Human-Automation Collaboration in Occluded Trajectory Smoothing

by

Jason M. Rathje

S.B. Aerospace Engineering
Massachusetts Institute of Technology, 2010

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics
at the
Massachusetts Institute of Technology

June 2010

Signature of Author: ______________________________
Jason M. Rathje
Department of Aeronautics and Astronautics
May 21, 2010

Certified by: ______________________________
M.L. Cummings
Associate Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by: ______________________________
Eytan H. Modiano
Associate Professor of Aeronautics and Astronautics
Chair, Committee of Graduate Students

This Page Left Intentionally Blank

Human - Automation Collaboration in Occluded Trajectory Smoothing

by

Jason M. Rathje

Submitted to the Department of Aeronautics and Astronautics
on May 21, 2010, in partial fulfillment of the requirements for the degree of
Master of Science in Aeronautics and Astronautics

ABSTRACT

Deciding if and what objects should be engaged in a Ballistic Missile Defense System (BMDS) scenario involves a number of complex issues. The system is large and the timelines may be on the order of a few minutes, which drives designers to highly automate these systems. On the other hand, the critical nature of BMD engagement decisions suggests exploring a human-in-the-loop (HIL) approach to allow for judgment and knowledge-based decisions, which provide for potential automated system override decisions.

This BMDS problem is reflective of the role allocation conundrum faced in many supervisory control systems, which is how to determine which functions should be mutually exclusive and which should be collaborative. Clearly there are some tasks that are too computationally intensive for human assistance, while other tasks may be completed without automation. Between the extremes are a number of cases in which degrees of collaboration between the human and computer are possible. This thesis motivates and outlines two experiments that quantitatively investigate human/automation tradeoffs in the specific domain of tracking problems.

Human participants in both experiments were tested in their ability to smooth trajectories in different scenarios. In the first experiment, they clearly demonstrated an ability to assist the algorithm in more difficult, shorter timeline scenarios. The second experiment combined the strengths of both human and automation to create a human-augmented system. Comparison of the augmented system to the algorithm showed that adjusting the criterion for having human participation could significantly alter the solution. The appropriate criterion would be specific to each application of this augmented system. Future work should be focused on further examination of appropriate criteria.

Thesis Supervisor: M.L. Cummings
Title: Associate Professor of Aeronautics and Astronautics

ACKNOWLEDGEMENTS

This thesis was not by any means a solo effort. From my two advisors to family and friends, I would not have been able to complete this without you.

First, a big thanks to both of my advisors. Professor Missy Cummings, for taking me on as a UROP only three years ago and allowing me to adopt my own research as a Senior. Thank you for trusting me and having the faith that I would accurately and completely accomplish this research.

Dr. Lee Spence, if it wasn't for you, I would not have had this amazing opportunity. You trusted Prof. Cummings and put this opportunity on my plate, and you have helped me every step of the way. We've had a great few years working on this project, and I'm still waiting for that boat ride!

To Lincoln Laboratory and all their employees. First, with blessing HAL and myself with this opportunity by funding this research and secondly, with putting up with our long and somewhat monotonous experiments! A special thank you to Allison Loftin for all your help with scheduling the experiments, making sure the documents where squared away, and cleaning up my thesis. To Michele Andriolo for helping me with the thesis figures and picking me up when my car broke down! To Hody Lambert for explaining to me how LMTS worked and running the test cases and Crystal Reichert for working through those long days making sure my experiment worked. And another special thank-you to the staff and librarians, for your immense help from getting all the security paper work done to assisting me in the direction of my research, which has been invaluable.

To Birsen Domez, the stats queen, I would not have been able to format my results and use the appropriate analysis without you. To Sylvain Bruni, if it was not for MUPPETS, I would have never been in this lab. To Carl, Brian, Farzan, Kris and Luca, you all were a blast working with and kept me sane during the hardest, most frustrating times.

To Yves and Ryan, we had a great time from Ryan's wedding to CrossFit to partying on HAL's budget! To the other HALiens, Christin, Andrew, Thomas, Tuko, Jackie, Armen, Jason, Americo, Jodyann, Anna, Geoff, Dave and Dave, thanks for all the good times.

To LCA, 1038, and all of my friends through out my MIT years. I want to give some extra special love to my class, LCA '09, in order: Jco, Bucket, Loper, Chotch, TFrank, Old Bill, JB, El Commodore, Hdog, BenG, Coffey, and in loving memory, Zach Morris. From the boat cruise to AC, we lived the dream. To Pat Shannon, thanks for the great times, you were my most trusted confidant through out these years. It's only going to get better.

To my best friend, Matt, and the rest of the crew, Chirag, Kaushal, Julie, Andy, Marc, Mike, Jon, Lodics, Leone, and all you other Jersyians. I know these months before the Air Force are going to be epic. To my fellow Air Force Cadets, my class of Buko, LG, KL, Henry, McKinney, and the one and only Eric Lanser. Now that I finally have my butter bars, you all are going to get promoted. I'll still salute!

To my family, who have supported me throughout my time at MIT. My grandparents, MomMom and PopPop, Grandma and Grandpa, who sent me care packages and letters to keep me motivated. To my Aunts and Uncles, especially my Aunt Lisa who would always give me that extra call to make sure I was doing OK!

To my parents, by far my biggest fans and platinum level supporters. I would not have been able to make it through MIT, high school, middle school, grammar school, or pre-school without you. Your support means the most to me. I love you both.

To my favorite brother and sister, Chris and Elizabeth. You two are growing up way too fast, but it is still one of the best feelings getting to come home and hang out with you two.

To Kathryn, you've been by far my best friend these past four years. I need say no more, except thank you for being the first person I call for both the good and the bad, and accepting me for me.

Most importantly, I thank God. The reason for all the good things in my life, it is He who brought me to MIT and He who got me through it. I'm looking forward to what You have lined up for me next.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

BMDS: Ballistic Missile Defense System

CI: Confidence Interval

HAL: Humans and Automation Laboratory

HIL: Human-In-the-Loop

HuGS: Human-Guided Search

LMTS: Lincoln Multi Target Smoother

MIT: Massachusetts Institute of Technology

MSE: Mean Squared Error

RMSE: Root Mean Squared Error

TAL: Time After Launch

UROP: Undergraduate Research Opportunities Program

1. Introduction

This chapter addresses the motivation for research into occluded trajectory smoothing. In this thesis, occluded trajectories, or incomplete trajectories that contain gaps, will be considered in the context of radar tracks in the Ballistic Missile Defense System (BMDS). However, such trajectories can be encountered in a number of other areas of tracking problems such as satellite tracking or air traffic control. This chapter first addresses particular issues involved in the BMDS that motivated this research and how these issues might be addressed by a collaborative effort between humans and automation. Next it presents the problem statement and the four research objectives addressed in this thesis. A brief description of the subsequent chapters concludes the introduction.

1.1. Motivation

The basic function of the BMDS is to protect a designated area, such as the continental United States, against ballistic missile attacks. These attacks can be characterized by scenarios that give launch points, numbers of objects launched, and targets. However, the number of objects launched may be much larger than the number of available missiles to defend against them so deciding if and which objects should be engaged becomes a more complex issue. The system is very large as it has many interconnected elements and is physically spread over an area that is a significant fraction of the Earth. The information for such decisions may be incomplete and/or inconclusive, and, given the enormity of the decision-making task, the timelines may be extremely short. For an intercontinental ballistic missile, the flight time is on the order of half a hour [1]. The magnitude of the task and the short timelines drive designers to highly automate these systems because of the computational speed, repeatability, and high consistency that automated systems provide. On the other hand, the grave nature of BMDS engagement decisions suggests exploring a human-in-the-loop (HIL) approach to exploit the judgment and knowledge-based decisions that humans can provide [2], which could allow the humans to override automated system decisions when anomalies are apparent.

This BMDS problem is representative of the role allocation conundrum faced in many supervisory control systems, which is how to determine which functions should be mutually exclusive between the human and automation, and which should be collaborative [3]. There are some tasks that are simply too fast or computationally intensive for humans to make useful contributions, especially in time-pressured environments. An example of such a task in the BMDS is real-time target tracking computations. On the other hand, there are tasks that are tractable by humans in the available time, and these may be completed without automation or with basic computer assistance. An example of such a task in the BMDS is in setting defended target priorities. Between these extremes are a number of cases in which degrees of collaboration between the human and computer are possible.

Because humans can reason inductively and generate conceptual representations based on both abstract and factual information, they also have the ability to make decisions based on qualitative and quantitative information [4]. In addition, allowing operators active participation in decision-making processes provides not only safety benefits, but promotes situational awareness and also allows a human operator, and thus a system, to respond flexibly to uncertain and unexpected events (as opposed to the brittleness of many algorithms). Thus, decision support systems that leverage the collaborative strength of humans and automation in supervisory control planning and resource allocation tasks could provide substantial benefits, in terms of both human and computer impacts on system performance.

To make these notions of human-automation collaboration more concrete, this thesis motivates and discusses two experiments that quantitatively investigated human/automation role allocation tradeoffs in the specific domain of trajectory smoothing. The need for smoothing occurs when, due to normal processing errors, trajectories do not appear as continuous curves but as segments and thus create ambiguity about how the segments should be connected. Computers running predictive algorithms could be relied upon for proper connection of such track segments. However, this task has clear vision-based pattern recognition elements, so it is possible that human operators could perform as well, or better than, the automation. The possibility that humans can make contributions to these problems has support in a number of studies which investigated the ability of humans to perceive lines in data that is incomplete or has the appearance of being occluded [5]. The experiments assessed how well humans and a specific algorithm, known as the Lincoln Multi Target Smoother (LMTS), described in Chapter 2, compare in the task of correlating track segments that offer cases of varying difficulty. The ultimate goal is to determine an empirically-based rationale for a collaborative human-automation track correlation decision support system.

1.2. Problem Statement

The problem is to determine the best way to perform the trajectory smoothing task. Both human and the LMTS, collectively referred to as the decision sources in this thesis, each have a set of comparative strengths that can potentially be combined to output a better solution in the track smoothing environment than either acting alone.

1.3. Research Objectives

In order to address the problem statement, the goal of this research is to understand both human and algorithm strengths and explore the possible areas of collaboration. This goal is addressed through the following objectives.

- Objective 1. **Study the various ways in which LMTS and human participants smooth trajectories.** In order to achieve this objective, the way in which LMTS operates is investigated. Next, the prior research into the ability for humans to perceive and interpolate occluded contours was reviewed. This information is described in detail in Chapter 2.

- Objective 2. **Study and assess if human participants can outperform the LMTS algorithm in smoothing occluded trajectories.** Experiment One, discussed in Chapter 3, addresses whether a human operator, presented with the same data as the LMTS, can give more accurate, smooth trajectories than the algorithm. The results from that experiment are given in Chapter 4.

- Objective 3. **Analyze to what degree participant confidence correlates with accuracy in trajectory smoothing.** Experiment One also addresses the confidence that the subjects had in their solutions by asking them to estimate confidence for each smoothed trajectory. Since confidence may play an important role in an actual application of this research, it is useful to determine if participant performance would correlate with confidence. These results are also discussed in Chapter 4

- Objective 4. **Study a collaborative effort between human participants and the LMTS**. Experiment Two, discussed in Chapter 5, was designed to investigate whether a collaborative effort can produce a better solution than LMTS acting alone. This experiment builds on the results of Experiment One. The subsequent results and evaluation of the second experiment are discussed in Chapter 6.

1.4. Outline

This thesis is organized into seven chapters:

- Chapter 1, *Introduction*, provides the motivation for this research, the problem statement, and the research objectives.

- Chapter 2, *Background*, addresses objective one by providing a summary of the information used to create and interpret the two experiments. It explains Lincoln's Multi Target Smoother (LMTS) algorithm, discusses how the human visual system interprets and connects occluded contours, and details possible areas for collaboration between humans and automation.

- Chapter 3, *Experiment One Design*, explains the initial experiment design and addresses the second and third research objectives. It outlines the hypotheses, participants, apparatus, experimental design, and subsequent testing.

- Chapter 4, *Experiment One Results and Discussion*, addresses the results of the initial track smoothing experiment and discusses their implications. The results are addressed in four categories, missed and false trajectories, accuracy of correct trajectories, total performance, and confidence. The discussion addresses the two hypotheses stated in Chapter 3 in the context of the results of the experiment.

- Chapter 5, *Experiment Two Design*, explains the second experiment's design and addresses the fourth research objective. It outlines the hypothesis, participants, apparatus, experimental design, and subsequent testing.

- Chapter 6, *Experiment Two Results and Discussion*, addresses the results of the second track smoothing experiment and discusses their implications. The results are addressed in three categories: missed and false trajectories, accuracy of correct trajectories, and total performance. The summary addresses the hypothesis stated in Chapter 5 using the results of the experiment.

- Chapter 7, *Conclusions and Future Work*, examines results from both experiments to suggest ways in which human operators can contribute to the track smoothing task. This chapter summarizes the rationale of a collaborative effort by discussing the cost and benefits of invoking human assistance in the track smoothing task. Future work is also discussed.

2. Background

This chapter discusses the background information used to create and interpret the experiments in this thesis. The Lincoln Multi Target Smoother (LMTS) algorithm, which is the algorithm that was used in this research effort for connecting and smoothing contours, is presented first. A contour is defined as a "continuous perceived boundary between regions of a visual image [6]," which in this case is defined by trajectories. The chapter then discusses how the human visual system interprets and connects occluded contours. While the occluded vision field of study is diverse, this research will only deal with the particular areas that apply to occluded track smoothing problems in supervisory control domains. Possible areas for collaboration between humans and automation are discussed in the final section.

2.1. The LMTS Algorithm

Many radars process both signature and metric data. Signature information is dependent on the tracked object's inertial properties, such as spin about its center of mass, while metric data are dependent upon kinematic data such as the object's trajectory. The metric data usually have six dimensions (position and velocity in three dimensions). In many current algorithms, data are processed sequentially on a pulse-by-pulse basis as they are received. While these algorithms tend to work well with widely-spaced targets in multi-target scenarios, noise and interference among closely spaced targets often disrupt pulse-by-pulse algorithm performance [7]. Sequential algorithms by design cannot backtrack to connect information that may be missed by poor tracking. In order to overcome this limitation, a batch model algorithm, the LMTS used in this experiment, was created in 2007.

Batch mode algorithms collect data pulse-by-pulse just as sequential algorithms would, but they store some set of data points for a certain predetermined period of time prior to processing. In this manner, a batch mode algorithm can collect a large amount of information that enables it to connect trajectories that may be separated in distance and time. LMTS operates by connecting partial trajectory segments through a series of ballistic fits [7]. Even though it functions at a high percentage of accuracy, some researchers at MIT Lincoln Laboratory noticed that in certain instances, the algorithm was not able to match all of the trajectories. One main source of failure is due to the step-by-step processes the algorithm uses. The algorithm connects partial trajectories only if they meet a chi-square test of fit to a ballistic trajectory. One downfall of this procedure is that if the partial trajectories fall within the chi-squared distribution, they are connected with no further consideration of their accuracy compared to other possibilities. The best fits are matched first, and then successively worse fits are matched. This difficulty is illustrated in Figure 2-1, which shows a plot of occluded radar information that could be processed by LMTS. The dashed box highlights an occluded region where the algorithm has calculated multiple, high probability fits to the partial trajectories on either side of the occlusion. In such cases, LMTS will always choose the highest probability. For example, if two possible trajectories can be calculated, say the probability of smoothing Trajectory 1 is $p_{T1} = .95$ and to smooth Trajectory 2 is $p_{T2} = .92$, Trajectory 1 will be formed instead of Trajectory 2.

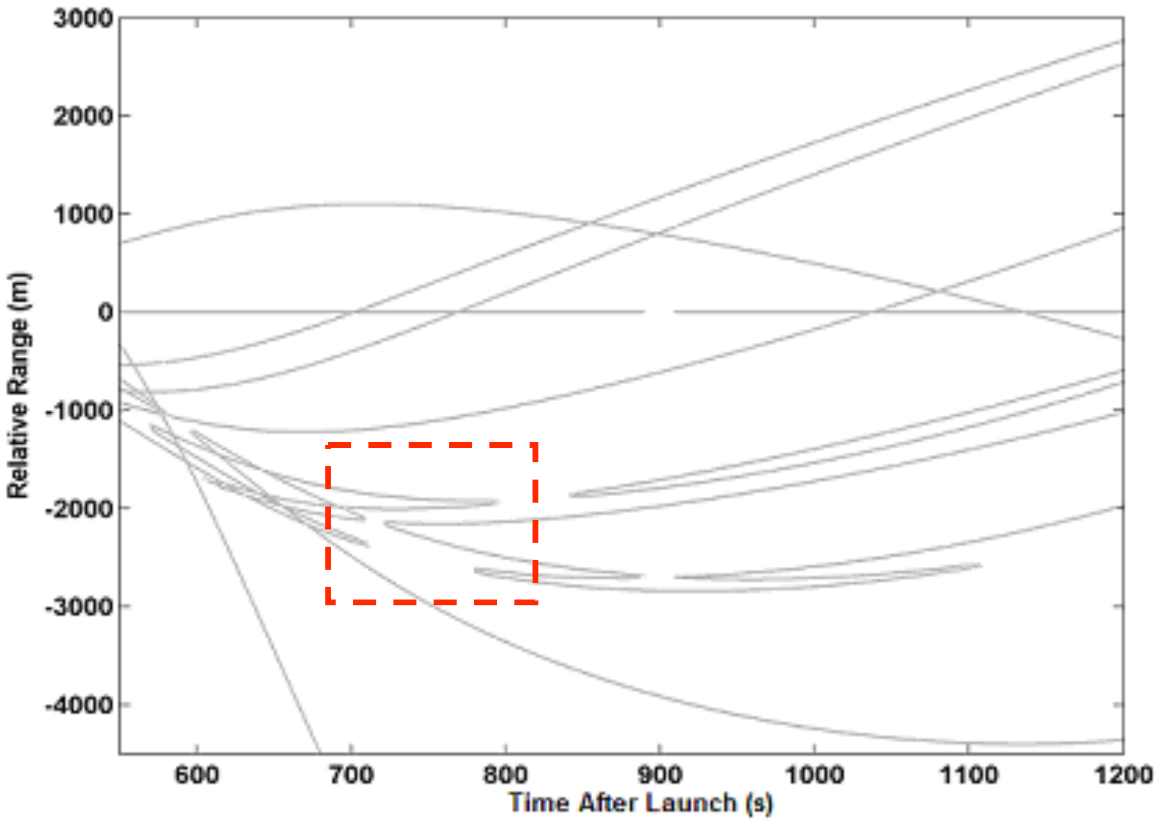
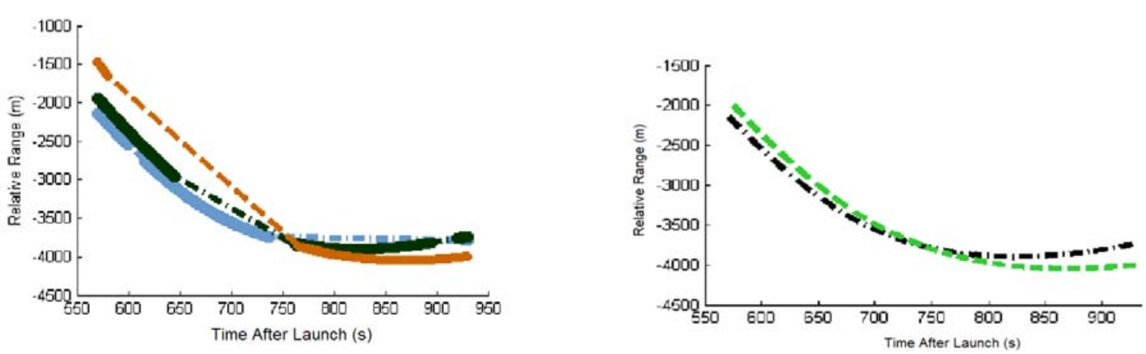


Figure 2-1: Trajectories occluded by a data drop out.

However, if the standard error of fit is large enough, the probability of Track 1 existing may not be significantly different from Track 2. In multiple test cases, it was shown that frequently two almost-equivalent trajectories were incorrectly plotted, one with a small amount of error and one with a large amount of error, as shown in Figure 2-2. These cases were most prominent for trajectories which were occluded over a point of crossing.



(a) *(b)*

Figure 2-2: Figure 2-2a depicts multiple partially occluded trajectories, in which the different colors represent the different connected trajectories by LMTS. The algorithm mistakenly connected the trajectories, which are shown correctly connected in Figure 2-2b.

These errors, while few, were obvious to the LMTS developers when the LMTS results were plotted. While it would likely take a complicated and computationally expensive algorithm to correct these errors and output a better solution, there is no guarantee it would be 100% accurate.

2.2. Human Performance in Track Smoothing

Because of potential errors in LMTS, the possibility that a human operator could assist to improve overall performance was considered in this thesis. This will be discussed in this section in terms of Gestalt theory and extrapolation, as they relate to occluded contour perception.

2.2.1. Gestalt Theory and Occluded Contour Interpolation

Trajectory smoothing can be interpreted as occluded contour interpolation given trajectories with incomplete segments. Gestalt theory studies human perception of occluded contours and contains specific theories on how humans connect, or in this case smooth, those contours. There are a number of specific areas such as similarity, proximity, good continuation, and closure that are applicable to the smoothing problem. While these features do not prescribe the specific steps for the interpolation of occluded contours, they give some insight into how the human visual system works [8].

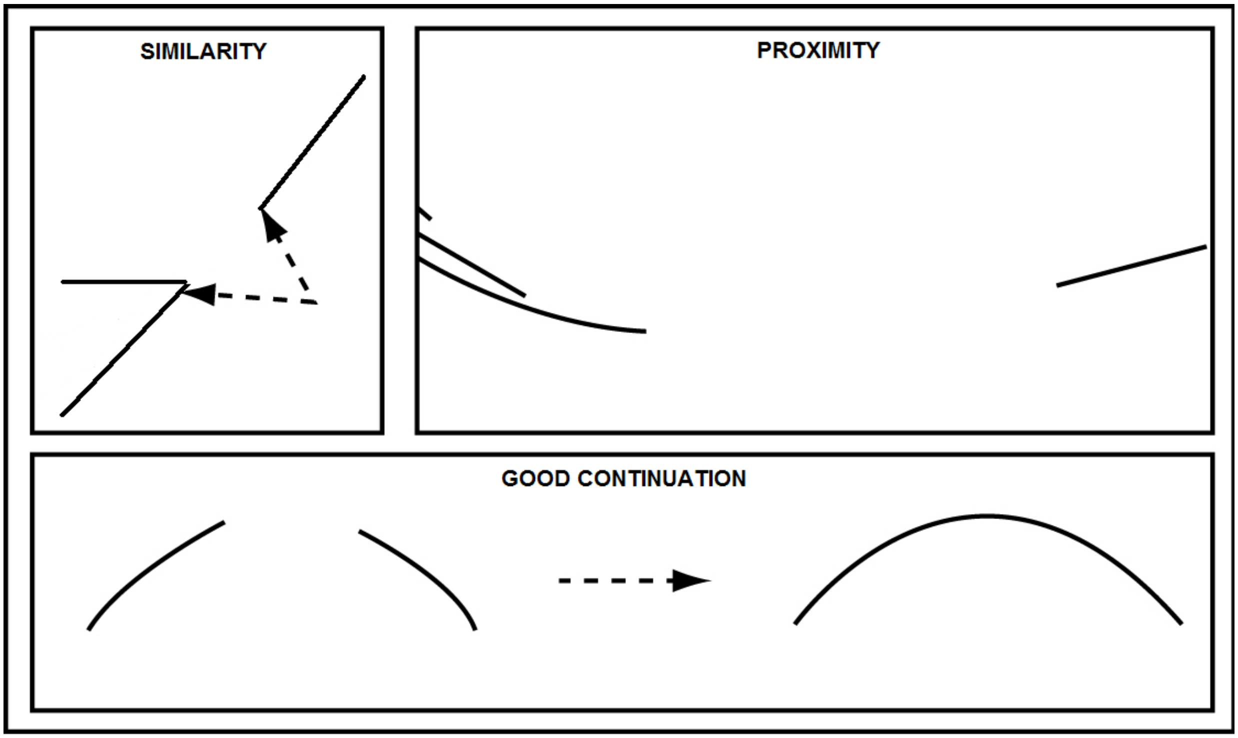


Figure 2-3: A representation of occluded trajectories. The 3 sections are representations of similarity, proximity, and good continuation (left, right and bottom section respectively).

Similarity implies that contour shape and orientation leads to a grouping together of patterns. In the case of trajectory smoothing, the similarity of partial trajectories allows the human to accurately interpolate them. The upper left section of Figure 2-3 shows how similarity plays an important role in interpolating the correct trajectory, which is a shown by the two arrows. Proximity maintains that items close together will be grouped more frequently than those farther apart. The upper right section of Figure 2-3 displays how the partial trajectories are much more difficult to interpolate when they are farther apart. Quantitatively, there may exist some distance over which proximity can filter out wrong choices in interpolation [9]. Furthermore, that range may exist but be substantially different depending upon whether a human or LMTS is trying to interpolate. Good continuation is the name of a Gestalt principle which states that objects arranged along lines will be visually grouped together. It can be seen as a combination of similarity and proximity, as it represents how patterns are simply a separation of whole figures. For example, a disconnected parabolic contour, which is representative of partial trajectories, does not appear as a two segments but as one continuous curve, as shown in bottom section of Figure 2-3. This effect is, however, dependent upon the proximity (they are close together) and similarity (they are mirror images of each other) of the segments.

Good continuation has been tested using occluded contours [5, 10, 11]. Kellman and Shipley showed that a disrupted line segment would be perceived as whole if 1) the linear extension of the two edges intersects and 2) the turning angle, or the angle at which the two edges intersect, does not exceed 90°. More recent research has challenged the last constraint by showing there may be no "hard cutoff" in the turning angle criteria, rather this results in a decrease in precision of interpolation [5]. This loss in precision could be remedied by ensuring that neighboring segments are co-circular (tangent to the same circle) [11]. Additionally, there has been a limited amount of research showing that parabolas, similar to ballistic trajectories, may play an important role in connection [12, 13]. Parabolas can act as a connection tool, where the visual system projects parabolas, independent of the perceived crossing angle, to interpolate occluded line segments. This often leads to the interpolation of a smooth rather than a sharp, discontinuous connection.

The type of connections humans make during interpolation has been investigated in an experiment where the subjects were asked to place a dot to estimate the point of connection between two equally sloped projected surfaces [14]. It was found that the subjects predicted a smooth connection rather than a discontinuous one, such as the linear extension of the edges of the surfaces, in about 90% of the trials. While only three subjects participated in this experiment, the results suggest that humans tend to favor smooth connections in the trajectory smoothing case, particularly for parabolic connections.

Closure is a Gestalt property that addresses the fundamental tendency of humans to want to close gaps. Humans can perceive inside and outside space between contours, allowing them to fill in the boundary that should, in their mind, exist [6]. Since partial trajectories can be defined as occluded trajectories (for which closure will be applicable), this relationship allows the closure property to be useful in trajectory smoothing. As long as a partial trajectory can be identified as being one trajectory the human will tend towards closing the gap.

2.2.2. Extrapolation

Extrapolation, defined as the extension of a contour beyond itself, is another possible theory used to interpret occluded contours. Some work has shown that the completion of discontinuous curves is sometimes the extension of one curve rather than interpolation, which is the connection of two pieces [15]. Furthermore, some partial trajectories do not have connecting information, and in those cases extrapolation would be the only way to smooth that trajectory. LMTS does not have any capability to extrapolate, and therefore the human ability to extend trajectories may provide further utility in the trajectory smoothing task. Figure 2-4 shows that the bold partial trajectory can be interpolated to the left of the vertical line. However, it must be extrapolated to the right where there are not any possible connecting segments. The extrapolation will most likely follow the same shape as its surrounding elements. This means that while LMTS would not be able to fully interpret the occluded trajectory, human extrapolation may be able to accurately predict the extension of the bold segment. While interpolation will be the primary focus of this research, the ability to extrapolate may provide additional capability that human participants could add to the trajectory smoothing task.

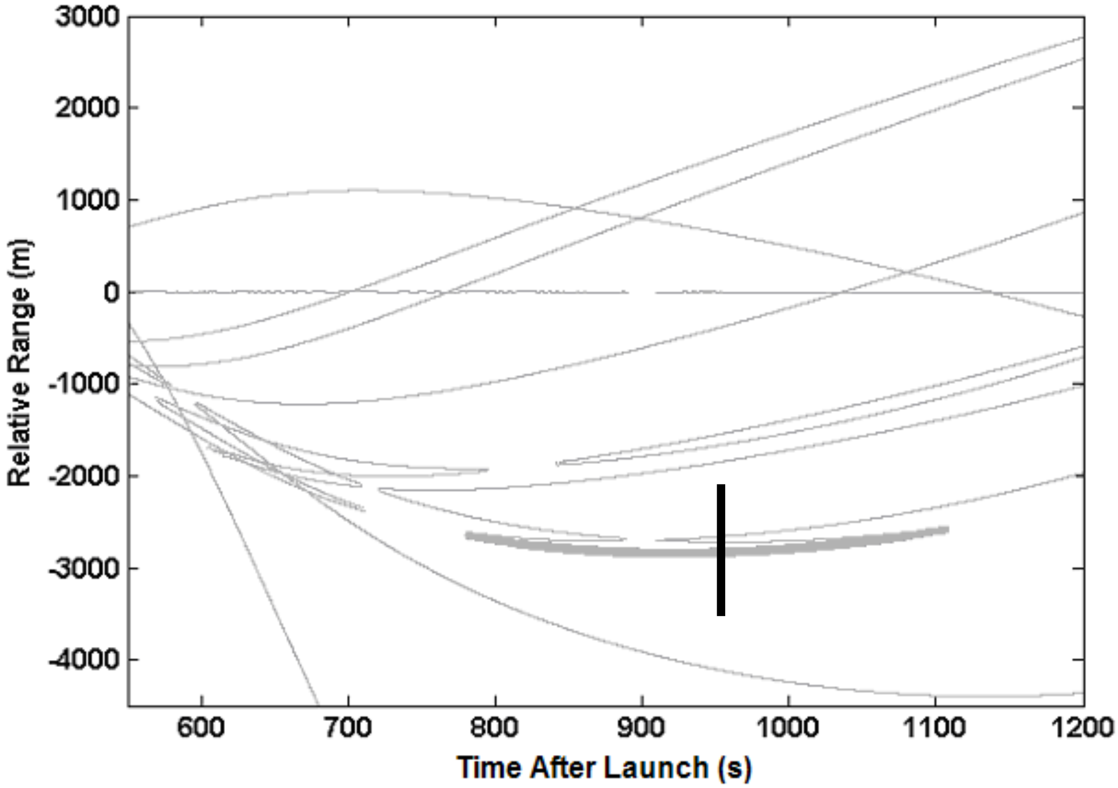


Figure 2-4: The bold trajectory can only be extrapolated to the right of the vertical line.

