldiers were required to reach every inaccurate waypoint, the mere presence of inaccurate waypoints would confound the experiment's metrics of Soldier performance; forcing Soldiers to reach inaccurate waypoints would always increase the distance Soldiers were required to travel and would sometimes force them to leave the prescribed route.

Skipping a waypoint had the same effects as reaching one: the skipped waypoint was deleted from the overlay, the second-nearest waypoint became the nearest waypoint, and information from the next upcoming waypoint was added to the display. Soldiers could skip the nearest waypoint, but not the second-nearest waypoint. To skip a waypoint, Soldiers had to orient towards a waypoint (so that it appeared in their FOV) and pressed a button on their controller. Although this feature was added to primarily help Soldiers mitigate any unrealistic effects of working with imperfect waypoints, Soldiers were allowed to skip a waypoint whenever they found it useful, including when the system was providing perfect guidance.

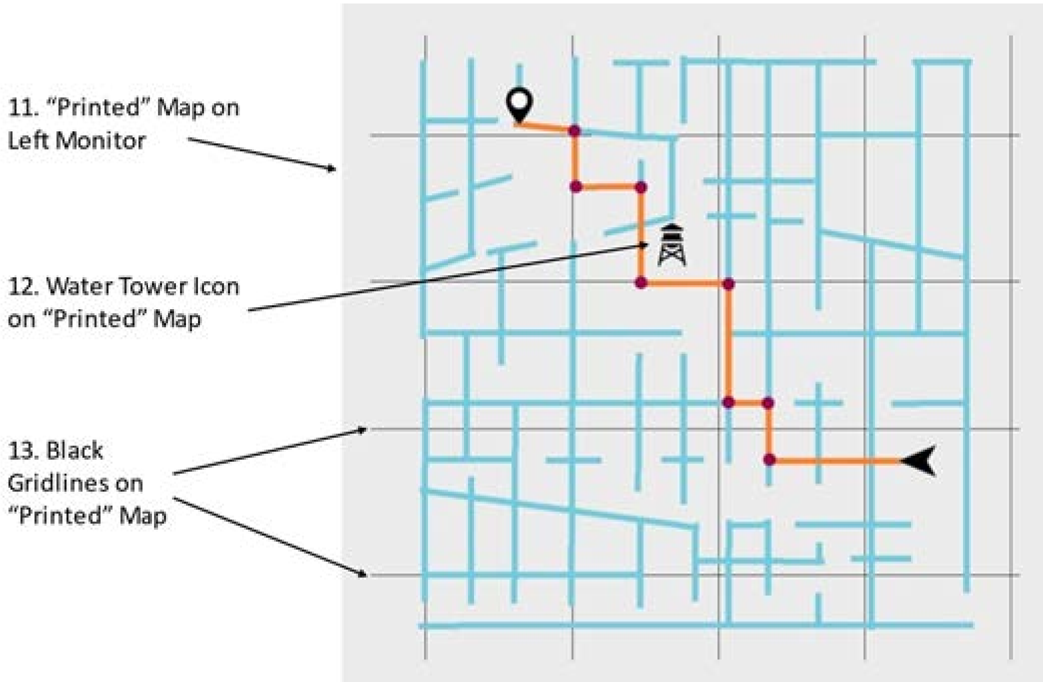

Figure 1-1: Example of "Printed" Map displayed to Soldiers.

Figure 2-1: Experimental set-up. A Soldier completes a navigation route by using two computer screens. The primary computer screen displayed the NVD sensor feed, and an adjacent, secondary computer screen displayed the "printed" map.

2.3.2 Test Procedures

The test was administered to 40 Soldiers; Soldiers were tested in groups of 10 across four different testing events. Each Soldier took the test over a two-day period to comply with schedule restraints. All testing took place in a dark room where each Soldier had their own individual workstation. On the first day, the Soldiers began by receiving a brief PowerPoint presentation covering the capabilities, functionality, and controls of the AR overlay navigation system. After completing the PowerPoint presentation, Soldiers were given the opportunity to train with the system prior to beginning the experiment. For practice and to gain efficiency using the system controls, Soldiers navigated through 12 different routes with and without the AR overlay. Their first three training trials were with perfect AR and the next three training trials were completed with no AR overlay. The remaining training trials presented imperfect AR waypoints, with the first half showing Low Waypoint Accuracy and the final three trials displaying High Waypoint Accuracy. All five map grids were used during training, distributed approximately evenly across the twelve trials. These training trials allowed Soldiers to use the system, learn the controls, and gain familiarity with the system and its features prior to beginning the experimental trials. It also gave Soldiers the opportunity to practice navigating routes and ensured each Soldier experienced the full range of possible system performance before beginning the experimental trials.

The experimental trials were broken into four separate blocks and were completed over two days. The first two blocks were completed on the same day as the initial training and the next two blocks were completed the following day. Each block consisted of twelve total trials presented in a random order; three with no AR, three with perfect AR, and six with varying levels of imperfect AR (*e.g.*, High and Low Waypoint Accuracy). In order to prevent Soldiers from learning a particular simulated environment, a total of five different map grids were used throughout the training and experimental trials. Each map grid was similar in overall structure in that they all consisted of a 12 by 12 block grid of roads. However, the specific layout, in terms of the location of dead ends and turns, was unique to each map grid such that each of the five map grids was distinct from one another (see **Figure 3-1** for two example map grids). Trials with different types of AR information (*i.e.*, no AR, perfect AR, and imperfect AR) were distributed evenly across the various underlying map grids so that participants could not predict the accuracy of the AR based on the underlying map grid. Map grids were also distributed across blocks of trials such that each of the five map grids appeared (approximately) equally frequently within each block of 12 trials; this was done to make it harder for Soldiers to learn the layout of the map grids.

2.3.3 Description of Maps, Environments and Routes

A total of five different map grids were created for training and testing. Each map began as a full 12 by 12 grid of roads and was then adjusted to create dead ends, diagonal roads, etc. so that each map was unique. Twelve unique routes were then created for each of the five maps. To create a route, a random number generator was used to provide random start and end points. The random number generator would provide two numbers between 1-12 for the start point, and again for the end point. For example, if the given start point was (2, 6), the start point would be located two blocks up and 6 blocks to the right of the bottom left corner of the map. Once these start and end points were plotted on the map grid, a route was drawn on the map connecting these two points. In order to improve realism, the routes were drawn such that they were the most direct route (*i.e.*, shortest distance) to get from the beginning to the end point. If the drawn route seemed too simple, as determined by the research team (*i.e.*, it contained too few turns or was too short), the start or end point was moved to either add more turns or increase the distance travelled. On average, there were 6.0 turns per route (range = 4-8) and approximately 13.7 grid blocks travelled (range = 10-19). These averages and ranges were approximately equal between the routes on all five map grids. For some underlying grid structures, route dead ends were intentionally placed adjacent to prescribed routes, while for other grid structures, dead ends were randomly placed before the routes were created; this resulted in routes that varied in difficulty. See **Figure 3-1** for an example of two created map grids.

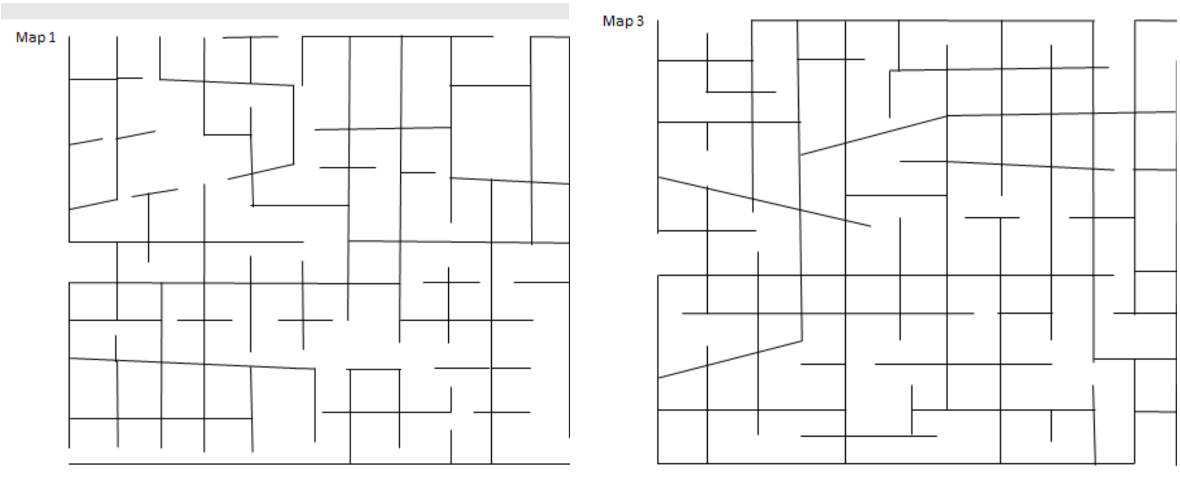

Figure 3-1: Two example map grids illustrating similarities and differences.

2.3.4 Simulation Construction

The simulation scenarios were constructed using government-owned software. A 32-inch (29.5" x 17.8") 4K monitor displayed the sensor FOV, while the map was displayed on a smaller 1600 by 1200 resolution monitor. Soldiers navigated through the scenarios by using a commercially available gaming controller.

2.5 Measures

2.5.1 Human Performance

Human performance data metrics were used to evaluate the speed of Soldier routes, the accuracy of their navigation paths, and their dismissal of waypoints. The *route completion time* metric was defined as the number of seconds from the beginning of a trial to the time they reached the end point of the route. Two metrics were calculated to quantify route deviations: the *amount of route deviation* and *percentage of time off route*. As the Soldier progressed through the virtual scenario, their position was continually logged and checked against the prescribed path. The *amount of route deviation* metric describes the distance in meters from the center of the prescribed path. The scenario roads all had a width of 10 meters, so Soldiers could walk on either side of the invisible line used for scoring; small deviations indicated the user was on the correct path (but not traveling in the center of the road) and larger deviations indicated the Soldier had taken a wrong turn. Soldiers were considered "off route" when their deviance from the prescribed path exceeded the width of the road (*i.e.*, deviations of greater than 5.5 meters from the center of a road 10 meters wide). Consequently, Soldier position was continuously categorized as either correct or incorrect and the *percentage of time off route* variable reflects the percentage of time Soldiers were classified as "off route" during a given trial.