2.3. Barriers to Interpolation

Noise, or possible distractions, is a well-researched area of contour perception [16-19]. This section will define noise and relate it to the way in which trajectory information is presented. Specifically, each incomplete trajectory can be considered as an occluded contour, and other trajectories present in the display act as noise, potentially interfering with the specific contour a human will try to connect. The bold contour displayed in Figure 2-4 is an example of this. If one tries to interpolate this contour to the left, determining which contours to connect becomes a difficult problem, as the correct contour is obscured by the contours surrounding it. This section will discuss several aspects of contour interpolation in noise, including contour detection and how noise may hinder interpolation, as well as the effects of orientation in noise.

2.3.1. Contour Interpolation in Noise

Noise can be defined as the surrounding information in which the target curvature, i.e. the curvature that a human is trying to interpolate, is presented [17]. Specifically, noise can create illusory effects, which in turn alters the perception of the contours. In the trajectory smoothing case, this means the perception of the actual curvature may be altered, which impacts how well the human will be able to interpolate trajectories. Subsequent sections discuss target contour detection in noise and how noise may act as an inhibitor to contour detection.

2.3.2. Target Contour Detection in Noise

Target contour detection is the ability to focus on a single partial trajectory in noise. A common approach to detection begins with filtering the information to determine edges in a figure [19]. Edge detection is important because it allows the observer to distinguish contours in possibly noisy, densely packed backgrounds. It has also been shown that edge detection is no different for straight lines (such as chevrons) or curvatures [16]. Therefore in the case of trajectory smoothing, all types of trajectories should be able to be detected in noise irrespective of their actual shape.

2.3.3. Noise as an Inhibitor

Perception of a curvature (which occurs after detection) is the ability of the visual system to correctly identify how convex or concave a curvature is. It has been shown that human interpolation may have a "specific sensitivity to contour curvature" [18]. There may be aspects of the visual system that readily grasp different curvatures. Therefore, it is important to understand how the sensitivity to target curvature would vary dependent upon different noise factors. Recognition of curved lines in straight-line noise was shown to be independent of numbers of "distracters," which are all the segments besides the target curvature. This is particularly applicable in the case of trajectory smoothing, as the noise contains surrounding trajectories that could be any shape from straight lines to steeply curved trajectories. However, it has also shown that the perception of straight lines can be altered, and is dependent upon the number of curved contours that exist as noise [18]. So, while curvature is more easily deciphered in the presence of straight-line distracters, it is also true that straight lines can appear to be curved when surrounded by other curved lines. Figure 2-5 shows how surrounding a target curvature with noise could alter its perception.

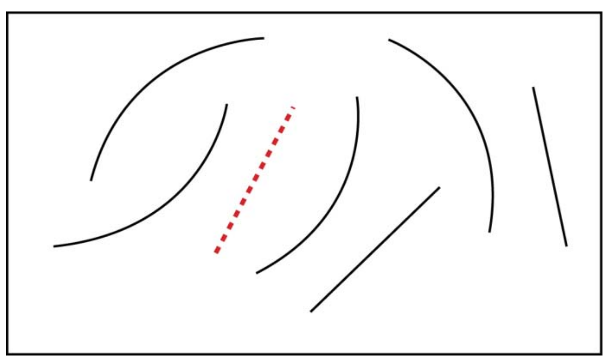


Figure 2-5: The dashed target curvature, a straight line, could be perceived to be curved depending upon the degree of curvature of surrounding contours.

Foster showed that humans' responses to different curvatures from straight lines to partially curved lines (0 to 20′ arc) can be grouped into categories of straight, just curved and more than just curved that he defined as a "discrete encoding process" [17]. This process allowed him to place curvature detection into various categories in which an observer would be more or less sensitive to multiple contour curvatures. While the exact numbers for the categories in this process are not applicable to this thesis because they apply to research done on angle changes on the order of minutes, this research shows that there may be a separate discrete encoding process for the varying curvatures in the discontinuous trajectories that allows the human to readily perceive and connect them.

2.3.4. Orientation of Noise

The presence and orientation of distracting contours is important in perceiving and interpolating contours of interest. There is evidence that contour detection is achieved through the use of separate contour filters that are in effect tuned to different contour orientations [20]. Moreover, there exists some relationship between orientation and the observer's viewing condition in the ability to decipher symbols. This is interesting as many of the "symbols" used to determine those contour filters are similar in shape to the crossings found in track smoothing/identification data. Furthermore, there is evidence that the shape of curvature is distinguished by the same visual devices that recognize orientation [16]. Therefore in the case of trajectory smoothing, altering the orientation of noise may affect the perception of partial trajectories.

Secondly, integration of various contour properties suggests that the visual system has filters that preferentially perceive certain orientations [21-23]. For example, certain orientations may become visually salient to the human before others. Furthermore, some recent experiments have suggested that human perception may be more accurate in certain orientations. In a study using various types of displays, it was shown that performance of contour interpolation in 3D decreased slightly by "inducing surfaces with a vertical rather than a horizontal tilt direction" [24]. This suggests that a horizontal display of information may provide superior results than a vertical orientation.

2.4. Human-Automation Collaboration

LMTS was developed to solve the problem of trajectory smoothing in the BMDS. Like many other automated tools used to assist in problem-solving tasks, it has been shown that LMTS may not function in situations that were not anticipated during its creation. For example, LMTS may function poorly when dealing with smaller amounts of data as it was designed to smooth a completed, long term data set. To deal with this in other areas, studies have suggested that cooperative problem-solving systems should be considered [25]. Given that humans have been shown to have some track smoothing abilities, a collaborative effort could possibly produce the best solution for the track smoothing application.

Similar to this smoothing application, path planning research deals heavily with perception. The ability for collaboration in these studies depends on a unified understanding of the search space, and thus perception of that space is vital. The Human-Guided Search (HuGS) platform [26], developed as a tool for solving optimization problems, focused on involving people in a heavily automation-dominated field. The HuGS platform requires people to alter selection criteria by presenting solutions in a simple step by step process. This allows the human-user to gain different views of the search space, altering perception of the problem and therefore arriving at alternative human-aided solutions that can be evaluated.

Other work has recommended that if technology cannot fully solve a complex issue, the design of a support system can influence alternative solutions [25]. Multiple visualizations of an automated solution could be given to users to see if a difference in perception would alter overall performance [27]. In the trajectory smoothing case, a possible application of this work would be to provide human insight into areas that are problematic for LMTS, which could in turn allow LMTS to generate alternative, better solutions.

Fitts' List [28] has been adapted to illustrate which traits of each decision source may prove helpful in the trajectory smoothing task. However, since this is the first time the track smoothing task has been performed by human operators, certain assumptions have been made about which decision source, human or algorithm would be better at certain tasks. This Fitts' List adaptation simply defines possible superior traits, rather than presenting a "who is better at what" table.

Table 2-1:
Fitts' List [28] of Possible Superior Traits From Each Decision Source Adapted to LMTS Settings

Humans Are Better At	***LMTS Is Better At***
Resolving uncertainty to improvise fits to trajectories	Using strict rules to evaluate likelihood of a match
Extrapolating trajectories	Consistently producing results, such that each application of the algorithm to the same data will provide the same output
Viewing data thought knowledge-based filters to resolve uncertainty	Defining the curvature of the various partial trajectories
Recalling previous similar pattern-matching experiences	Quick and efficient computation

2.5. Summary

Previous research has shown that either automation or a human acting alone often does not generate an acceptable end result [25-27]. It has been suggested that the automation can utilize its computational speed to search the solution space while keeping the human involved [25]. Leveraging the human ability to intuitively assess the LMTS solutions and modify automated solutions allows for use of both parties' strengths.

This chapter has presented the operation and potential weaknesses of LMTS and how a human operator may be able to assist in those areas. The potential for human contribution to the track smoothing task is supported by research that has taken place in a wide variety of fields and affords an understanding of how humans interpret occluded contours. Gestalt theory provides some proposed bounds on this capability while other research shows that depending on the application, these boundaries may be altered or may not exist at all. Finally, the possibility of a collaborative solution through reliance on human perception was discussed.

3. Experiment One Design

An experiment was designed investigate whether a human could better connect trajectories than the LMTS algorithm, and the conditions under which the human outperformed LMTS most dramatically. The subsequent sections will describe how that information was collected and analyzed. The hypotheses, participants, apparatus, experimental design, and testing will be discussed.

3.1. Hypotheses

The data used for this experiment are simulated outputs of a hypothetical radar's real-time, multi-target tracker an example of which is shown in Figure 3-1. The data contains incomplete segments similar to those discussed earlier. As discussed in the last chapter, the amount of information available to the decision source plays an important role in perception and interpolation. In the case of trajectory smoothing, there are distinct cases in which LMTS performs better or worse, which can be used to measure the degree of difficulty of a data set. Human and algorithm performance are expected to vary with both the degree of difficulty and the amount of information present. The confidence of human participants is also expected to vary. The following hypotheses capture the expected decision source performance.

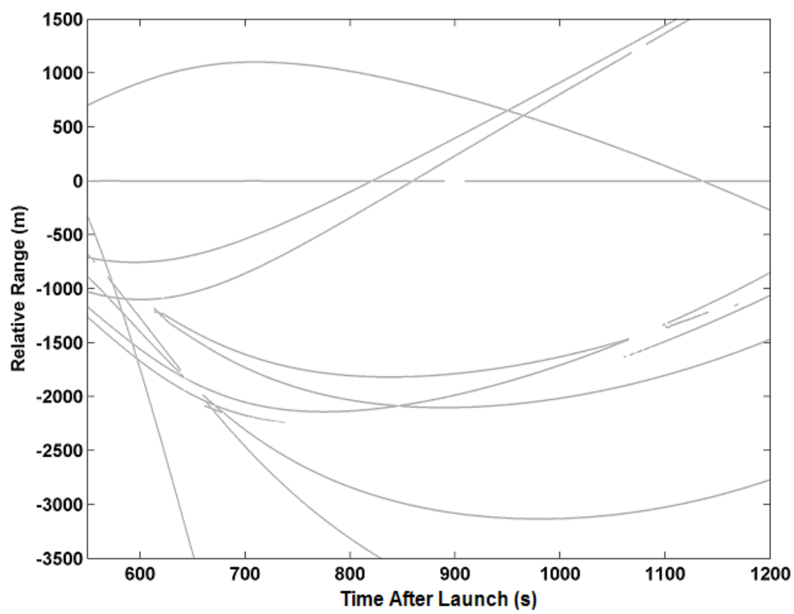


Figure 3-1: Occluded radar data as used in the experiment.

3.1.1. Hypothesis One

Human and algorithm performance will deteriorate with shorter data spans and narrower crossing angles, as there is less information and trajectories are harder to interpolate.

3.1.2. Hypothesis Two

Human participants will be less confident with shorter data spans and narrower crossing angles, as there is less information and trajectories are harder to interpolate.

3.2. Participants

A total of 29 participants participated in the initial experiment, 20 male and 9 female. The participants were all Lincoln Laboratory employees who received no additional compensation for their participation. The subject pool ranged in age from the early 20s to mid 70s, and included a sampling of individuals with diverse backgrounds. The participants' ages can be split into 4 groups as follows: 38% were over 50, 31% were ages 35-50, 28% were 25-35, and 3% 18-25. The population spanned multiple fields of work including engineers, scientists, librarians, executive assistants, and support staff. The complete subject demographic information is listed in Appendix A.

A pre-experiment survey asked participants the amount of time spent drawing on a computer (Appendix A). Out of the 29 participants, 17 had computer drawing experience. The amount of experience was split into three categories where 7 drew on a yearly basis, 5 drew on a monthly basis, and 5 drew on a weekly basis. It was considered important to acquire such a diverse sample population in the initial experiment to explore the overall human ability to interpolate incomplete contours.

3.3. Apparatus

This section outlines the interface and the equipment used by the human participants to smooth trajectories.

3.3.1. Track Smoothing Interface

In order to conduct the experiment to investigate human operator performance in track smoothing, an interface was designed to allow a human to interpret and interpolate the radar plots (Figure 3-2). The design of this interface was guided by principles that direct that effective displays should allow users to have appropriate control, both in solution creation and editing and provide them with error correction and appropriate feedback, while keeping the interface as simple as possible, especially for a time-constrained task [29]. While the interface was not designed to be an actual operational interface, it was designed to evaluate concepts for an operational interface insofar as possible. In order for the user to efficiently produce best-fit tracks, simple graphing ability, including an editing capability, was vital. From an operational standpoint, to keep the participant aware of the time available for fitting the tracks, a timer was placed in the interface. The resulting interface has four major components: plotting area, interaction panel for track fitting including the capability for the user to express his/her confidence in the fit, plot appearance, and timer. These are discussed below.

The plotting area, which constitutes the majority of the interface in Figure 3-2, displays the simulated radar plot that is presented to the subjects. This is the working area for the subjects to select their best fit for each track. The radar track data is displayed in gray so as not to prejudice subjects about possible connections. The option of utilizing color-coded data in a subsequent experiment may be considered for future work.

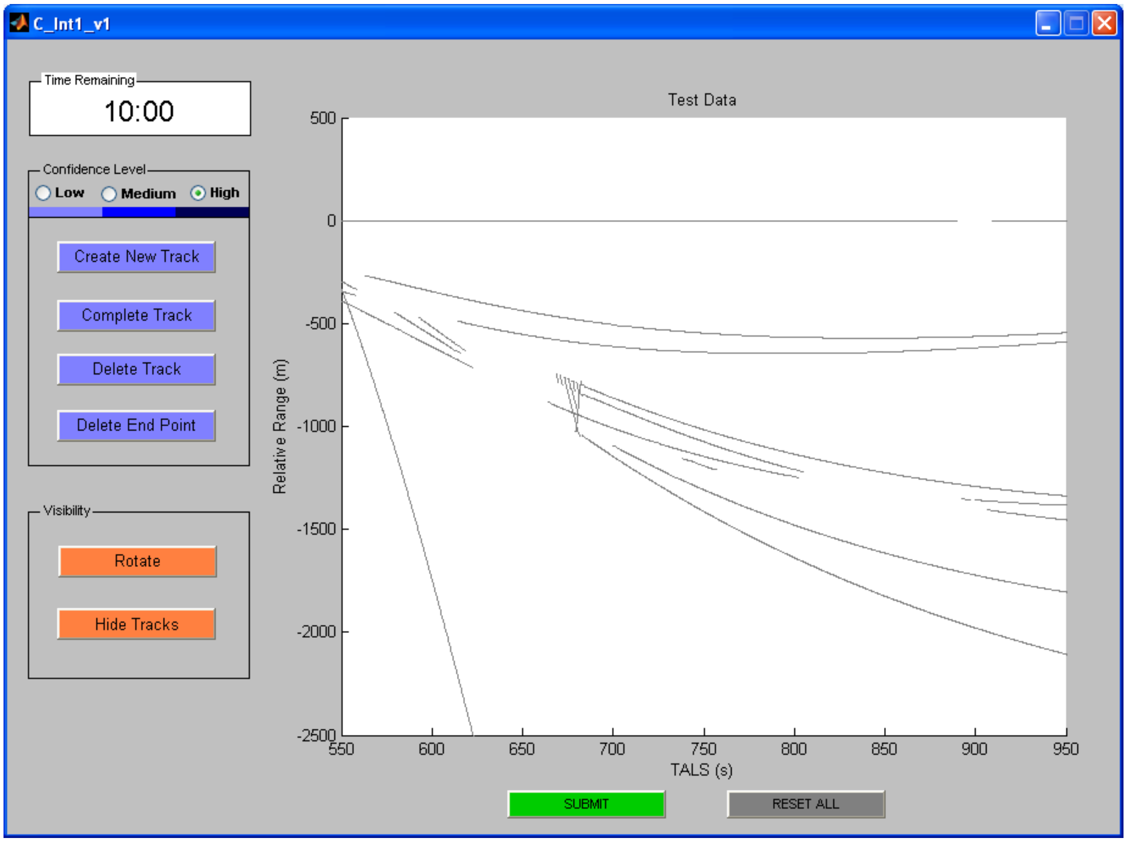


Figure 3-2: Track Smoothing Interface

The panel of buttons at the top left of the interface in Figure 3-2 is used for plotting the fitted trajectories and manipulating them. The user can initiate, complete, and edit tracks using this panel. To initiate a trajectory, the user clicks "Create New Track" which gives the operator the ability to plot. The user clicks then points along the proposed trajectory until the trajectory is completed, which is finalized by clicking the button "Complete Track". The trajectory is then plotted using a spline fit, with a segment fitted in between each of the plotted points. After completion a user can adjust or delete trajectories by selecting the trajectory to be edited, then choosing "Delete Track" or "Delete End Point".

In addition, this panel allows the user to select a confidence that he/she feels appropriate for each segment of each track. This estimate is color-coded along the track, indicating when confidence estimates change from point to point along the track. Appendix B explains the operation of the interface in detail.

The panel of buttons in the lower left of the interface in Figure 3-2 allows the user to change the plot appearance during the experiment. One button allows the user to select a “right/left” or “up/down” data display and allows for adjustment for individual preferences as well as allowing operators to observe a different orientation to spot potential patterns. A second button allows the user to inhibit the display of previous tracks that may be interfering with a current selection; the remaining data is then darkened to alert the user to the fact that visibility of previous selections has been turned off.

Lastly, because the tasks must be completed in a specified time, a timer is included in the interface to keep the user aware of the time available for the task. The timer’s size and font were chosen to be salient while not taking away from the task at hand. In the present experiment the users are given ten minutes to complete each scenario.

3.3.2. Test Bed

The experiment was run on a Dell Precision 670 computer with a Intel Xeon CPU 3.2 GHz processor and a NVIDIA Quadro FX 3400 graphics card. The monitor was a Dell 2001FP with a resolution of 1280 x 1024. All experiments were run in a study room in the library at Lincoln Laboratory.

3.4. Experimental Design

The independent and dependent variables are outlined in this section.

3.4.1. Independent Variables

The first independent variable is decision source, as both LMTS and human participants were tested. It has been shown that over a long time period, the algorithm performs well; however, at shorter time spans the algorithm’s performance degrades [7]. Therefore the second independent variable, the data span, was selected to test cases that represent temporal effects. For the experiment, data spans of 30%, 60% and 100% of the interval over which radar data were available (which is represented as Time After Launch or TAL) were investigated. For the data used here, these data spans were on the order of 3, 6, and 9 minutes, which represent operational scenarios.

The data were also sorted by degree of difficulty. Like the data span, the degree of difficulty affects the algorithm’s ability to accurately smooth trajectories. Observations show that cases with shallow crossing angles and high track density are the most difficult cases [7]. Therefore the third independent variable, degree of difficulty, was split into two categories, easy and hard. Based on previous LMTS experience, the hard degree of difficulty was defined as trajectories crossing at <15° degrees, shown in Figure 3-3, and the easy degree of difficulty was defined as crossings at >15°, shown in Figure 3-4.

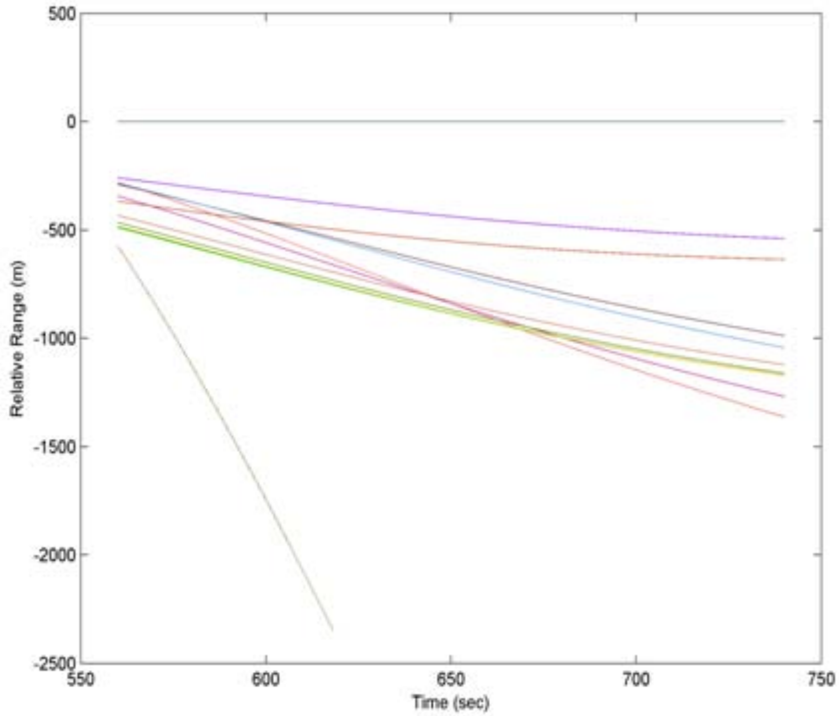


Figure 3-3: Hard Degree of Difficulty

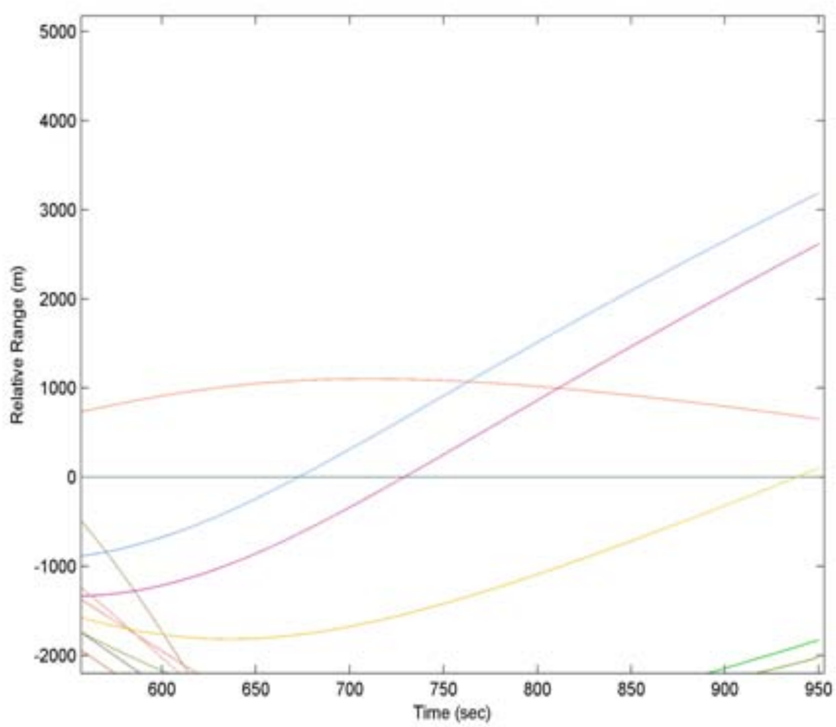


Figure 3-4: Easy Degree of Difficulty

The experiment was conducted by presenting the subjects with representative data segments on an interface and asking them to fit lines to the segments that they perceive as actually connected. In parallel with the human subject testing, the same data presented to the subjects were also presented to the LMTS algorithm. The results for both cases were scored using truth data that were generated as part of the simulation. The independent variables for the experiment are summarized in Table 3-1. Given the three independent variables, this is a 2x2x3 fully crossed within-subjects experiment.

Table 3-1:
Independent Variable Breakdown

Variable	*Levels*
Decision source	Human or Computer
Degree of difficulty	Hard or Easy
Data span	30, 60 or 100% of the available data interval

3.4.2. Dependent Variables

Trajectory detection and trajectory fit were the primary dependent variables used to score performance. A measure of the confidence of each subject on each trajectory, defined below as confidence density, was used to answer the third research objective. These variables are defined below.

Trajectory Detection

Each decision source was tasked with correctly detecting each trajectory present in the data. Both the humans and LMTS could miss some trajectories (Missed Trajectory) or plot additional trajectories (False Trajectory) as neither decision source had previous knowledge of the number of truth trajectories that existed. These scores were tallied for each plot using a matching algorithm, which uses the relative range from the plotted to the truth trajectory to calculate which user trajectory should be matched with a truth trajectory, as displayed in Figure 3-5.

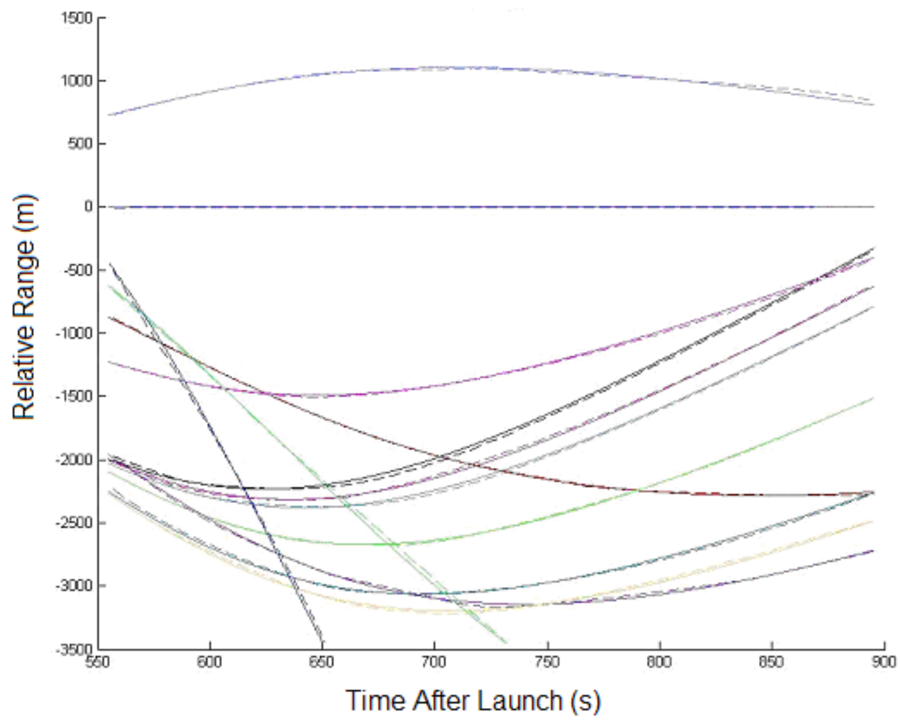


Figure 3-5: The dotted lines show the user plotted trajectories and the solid lines the corresponding truth.

Trajectory Fit

In addition to detection, it is important to quantitatively determine how accurate each decision source was. The closer the human or algorithm matches the corresponding truth trajectory, the better the score. This score is calculated from goodness-of-fit measures used in linear regression modeling. The truth trajectory is assumed to be the model (regression fit), while the line plotted by the participants or the algorithm is assumed to represent the observations. The mean squared error (MSE) and the root mean squared error (RMSE), shown in Equation 3-1 and Equation 3-2 respectively, are calculated for each plotted track [30]. The number of samples along the line was set at N = 500. This was selected by observing that, while the results changed significantly from as N was increased from 100 to 500, there was little change beyond this.

$$MSE = \frac{\sum_{i=1}^{N}\left(Y_{estimate_i} - Y_{truth_i}\right)^2}{N}$$

Equation 3-1: Mean Squared Error

$$RMSE = sqrt(MSE)$$

Equation 3-2: Root Mean Squared Error

The resolution of the monitor is another important factor in scoring. The display area horizontal dimension, D_H, is about 41 cm (1280 pixels) and the vertical dimension, D_V, is about 31 cm (1024 pixels). When performance is scored, the difference between results and truth is calculated in meters. However, the difference between human results and truth is ultimately limited by the ability of the human to resolve points on the screen. It has been estimated that the human eye can resolve points to about a minute of arc [6]. At a nominal distance from a monitor of 50 cm, this is a point separation of about 0.015 cm or 0.15 mm. On the other hand, the extent of a pixel on the screen can be approximated as $D_V/1024$ = 0.03 cm or 0.3 mm. Clearly the pixel extent and not the resolution of the eye dominates the observer's power of resolution. Thus it is assumed that, if a human performance is within one pixel of the truth, that is the best that can be achieved, and the human should be given credit for a "perfect" performance when this occurs.

Confidence Density

The last dependent variable is the measurement of operator confidence. Confidence density was calculated as shown in Equation 3-3.

$$\frac{\sum_{i=1}^{N=13}\sum_{j=1}^{M=500} x_j}{N * M} \quad in\ which\ \ x \in 1,2,3$$

Equation 3-3 Confidence Density

N represents the number of total plotted trajectories and M represents the number of iterations that the color was evenly sampled along each trajectory. M = 500 was used as a limit to be consistent in the measurement bounds used in the Trajectory Fit variable. At each iteration the confidence was recorded (Appendix C). The value 1 corresponded to a high confidence level, 2 a medium level and 3 a low confidence level. The confidence was thus polled by the algorithm at each iteration, and therefore the average confidence per scenario (see Section 3.5.3) could be calculated, providing an overall confidence measure.

3.5. Testing

Testing involved pre-experiment activities, a training session, and a test session.

3.5.1. Pre-Experiment Activities

Each participant was first introduced to the experimental setup. Next, each participant read and signed the Consent to Participate Form (Appendix C) which discussed the purpose of the experiment, the compensation policy, and the experimental aspects that the participants would be asked to complete. Then the participants filled out a pre-experiment survey.

3.5.2. Training Session

After filling out the required information, a tutorial instructing participants on the intricacies of the interface was presented. The tutorial started with an overview of the BMDS scenario and the motivation for the trajectory smoothing task (Appendix B). Then the interface was presented in detail. The participants were given instructions on how to smooth the trajectories and manipulate the interface to indicate their desired confidence. After the tutorial, Camtasia® software was turned on to record all interface activity during each experiment.

Three practice scenarios were given to each user. The three practice scenarios were chosen to familiarize the user with both the interface and the interpolation task they were asked to perform. The first scenario had 4 trajectories to interpolate, and the scenarios became progressively harder until the last practice scenario, which was as difficult as any data set the participants would be asked to interpolate.