For each of these dependent variables, the research team compared performance in the Perfect, High, and Low Waypoint Accuracy conditions against the No AR condition; this allowed the research team to assess whether the AR system improved Soldier performance at each accuracy level. The imperfect waypoint accuracy conditions (*i.e.*, High and Low Waypoint Accuracy), were also compared to the Perfect Accuracy condition via contrast analyses in order to assess whether inaccuracies caused a loss of potential system benefits. Furthermore, in order to approximate Soldier performance with a system that varies over time (as a real system might), subsequent analyses pooled on trials where the AR system was on and compared the aggregate result against the No AR condition. This aggregate analysis represents a system that performed with Perfect Waypoint Accuracy a third of the time, High Waypoint Accuracy a third of the time, and Low Waypoint Accuracy a third of the time.

In order to conduct analyses of all AR waypoint conditions against the No AR condition, a linear mixed-effects model was used. A second linear mixed-effects model was used to compare performance in the two imperfect AR conditions against performance in the Perfect Waypoint Accuracy condition. A final model was used to test the aggregated AR conditions against the No AR condition. Main effects were tested with nested-model comparisons.

The final metric described Soldier interactions with the waypoints. The *percentage of waypoints skipped* describes the number of waypoints a Soldier skipped out of the total number of possible waypoints contained within a collection of scenarios (including the waypoints marking the end of the scenario). More specifically, the total number waypoints skipped was computed for each Soldier for each scenario to derive average percentages per experimental condition. Although this metric describes human use of the presented navigation information, the metric is somewhat difficult to interpret in that some routes with imperfect AR required Soldiers to skip a waypoint (because it was unreachable). Additionally, because Soldiers were required to press a button for the system to register a "skipped" waypoint, it is also possible that Soldiers ignored or circumvented some waypoints in other ways that are not reflected in the logged data (*e.g.*, getting lost and subsequently going straight to the end of the scenario instead of interacting with the waypoints). Given that waypoints were not present in the No AR condition (and thus could not be "skipped"), analysis of the *percentage of waypoints skipped* followed similar data analysis procedures except only a single linear mixed-effects model was used. This model compared performance in the two imperfect AR conditions to performance in the Perfect Waypoint Accuracy condition.

3.0 RESULTS

3.1 Human Performance

3.1.1 Route Completion Time

The level of accuracy with which the AR overlay displayed the waypoints significantly affected the amount of time Soldiers took to complete the routes, F (3, 1825.5) = 67.27, p < .001 (**Figure 4-1**). Soldiers completed the routes significantly faster in each of the conditions in which they received AR guidance (Low Waypoint Accuracy, High Waypoint Accuracy, and Perfect Waypoint Accuracy) compared to when the AR overlay was off (all p's < .001, **Table 2-1**). Perfect Waypoint Accuracy resulted in a 19.1% reduction in the time it took for Soldiers to complete the average route, whereas High and Low Waypoint Accuracy trials showed a 14.4 and 7.1% reduction, respectively. Additionally, comparing the two imperfect AR conditions to the Perfect Waypoint Accuracy condition revealed the observed increases in Soldiers' completion times as the accuracy of the AR waypoints decreased were statistically significant (all p's < .001, **Table 3-1**). Comparing the aggregated performance of all three AR conditions to the No AR condition revealed that Soldiers took a significantly longer time to complete the route when the AR overlay was off (M = 107.54, SD = 36.26) compared to when they received guidance of varying quality from the AR system (M = 92.97, SD = 21.39), B = -14.89, *95% CI* [-17.50, -12.29], p < .001 (**Figure 5-1**).

Table 1-1: Means (M), standard deviations (SD), and percent (%) change of the route completion times (in seconds) aggregated across participants and routes for each AR condition.

AR Level	*M*	*SD*	*% Change*
NO AR	107.54	36.26	N/A
LOW ACCURACY	99.95	24.60	-7.1%
HIGH ACCURACY	92.10	20.34	-14.4%
PERFECT ACCURACY	87.00	16.55	-19.1%

Table 2-1: Model estimates from the linear mixed effect model comparing average route completion time in each of the AR accuracy conditions to the baseline No AR condition.

Parameter	*B*	*95% CI*	*t*	*df*	*p*
AR Level [LOW ACCURACY]	-7.85	[-10.96, -4.74]	-4.95	1864	<.001
AR Level [HIGH ACCURACY]	-15.83	[-18.93, -12.72]	-10.00	1864	<.001
AR Level [PERFECT ACCURACY]	-20.86	[-23.96, -17.77]	-13.23	1864	<.001

Table 3-1: Paired comparisons between each of the AR accuracy conditions.

Contrast	*Diff*	*95% CI*	*SE*	*df*	*t*	*p*
HIGH ACCURACY- PERFECT ACCURACY	5.04	[0.95, 9.13]	1.55	1827.12	3.26	<.001
LOW ACCURACY- HIGH ACCURACY	7.97	[3.86, 12.08]	1.56	1827.09	5.12	<.001
LOW ACCURACY- PERFECT ACCURACY	13.01	[8.91, 17.11]	1.55	1827.12	8.39	<.001

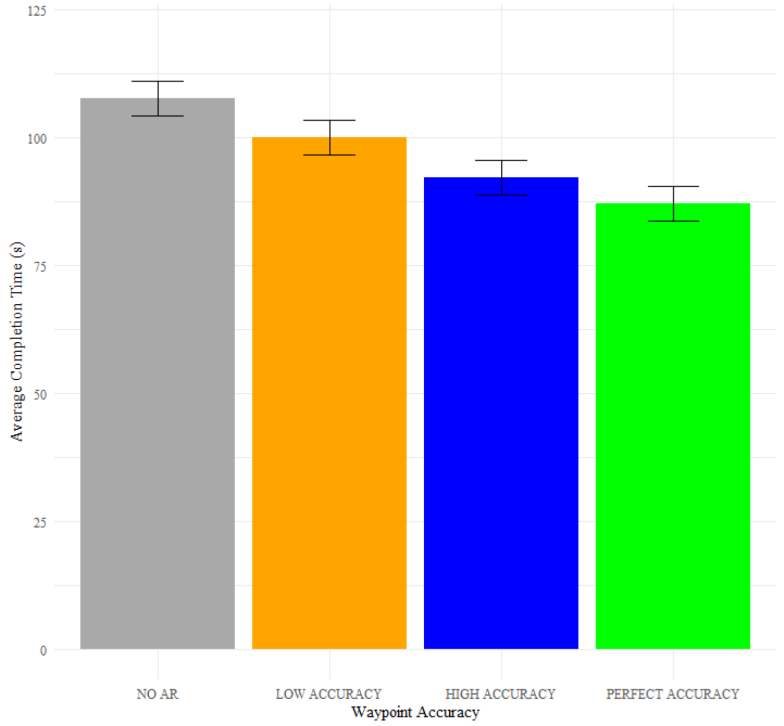

Figure 4-1: Bar plot displaying increasingly faster route completion times at varying levels of AR accuracy (Low, High, and Perfect Waypoint Accuracy) compared to the No AR condition.

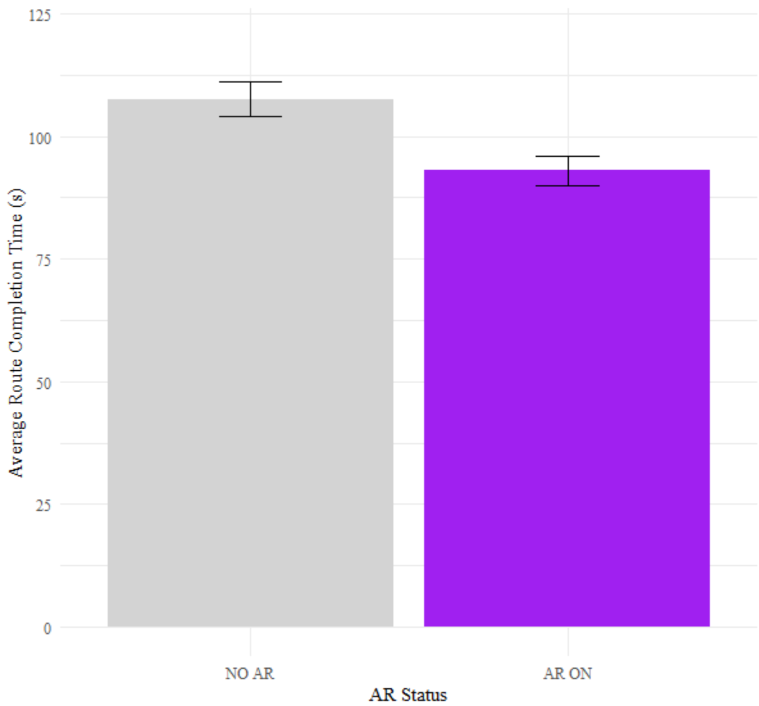

Figure 5-1: Bar plot displaying comparison of route completion time as a function of whether the AR overlay was on or off.

3.1.2 Route Deviation

3.1.2.1 Amount of Route Deviation. The level of AR accuracy had a statistically significant effect on Soldiers' measured deviance from the prescribed route ($p < .001$). When the AR overlay was off, Soldiers deviated significantly further from the prescribed route compared to each of the conditions in which they received AR guidance (all p's < .001; **Table 5**). The Perfect Waypoint Accuracy condition represented a substantial 81.9% reduction in measured deviance from the route, whereas the High and Low Waypoint Accuracy conditions resulted in a 77.7% and a 67.2% reduction, respectively. Again, the lost benefits due to system error were statistically significant, as the distance that Soldiers deviated from the prescribed route was significantly higher in each of the imperfect AR conditions compared to the Perfect Waypoint Accuracy condition (**Figure 6-1**, **Table 6-1**). Aggregating performance from the three AR accuracy conditions revealed participants deviated from the prescribed route significantly more in the No AR condition (M = 15.26, SD = 34.79) compared to when it was on (M = 3.77, SD = 13.44), B = -10.99, *95% CI* [-11.06, -10.92], $p < .001$. (**Figure 7-1**). It should be noted that the average amount of route deviation in the Low, High, and Perfect Waypoint Accuracy conditions does not actually reach the "off route" threshold as defined in the following section by the proportion of time off route variable (*i.e.*, 5.5 meters). These numbers simply reflect the amount of deviation *from the center of the correct path*.

Table 4-1: Means (M), standard deviations (SD), and percent (%) change in the amount of route deviation (in meters) aggregated across participants and routes for each AR condition.