3.5.3. Test Sessions

After the three practice sessions, all participants were asked to complete six scenarios, which were derived from crossing the three data spans and two degrees of difficulty. All the test scenarios given to each subject are presented in Table 3-2. While the practice scenarios were given in a specific order, all test scenarios were randomized such that there was no specific order or presentation. All users were allowed to ask questions throughout the entire experiment regarding the interface and its capabilities.

Table 3-2:
Scenarios Seen by Participants

Scenario	Crossing Angle	Data Span
Practice Scenario	Easy	100%
Practice Scenario	Easy	100%
Practice Scenario	Hard	100%
Scenario 1	Easy	30%
Scenario 2	Easy	60%
Scenario 3	Easy	100%
Scenario 4	Hard	30%
Scenario 5	Hard	60%
Scenario 6	Hard	100%

3.6. Summary

Experiment One, the initial track smoothing experiment, was designed to answer the primary research question of which decision source (human or automation) could best smooth trajectories. Independent and dependent variables were created to measure the necessary information to determine which decision source was better and why. The Track Smoothing Interface was designed to best capture the operator's interpolation of the occluded contours of the radar plot. Subjects received background and training prior to the experiment and each subject completed six text scenarios. All information was gathered in real time and recorded for analysis, which will be detailed in the next chapter.

4. Experiment One Results and Discussion

This chapter addresses the results of the first track smoothing experiment and discusses their implications. The chapter then addresses the two hypotheses outlined in Section 3.1 of this thesis.

4.1. Results

The results are addressed in four categories: 1) Missed or false trajectories, 2) Accuracy of correct trajectories, 3) Total performance tables, and 4) Confidence.

4.1.1. Missed or False Trajectories

In Experiment One, each of the 29 tested participants tried to plot 77 trajectories. There were actually 78 true trajectories but, due to noise in the data, one trajectory could not be seen by the participants. Combining all participant data, there were a total of 2,233 possible trajectories to plot. Human participants missed a total of 144 trajectories and predicted 12 false trajectories. This averages to 3.97 (1.40 std. dev.) missed and 0.41 (0.73 std. dev.) false trajectories predicted per participant. The algorithm missed a total of 2 trajectories and predicted 5 false trajectories. The number of missed and false trajectories for all participants is listed in Appendix E.

4.1.1.1. Algorithm-Missed Trajectories

Table 4-1 shows the first step in the analysis, which was to research the 2 trajectories the algorithm missed and determine if any human participants detected those trajectories. The algorithm only missed trajectories in the hard degree of difficulty, 30% data span case, which was Scenario 4. The specific trajectories the algorithm missed are designated by the dashed arrows in Figure 4-1.

Table 4-1:
The Number of the Human Correctly Plotted Trajectories vs. the Number of Missed Trajectories for the 2 Trajectories Missed by the Algorithm

Scenario, Track, Factor Level	*Scenario 4, Trajectory 1, Hard 30%*	*Scenario 4, Trajectory 6, Hard 30%*
Human-Correct Trajectories	4	6
Human-Missed Trajectories	25	23

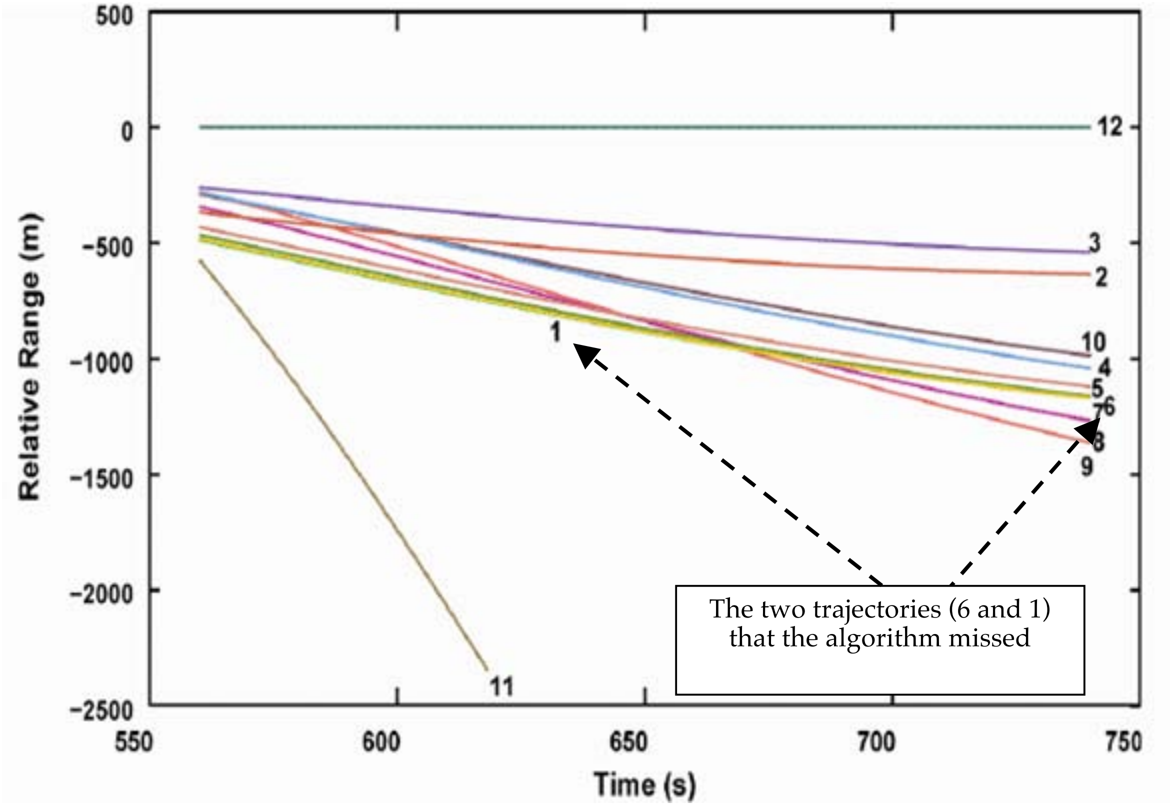
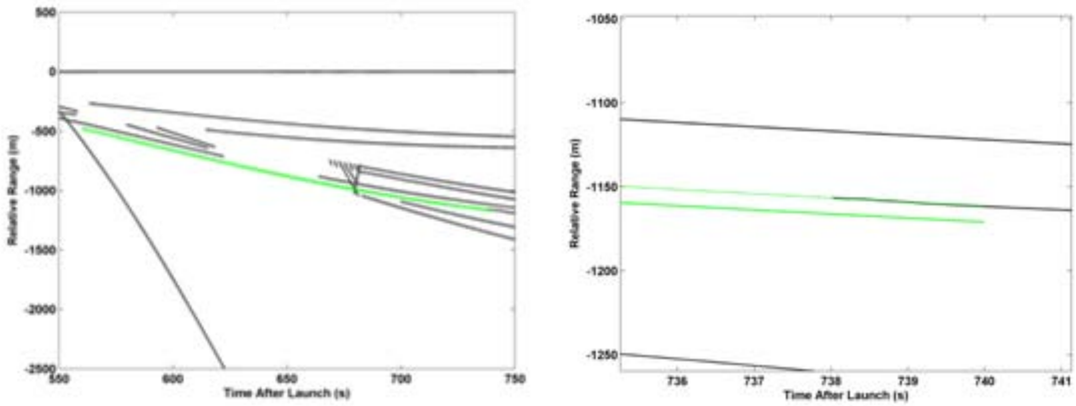


Figure 4-1: The 12 possible trajectories for Scenario 4 (Hard, 30%) that could have been identified by either decision source.

Figure 4-2 shows the two cases where some humans managed to form trajectories when the algorithm could not. In the area designated by the dashed green lines in Figure 4-2a, each decision source should have plotted four trajectories. However, both the majority of human participants and the algorithm only plotted two. The green dashed trajectories are the two trajectories that the majority of the participants and the algorithm missed (Trajectories 1 and 6). It is easy to see that these trajectories are difficult to plot. They have been magnified in Figure 4-2b to highlight the difference between the two. While the number of humans who managed to accurately identify those trajectories is relatively small, 14% for trajectory 1 and 21% for trajectory 6, it still demonstrates that there were at least 14% of the participants who were able to identify trajectories when the algorithm could not.



(a) *(b)*

Figure 4-2: Test data (for Scenario 4) as seen by both decision sources prior to processing. The location where both trajectories 1 and 6 should be plotted is shown by the green dashed trajectories, which are magnified in (b).

4.1.1.2. Human-Missed Trajectories

The next step in the analysis was to investigate the cases in which the human missed trajectories. All participants missed at least 2 trajectories during the course of the experiment. The 4 trajectories that approximately half or more of the humans missed are listed in Table 4-2. A full list of these results is given in Appendix E.

Table 4-2:
The 4 Most Missed Trajectories by Humans

Scenario, Trajectory, Factor Level	***Scenario 3, Trajectory 1, Easy 100%***	***Scenario 3, Trajectory 7, Easy 100%***	***Scenario 4, Trajectory 1 Hard 30%***	***Scenario 4, Trajectory 6 Hard 30%***
Human-Correct Trajectories	3	13	4	6
Human-Missed Trajectories	26	17	25	23
Algorithm Missed (Yes or No)	No	No*	Yes	Yes
Human-Missed Trajectory %	90%	59%	86%	79%

* LMTS was only able to smooth a small portion (30%) of this trajectory

Table 4-2 shows that the two trajectories which the algorithm did not detect (Scenario 4, Trajectories 1 and 6) were also difficult for the human. However, as discussed next, the algorithm was able to identify some trajectories with which humans had trouble.

Figure 4-3 shows the missed trajectories as a function of the degree of difficulty and data span. The family-wise alpha for these tests was set to $\alpha = .02$. Overall, the algorithm and the human are significantly different (Wilcoxon, Z = -6.814, p = 0.000+). There is also a difference between decision sources in the easy degree of difficulty case (Wilcoxon, Z = -4.88, p = 0.00+), showing algorithm superiority in this area. Figure 4-3 shows that for the easy crossing degree of difficulty factor, the algorithm correctly identifies all trajectories, while the human performance starts to degrade for the 100% data span case (where human participants only match 87% of trajectories).

However, there is no difference between decision sources in the hard degree of difficulty factor (Wilcoxon, Z = -1.507, p = .132). This suggests that the human participants may be able to best assist the trajectory smoothing task in this area. The two trajectories the algorithm missed (Scenario 4, Trajectories 1 and 6) are in the hard degree of difficulty factor, and at least 14% of human participants correctly predicted those trajectories, which demonstrates the potential for human improvement over the algorithm.

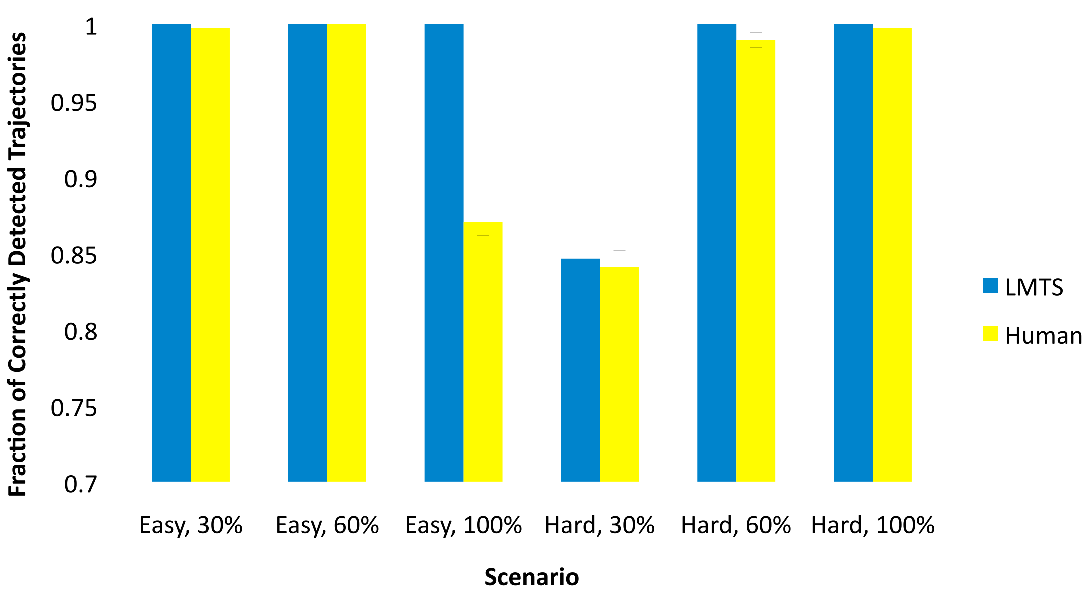


Figure 4-3: Fraction of correctly detected trajectories by both algorithm (first bar, blue) and humans (yellow) separated by Scenario.

Of the human-missed trajectories, Scenario 3, Trajectory 7, presents the most interesting case. Only a slight minority (13) of the participants were able to accurately detect it. Figure 4-4 shows the truth Trajectory 7 as the dashed, green trajectory. Since the human participants were instructed to plot from start to finish of the data span presented to them, in order to detect this trajectory, each participant had to connect partial trajectories spanning an information gap that was 66% of the data span. Perceiving a connected trajectory over this gap was extremely difficult.

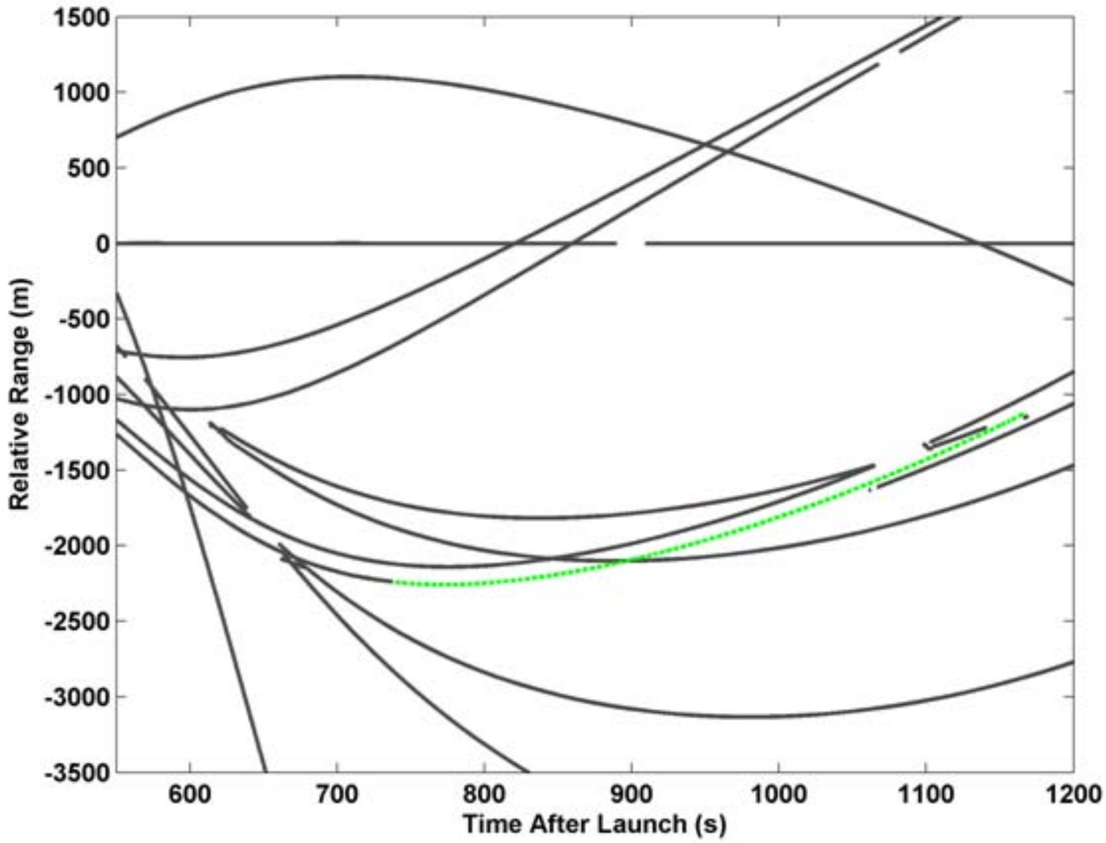


Figure 4-4: Scenario 3 in which Trajectory 7, demarcated by the green, dashed line, can only be correctly detected if the user connects partial trajectories over a span of 66%.

On the other hand, the algorithm lacked the ability to extrapolate, and therefore detected only the first 30% of Trajectory 7. So while LMTS detected a partial trajectory, it was unable to associate it over the large gap to any other partial trajectory. That large information gap is suspected to be the cause of the failure to connect the trajectory by both the algorithm and the majority of the participants. However, 13 participants were able to smooth Trajectory 7 and all 13 of those participants were able to complete the trajectory over the information gap of 66%. It is important to realize that that failure of LMTS to detect a full trajectory over large gaps could limit its effectiveness while it is equally important to realize that some human participants have detection ability, in the form of extrapolation, which could be exploited. This result is revisited in the accuracy section of this chapter.

4.1.1.3. False Trajectories

Unlike missed trajectories, there is a major difference between the decision sources in the prediction of false trajectories. Out of the 174 possible scenarios, human participants predicted 12 false trajectories. However, out of the 6 possible scenarios the algorithm processed, it predicted 5 false trajectories. Broken down by individual test scenario, the algorithm predicted 0.83 false trajectories per test scenario while the aggregate human participant predicted 0.069 false trajectories per test scenario. Averaging over all independent variables, there was a significant difference between the algorithm and the average human participant in predicting false trajectories (Wilcoxon, $Z = 7.819$, $p = 0.00+$, $\alpha = .05$).

The number of false tracks predicted by both decision sources as well as the percentage of false tracks predicted by human participants is listed by independent variables in Table 4-3. The two numbers that should be compared are the ratio of human false tracks (averaged per participant) vs. the number of false tracks predicted by LMTS. Comparison of these two numbers demonstrates the ability for the average human to outperform the algorithm in all cases except for the two shortest data spans, easy degree of difficulty scenarios. However, even in those cases the vast majority of participants did not predict a false trajectory. A table of all participants and the false trajectories predicted for each is given in Appendix E.

Table 4-3:
The Ratio of False Tracks Predicted Per Scenario.

(Decision Source, Scenario)	Easy, 30%	Easy, 60%	Easy, 100%	Hard, 30%	Hard, 60%	Hard, 100%
(# Human false tracks/ # total participants)	.103	.035	0	0	.172	.103
LMTS #	0	0	1	0	1	3

The bold lines in Figure 4-5 show an example of how the algorithm could predict false trajectories. First, the algorithm incorrectly connected a trajectory, which is shown by the complete bold trajectory with the discontinuity. Because the algorithm plotted the incorrect trajectory first, the two bold partial trajectories that are designated by the double arrow were calculated to be their own separate trajectories by the algorithm, which were false representations.

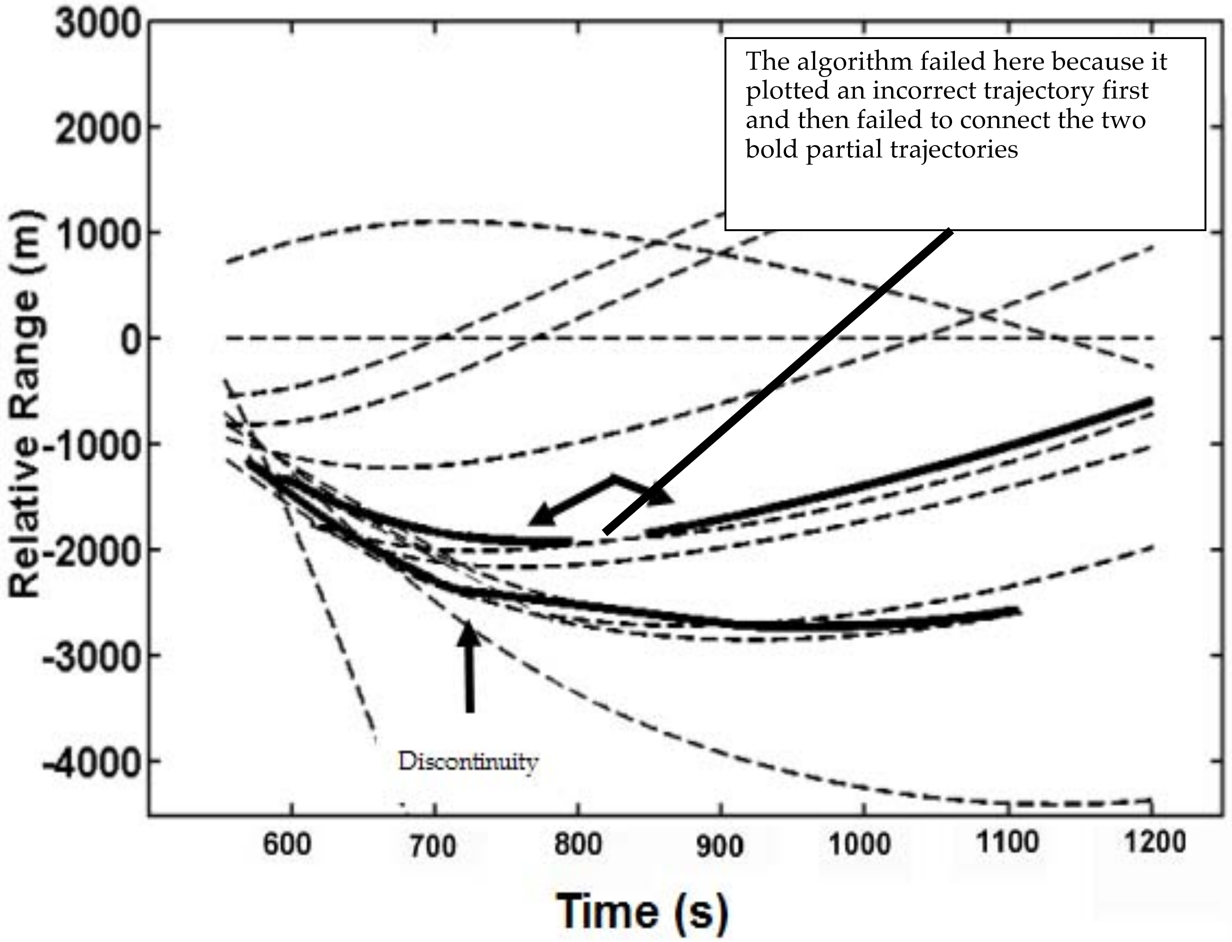


Figure 4-5: Hard, 100% (Scenario 6) Factor-Level crossing as plotted by LMTS

4.1.2. Accuracy of Correctly Plotted Trajectories

The next step in the analysis was to measure the accuracy of each decision source. For all the trajectories correctly detected (i.e. not missed or falsely identified), Equation 3-2 was used to calculate the RMSE. Figure 4-6 shows the respective RMSE averages for both decision sources in each scenario.

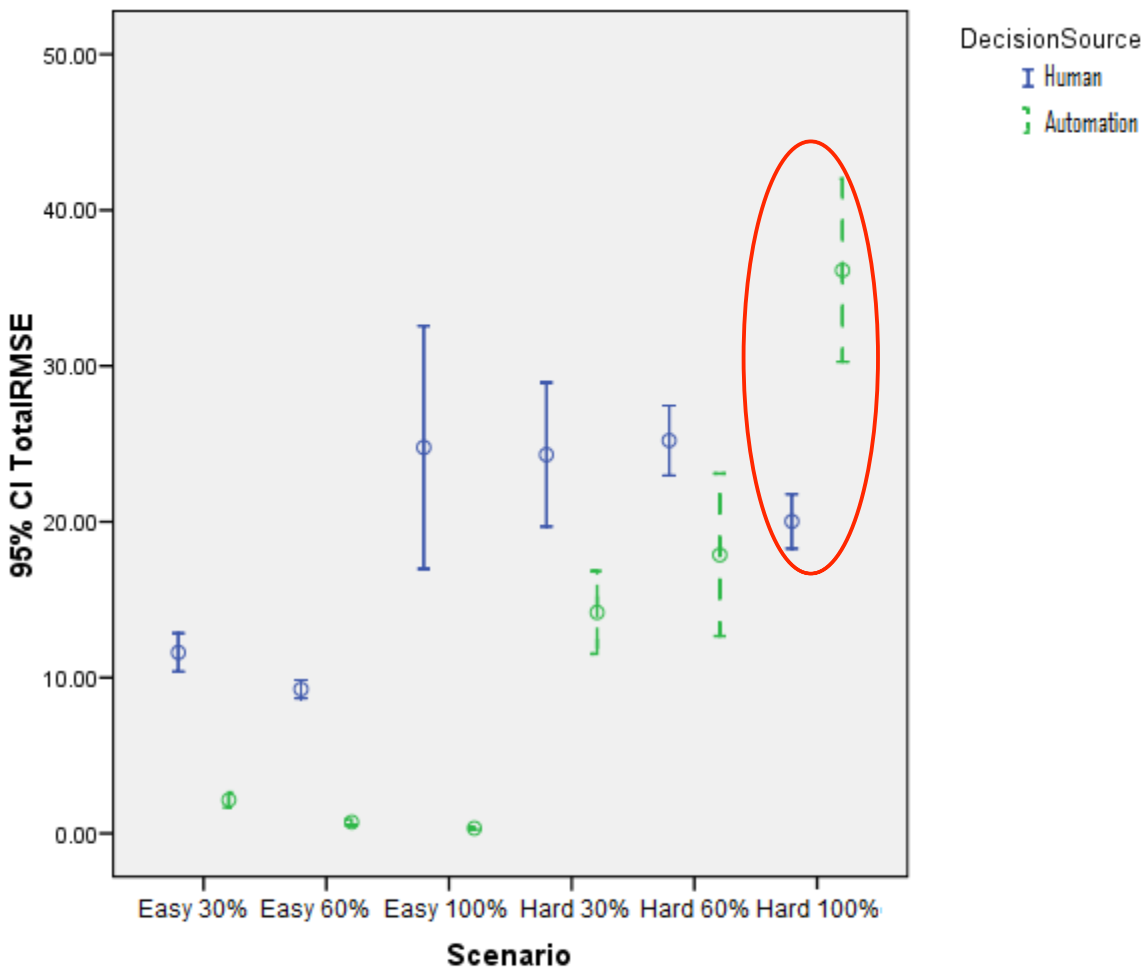


Figure 4-6: Error plot for both decision sources

The results in Figure 4-6 depict varying accuracy for both decision sources. The family wise error was set to α = 0.01. Using the RMSE of all trajectories plotted, there is a significant difference between decision sources (t = 6.704, p=0+), between degrees of difficulty (15.167, p=0+) and between the 30%-100% (t = 5.526, p=0+) and 60%-100% (t = 4.564, p=0+) data spans. There was no significant between the 30%-60% (t = 1.02, p=0.303) data span crossing, which provides evidence that something may be happening in the 100% data span case that would alter accuracy. Figure 4-6 depicts that the human and automation exhibited differences in ability to accurately plot trajectories. The algorithm (mean 1.06 m, std dev 3.09) clearly outperformed the human (mean 14.87 m, std. dev 37.5) in the easier degree of difficulty cases. However, the algorithm (mean 23.44 m, std dev 49.05) and human (mean 23.17 m, std dev 26.63) were a closer match in the harder degree of difficulty cases. The humans also seemed to have a relatively constant ability to accurately plot trajectories for the last four scenarios, while the algorithm has a definite decrease in ability with increasing data span in the hard degree of difficulty factor scenarios.

Scenario 3 (Easy, 100%) presents an interesting case as it has the largest difference in RMSE between algorithm and human participants. In Section 4.1.1.2 it was shown that

Trajectory 7 of this scenario was smoothed by 13 participants while the algorithm could detect, but not smooth, the same trajectory. Those 13 participants plotted Trajectory 7 with an average RMSE of 97m, which in relation to the averages shown in Figure 4-6 is extremely large (about 80m greater than average RMSE for the Easy degree of difficulty). While trajectory smoothing is possible over the large information gap in this scenario, the smoothing requires extrapolation rather than interpolation. Therefore, since LMTS cannot extrapolate, future experiments should take this into consideration when comparing accuracy between human participants and automation.

The most interesting case is Scenario 6 (Hard, 100%), circled in red on Figure 4-5, where the human (mean 20.02 m, std. div. 16.02), on average, was superior to the algorithm (mean 37.13 m, std. div. 57.96). This is due to the fact that the algorithm missed a trajectory crossing (i.e. connected two trajectories erroneously) and thus predicted false trajectories as shown in Figure 4-5. Therefore, it had a large error over a long data span.

Analyzing the RMSE results for the hard degree of difficulty factor level, it was found that errors occurred due to missed crossings, and all missed crossings occurred over information gaps. Similarly, those information gaps were all large compared to the gaps over which the trajectories where correctly smoothed. In Scenario 6 that gap was ~34% of the data span. The average gap for missed crossings was 29% while the average gap for correct connection was ~7%. On the other hand, the humans' ability to perform with relatively constant accuracy allowed them to be more accurate, in comparison with the algorithm, as difficulty increased.

Demographically, only the age group proved to be correlated with the data (Spearman $\rho = .104$, $p=0+$). While this is a significant correlation, it is weak, and therefore any future research would have to address this question in more detail.

4.1.3. Factored Performance Tables

In comparing decision sources, it is necessary to compare performance on the individual trajectories. Out of 77 possible trajectories, the average human outperformed the algorithm 10 times and did equally well 3 times. This means that 17% of the time, the average human participant did better than or the same as the algorithm. Furthermore, at least one person did better than or the same as the algorithm for a total of 45 trajectories (Table E-4, Appendix E). That means the algorithm did better than all participants only 42% of the time

Additional analysis was conducted to determine when to best rely on algorithm or human input. First for each trajectory the difference between the human and algorithm RMSE was calculated. Then differences were tallied for superior performance by either decision source or a tie, which occurred when the difference in trajectories was within one standard error for that trajectory. In addition, if both decision sources missed a trajectory, that score was tallied as a tie. All of the result tables are in Appendix E. The tables show trends in how performance varied, i.e. which decision source was better as a function of the dependent variables. The tables also show the effects of mutually missed tracks on overall performance. Table 4-4, which shows the number of superior trajectories by dependent variable, illustrates that the automation outperforms humans by a wide margin. The dark gray boxes show the LMTS's best performance with respect to the humans, in which the ratio of algorithm to human performance shows that the algorithm was superior in 92% of the easy, longer (60%, 100%) data span cases.