AR Level	*M*	*SD*	*% Change*
NO AR	15.26	34.79	N/A
LOW ACCURACY	5.00	16.43	-67.2%
HIGH ACCURACY	3.40	12.34	-77.7%
PERFECT ACCURACY	2.76	10.28	-81.9%

Table 5-1: Model estimates from the linear mixed effect model comparing average route deviation in each of the AR accuracy conditions to the baseline No AR condition.

Parameter	*B*	*95% CI*	*t*	*df*	*p*
AR Level [LOW ACCURACY]	-9.81	[-9.90, -9.73]	-227.17	1765196	<.001
AR Level [HIGH ACCURACY]	-11.48	[-11.57, -11.40]	-259.43	1765196	<.001
AR Level [PERFECT ACCURACY]	-11.81	[-11.90, -11.72]	-263.89	1765196	<.001

Table 6-1: Paired comparisons between each of the AR accuracy conditions.

Comparison	*Diff*	*95% CI*	*SE*	*df*	*t*	*p*
HIGH ACCURACY- PERFECT ACCURACY	0.33	[0.20, 0.45]	0.05	1999920.27	7.00	<0.001
LOW ACCURACY- HIGH ACCURACY	1.67	[1.55, 1.79]	0.05	1557958.38	36.61	<0.001
LOW ACCURACY- PERFECT ACCURACY	2.00	[1.88, 2.12]	0.05	1765160.91	43.41	<0.001

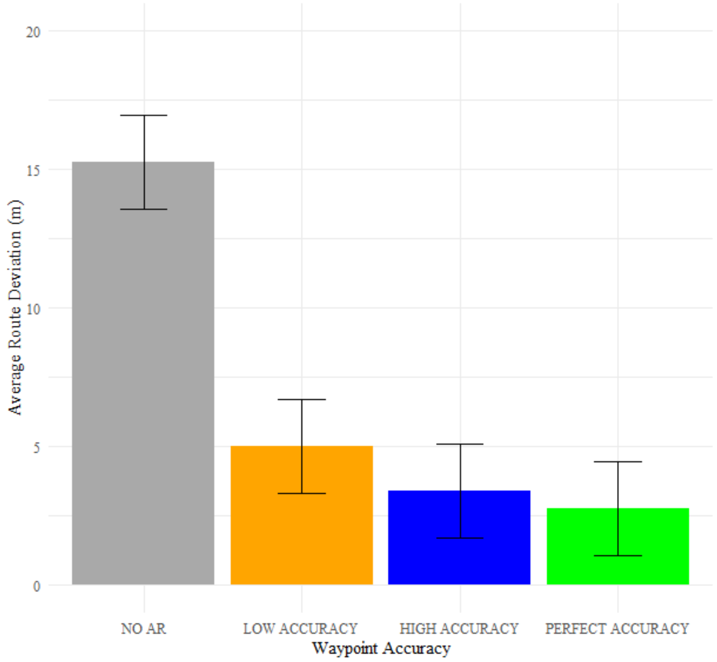

Figure 6-1: Bar plot displaying the average number of meters off the prescribed route Soldiers deviated by AR condition.

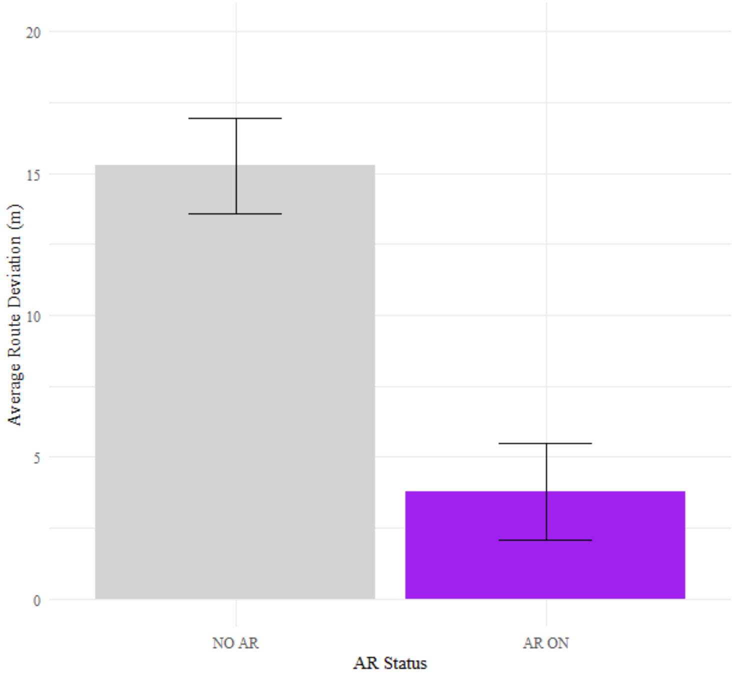

Figure 7-1: Bar plot displaying the average number of meters off the prescribed route Soldiers deviated as a function of whether the AR overlay was on or off.

3.1.2.2 Percentage of Time Off Route. Because the width of the path was 10 meters, participants were considered "off route" if they deviated from the center of the prescribed route by 5.5 meters or more. Compared to the No AR condition, Soldiers spent significantly less time deviated from the prescribed route during each of the conditions in which they received guidance from the AR system (all p's < .001, **Table 8-1**). Perfect Waypoint Accuracy resulted in a 66.1% reduction in the PERCENTAGE of Time Off Route, compared to smaller but similar reductions in the High and Low Waypoint Accuracy conditions (48.0% and 46.1% reductions, respectively). The PERCENTAGE of time spent off the prescribed route in the two imperfect accuracy conditions was also significantly different from performance in the Perfect Waypoint Accuracy condition, indicating that these reduced benefits due to system error are unlikely to be due to chance variance in Soldier performance (**Figure 8-1**, **Table 9-1**). Further, as expected, Soldiers spent significantly more time off the prescribed route in the No AR condition (32.27%) compared to the aggregated performance observed in all three AR conditions, $B = -19.68$, *95% CI* [-19.72, -19.64], $p < .001$ (Figure 10). When receiving AR guidance, participants spent an average of only 13.08% of the time off route (collapsed across AR accuracy conditions), compared to 32.27% when the AR system was off.

Table 7-1: Means (M), standard deviations (SD), and percent (%) change in the percentage of time Soldiers spent off the prescribed route.

AR Level	*M*	*SD*	*% Change*
NO AR	32.27%	14.38%	N/A
LOW ACCURACY	17.41%	11.41%	-46.1%
HIGH ACCURACY	16.89%	16.76%	-48.0%
PERFECT ACCURACY	11.05%	7.13%	-66.1%

Table 8-1: Model estimates from the linear mixed effect model comparing average percentage of time off route in each of the AR accuracy conditions to the baseline No AR condition.

Parameter	*B*	*95% CI*	*t*	*df*	*p*
AR Level [LOW ACCURACY]	-14.27	[-14.34, -14.21]	-440.90	206099	<.001
AR Level [HIGH ACCURACY]	-18.99	[-19.08, -18.91]	-447.85	206099	<.001
AR Level [PERFECT ACCURACY]	-20.89	[-20.99, -20.80]	-450.85	206099	<.001

Table 9-1: Paired comparisons between each of the AR accuracy conditions.

Contrast	*Diff*	*95% CI*	*SE*	*df*	*t*	*p*
HIGH ACCURACY- PERFECT ACCURACY	1.90	[1.75, 2.05]	0.06	206064.73	33.59	<0.001
LOW ACCURACY- HIGH ACCURACY	4.72	[4.60, 4.85]	0.05	206064.82	100.03	<0.001
LOW ACCURACY- PERFECT ACCURACY	6.62	[6.49, 6.76]	0.05	206067.61	130.31	<0.001

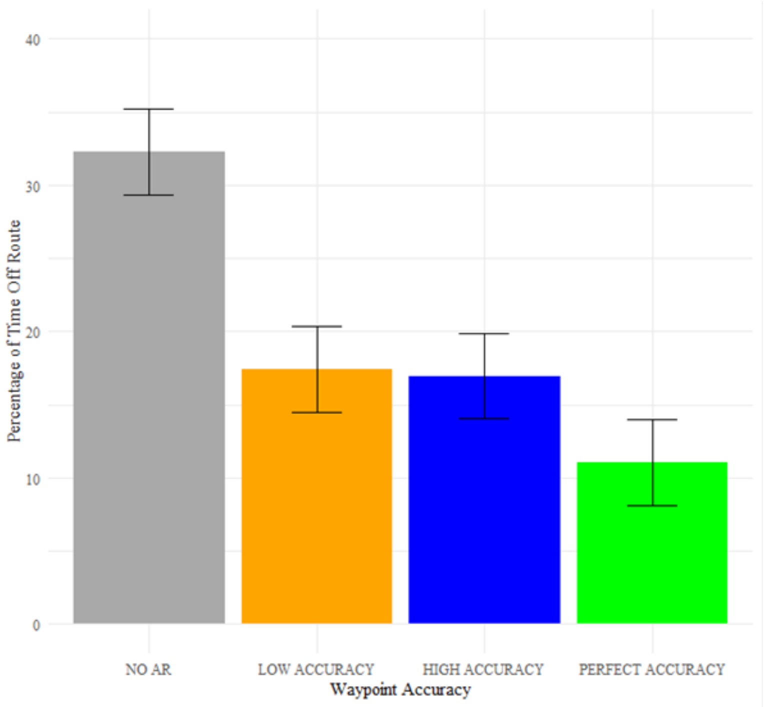

Figure 8: Bar plot displaying the average percentage of time Soldiers spent off the prescribed route for each AR condition.