However, humans showed an increase in performance in the harder degree of difficulty cases. The light gray boxes in Table 4-4 depict the best human performance in which the algorithm was superior to the human in only 57% of the cases. Table 4-4 also shows humans performing much better in comparison with the algorithm in the Easy, 30% case and the Hard 60% and 100% cases as compared to those in the dark gray boxes. This implies that humans have the opportunity to contribute the most when the algorithm has the least amount of information and/or the crossing angles are difficult.

Table 4-4:
Performance Results as a Function of Decision Source

Superior Decision Source		*Degree of Difficulty*		*Total*
Human and Tie	**Data Span**	**Easy**	**Hard**	
	30%	90	144	263
	60%	26	69	95
	100%	21	118	139
Algorithm	**30%**	270	188	458
	60%	261	284	545
	100%	308	213	521

To analyze the effect of jointly missed trajectories had on performance, Table 4-5 was tabulated, removing the jointly missed trajectories from consideration. Figure 4-7 was created in order to compare the results from both tables. The Hard, 30% data span case was significantly different in both cases, ($\chi 2 = 5.381$, $p=0.016$, $\alpha = 0.05$) before and ($\chi 2 = 26.472$, $p=0+$, $\alpha = 0.05$) after, even though approximately 30% of the cases in this factor-level crossing were jointly missed trajectories. The difference in the Hard, 30% factor level shows that the interpretation of missed trajectories as ties could change how to interpret human performance. While both trajectory detection and trajectory accuracy play a significant role in determining overall performance, it is important to show that they should be analyzed separately.

Table 4-5:
Performance Results Without Including Joint Missed Trajectories as a "Tie"

Superior Decision Source		*Degree of Difficulty*		*Total*
	Data Span	**Easy**	**Hard**	
Human and Tie	**30%**	89	86	175
	60%	26	69	95
	100%	21	118	139
Algorithm	**30%**	263	168	431
	60%	257	276	533
	100%	255	206	461

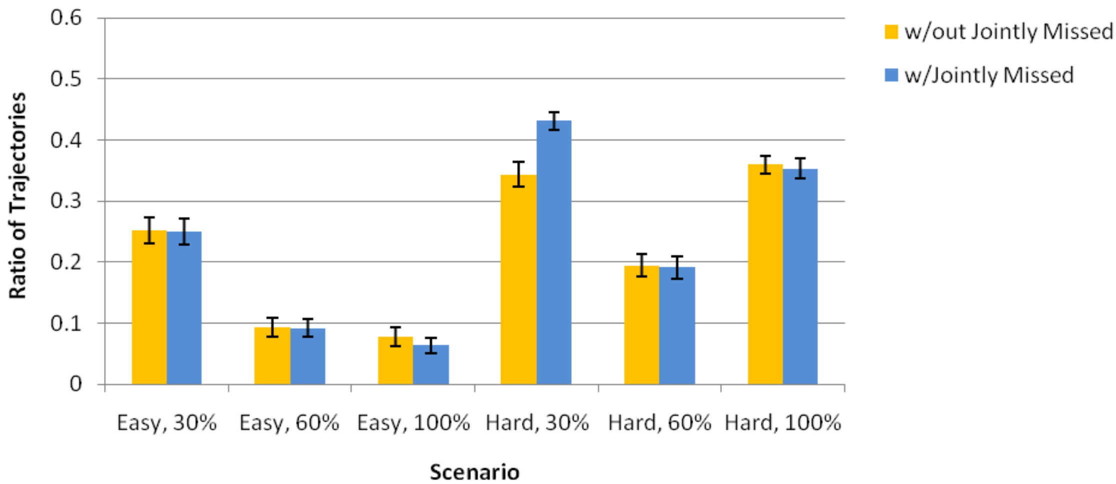


Figure 4-7: Ratio of trajectories which humans are superior to algorithm

It is instructive to also consider the superior participant's performance. Out of 29 participants, participant 8 was more than 2 standard deviations above the average of the number of superior human tracks per participant. She was superior to the algorithm 23 times and tied it 6 times out of the 77 possible trajectories. The next best user only outperformed the algorithm in 14 of the trajectories, and the average of all users was approximately 7. This participant's performance, broken down by the dependent variables, is shown in Table 4-6. Just as with the aggregate human performance, the ratio of human to algorithm superiority increases for the hard degree of difficulty and short time span cases. Figure 4-8 shows participant 8's performance with respect to the aggregate score from the previous figure. It highlights that the best performer outperforms the aggregate user in four of the six cases.

Table 4-6:
Performance table for Participant 8

Superior Decision Source		***Degree of Difficulty***		***Total***
	Data	**Easy**	**Hard**	
Human and Tie	**30%**	5	3	8
	60%	1	6	7
	100%	3	7	10
Algorithm	**30%**	7	6	13
	60%	11	7	18
	100%	9	6	15

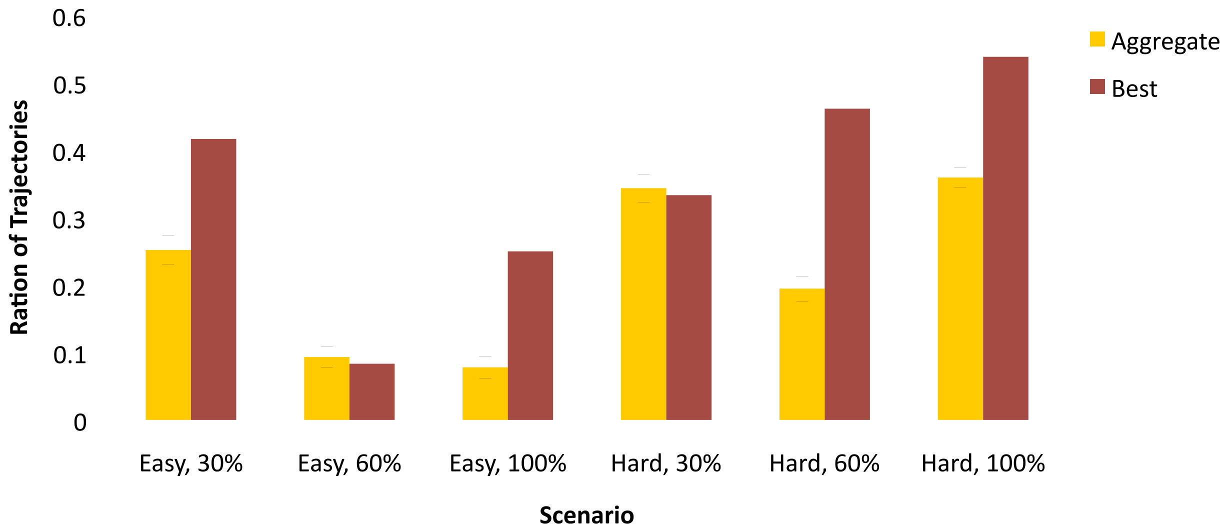


Figure 4-8: Ratio of trajectories, by factor level crossing, of the best participant (the red bars) as compared to the aggregate.

4.1.4. Confidence Measurement

To address the second hypothesis, which stated, "Human participants will be less confident at the short data span and narrower crossing angles, as there is less information and trajectories are harder to interpolate," participants were asked to color code the trajectories with a confidence level. Confidence levels were coded with a score, where a score of "1" corresponds to a "High" confidence, a "2" represents a "Medium" confidence, and a "3" represents a "Low" confidence. The results of averaging those values are shown in Figure 4-9.

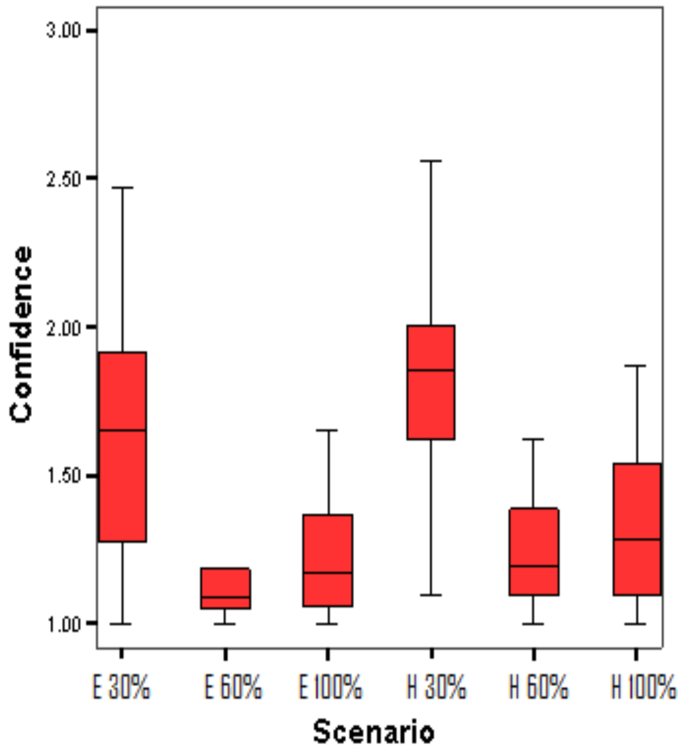


Figure 4-9: Averages of confidence measurements

It should be noted that while all participants were instructed to update their confidence levels depending upon perceived ability, the confidence measurement tool was sometimes ignored. Since the default confidence level was "High", all of the confidence levels in Figure 4-9 may be biased towards a higher confidence level than may have existed.

Pairwise tests (dependent Mann-Whitney U) were run for 9 factor-level crossings. A full table of those results can be found in Appendix E. Results from these tests show significant differences for the following comparisons (degree of difficulty by data span): Easy 30% - Easy 60%, Easy 30% - Easy 100%, Hard 30% - Hard 60%, Hard 30% - Hard 100%. There is also a significant difference between the 30%-60%, 30%-100% data spans. There is no significant difference between the 60% and 100% data spans or between degrees of difficulty (Figure 4-10). From the human perspective, it appears that only the data span played a role in determining the confidence of the participant, as there was no difference between easy and hard factor levels. Interestingly, the scenarios where the participants lacked confidence were those ones that they provided the most benefit in terms of correcting the automation.

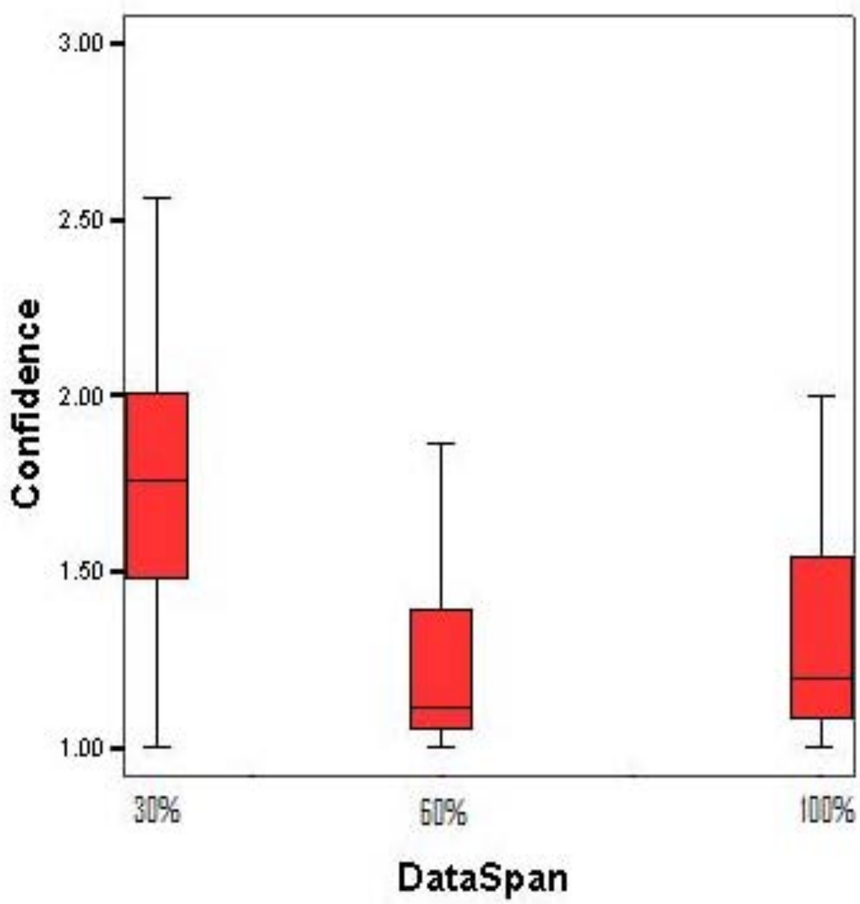


Figure 4-10: Confidence per data span box plots.

4.2. Discussion

This section discusses the results of Experiment One as they relate to the hypotheses listed in Chapter 3. The second and third research objectives, which aim to discover decision source performance and human confidence, are addressed through discussion of the hypotheses. The first hypothesis states:

> *Human and algorithm performance will deteriorate with the shorter data span and narrower crossing angle cases, as there is less information and the trajectories are harder to interpolate.*

In support of this hypothesis, the following findings are offered. Overall, human users missed more trajectories, while the algorithm predicted more false trajectories. Based on average RMSE, both human and LMTS performance deteriorated with the narrower crossing angle (hard degree of difficulty) cases. Human performance increased as compared with the algorithm for shorter data spans and harder degrees of difficulty.

Comparison of decision source factor levels showed where there may be opportunities for improvement through collaboration. Using the trajectory performance tables, the algorithm is the strongest at the easy, 60% and 100% data span factor level crossings and significantly outperformed the human in those scenarios. Human input became important at the easy, 30% and hard, 30%, 60%, and 100% factor level crossings where 20%-40% of the trajectories smoothed by the participants were superior. Therefore, there is a higher likelihood of collaboration being beneficial in the shorter data span and hard degree of difficulty where the human could best assist the algorithm. This suggests updating the first hypothesis to state:

> *Collaboration between the human and algorithm will produce the greatest benefit with the shorter data span and narrower crossing angle cases.*

It was also shown that missed and falsely predicted trajectories and overall accuracy should be analyzed separately. This is important because it means that increasing accuracy may not decrease the amount of missed or falsely predicted trajectories, and both should therefore be studied. Furthermore, the best participant did substantially better than the aggregate user, which shows that individual performance matters.

The second hypothesis from this experiment states:

> *Human participants will be less confident at the short data span and narrower crossing angle cases, as there is less information and the trajectories are harder to interpolate.*

The confidence results show that the only significant differences lie between the 30%, 60% and 30%, 100% factor level crossings. The most important inference gained from this is that the human user is significantly less confident in cases that it can best assist the algorithm. While there is less confidence at the short data span scenarios, there is no difference in confidence between degrees of difficulty, so the hypothesis was not accurate. The updated second hypothesis states:

> *Human participants will be less confident at the short data span cases as there is less information to interpolate.*

While it was not hypothesized, it has been shown that the size of information gap plays an important role in the correct connection of partial trajectories. This is important, especially for the hard degree of difficulty factor, as it creates an increase in missed and false trajectories and in RMSE. However, it has also been shown that both LMTS and the human participants respond to this gap differently, which will need further study to understand.

4.3. Summary

An experiment was conducted to address the second and third research objectives listed in Chapter 1. Two hypotheses were created to address those objectives. The first hypothesis was supported as algorithm and human accuracy decreased in the short data span and hard degree of difficulty scenarios. These results indicated that humans could possibly add value for the short data spans, especially with narrow crossing angles and in extrapolation. It was concluded that both humans and automated algorithms can contribute to the track smoothing task. The second hypothesis was not accurate as results showed that confidence is dependent solely on data span. The second hypothesis was restated to express the results that humans will be less confident when their input is most important. The next step in this investigation was to use these results to predict cases of algorithm failure and exploit human augmentation of those cases for a better result.

5. Experiment Two Design

The second experiment was designed to address the fourth research objective, to study a collaborative effort between human participants and the LMTS algorithm. It was motivated by the results of Experiment One, which suggest the creation of human-augmented system to achieve superior system performance. The subsequent sections of this chapter will describe how the information regarding that collaborative effort was collected and analyzed. The hypothesis, participants, apparatus, experimental design, and testing will be discussed.

5.1. Hypothesis

The results of Experiment One showed that for four distinct factor level crossings, a collaborative effort is possible. In order to exploit areas of possible improvement, the algorithm results from Experiment One were analyzed to find the areas with the highest probability of algorithm error. Experiment One showed that the algorithm had the worst accuracy when it incorrectly connected partial trajectories. Generally the incorrectly connected trajectories occurred when LMTS connected partial trajectories over visually large occlusions. This was shown in the Section 4.1.2, where the information gap was shown to limit the ability of LMTS to correctly connect trajectories. This is logical as the algorithm would be expected to function less well if information to be connected was more greatly separated. Also shown in Section 4.1.2 was that the information gap over which the algorithm failed was, on average, 29% of the trajectory on the Time After Launch axis, or X-axis.

In order to create a human-augmented effort that allows the human participant to input information into the cases of most likely algorithmic failure, a criterion was created for the invocation of human performance. As was shown, the algorithm functioned the worst over large occlusions. While the average of these cases was 29% of the trajectory, the smallest gap was 22%. Since the exact percentage of the range over which the algorithm is more likely to fail could not be quantified statistically, a gap of 20% will be used as a conservative criterion for invoking the human in the smoothing task.

The process by which trajectories were identified for human assistance is as follows:

1. LMTS first solves the entire scenario and saves the results. An example of its performance is shown in Figure 5-1.

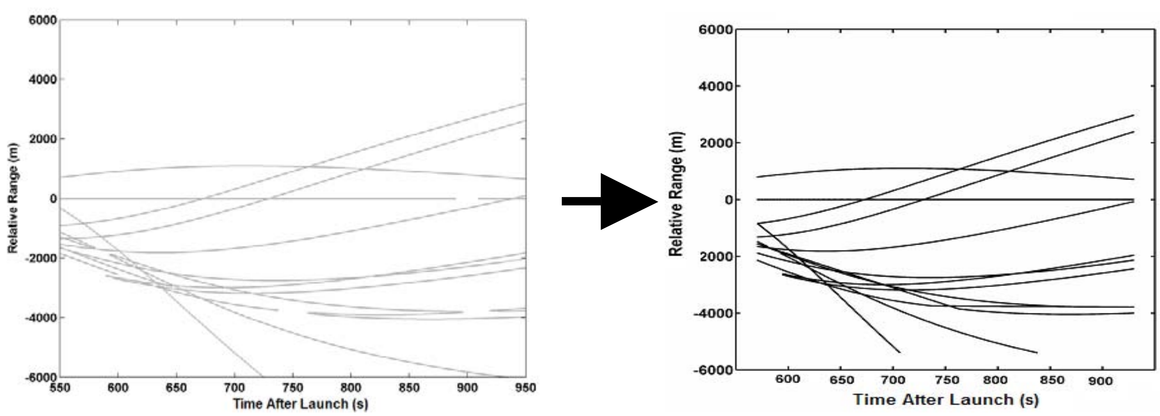


Figure 5-1: Algorithm solution (right) of a test scenario (left)

2. Any smoothed trajectory in the results that meets the conditions previously described (crosses over an information gap of greater than or equal to 20%) is considered an "incorrect trajectory". Such a trajectory is shown as the dashed trajectory in Figure 5-2 .

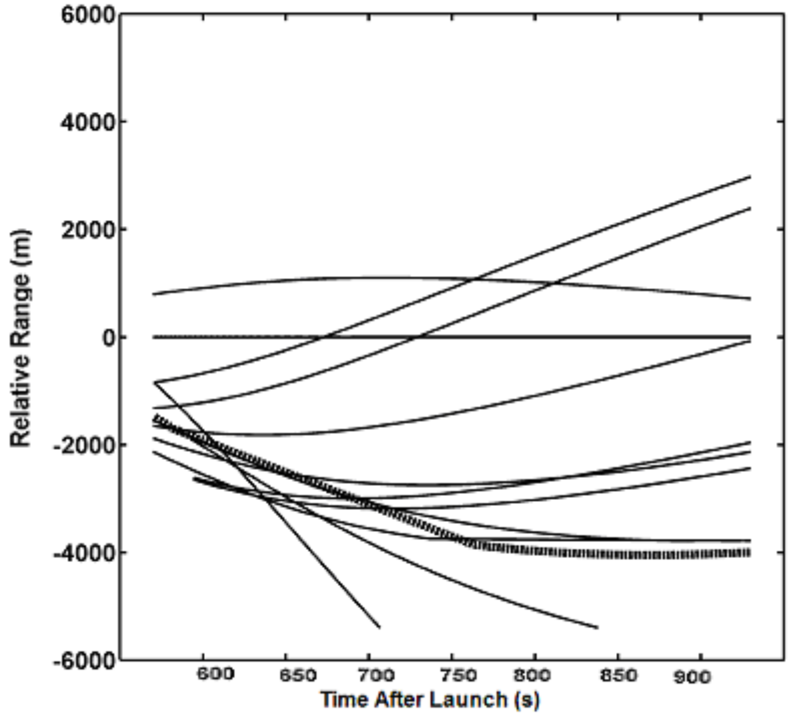


Figure 5-2: In the algorithm solution, the bolded, dashed line highlights the trajectory that connected segments which were greater than 20% of the data span apart from each other

3. All smoothed trajectories that cross the incorrect trajectory at any point are also considered possibly incorrect trajectories. While they might be correct, there is no way of knowing during the experiment, and since these trajectories have been shown to be incorrect in the past, they will be assumed to be incorrect for the human-augmented case. These trajectories are shown in bold in Figure 5-3.

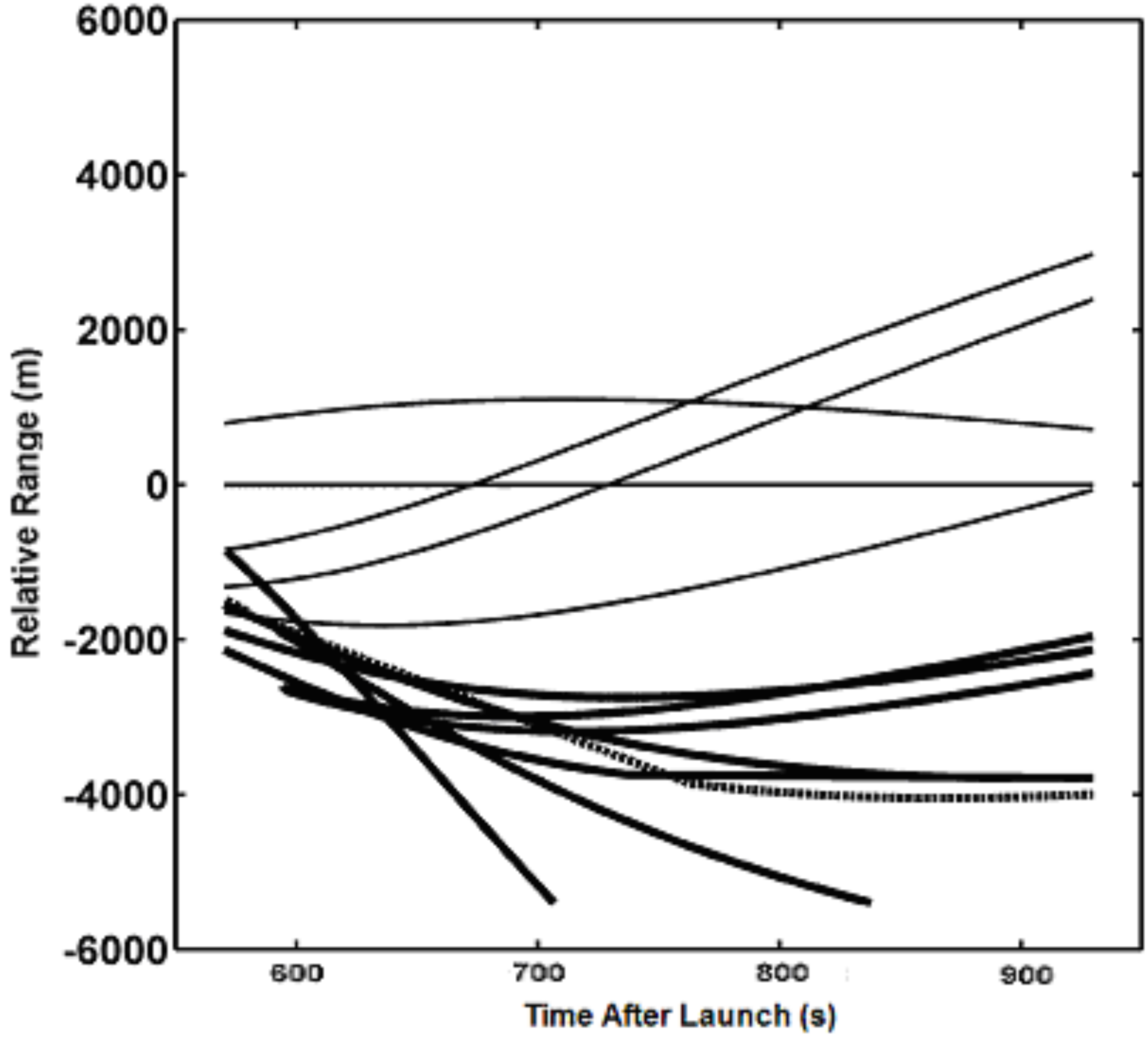


Figure 5-3 - All bolded lines meet the criterion in step 3

4. Since the track that met the 20% criterion connected multiple track segments, any (or all) of those segments could actually be parts of different trajectories. Therefore, any crossing trajectory (which may or may not have met the standards listed above) has some probability of containing the correct segments of the track (and vice versa). All segments are then considered to be possibly incorrectly connected and require operator attention.

5. Track segments used in the incorrect trajectories are then presented to the human operator to smooth

In summary, the criterion of a 20% gap acts as a flag for invoking the human review of the trajectories that were most prone for algorithm error. On the basis of this observation the following hypothesis was made to address the fourth research objective:

5.1.1. Hypothesis Three

The human-augmented decision source will produce a superior solution to the automation acting alone.

5.2. Participants

The participants were selected insofar as possible from the best participants in Experiment One – 12 in all. Best is defined as anyone who performed average or better (Appendix E). The 12 participants chosen, 6 male and 6 female, were all Lincoln Laboratory employees. The age group percentages were similarly spread as compared to Experiment One, 25% over 50, 33% were ages 35-50, 33% were ages 25-35, and 8% 18-25. Only 7 of the participants had computer-based drawing experience, which is similar to the first experiment. The complete subject demographic information is listed in Appendix F.

5.3. Apparatus

The apparatus and test bed used by the participants in Experiment Two was the same interface as used in Experiment One and described in Chapter 3, with one minor alteration. The hypothesis previously stated does not require a confidence measurement, so the confidence panel was removed. The interface used in this experiment is shown in Figure 5-4.

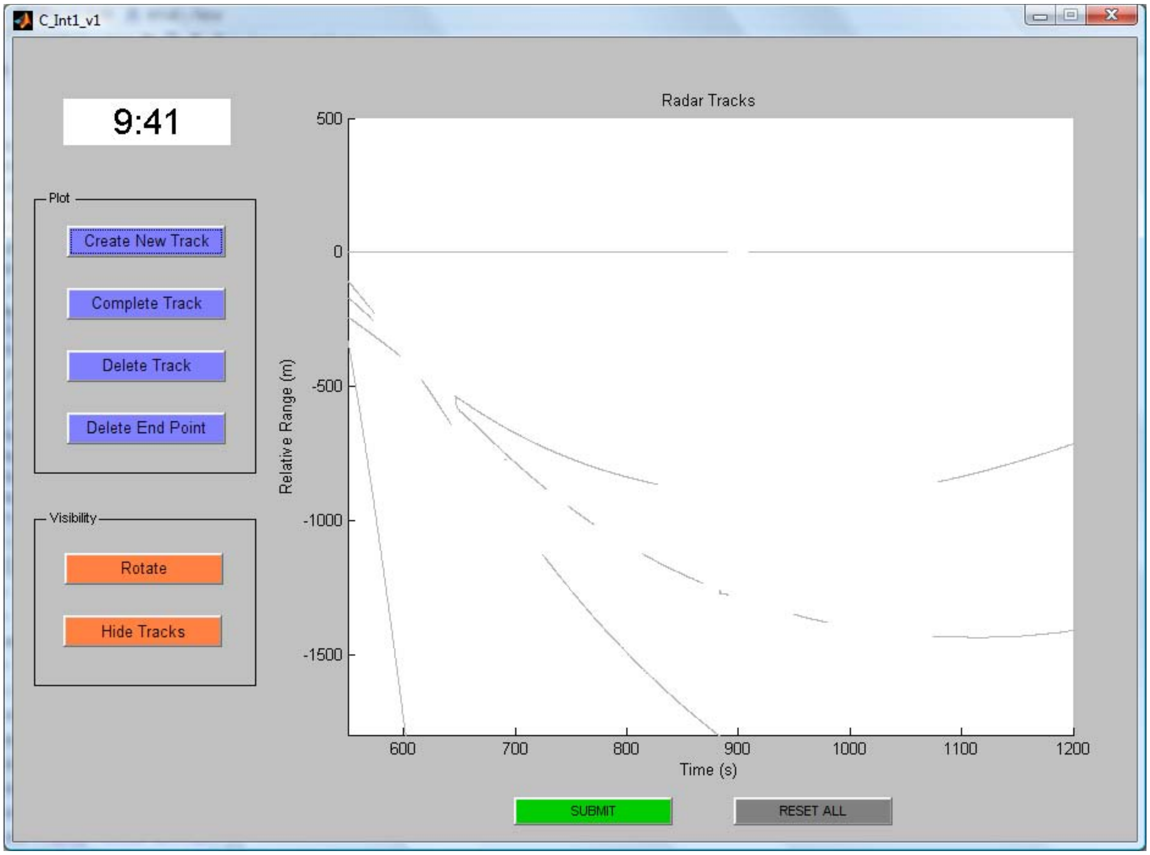


Figure 5-4: Experiment Two Interface

5.4. Experimental Design

The independent and dependent variables will be outlined in this section.

5.4.1. Independent Variables

The independent variables used in this experiment were the same as the variables used in Experiment One. The decision source for this experiment was either the LMTS algorithm or the augmented human decision source detailed in Section 5.1. Degree of Difficulty (Hard or Easy) and Data Span (30%, 60% 100%) were reused in this experiment. The factor level crossings that were examined, however, were limited only to the cases in which a collaborative effort had the best opportunity to be successful (Figure 4-7). Therefore, only the Easy 30%, Hard 30%, Hard 60%, and Hard 100% scenarios were examined.