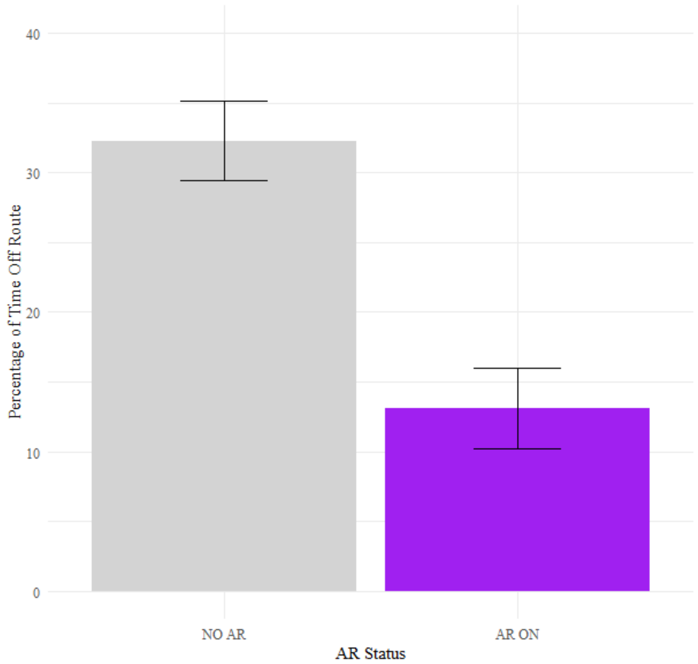

Figure 9-1: Bar plot displaying the average percentage of time Soldiers spent off the prescribed route as a function of whether the AR overlay was on or off.

3.1.3 Waypoints Skipped

3.1.3.1 Percentage of Waypoints Skipped. Consistent with the previous patterns, the level of AR waypoint accuracy (Low, High, or Perfect Waypoint Accuracy) had a statistically significant effect on the percentage of waypoints Soldiers skipped to reach the end of the route, F (2, 3731.51) = 180.25, $p < .001$ (**Figure 10-1**). Compared to the Perfect Waypoint Accuracy condition, the percentage of skipped waypoints was significantly higher in the Low Waypoint Accuracy condition, B = 13.57, *95% CI* [11.78, 15.36], $p < .001$. Similarly, Soldiers skipped significantly more waypoints in the Low Waypoint Accuracy condition compared to the High Waypoint Accuracy condition, B = 12.24, *95% CI* [10.45, 14.03], $p < .001$. The High Waypoint Accuracy and Perfect Waypoint Accuracy conditions did not significantly differ from one another, B = 1.32, *95% CI* [-0.54, 3.19], p = .164. See **Figure 10-1** for a visual depiction of these comparisons.

Table 10-1: Means (M), standard deviations (SD), and percent (%) change in the percentage of waypoints Soldiers skipped in each AR accuracy condition.

AR Level	*M*	*SD*
LOW ACCURACY	66.38%	17.48%
HIGH ACCURACY	53.17%	19.03%
PERFECT ACCURACY	45.77%	22.17%

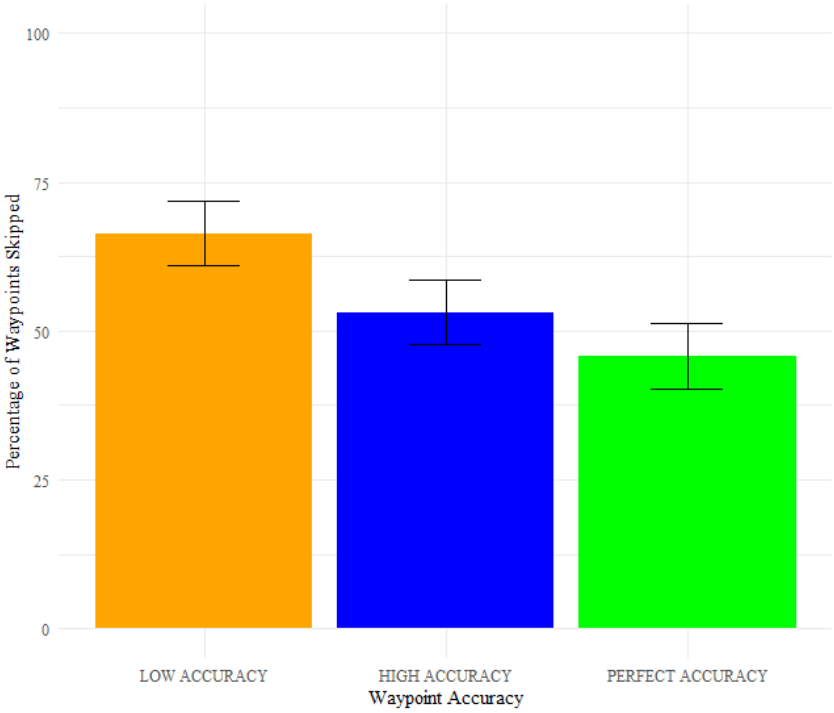

Figure 10-1: Bar plot displaying the average percentage of skipped waypoints for each AR condition aggregated across participants and routes.

4.0 DISCUSSION

Considering all the objective performance data, it is notable that all measures of Soldier performance changed as function of the AR overlay information that was or was not provided. The Soldier paths through the virtual environment were slowest and the least accurate when Soldiers were unassisted and were forced to navigate with only a map; unsurprisingly, their performance was best with the guidance of perfectly accurate waypoints. Compared to navigating without AR assistance, perfect AR waypoints reduced the amount of time it took for Soldiers to complete the routes by 19.1%. Likewise, with perfect waypoint guidance, the physical distance Soldiers deviated from the intended route decreased by 81.9% and the percentage of their navigation time where they were considered off the intended path decreased by 66.1%. These improvements to human performance are substantial and indicate that meaningful improvements to human performance are possible in situations resembling our simulated scenarios.