5.4.2. Dependent Variables

Trajectory detection and trajectory fit were the primary dependent variables used to score performance. They are the same variables as used in Experiment One. The scoring of trajectory fit did not consider the trajectories that the algorithm could not extrapolate. The issues with these trajectories were discussed in Section 4.1.2. Since Experiment Two contains more difficult scenarios, it is expected to have a number of trajectories the algorithm cannot extrapolate. Therefore, to fairly compare trajectory accuracy between decision sources, only interpolated (i.e. completely plotted by LMTS) trajectories were evaluated for trajectory fit. However, if the augmented human decision source is forced to extrapolate, those trajectories will be individually evaluated to see how well they did as compared to the other interpolated trajectories.

5.5. Testing

The testing was altered from Experiment One because it only tested the 12 best performers from Experiment One, and each user smoothed eight scenarios. There were 13 initial trajectories per scenario. The following section will detail the pre-experiment activities, a training session, and a test session.

5.5.1. Pre-Experiment Activities

Each participant was first introduced to the experimental setup. Next, each participant read and signed the Consent to Participate Form (Appendix G).

5.5.2. Training Session

After filling out the required information, a tutorial instructing them on the intricacies of the interface was presented. As a refresher from the previous experiment, the tutorial started with an overview of the BMDS scenario and the motivation for the trajectory smoothing task (Appendix B). The same tutorial was used for both Experiment One and Two, however the participants were told to ignore all information regarding the confidence measurement in the second experiment. After the tutorial, Camtasia® software, which recorded all interface activity during each experiment, was turned on. Finally the interface was started.

Next, to provide initial training, three practice scenarios were given to each user. The three practice scenarios were chosen to familiarize the user with both the interface and the interpolation task they were asked to perform. The practice scenarios were given in a fashion that would best facilitate training. The first scenario had only 4 trajectories to interpolate, and the scenarios became progressively harder until the last practice scenario, which was as difficult as any data set the participants would be asked to interpolate in the test sessions.

5.5.3. Test Session

After the three practice sessions, all participants were asked to plot eight scenarios, which were derived from 2 scenarios of each of the studied factor level crossings previously mentioned (Table 3-2). While the practice scenarios were given in a specific order, all test scenarios were randomized such that there was no specific order or presentation. All users were allowed to ask questions throughout the entire experiment regarding the interface and its capabilities.

Table 5-1: Scenarios Seen by Participants

Scenario	Crossing Angle	Data Span

Practice Scenario	Easy	30%
Practice Scenario	Easy	60%
Practice Scenario	Hard	100%
Scenario 1	Easy	30%
Scenario 2	Easy	30%
Scenario 3	Hard	30%
Scenario 4	Hard	30%
Scenario 5	Hard	60%
Scenario 6	Hard	60%
Scenario 7	Hard	100%
Scenario 8	Hard	100%

5.6. Summary

The second track smoothing experiment was designed to evaluate the third hypothesis developed from Experiment One, which was that a collaborative decision source could better smooth trajectories than the LMTS algorithm alone. Independent and dependent variables were created to accurately measure the necessary information to determine which was better and why. Subjects received background and training prior to the experiment and each subject completed eight test scenarios. All information was gathered in real time and recorded for analysis. The results will be presented in the next chapter.

6. Experiment Two Results and Discussion

This chapter addresses the results of the second track smoothing experiment and their implications. The summary addresses the third hypothesis which was presented in Section 5.1.1 of this thesis.

6.1. Results and Implications

The results are addressed in three categories: 1) Missed or false trajectories, 2) Accuracy of correct trajectories, and 3) Total performance tables.

6.1.1. Missed or False Trajectories

In Experiment Two, there were a total of 104 true trajectories. Thirty-two trajectories were found to be accurately plotted by the algorithm on the basis of the previously defined 20% criterion, and therefore all data associated with them were removed from further consideration. The 12 tested participants were then presented with 72 trajectories each (which was relatively close to the 77 trajectories plotted in the first experiment). Of these, LMTS missed a total of 2 trajectories and predicted 13 false trajectories. The combined augmented users missed a total of 42 trajectories and predicted 31 false trajectories. This averages to 3.5 (1.17 std dev) missed and 2.58 (2.31 std dev) false trajectories predicted per augmented decision source. The means and standard deviations for all participants are listed in Appendix I. The implications of these results for the third hypothesis will now be considered in detail.

6.1.1.1. Algorithm-Missed Trajectories

The first step in the analysis was to research the 2 trajectories the algorithm missed and determine if any human participants detected those trajectories. The algorithm only missed trajectories in the 30% time span case, one in both Scenario 1 (Easy, 30%) and Scenario 3 (Hard, 30%). The specific trajectories the algorithm missed are designated by the arrows in Figure 6-1 and Figure 6-2. The areas designated in Figure 6-1 and Figure 6-2 show that the algorithm missed trajectories due to a lack of information. Table 6-1 shows that there were two cases where some humans managed to form trajectories when the algorithm could not. These cases were similar to those found in Experiment One where two trajectories should have been formed in the location where only one was plotted. The results in Table 6-1 show that some participants were able to detect these trajectories, however in no instance was the augmented decision source able to detect the existence of two side-by-side trajectories.

Table 6-1:
The Number of the Augmented User Correctly Plotted Trajectories vs. the Number of Missed Trajectories for the 2 Trajectories Missed by the Algorithm

Scenario, Track, Factor Level	*Scenario 1, Trajectory 5, Easy 30%*	*Scenario 3, Trajectory 5, Hard 30%*
Augmented-Correct Trajectories	2	2

Augmented-Missed Trajectories	10	10

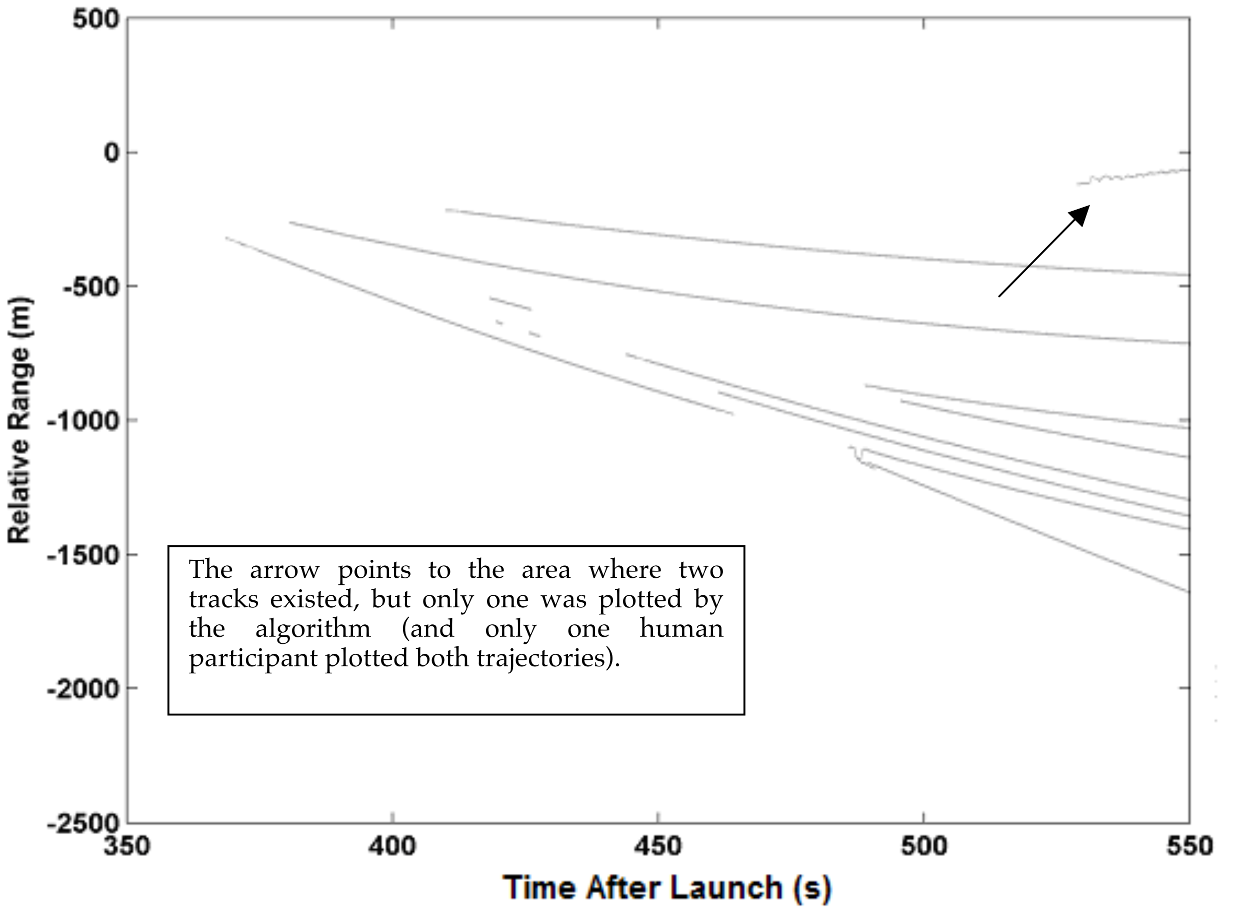


Figure 6-1: The raw data for Scenario 1 (Easy, 30%).

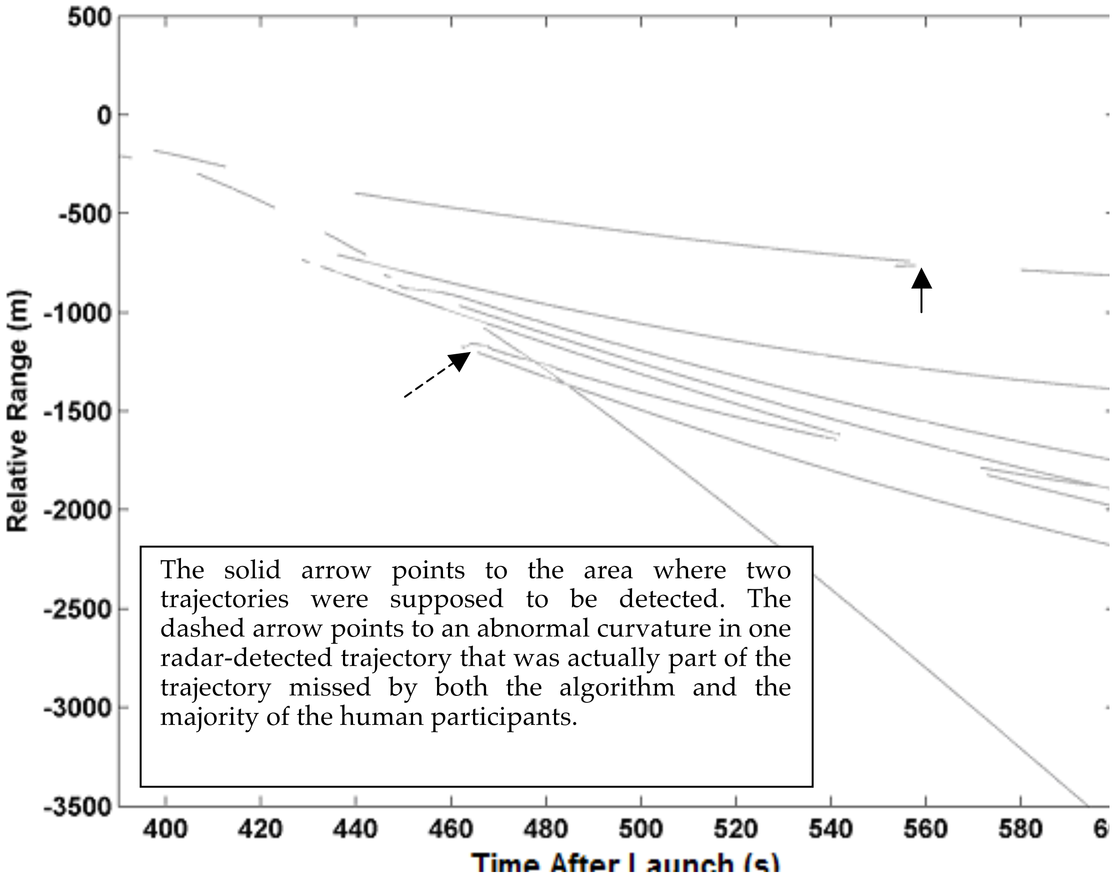


Figure 6-2: The data for Scenario 3 (Hard, 30%).

6.1.1.2. Augmented Decision Source-Missed Trajectories

The next step in the analysis was to look into the cases in which the human-augmented decision source missed trajectories. There were three trajectories (4.2%) that the majority (10) of the augmented participants missed. All participants missed at least 2 trajectories during the course of the experiment. The 3 trajectories that approximately half or more of the humans missed are listed in Table 6-2, which is an expansion of Table 6-1. The misses correlate well with the 2 trajectories missed by LMTS. A full list of these results is given in Appendix I.

Table 6-2:
The 3 Most Missed Trajectories by the Augmented Decision Source and the Percentage of Augmented Participants Who Missed These Trajectories

Scenario, Trajectory Factor Level	*Scenario 1,*	*(Scenario 3,*	*(Scenario 3,*

	Trajectory 5, Easy 30%)	Trajectory 1, Hard 30%)	Trajectory 5, Hard 30%)
Augmented-Correct Trajectories	2	2	1
Augmented-Missed Trajectories	10	10	11
Algorithm-Missed (Yes or No)	Yes	No	Yes
Missed Trajectory %	83%	83%	92%

Table 6-2 shows that both LMTS and the augmented user missed trajectories in the 30% data span case, and the augmented user missed the most trajectories in the Hard, 30% case (which is the same as in the first experiment). Figure 6-3 shows the missed trajectories as a function of the four factor levels. The family wise error was set at $\alpha = .01$. Overall, the algorithm and the augmented decision source are significantly different (Wilcoxon, $Z = -3.84$, $p = 0.000+$). However, testing all factor levels, the only significant difference is at the Hard, 30% case (Wilcoxon, $Z = -2.71$, $p=0.007$). The Wilcoxon scores for all factor levels can be found in Table 6-3.

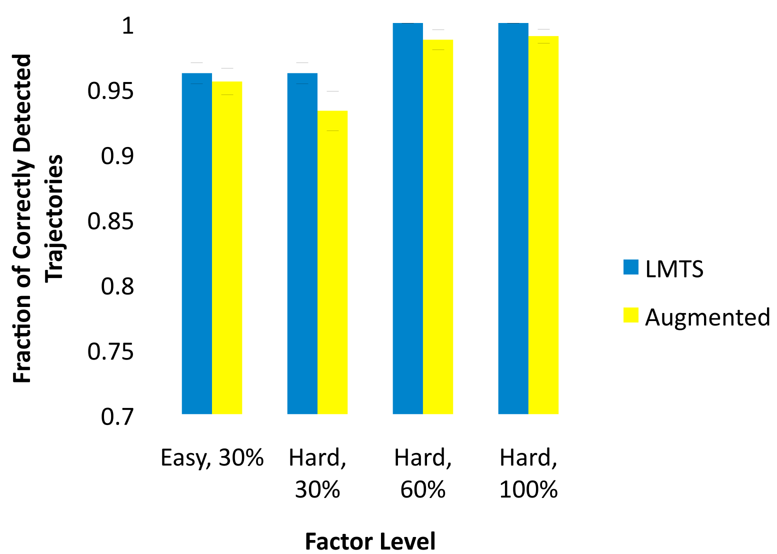


Figure 6-3: Fraction of correctly detected trajectories for both algorithm and the human-augmented decision source.

Table 6-3: Wilcoxon Scores for all Factor Levels

Factor Level	Easy, 30%	Hard, 30%	Hard, 60%	Hard, 100%
Wilcoxon Z	-1.41	-2.71	-1.63	-1.73
P	.157	.007	.102	.083

6.1.1.3. False Trajectories

Experiment Two showed similar false trajectory performance as the first experiment. The 13 trajectories the algorithm falsely predicted greatly outnumber the augmented decision source average of 2.58. Out of the 8 possible scenarios, the algorithm on average predicted 1.65 false trajectories per scenario, and the comparison between LMTS and the augmented decision source is shown in Table 6-4. The two numbers that should be compared are the ratio of average collaborative false tracks per participant vs. the number of false tracks predicted by LMTS. Comparison of these two numbers demonstrates the augmented user outperforms the algorithm in all cases. The augmented user clearly outperforms LMTS and does not predict nearly as many false trajectories. Averaging over all dependent variables, there was a significant difference between the collaborative participant and LMTS (Wilcoxon $Z = 6.219$, $p = 0+$, $\alpha = .05$).

Table 6-4:
The Ratio of False Tracks Predicted Per Factor Level

(Decision Source, Factor Level)	Easy, 30%	Hard, 30%	Hard, 60%	Hard, 100%
(# Collaborative false tracks/ # total participants)	.66	.5	.76	.7
LMTS #	4	3	4	2

There are two reasons for such large number of false trajectories from the algorithm. First, performance is similar to LMTS results from the first experiment, in which the algorithm took information from two trajectories and created three. The augmented user minimized this behavior by allowing the human to connect trajectories over areas in which the algorithm had the highest probability of failure. Second, the partial trajectories were often not connected. If two partial trajectories were not connected on either side of a large gap, they would both qualify as detected trajectories. These false hits would increase the number of "objects" detected by the algorithm above the number that actually existed.

It can also be seen that the numbers differ slightly from the first experiment. Table 6-5 represents the two false trajectory tables normalized by scenario for LMTS. It can be seen that there is an overall increase of 2.5 trajectories (38%), however, there was a decrease in the Hard, 100% crossing angle case.. This overall increase is believed due to the increase in difficulty of overall difficulty in Experiment Two.

Table 6-5:
Number of False Trajectories by LMTS Normalized by Scenario

(Experiment, Scenario)	Easy, 30%	Hard, 30%	Hard, 60%	Hard, 100%
Experiment One	0	0	1	3

Experiment Two	2	1.5	2	1

When looking at the same comparison for the individual human user in the first experiment to the augmented decision source in Experiment Two, there is still a similar increase. Table 6-6 presents the two false trajectory tables normalized by scenario for the augmented decision source. It shows an overall increase of .932 trajectories for the four factor levels in Experiment Two. This is a 72% increase in false hits, which is about 30% larger than the 42% increase by LMTS.

Table 6-6:
Number of False Trajectories by the Augmented Human Normalized by Scenario

(Experiment, Scenario)	Easy, 30%	Hard, 30%	Hard, 60%	Hard, 100%
Experiment One	.103	0	.172	.103
Experiment Two	.33	.25	.38	.35

The most likely reason for the increase in the number of overall false trajectories is that the second experiment is only testing cases in which the algorithm and the human user were most likely to predict false trajectories. The subsequent increase in overall difficulty between the two experiments would contribute, in part, to the overall increase in the number of false tracks predicted by both the algorithm and the human. However, the human still significantly outperforms LMTS predicting less false trajectories, 5 to .41 in the first experiment and 13 to 2.58 in the second.

6.1.2. Accuracy of Correctly Plotted Trajectories

In addition to missed or false trajectories, the accuracy of each decision source is an important measure of performance. For all the trajectories correctly detected (i.e. not missed or falsely identified), Equation 3-2 was used to calculate the RMSE. This section first compares the RMSE from each decision source, not including extrapolated trajectories. It then compares the accuracy of the extrapolated trajectories by the human to the average accuracy in those scenarios.

6.1.2.1. Accuracy Comparison

Figure 6-4 shows the RMSE averages for both decision sources at each factor level. This figure shows that the collaborative effort significantly increases the ability for humans to contribute, especially at the 30% data span levels. In order to fairly compare the RMSE for both LMTS and the augmented decision source, this information does not include any trajectories that the algorithm failed to extrapolate.

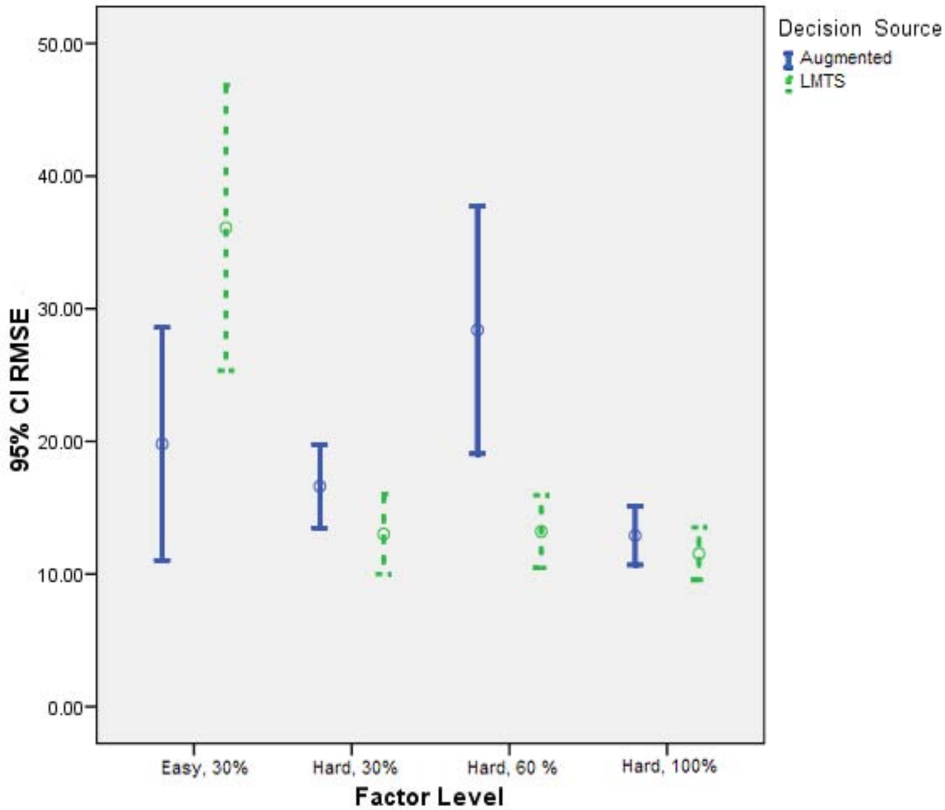


Figure 6-4: Error plot (w/o extrapolated trajectories) for both decision sources

To find how effective the algorithm was in comparison to the collaborative effort, the data were first broken down to individual scenarios, as shown in Figure 6-5. This provides a more in-depth view of the individual effort per scenario.

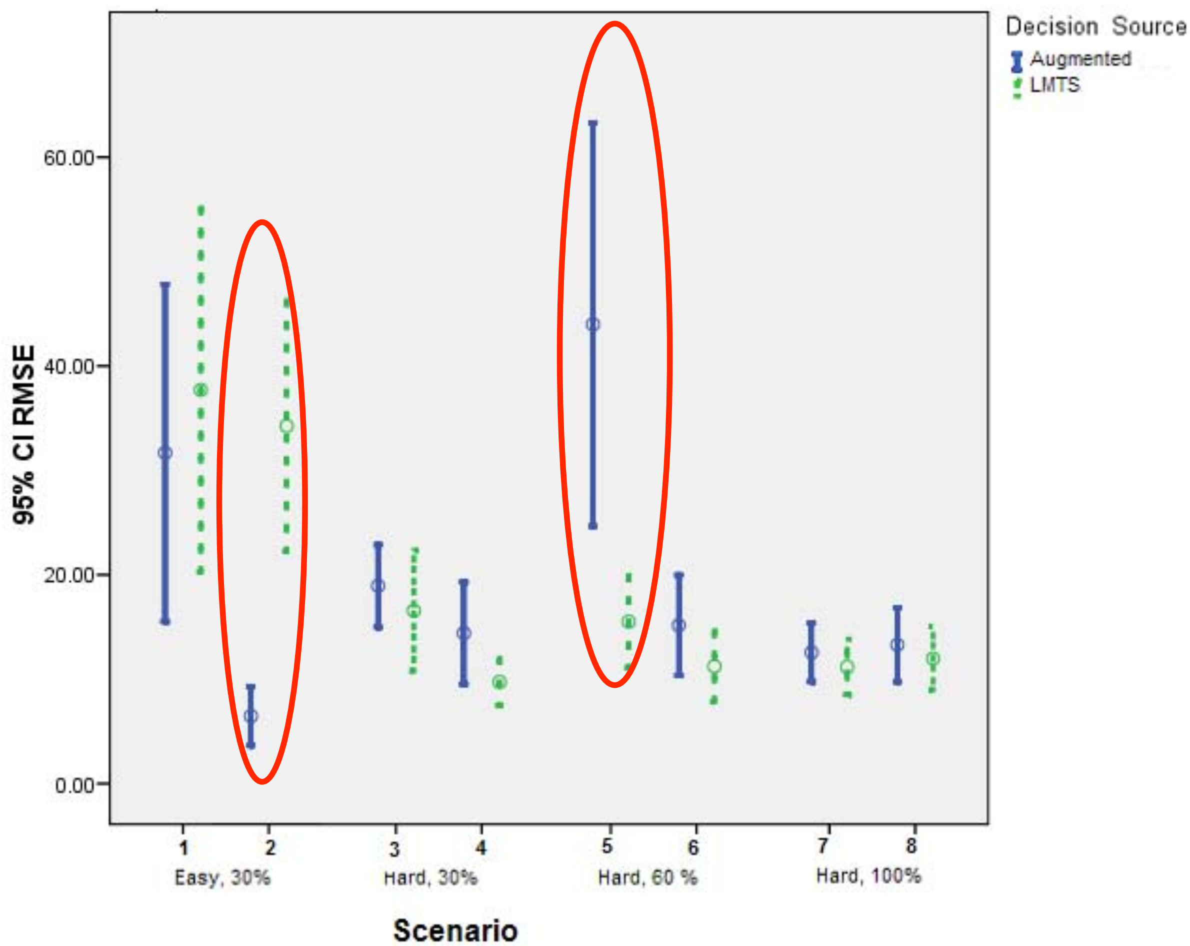


Figure 6-5: Error plot (w/o extrapolated trajectories) for both decision sources broken down by scenario

Figure 6-5 shows some interesting results. The family-wise error was set to α = .005. First, there was no overall significant difference between the algorithm and augmented user (t= -1.427, p=0.154). Further analysis showed that the augmented user significantly outperforms the algorithm in Scenario 2 (t= 4.482, p=0+). However, the algorithm outperforms the augmented decision source in Scenario 5 (t= -2.844, p=0.005). In both of those cases, the average mean difference was greater than 20 meters. However, the other cases (6 out 0f 8 scenarios) show no significant differences between algorithm and augmented user (Appendix I). The average difference between both decision sources in those cases is 3.41 m, an order of magnitude less than the two scenarios that showed significant differences. This information shows a difference from Experiment One where the average accuracy of the algorithm was significantly better, overall, than the human user. In this experiment, accuracy is significantly closer, and statistically there is no difference between algorithm and augmented user.

Figure 6-6 depicts RMSE error plots from both experiments. The red line shows the 20m accuracy level. Comparison of these charts shows that the average RMSE stays relatively constant between both experiments, 16m in Experiment One and 19m in Experiment 2. When comparing the error for the same factor levels, the average RMSE increases to 19m for Experiment One. So while there were more false trajectories predicted by each decision source in Experiment Two, the missed trajectories and accuracy stayed relatively constant.

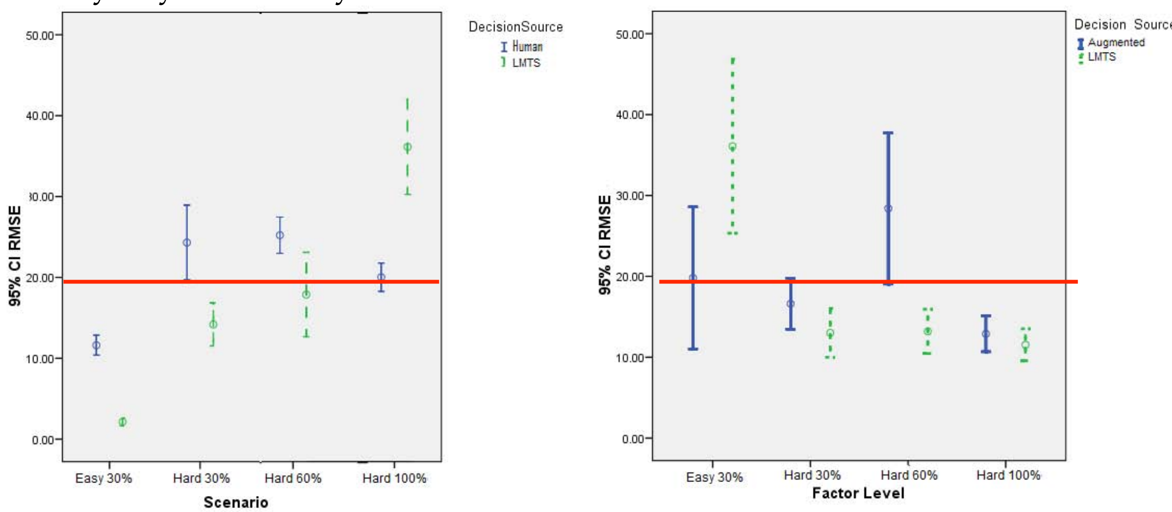


Figure 6-6: Comparison of error plots from Experiments One (left) and Two (right)

6.1.2.2. Extrapolated Trajectories

There were 14 trajectories that could not be extrapolated by the LMTS algorithm. On average 7.9 human participants (std dev 4.9) detected these extrapolations. In order to see if human contributions would be beneficial in this task, the average RMSE of those trajectories was calculated and found to be 95m, std. dev. 165m. This is approximately 80m more than the average RMSE of all other trajectories smoothed by the human augmented decision source. This shows a general lack of ability of the human to accurately extrapolate the trajectories over long distances. However, the large standard deviation is an indication that some participants may be able to accurately extrapolate. So while there is not enough information to do a full statistical analysis, it suggests that the lack of the capability to extrapolate by LMTS creates a void that can be somewhat filled by the human augmented user. However, extrapolation seems to be a problem for both decision sources and, therefore, should be a focus of future research.

6.1.3. Factored Performance Tables

In comparing decision sources, it is also necessary to compare accuracy of performance on the individual trajectories. Since the comparison is now strictly comparing human to algorithm results, only the trajectories that the algorithm and the human both smoothed were examined. Out of the 72 possible trajectories, the average human outperformed the algorithm 15 times and did equally well 4 times. Thus in 26% of the trajectories, the average human participant did better than or the same as the algorithm. The algorithm only outperformed every human participant in 20 trajectories, meaning that at least one human user did better than or the same as the algorithm in 72% of the trajectories.

Further analysis was directed to determining if the collaborative effort was the best place to exploit human input. Using the same factor performance tables found in Chapter 4, and using only the interpolated trajectories (recall the algorithm cannot extrapolate), all ties and superior human trajectories were summed as were all superior algorithm trajectories as shown in Table 6-7.

Table 6-7:
Number of Superior Trajectories as a Function of Decision Source Including Ties

Decision Source	Factor Level				Total
	Easy 30%	Hard 30%	Hard 60%	Hard 100%	
Human (Tie or Better)	53	57	105	83	298
Algorithm	45	119	88	107	359

To analyze if the humans outperformed the algorithm, strictly superior performance, not including ties, was examined in Table 6-8. It was important to see if there were any cases for which both human and LMTS plotted with the same accuracy. It is shown in Figure 6-7 that there are similar results to Experiment One, in which there were a large number of trajectories that are ties for which both decision sources plotted accurately. Furthermore, the figure shows that the collaborative effort performed as expected. The criterion for invoking human performance was set to allow the human sufficient opportunities to alter the algorithm's results. In this experiment the human user was asked to plot 70% of the trajectories, which averages to 9 out of every 13 in each scenario.

Table 6-8:
Performance Results Without Including Ties

Decision Source	Factor Level				Total
	Easy 30%	Hard 30%	Hard 60%	Hard 100%	
Human	39	25	47	54	165
Algorithm	45	119	88	107	359

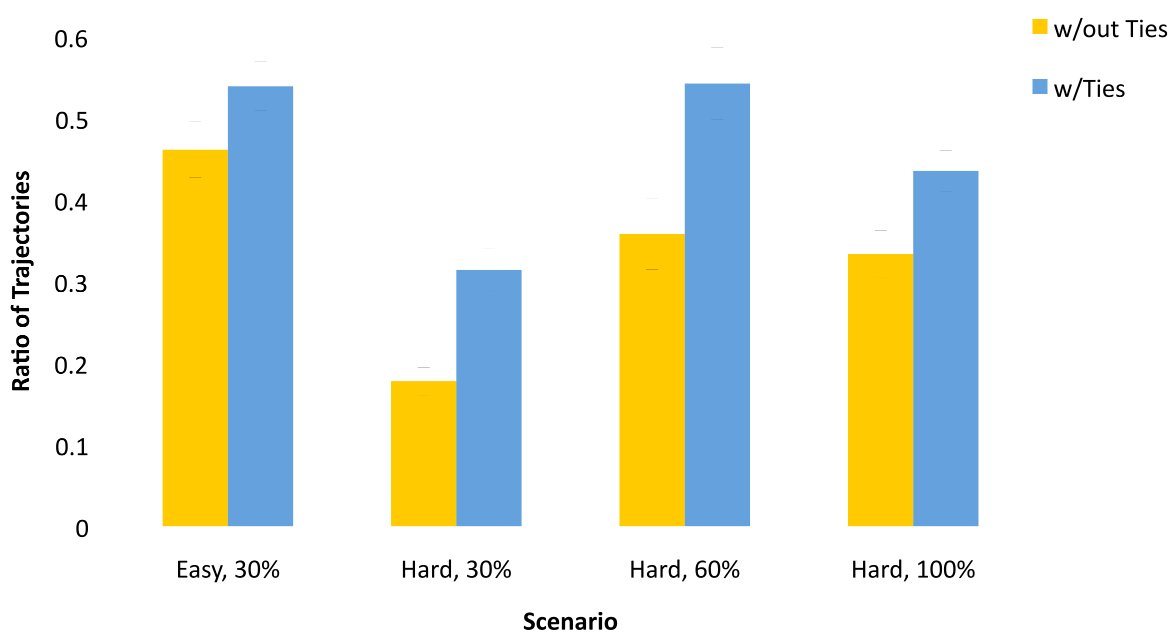


Figure 6-7: Ratio of trajectories the human either tied or was more accurate than the algorithm with and without including ties.

The performance increase between experts can be seen in Figure 6-8 which shows the difference in superior trajectories by the human in Experiment One and Experiment Two. There is an increase in superior trajectories in Experiment Two. The only factor level that does not experience an increase is the Hard, 30% Scenario, which is interesting as this was one of the best cases for the human in Experiment One. These show that overall the participants in Experiment Two did better than those of Experiment One, even with more difficult scenarios.

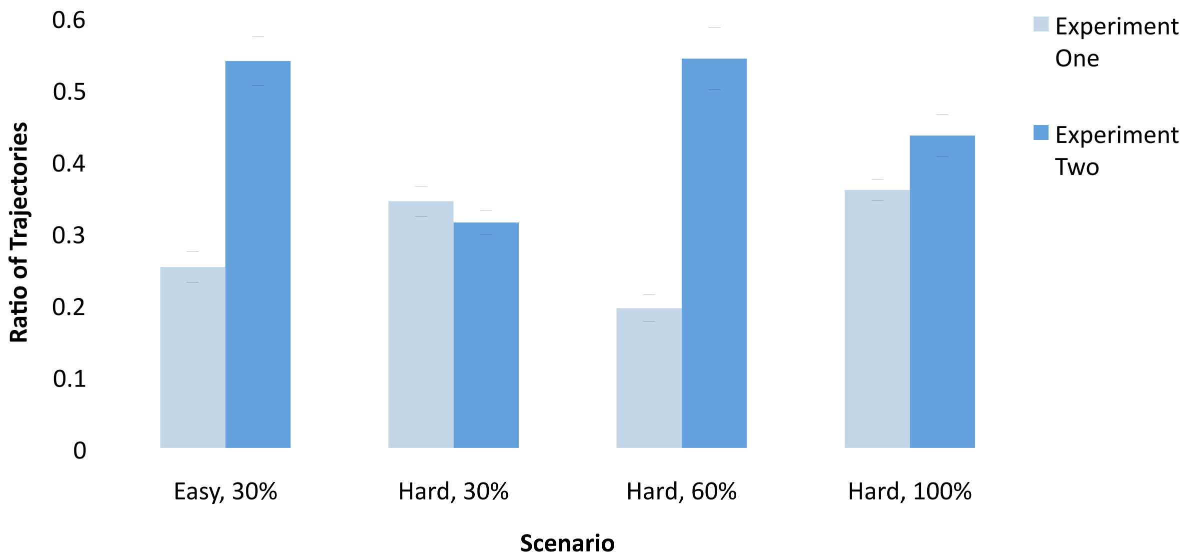


Figure 6-8: Ratio of trajectories the human either tied or was more accurate than the algorithm with out jointly missed trajectories

It is instructive to also consider superior participant performance. Participant 11 bested the algorithm 17 times and tied it 9 times. Participants 11's performance for superior trajectories is broken down by factor level in Figure 6-9. The resultant graph is similar to Figure 6-7, which is different from Experiment One. While one participant was numerically the best, there was no clearly superior participant. The average number of superior trajectories was 14 per participant, with a standard deviation of 3. While participant 11 did the best with 17, 4 other participants were superior in 16 trajectories, with 8 participants scoring better than average (Appendix I). The evidence shows that the superior performer is now not that much different from the aggregate, which suggests that the user pool is beginning to converge to the best possible performers. This further supports the evidence that the participants in Experiment Two outperformed those in Experiment One and validates that an expert user pool can really increase overall performance.

Table 6-9:
Performance Table for Participant 11

Superior Decision Source	Factor Level				Total
	Easy 30%	Hard 30%	Hard 60%	Hard 100%	
Human (Tie or Better)	6	5	8	7	26
Algorithm	3	11	9	6	29

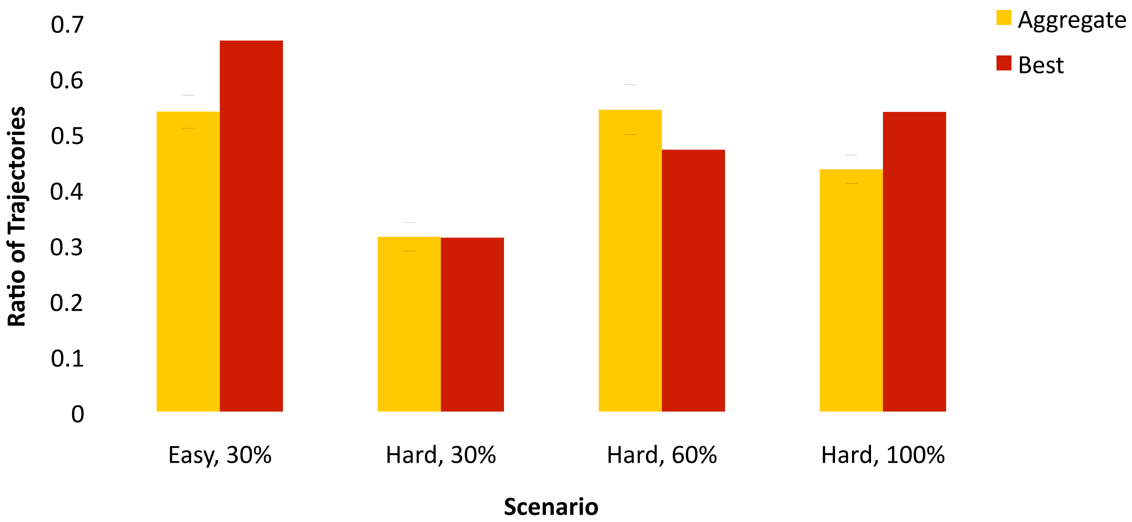


Figure 6-9: Ratio of trajectories, by factor level, comparing aggregate vs. best participant superior trajectories

Figure 6-10 depicts the best performer in both experiments. This provides similar evidence to Figure 6-9, which shows that the best performers in both experiments have similar performance in comparison with the algorithm. The only large difference is in the Easy, 30% case which is most likely due to the increase in overall performance by the human decision source in this factor level in Experiment Two. This data also suggests that as the expert user pool becomes narrower, the superior performance charts will begin to converge.

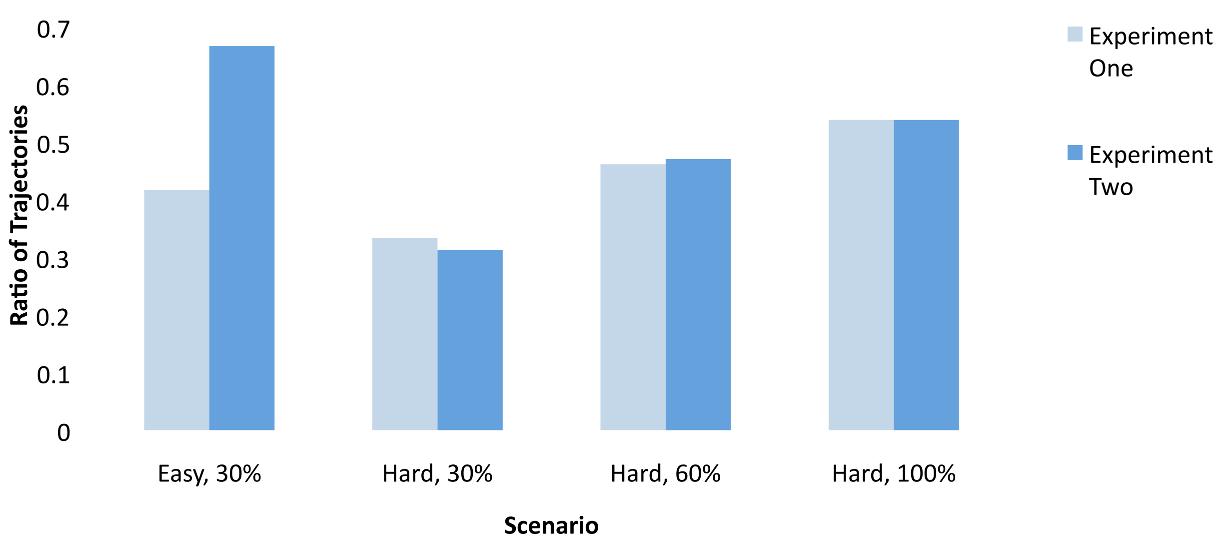


Figure 6-10: Ratio of trajectories, by factor level, comparing best users from Experiment One and Experiment Two.

6.2. Summary

An experiment was conducted to address the fourth research objectives as listed in Chapter 1. A hypothesis was created to address that objective. That hypothesis was:

Human and algorithm collaboration will produce a superior solution to the automation acting alone.

That hypothesis was shown to be partially correct. While the human augmented solution did not produce a superior solution overall, it produced a solution that in terms of accuracy was no different than the algorithm working alone. However, it was far superior in terms of fewer false trajectories and slightly worse in failed detection of actual trajectories. The hypothesis is amended to state:

Human and algorithm collaboration will produce a superior solution to the automation acting alone, depending upon the resources of the system and the needs of the operator.

7. Conclusions and Future Work

This thesis has presented the results of two experiments that investigated potential human contributions to a track smoothing task that had previously been done only by the LMTS algorithm. The results from both experiments demonstrate ways in which humans can contribute to the task. This chapter will first summarize how human augmentation of the LMTS algorithm can generate solutions superior to those generated by either the human or algorithm acting alone. It will then discuss potential tradeoffs in implementing that augmentation. Some results are then generalized to a wider domain of human-automation collaboration. Finally, possible future work is discussed.

7.1. Augmentation to Automation

Experiment One showed that both humans and automation have distinct areas of superior performance. On average, the algorithm missed fewer trajectories and was more accurate in making connections. However, the humans predicted significantly fewer false trajectories and were more accurate in one factor level crossing (Scenario 6, Hard 100%). Experiment Two revisited four factor level crossings from Experiment One using 12 of the best participants from the previous experiment. In Experiment Two, the participants were asked to smooth only those cases where the results of Experiment One suggested that the LMTS algorithm would have difficulty. The results showed that the human-augmented system matched the algorithm in terms of accuracy with no significant difference for six of the scenarios. The results further showed that the humans missed, on average, one more trajectory than the algorithm alone but predicted significantly less false trajectories than the algorithm alone. The Fitt's List from Chapter 2 has been slightly updated as shown in Table 7-1 to highlight these differences.

Table 7-1:
Fitts' List [28] for the Track Smoothing Task

Humans Are Better At	***LMTS Is Better At***
Resolving uncertainty to improvise fits to trajectories	Using strict rules to evaluate likelihood of a match
Extrapolating trajectories	Consistently producing results, such that each application of the algorithm to the same data will provide the same output
Not adding false trajectories	Detecting all partial trajectory information so less trajectories are missed
Recalling previous similar pattern-matching experiences	Quick and efficient computation

The differences in results from Experiment One to Experiment Two show that augmented automated systems can mitigate missed or false outcomes without decreasing system accuracy. While accuracy was constant between LMTS and the human-augmented system, human participation dramatically decreased the number of false trajectories as compared to LMTS alone. LMTS, on the other hand, significantly decreased the number of missed trajectories compared to the human acting alone. These benefits came about because LMTS was not forced to choose trajectories when there was a small probability of fit, and likewise, potentially confusing trajectories were not presented to the human.

7.2. Possible Role Allocations

The criterion for invoking augmentation in Experiment Two was to remove the data for any trajectory that was connected over a gap that was smaller than 20% of the data span, provided that trajectory did not connect with any other trajectory whose gap was greater than 20%. This resulted in a human-dominated augmented system, since the

20% criterion resulted in the human addressing 70% of the true trajectories. The accuracy of the trajectories processed by humans was found to be equivalent in accuracy to LMTS. This section considers the costs and benefits that changing the 20% gap criterion would have on performance in the trajectory smoothing task. The discussion will focus on missed trajectories, false trajectories, and accuracy of smoothed trajectories.

The humans averaged 1-2 more missed trajectories than the algorithm. While the algorithm was clearly superior in not missing trajectories, the 20% gap criterion and use of the best performers from the previous experiment clearly resulted in better system performance than the humans acting alone. False trajectories, on the other hand, were a problem for LMTS. LMTS's inability to extrapolate resulted in having many partial trajectories remaining, which were not complete predictions and thus labeled as false in these experiments. The algorithm predicted 5 false trajectories in the first experiment and 13 in the second experiment while the average human predicted 0.41 and 2.58 trajectories, respectively. This demonstrates that the 4 factor level crossings (Easy 30%, Hard 30%, Hard 60%, Hard 100%) are indeed much harder since both decision sources predicted more false trajectories. It also shows that the human and human-augmented system were superior to the algorithm in mitigating false trajectories.

As discussed above, accuracy increased in the human-augmented system to the point that there was no statistical difference between the augmented system and LMTS. Further adjustment of the criterion will likely result in a trade-off between missed/false trajectories and accuracy of the solution, similar to the tradeoff that occurred from Experiment One to Experiment Two. This is further discussed in Section 7.4.

7.3. Generalization of Results

While the results of this thesis are specific to the track smoothing application, extrapolation difficulties, failure to predict trajectories that actually exist, and the results of using an expert pool of users suggest generalization to other applications in human-augmented automated systems.

Extrapolation, as previously stated, can be thought of as prediction of future states. The fact that the algorithm was not programmed to extrapolate forced the human participant to have to do so, and in most cases the human performed poorly with low accuracy. The result shows that automated system designers must consider human limitations in extrapolation if it is necessary for successful system operation.

Previous research into interpolation shows that there are factors that limit the human's ability to perceive whole contours. In addition, the results of Experiments One and Two showed that both the human and LMTS had difficulties connecting correct segments due to the constraints data span and degree of difficulty of the trajectories. Thus, human and automation error may not be so different in these types of tasks, which should be a consideration in evaluating future automation of techniques that also involve human perception.

There are clear benefits of using experts. Experts, in this experiment, were identified only after an hour of operation. After the second experiment, the field had narrowed even further, but the variation was significantly less than the first experiment. After two full hours of experience, the expert users were a distinct twelve individuals. Their expertise resulted in increased mean accuracy, in comparison to the first experiment.

7.4. Future Work

This thesis suggests two topics in particular for further work: better extrapolation of trajectories and adjustable cost-benefit criterion. Extrapolation, which is essentially prediction, is needed in these track smoothing tasks to understand the past, not just the future. Understanding the past cannot only assist in prediction, but more importantly increase the likelihood of determining correct trajectories. LMTS had no ability to extrapolate. Human extrapolation, which sometimes was useful, showed poor accuracy and therefore future work should look into better techniques that may help in extrapolation.

As discussed earlier, adjusting criterion for collaboration could alter the type of smoothed trajectories. While it cannot be determined from these experiments, the alteration of the 20% criterion could also affect missed and false trajectories. For example, a criterion that would invoke smoothing on trajectories with a 10% gap might decrease the augmented missed trajectory average to approach the algorithm's performance. Future experiments may be able to establish a relationship between the number of predicted trajectories and the accuracy of the truth trajectories, which could be crucial to maximizing the cost-benefit analysis of future augmented track smoothing systems.

Appendix A: Demographic Information

Participant	Gender	Age	Career	Served in Military	Country Served	Service	Years of Service	Drawing Experience	How Often
1	Male	25-35	Engineer	No				Yes	Weekly
2	Male	>50	Scientist	No				Yes	Monthly
3	Male	25-35	Engineer	No				Yes	Yearly
4	Male	>50	Scientist	No				No	
5	Female	25-35	Engineer	No				Yes	Yearly
6	Male	25-35	Engineer	No				No	
7	Male	35-50	Engineer	No				No	
8	Female	25-35	Staff	No				No	
9	Male	35-50	Engineer	No				No	
10	Male	>50	Staff	Yes	USA	Army	22	Yes	Yearly

11	Female	>50	Librarian	No				Yes	Yearly
12	Male	25-35	Engineer	No				Yes	Yearly
13	Female	35-50	Exec. Assistant	No				No	
14	Female	35-50	Secretary	No				No	
15	Male	>50	Engineer	No				Yes	Monthly
16	Male	>50	Staff	Yes	Taiwan	Air Force	1	No	
17	Male	18-25	Engineer	No				Yes	Yearly
18	Male	>50	Librarian	Yes	USA	Air Force	4	No	
19	Male	25-35	Engineer	No				Yes	Monthly
20	Male	>50	Engineer	No				Yes	Weekly
21	Female	>50	Editor	No				Yes	Monthly
22	Male	25-35	Electrical Engineer	No				Yes	Weekly
23	Male	>50	Mathematician	Yes	USA	Navy	3	No	
24	Male	35-50	Engineer	No				Yes	Yearly
25	Female	35-50	Radar Analyst	No				Yes	Weekly
26	Female	>50	Secretary	No				No	
27	Male	35-50	Research Staff	No				No	
28	Female	35-50	AE Engineer	No				Yes	Monthly
29	Male	35-50	Engineer	No				Yes	Weekly

Pre-Experiment Survey

- *Please indicate your sex:*
  - *Male*
  - *Female*

- *Please indicate your age:*
  - *18 – 25*
  - *25 – 35*
  - *35 – 50*
  - *>50*

- *Please indicate your occupation*

____

- *Are you currently or have you ever served in the armed forces of any country?*
  - *Yes*
  - *No*

  If yes,
  - *Country: ____________*
  - *Service: __Army __Navy __Air Force*
  - *Years of service:______________*

- *Do you have any experience with drawing on a computer?*
  - *Yes*
  - *No*

  If so, how often?
  - *Weekly*
  - *Monthly*
  - Yearly

Appendix B: Tutorial

Track Smoothing Experiment Introduction

MIT Humans and Automation Lab

Research funded by
MIT Lincoln Laboratory

Overview

Objects in space, that are not accelerating, have trajectories that are smooth and continuous. However, when such objects are tracked by a radar, sometimes the software produces paths that contain gaps or track segments that may be mislabeled as if there are multiple objects present on a single track.

In this experiment you are responsible for improving the trajectories for a set of these objects by plotting what you feel is the "best fit" to the trajectories presented to you.

The available information will be presented as object tracks plotted over time and relative distances. Some of the objects may cross in the data that you are given and some may contain gaps. In each of these cases mislabeling could occur. Also, the way the data has been produced, no object will sub-divide in the time span presented so you should make sure every plotted trajectory has a start and a finish. To clarify, every object starts and finishes outside the data shown in the plot.

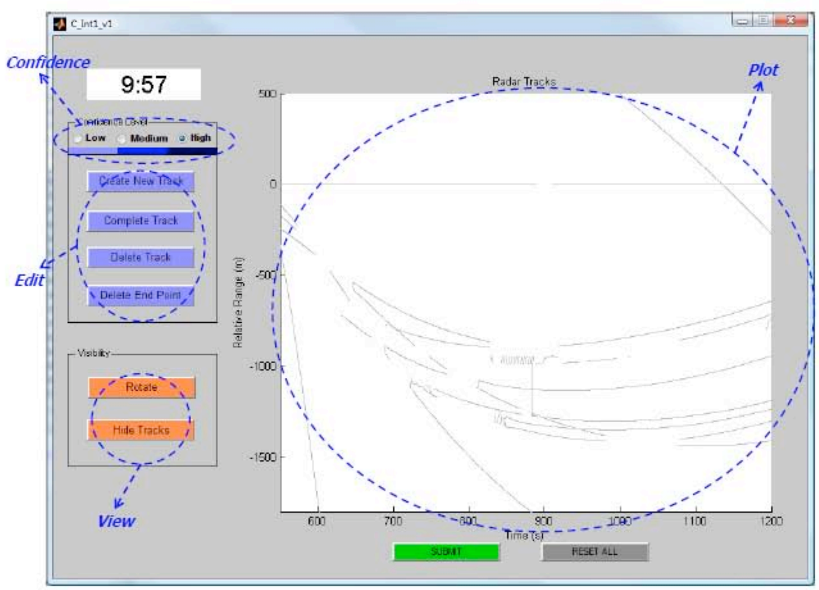
User Interface Introduction

You will interact with the data by means of a user interface. On the interface, there are two panels of buttons. The first panel deals with plotting the trajectory and manipulating it. The second panel deals with the plotted data itself. You can turn the visibility of the tracks on and off, and rotate the axes as needed to select the view that you find most useful.

In the course of the experiment you use the cursor to select the tracks that seem to you to fit best. Furthermore, you will be asked to indicate your confidence as you make these selections. You can change your confidence (High, Medium, Low) at any time both from track to track and along a particular track. The confidence will be denoted by the color change in the track, the darker the color the higher the confidence. You will be given ten (10) minutes to complete each data set.

Subsequent slides give more details about the experiment itself.

User Interface



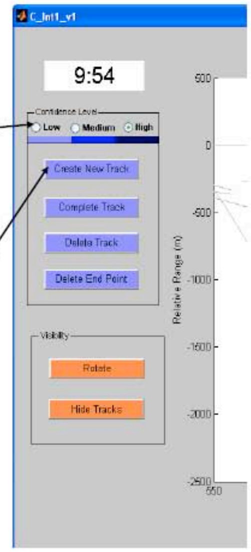
User Interface

Confidence Level

- Before plotting, you must select your desired confidence level (low, medium, high) which represents how sure you are about the track you are about to plot
- The color of the track will be reflect which level you select
- You can also alter confidence level while plotting
- However, you cannot change confidence level after you have completed the track

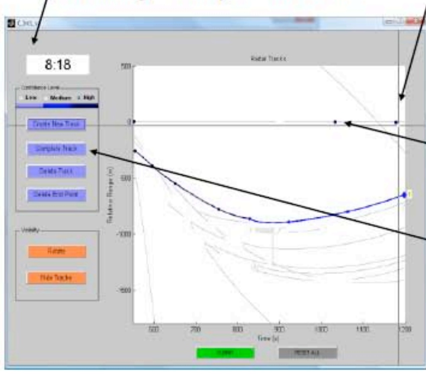
Create New Track

- To start plotting the trajectory click "Create New Track"
- This will activate a large cross hair, which then allows you to begin plotting points



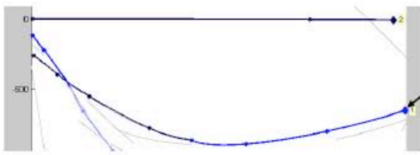
User Interface

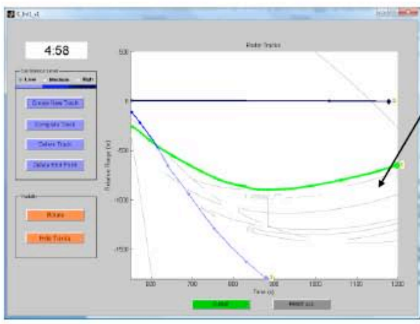
- After clicking on "Complete Track" you are in plotting mode
- The timer shows the time remaining to complete the tracks



- The cursor fills the screen so that you can tell that you're in plotting mode.
- After clicking "Create New Track," you can create points inside the plot area by clicking on the left mouse button. Outside the plot area the cursor will remain visible but you will not be allowed to plot points
- Each point will be plotted in the color selected by your confidence level
- To complete each trajectory click on the "Complete Track" button, which will complete the track at the current point.

User Interface





- After clicking "Complete Track", a line will be plotted through the points in the confidence color you selected
- Each track will be numbered in the order they were plotted.
- You can change the plot at any time. To do this:
  - First click on the track to highlight it
  - Click on the track a second time to continue plotting. You can continue the track in either direction by clicking on the end point you wish to work on (denoted by a diamond).
  - The large crosshair will appear to show that you are in plotting mode
  - If you double clicked by accident, click on "Complete Track" to leave the track as is

User Interface

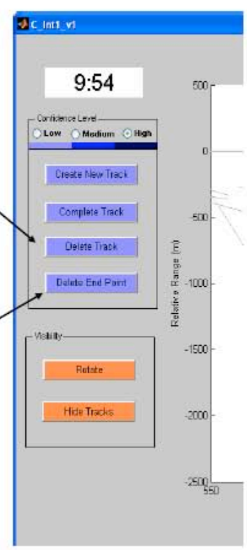
To edit a track, you must highlight it first

Delete Track

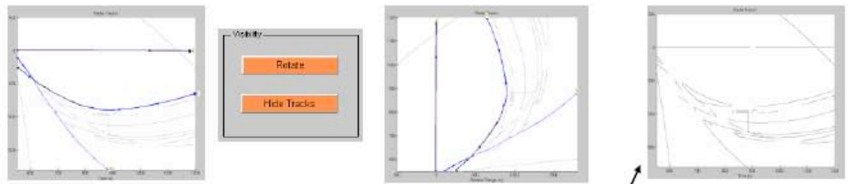
- This will delete the entire highlighted track
- It will not affect the numbering scheme

Delete End Point

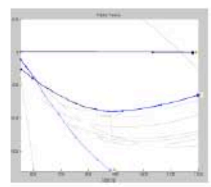
- Just like continuing the plot, you can change the trajectory by deleting the end point.
- This button will delete the current end point (denoted by a diamond) and the line between the end point and the "next" point
- Select the specific end-point by clicking on which side of the track you wish to edit
- Again, you can use this button only when a track is highlighted. The point deleted will be the one highlighted by a diamond



User Interface



- The lower panel contains the "Visibility" buttons, that control how the underlying tracks are displayed
  - The Rotate Button will rotate the axis into either a vertical or horizontal direction
  - The Hide Tracks button will hide all the tracks already plotted so that you can look at the radar data with out trajectory interference. It will also darken the data so that you can visually see that you have turned off the track visibility.
  - The Show Tracks button will then allow you to see all plotted tracks.



User Interface

Submit

- Will submit all tracks and complete the scenario
- You can not re-visit the scenario after submitting

Reset All

- Will delete all user plotted tracks and reset the interface
- Will also restart all numbering (as all tracks are deleted)

Things To Remember:

- All tracks go from start to finish. This means tracks have to start left side of the data plot and finish somewhere in the time span. For this reason you cannot plot outside the axis and no track can split
- You can continue plotting or delete points, but only from the end point, which can be selected and is highlighted by a diamond
- You cannot undo a track deletion, reset, or submit so make sure you want to perform this action before you click

Appendix C: Experiment One Consent to Participate Form

CONSENT TO PARTICIPATE IN
NON-BIOMEDICAL RESEARCH

Human Performance in Ballistic Missile Discrimination Scenarios

You are asked to participate in a research study conducted by Lee Spence, Ph.D., from the MIT Lincoln Laboratory Advanced Concepts and Technology Group. You were selected as a possible participant in this study because of your interest in improving human performance in ballistic missile defense scenarios. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

- **PARTICIPATION AND WITHDRAWAL**

Your participation in this study is completely voluntary and you are free to choose whether to be in it or not. If you choose to be in this study, you may subsequently withdraw from it at any time without penalty or consequences of any kind. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

- **PURPOSE OF THE STUDY**

Ballistic Missile Decision Support involves a number of very broad and complex issues. The system is very large, it has many interconnected elements, and it is physically spread over an area that is a significant fraction of the Earth. In addition the information on which to make decisions is often incomplete and/or inconclusive, and, given the enormity of the decision making task, the timelines are extremely short. Thus, the allocation of tasks automation and humans requires careful consideration of the areas where each can best perform. The general purpose of this research program is to investigate automation and human operator performance. To this end, this experiment will investigate how well humans and computers can assess missile crossing tracks.