However, the system's benefits to Soldier navigation performance were reduced as Soldiers received information that was progressively less accurate. With high waypoint accuracy, Soldiers experienced a more modest reduction in route completion time of 14.4%, and substantially reduced improvement of only 7.1% with low waypoint accuracy. However, the physical accuracy of Soldier navigation paths was still greatly improved with imperfect guidance, as measured Soldier route deviations were reduced by 77.7% and 67.2% with high and low accuracy waypoints, respectively. This suggests that the impact of spatial errors in navigation waypoints is not uniform. Although the study only tested three different levels of AR waypoint accuracy, it appears that as Soldiers received worse and ultimately no guidance, the time it took to complete the route increased linearly while their deviance from the intended route increased non-linearly.

In reality, system performance may vary across time due to a variety of changing factors. This experiment intentionally varied the performance of the system randomly from trial to trial so participants could not learn patterns in the accuracy of the AR system errors or how to unrealistically compensate for them. To approximate how Soldier performance with a system with varying levels of accuracy might compare, the presented analyses aggregated data from all trials with AR guidance and compared the aggregate result to unaided performance; this comparison approximates a system that provided an equal mix of perfect, high and low accuracy waypoints. For all measures of Soldier performance, aggregating all trials with varied levels of AR guidance still represented a statistically significant improvement over unaided performance.

In summary, although all forms of AR guidance resulted in statistically significant improvements over unaided performance, the two imperfect AR waypoint conditions always resulted in a reduction in benefits (compared to perfect guidance) that were also statistically significant. Thus, we can conclude that spatial registration waypoint errors similar to those portrayed in this study are likely to have a detrimental impact on Soldier navigation performance in environments in similar navigation scenarios. Designers of physical systems should endeavor to make those systems as accurate as possible and to test whether the level of accuracy practically achievable on the battlefield is sufficient to provide meaningful assistance.

Although this experiment provided many insights, all experiments have limitations. One limitation of the experiment was the fidelity of the simulation; navigating a virtual urban environment is an imperfect representation of real-world land navigation. The virtual environment did not contain all elements of the real world (*e.g.*, the virtual environment was silent) and the virtual environment presented limitations that do not exist in the real world (*e.g.*, not being able to move off road). Real world and virtual navigation tasks require imperfectly overlapping skills, and these skills are influenced by differences between environments (*e.g.*, the lack of perceptual and proprioceptive feedback experienced in a virtual environment). Finally, it is worth noting that the navigation task was long and repetitive, which may have impacted Soldiers' motivation and subsequent performance as the experiment went on. This may have caused Soldiers to use the system differently than they would have in real world scenarios. These are all reasons why the quantitative estimates of human performance derived in this simulation may not generalize perfectly to real world navigation tasks.

Despite these limitations, virtual environments provide many advantages. In addition to providing the research team with better control and consistency in the environment and its simulated scenarios, the virtual environment also facilitated easy data logging and scoring that would have been more challenging in the real world (*e.g.*, measuring path deviations). Furthermore, the research team was able to collect a large volume of performance data from a total of 40 Soldiers during the 48 unique simulated nighttime scenarios; this resulted in data from a total of 1,920 navigation paths. Although collecting data at this scale in the field at night is theoretically possible, it is rarely feasible. Such field collections can also contain significant sources of measurement error due to test constraints (*e.g.*, varied levels of lighting both within and across routes). Consequently, this experiment represents an important first step at evaluating the benefits of navigation assistance with varying levels of AR Waypoint accuracy. The use of simulation and our virtual prototype also enabled the evaluation of a proposed system in an idealized environment earlier in the design cycle. Although quantitative experiments leveraging virtual reality contain many methodological challenges, more immersive simulations may better approximate land navigations skills.

A further limitation of the experiment was the study design did not enable collecting subjective Soldier feedback for each presented level of AR Waypoint accuracy. For example, successful technology adoption requires the user to trust the system. A consequence of varying the system's performance randomly each trial was clear measures of Soldier trust in the system at various performance levels could not be collected. This experimental design varied system performance across trials randomly to prevent Soldiers from learning patterns in the AR errors and from unrealistically compensating for them, but assessing trust requires sustained interactions over time. Future research could have Soldiers consecutively interact with a series of trials displaying a consistent level of waypoint accuracy to see how trust is affected by system errors.

5.0 CONCLUSION

The U.S. Army DEVCOM C5ISR Center completed a quantitative experiment simulating the use of an AR overlay navigation system within an NVD. The purpose of the experiment was to understand how the accuracy of AR waypoints affected the speed and accuracy of Soldier navigation attempts. This experiment was a critical first step towards understanding the impact of AR waypoint accuracy on Soldier performance. Soldiers navigated routes more accurately and faster with all levels of waypoint accuracy compared to when they received no AR guidance. However, the benefits of AR navigation waypoints were meaningfully reduced as the waypoints became less accurate. For example, although perfect waypoint guidance resulted in a 19.1% reduction in the time it took Soldiers to complete the routes, low accuracy waypoints resulted in only a 7.1% reduction. Further, the impact of waypoint accuracy on Soldier performance varied by metric; physical route deviations were reduced substantially by 67.2% with even low accuracy waypoints, although greater waypoint accuracy still resulted in even fewer route deviations. While this experimental design allowed Soldiers to experience a system where performance varied trial to trial, future tests could utilize extended, continuous use of systems that differ in their level of performance to facilitate the collection of subjective measures of trust.