- **PROCEDURES**

If you volunteer to participate in this study, we would ask you to do the following things:

  - Participate in a 15 minute training session to familiarize yourself with the display and test conditions.
  - Assess automated track identification post track crossings in up to 10 test scenarios, each about 5 minute in length.
  - All of these steps will occur in the Lincoln Laboratory S-Building (South Laboratory), Room S1-346 and in laboratories in Building 33 on campus.

- **POTENTIAL RISKS AND DISCOMFORTS**

There are no foreseeable risks, discomforts, inconveniences in participating in this experiment.

- **POTENTIAL BENEFITS**

Your benefit in participation in this study is developing a better understanding of human and computer strengths and weaknesses in the track estimation task. In terms of benefit to society, this research will provide for better understanding of human versus computer capabilities in track estimation, which is applicable not only to ballistic missile defense but also the air traffic control and other domains.

- **PAYMENT FOR PARTICIPATION**

Participation in this experiment is strictly voluntary with no payment.

- **CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. Your performance in this study will only be coded by your subject number, which will not be linked to your name so your participation in this research is essentially anonymous.

- **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please feel free to contact Lee Spence at Group 32 – Advanced Concepts and Technology, MIT Lincoln Laboratory, 244 Wood St, Lexington MA 02240-9185 (781) 981-5043 or Professor Cummings at 77 Massachusetts Ave., 33-305, Cambridge, MA 02139 (617) 252-1512.

- **EMERGENCY CARE AND COMPENSATION FOR INJURY**

If you feel you have suffered an injury, which may include emotional trauma, as a result of participating in this study, please contact the person in charge of the study as soon as possible.

In the event you suffer such an injury, M.I.T. may provide itself, or arrange for the provision of, emergency transport or medical treatment, including emergency treatment and follow-up care, as needed, or reimbursement for such medical services. M.I.T. does not provide any other form of compensation for injury. In any case, neither the offer to provide medical assistance, nor the actual provision of medical services shall be considered an admission of fault or acceptance of liability. Questions regarding this policy may be directed to MIT's Insurance Office, (617) 253-2823. Your insurance carrier may be billed for the cost of emergency transport or medical treatment, if such services are determined not to be directly related to your participation in this study.

- **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143B, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253 6787.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

____
Name of Subject

____
Name of Legal Representative (if applicable)

____ ______________
Signature of Subject or Legal Representative Date

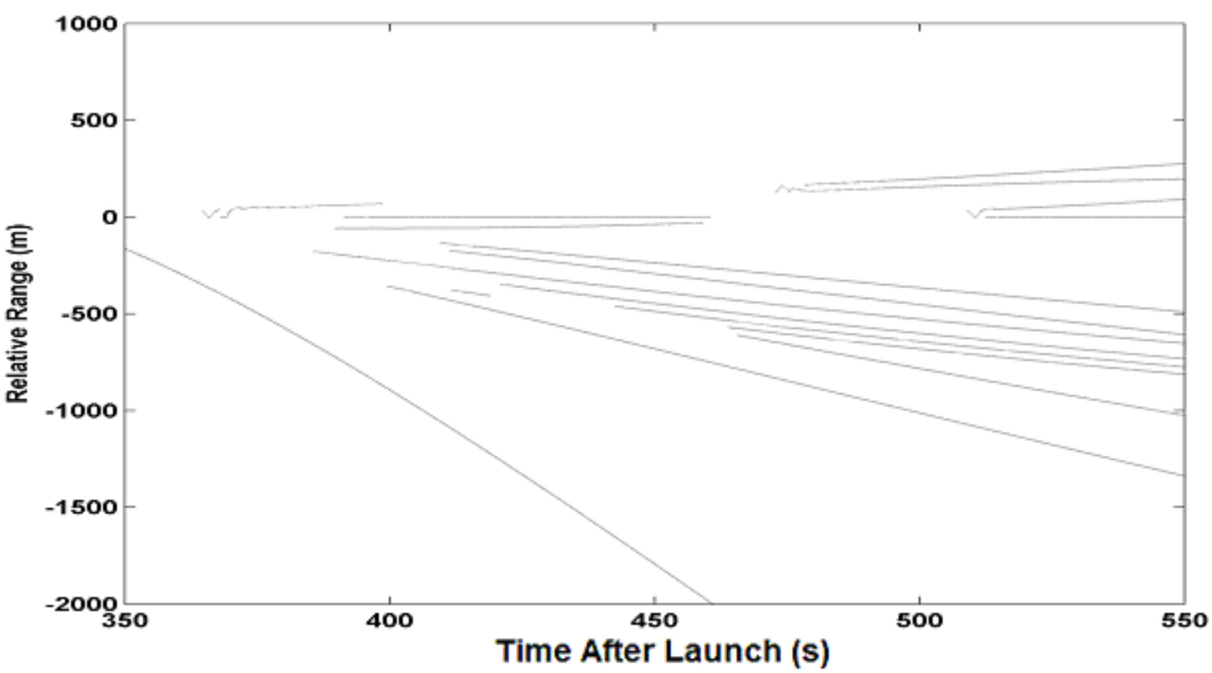
SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

____ ______________
Signature of Investigator Date

Appendix D: Experiment One Radar Data

The following data are broken into individual scenarios. The first figure represents the pre-processed radar data given to both decision sources. The second figure represents truth. The figures are listed in order of scenarios, one through six.



1000
500
0
-500
-1000
-1500
-2000
Relative Range (m)
350
400
450
500
550
Time After Launch (s)

Figure D-1: Scenario 1 (Easy, 30%)

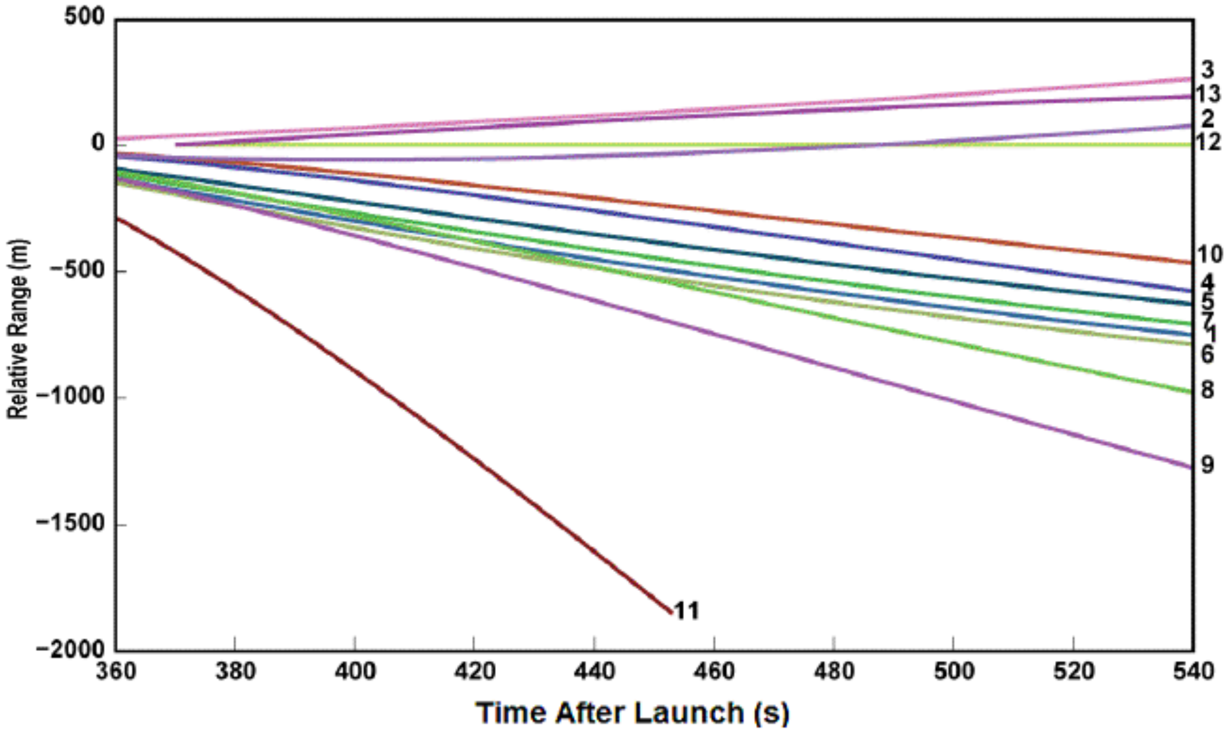


Figure D-2: Truth trajectories for Scenario 1 (Easy, 30%)

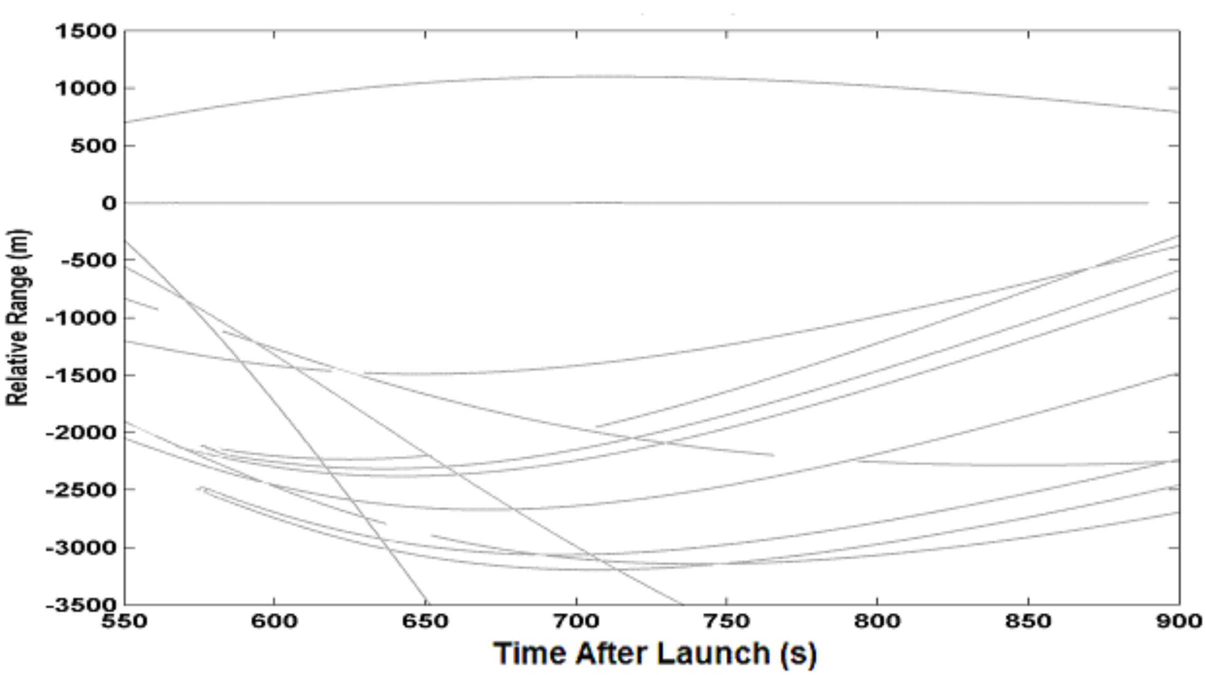


Figure D-3: Scenario 2 (Easy, 60%)

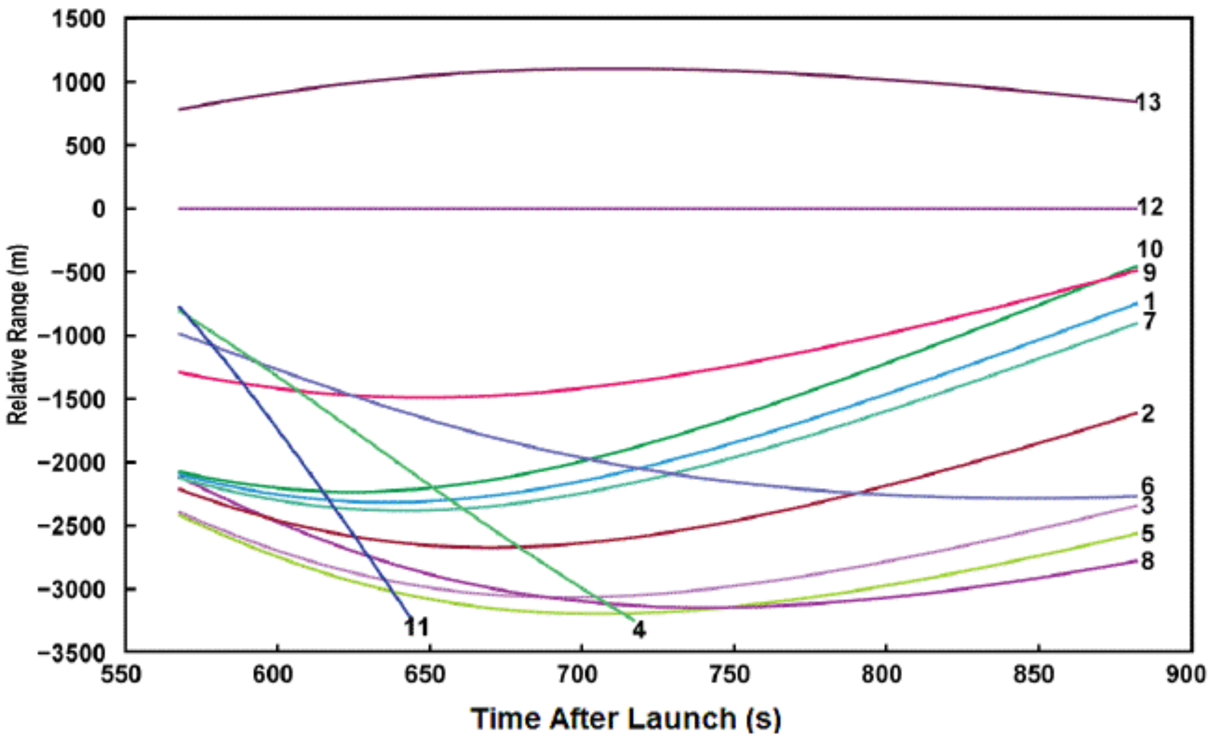


Figure D-4: Truth trajectories for Scenario 2 (Easy, 60%)

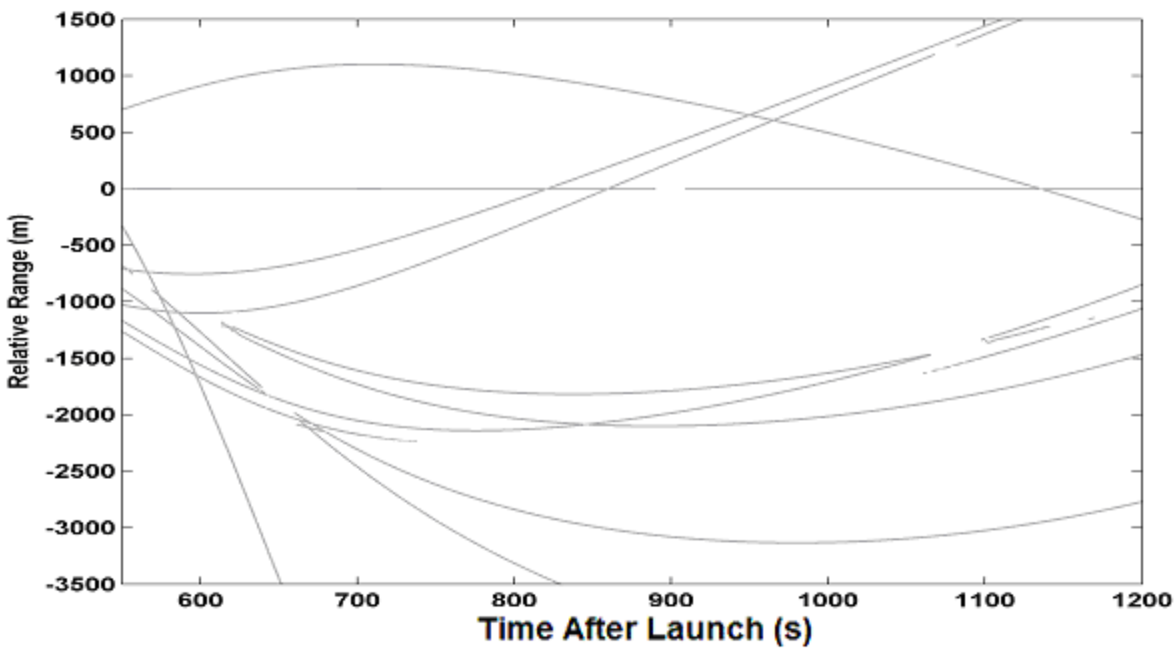


Figure D-5: Scenario 3 (Easy, 100%)

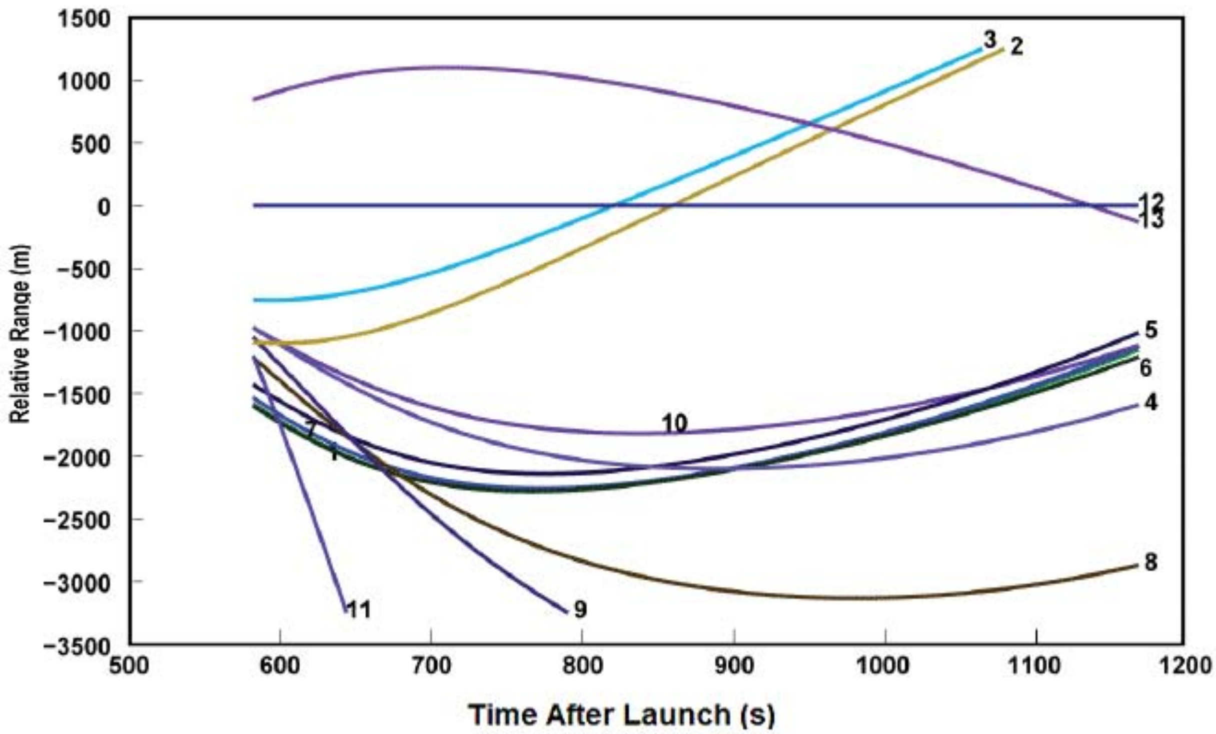


Figure D-6: Truth trajectories for Scenario 3 (Easy, 100%)

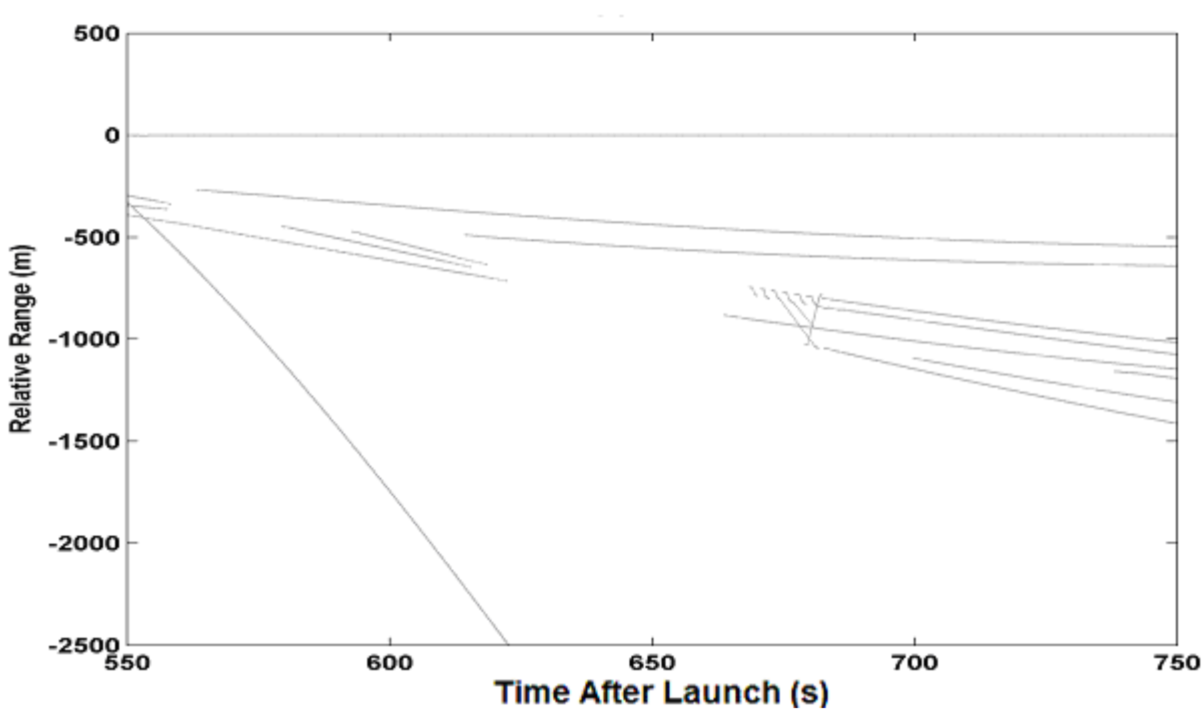


Figure D-7: Scenario 4 (Hard, 30%)

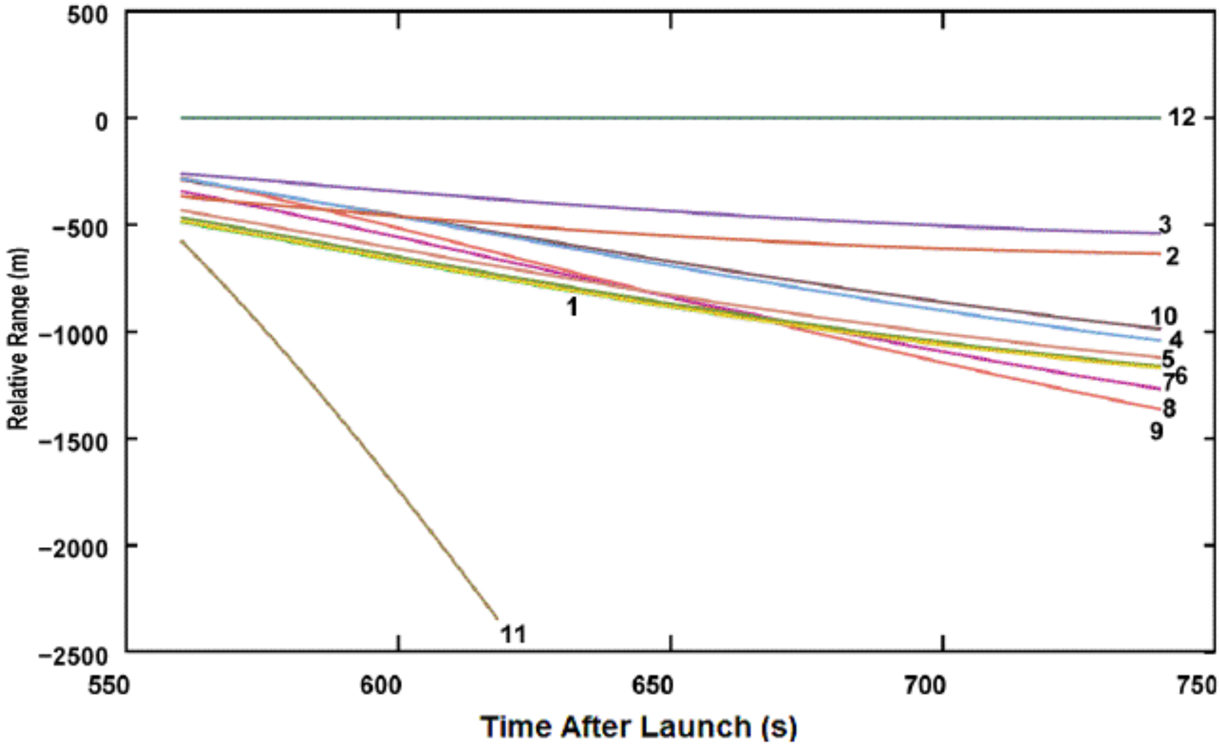


Figure D-8: Truth trajectories for Scenario 4 (Hard, 30%)

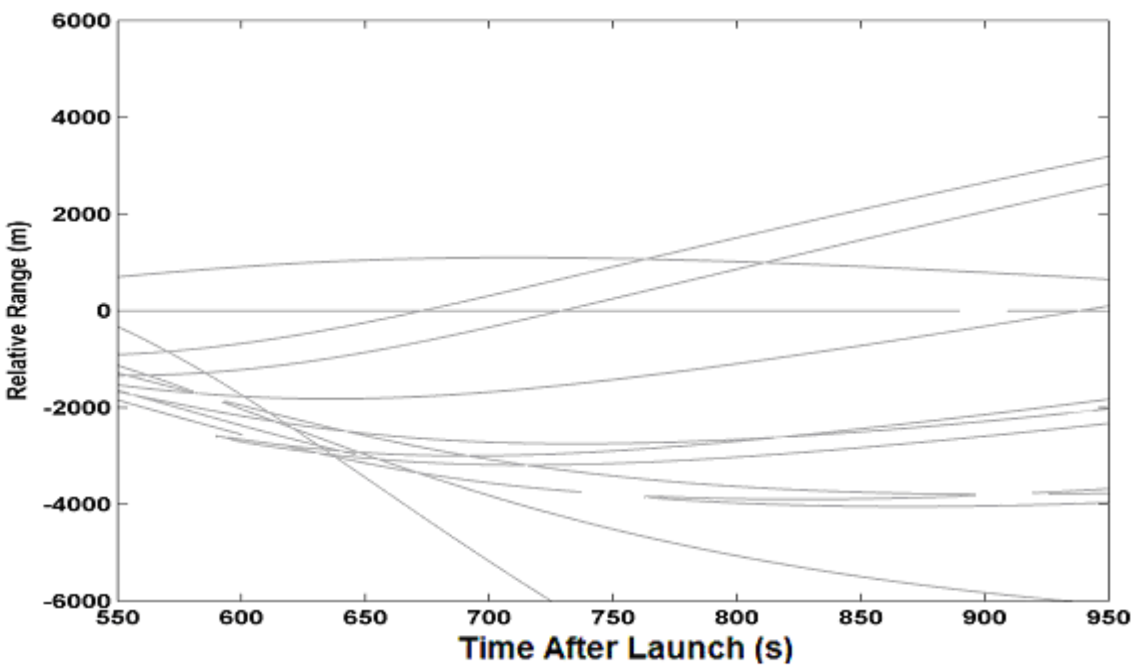


Figure D-9: Scenario 5 (Hard, 60%)

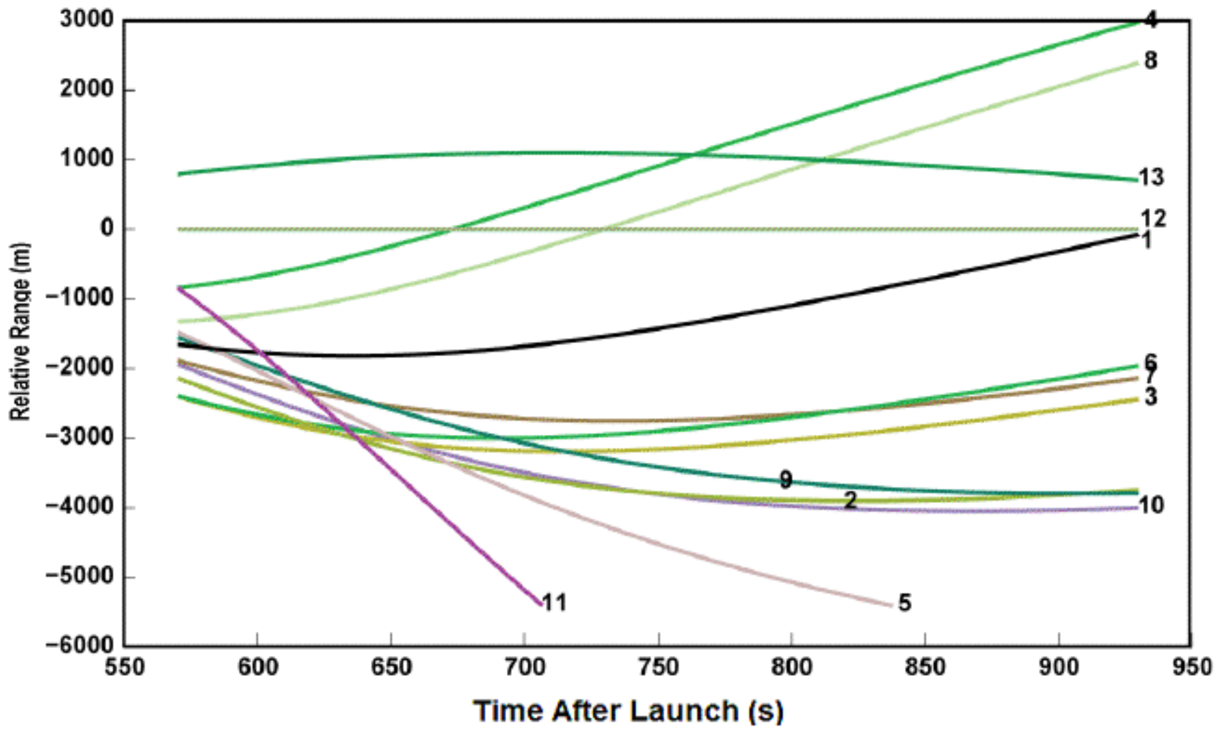


Figure D-10: Truth trajectories for Scenario 5 (Hard, 60%)

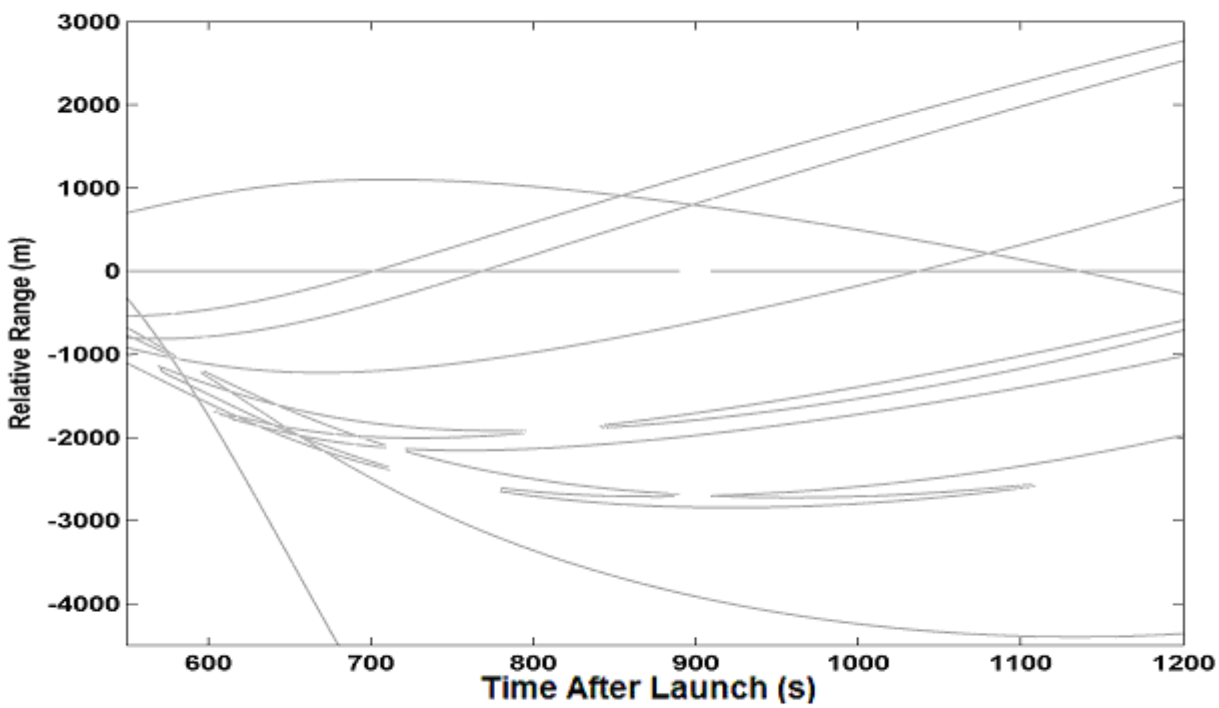


Figure D-11: Scenario 6 (Hard, 100%)

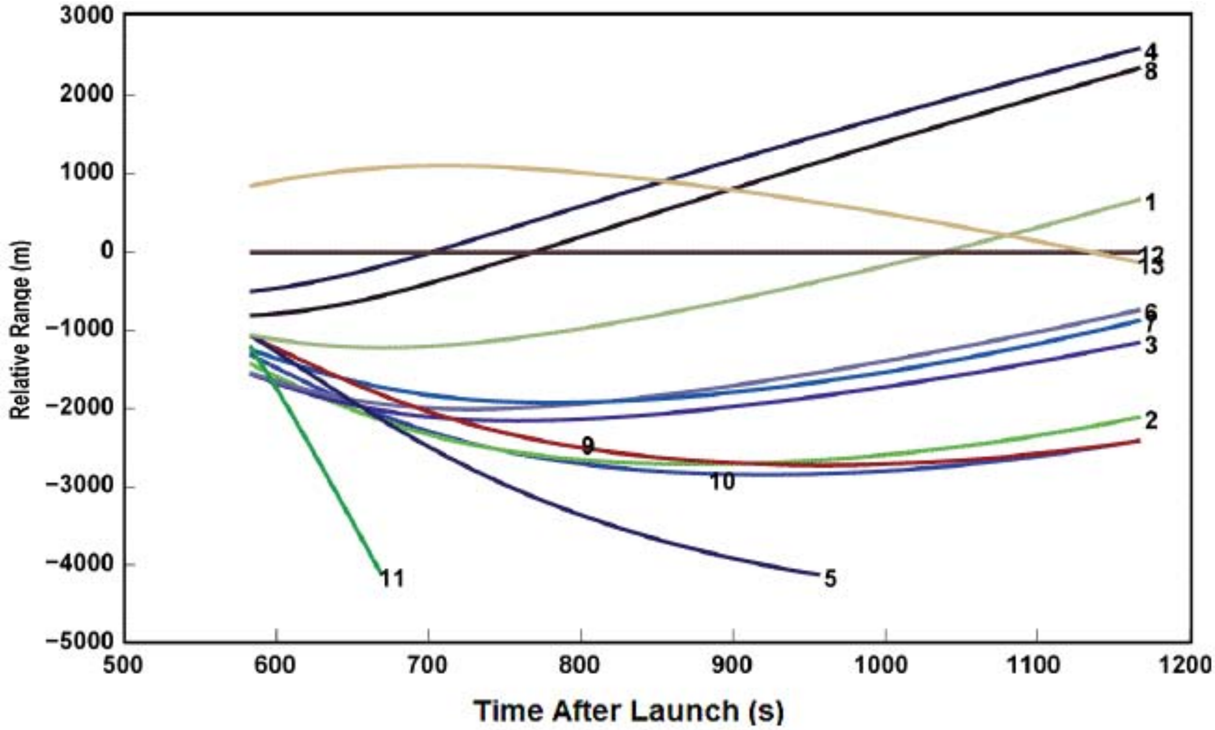


Figure D-12: Truth trajectories for Scenario 6 (Hard, 100%)

Appendix E: Experiment One Statistics

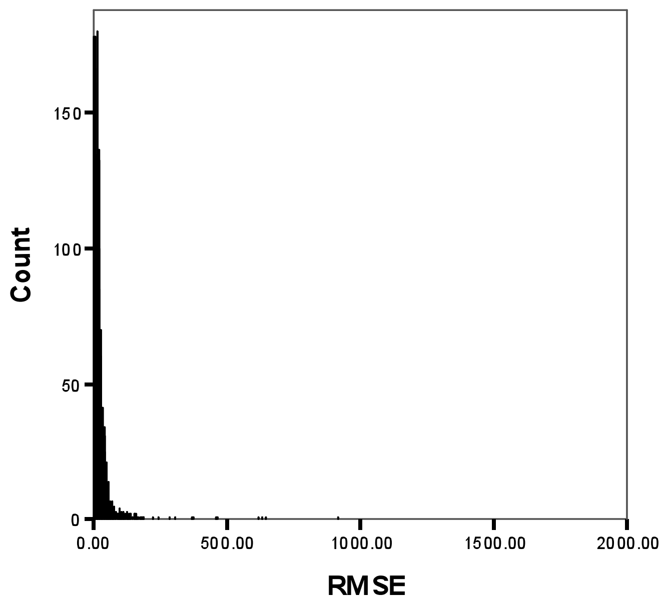


Figure E-1: Histogram of the RMSE for both decision sources. It's heavily weighted towards a "0" score which represents a perfect match

Missed and False Trajectories

Table E-1:
The Number Participants Who Missed Each of the Trajectories

Scenario/Trajectory	1	2	3	4	5	6
1	0	0	1	0	0	1
2	0	0	1	0	0	0
3	26	0	0	0	0	4
4	25	2	1	1	0	23
5	0	3	1	0	0	0
6	1	2	1	0	0	0

Scenario/Trajectory	7	8	9	10	11	12	13
1	2	1	0	0	0	2	1
2	1	1	0	0	1	0	0
3	17	0	0	2	2	1	1
4	13	0	0	0	1	2	N/A
5	1	0	0	2	1	0	0
6	0	0	1	1	0	0	1

Table E-2:
The Percentage of Missed Trajectories Per Scenario, Per Track.

(Scenario, Track)	1,3	1,6	1,7	1,8	1,12	1,13	2,3	2,7	2,8	2,11	3,1	3,6	3,7	3,10	3,11	3,12	3,13
H Missed Trajectories	1	1	2	1	2	1	1	1	1	1	26	4	17	2	2	1	1
Missed Trajectory %	3.4	3.4	6.9	3.4	6.9	3.4	3.4	3.4	3.4	3.4	90	14	59	6.9	6.9	3.4	3.4

(Scenario, Track)	4,1	4,6	4,7	4,11	4,12	5,2	5,3	5,7	5,10	5,11	6,1	6,2	6,3	6,9	6,10	6,13	
H Missed Trajectories	25	23	13	1	2	3	1	1	2	1	1	2	1	1	1	1	
Missed Trajectory %	86	79	45	3.4	6.9	10	3.4	3.4	6.9	3.4	3.4	6.9	3.4	3.4	3.4	3.4	

Please refer to Appendix D for the Actual Trajectories

Table E-3:
The Number of False Trajectories per Participant

Participant	1	2	3	4	5	6	7	8	9	10
# False Trajectories	0	0	1	2	0	0	0	0	0	0
Participant	11	12	13	14	15	16	17	18	19	20
# False Trajectories	1	0	0	0	1	0	0	1	0	0
Participant	21	22	23	24	25	26	27	28	29	
# False Trajectories	1	0	1	0	1	2	0	0	0	

Accuracy

The following family wise error correction was used to correct for Type-I error.

$$\alpha_{fw} = 1 - (1 - \alpha)^C$$

Equation E-1: Family wise error correction. C = # of comparisons

Performance Tables

Table E-4:
Number of Superior and Tied Trajectories by the Human Decision Source

Scenario/Trajectory	1	2	3	4	5	6
1	4	2	1	4	10	3
2	0	1	1	1	0	1
3	2	0	0	0	1	0
4	29	2	1	24	24	29
5	1	16	0	1	0	0
6	1	24	0	1	1	0

Scenario/Trajectory	7	8	9	10	11	12	13
1	20	1	1	5	12	20	7
2	0	0	0	2	2	14	4
3	1	1	0	0	1	15	0
4	1	0	26	0	1	7	0
5	1	0	2	25	5	13	5
6	26	0	25	25	1	14	0

Table E-5:
Performance Table Including Joint-Missed Trajectories as Ties.

Superior Decision Source		*Degree of Difficulty*		*Total*
	Data Span	**Easy**	**Hard**	
Human	**30%**	24	81	105
	60%	2	37	39
	100%	0	100	100
Tie	**30%**	66	63	158
	60%	24	32	56
	100%	21	18	39
Algorithm	**30%**	270	188	458
	60%	261	284	545
	100%	308	213	521

Table E-6:
Performance Table Not Including Joint-Missed Trajectories

Superior Decision Source		*Degree of Difficulty*		*Total*
	Data Span	**Easy**	**Hard**	
Human	**30%**	24	71	95
	60%	2	37	39
	100%	0	100	100
Tie	**30%**	65	15	80

	60%	24	32	56
	100%	21	18	39
Algorithm	**30%**	263	168	431
	60%	257	276	533
	100%	255	206	461

Table E-7: Number of More Accurate Trajectories and Number of Standard Deviations From Mean For All Participants
Overall Mean Number = 6.758, Standard Deviation = 4.556

Participant	1	2	3	4	5	6	7	8	9	
# of More Accurate Trajectories	7	6	7	10	14	5	2	23	5	
Std Dev from Mean	0.053	-0.166	0.053	0.711	1.589	-0.386	-1.044	3.564	-0.386	
Participant	10	11	12	13	14	15	16	17	18	19
# of More Accurate Trajectories	5	3	11	6	7	5	2	6	1	10
Std Dev from Mean	-0.386	-0.825	0.931	-0.166	0.053	-0.386	-1.044	-0.166	-1.264	0.711
Participant	20	21	22	23	24	25	26	27	28	29
# of More Accurate Trajectories	7	6	11	5	0	6	3	9	3	11
Std Dev from Mean	0.053	-0.166	0.931	-0.386	-1.483	-0.166	-0.825	0.492	-0.825	0.931

Confidence

Table E-8 lists the results for the 13 pairwise comparison tests, in which the significant differences are depicted with an asterisk.

Table E-8: Pairwise Comparisons (Mann-Whitney U)

Variables	***Two-Tailed Significance (α = .004)***
Easy, 30%-60%	Z = -4.986, p = .000*
Easy, 60%-100%	Z = -2.007, p = .045
Easy, 30%-100%	Z = -4.286, p = .000*
Hard, 30%-60%	Z = -4.286 p = .000*
Hard, 60%-100%	Z = -.955, p = .339
Hard, 30%-100%	Z = -3.969, p = .000*
30%, Easy-Hard	Z = -1.978, p = .048
60%, Easy-Hard	Z = -1.682, p = .093

100%, Easy-Hard	Z = -1.286, p = .199
30%-60%	Z = -4.625, p = .000*
60%-100%	Z = -1.141, p = .254
30%-100%	Z = -4.398, p = .000*
Easy-Hard	Z = -1.459, p = .145

Appendix F: Experiment Two Demographic Information

Participant	Participant (from Exp 1)	Gender	Age	Career	Served in Military	Drawing Experience	How Often
1	20	Male	>50	Engineer	No	Yes	Weekly
2	22	Male	25-35	Electrical Engineer	No	Yes	Weekly
3	19	Male	25-35	Engineer	No	Yes	Monthly
4	14	Female	35-50	Secretary	No	No	
5	4	Male	>50	Scientist	No	No	
6	21	Female	>50	Editor	No	Yes	Monthly
7	25	Female	35-50	Radar Analyst	No	Yes	Weekly
8	8	Female	25-35	Staff	No	No	
9	17	Male	18-25	Engineer	No	Yes	Yearly
10	13	Female	35-50	Exec. Assistant	No	No	
11	27	Male	35-50	Research Staff	No	No	
12	12	Male	25-35	Engineer	No	Yes	Yearly

Appendix G: Experiment Two Consent to Participate Form

CONSENT TO PARTICIPATE IN
NON-BIOMEDICAL RESEARCH

Human Performance in Ballistic Missile Discrimination Scenarios

You are asked to participate in a research study conducted by Lee Spence, Ph.D. from the MIT Lincoln Laboratory Advanced Concepts and Technology Group and Jason Rathje from the MIT Humans and Automation Laboratory. You were selected as a possible participant in this study because of your interest in improving human performance in ballistic missile defense scenarios. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

- **PARTICIPATION AND WITHDRAWAL**

Your participation in this study is completely voluntary and you are free to choose whether to be in it or not. If you choose to be in this study, you may subsequently withdraw from it at any time without penalty or consequences of any kind. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

- **PURPOSE OF THE STUDY**

Ballistic Missile Decision Support involves a number of very broad and complex issues. The system is very large, it has many interconnected elements, and it is physically spread over an area that is a significant fraction of the Earth. In addition the information on which to make decisions is often incomplete and/or inconclusive, and, given the enormity of the decision making task, the timelines are extremely short. Thus, the allocation of tasks automation and humans requires careful consideration of the areas where each can best perform. The general purpose of this research program is to investigate automation and human operator performance. To this end, this experiment will investigate how well humans and computers can assess missile crossing tracks.

- **PROCEDURES**

If you volunteer to participate in this study, we would ask you to do the following things:

- Participate in a 15 minute training session to familiarize yourself with the display and test conditions.
- Assess automated track identification post track crossings in up to 7 test scenarios, each about 10 minutes in length. Total time will be no longer than 1 hr 10 minutes.

- All of these steps will occur in the Lincoln Laboratory S-Building (South Laboratory), Room S1-425.

- **POTENTIAL RISKS AND DISCOMFORTS**

There are no foreseeable risks, discomforts, inconveniences in participating in this experiment.

- **POTENTIAL BENEFITS**

Your benefit in participation in this study is developing a better understanding of human and computer strengths and weaknesses in the track estimation task. In terms of benefit to society, this research will provide for better understanding of human versus computer capabilities in track estimation, which is applicable not only to ballistic missile defense but also the air traffic control and other domains.

- **PAYMENT FOR PARTICIPATION**

Participation in this experiment is strictly voluntary with no payment.

- **CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. Your performance in this study will only be coded by your subject number, which will not be linked to your name so your participation in this research is essentially anonymous.

- **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please feel free to contact Lee Spence at Group 32 – Advanced Concepts and Technology, MIT Lincoln Laboratory, 244 Wood St, Lexington MA 02240-9185 (781) 981-5043 or Professor Cummings at 77 Massachusetts Ave., 33-305, Cambridge, MA 02139 (617) 252-1512.

- **EMERGENCY CARE AND COMPENSATION FOR INJURY**

If you feel you have suffered an injury, which may include emotional trauma, as a result of participating in this study, please contact the person in charge of the study as soon as possible.

In the event you suffer such an injury, M.I.T. may provide itself, or arrange for the provision of, emergency transport or medical treatment, including emergency treatment and follow-up care, as needed, or reimbursement for such medical services. M.I.T. does not provide any other form of compensation for injury. In any case, neither the offer to provide medical assistance, nor the actual provision of medical services shall be considered an admission of fault or acceptance of liability. Questions regarding this policy may be directed to MIT's Insurance Office, (617) 253-2823. Your insurance carrier may be billed for the cost of emergency transport or medical treatment, if such services are determined not to be directly related to your participation in this study.

- **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143B, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253 6787.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

____

Name of Subject

____

Name of Legal Representative (if applicable)

____ ______________

Signature of Subject or Legal Representative Date

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

____ ______________

Signature of Investigator Date

Appendix H: Experiment Two Radar Data

The following data are broken into individual scenarios. The first figure represents the pre-processed radar data given to both decision sources. The second figure represents truth. The figures are listed in order of scenarios, one through six.

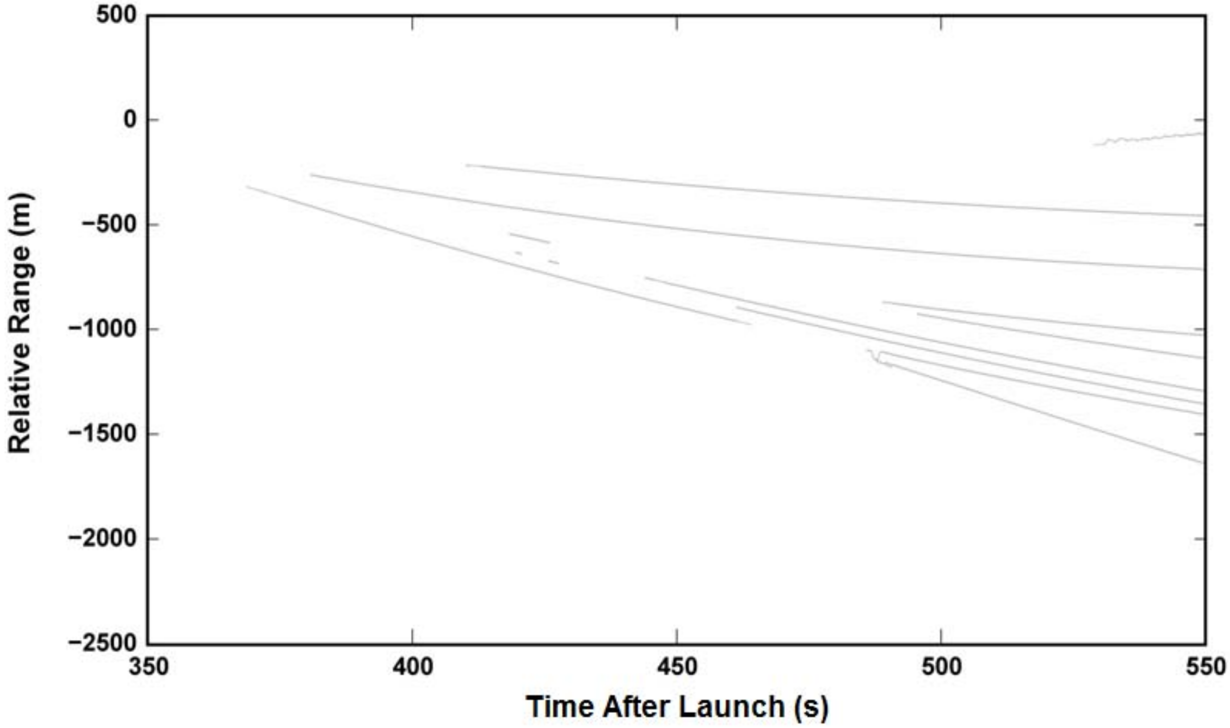


Figure H-1: Scenario 1 (Easy, 30%)

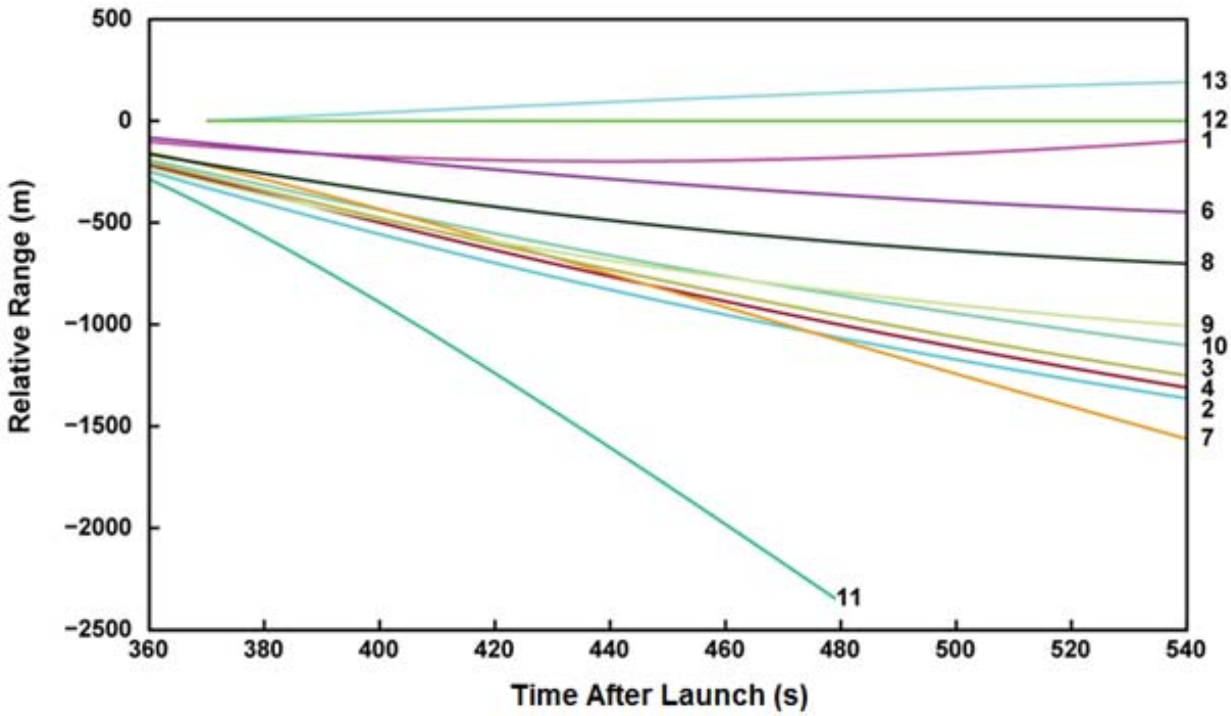


Figure H-2: Truth trajectories for Scenario 1 (Easy, 30%)

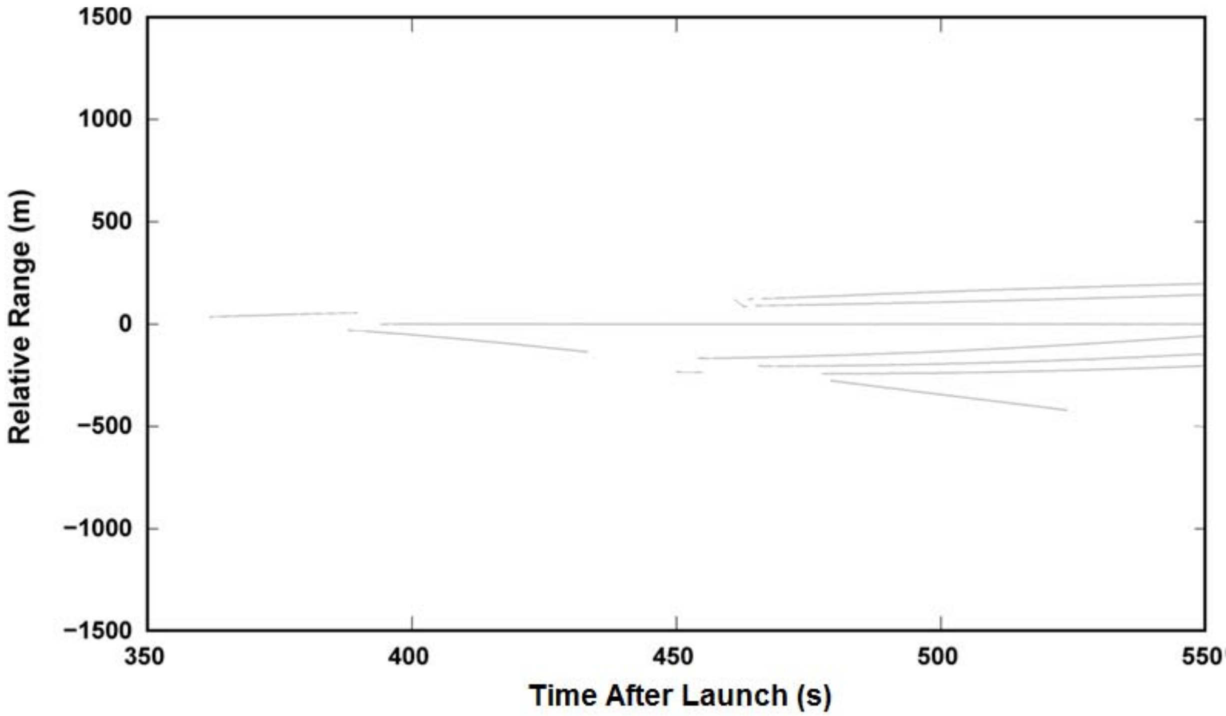


Figure H-3: Scenario 2 (Easy, 30%)

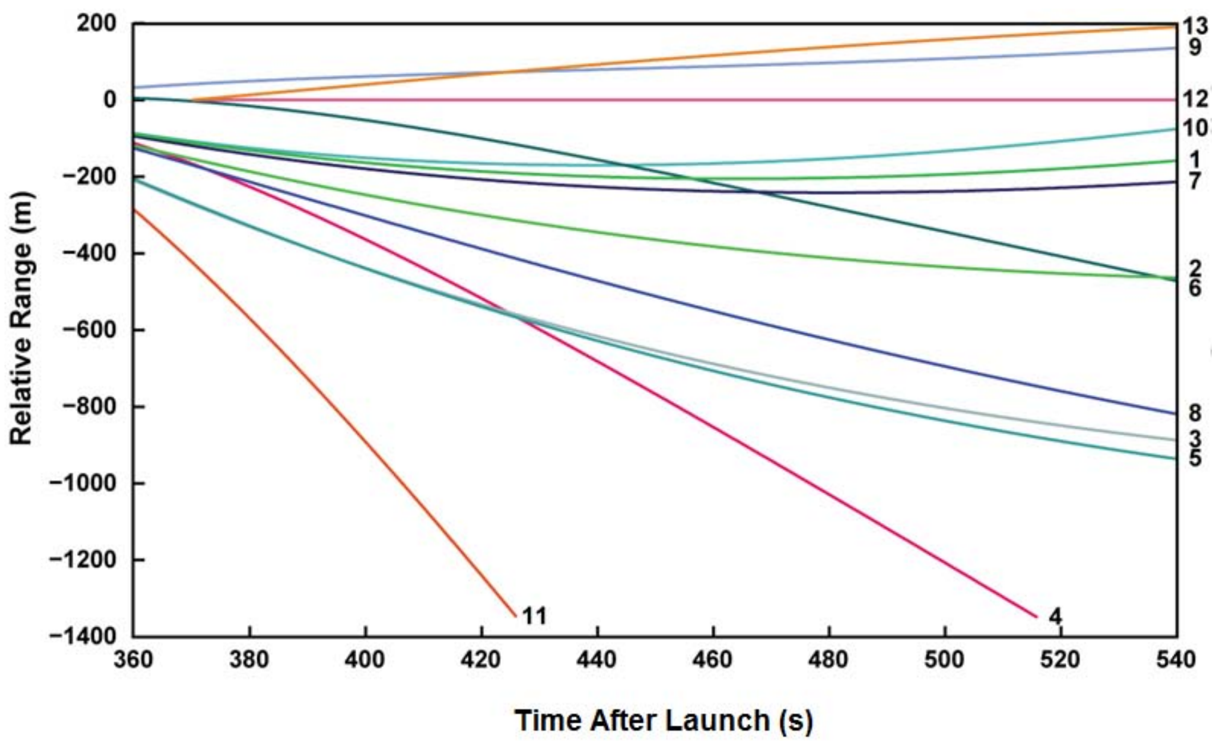


Figure H-4: Truth trajectories for Scenario 2 (Easy, 30%)

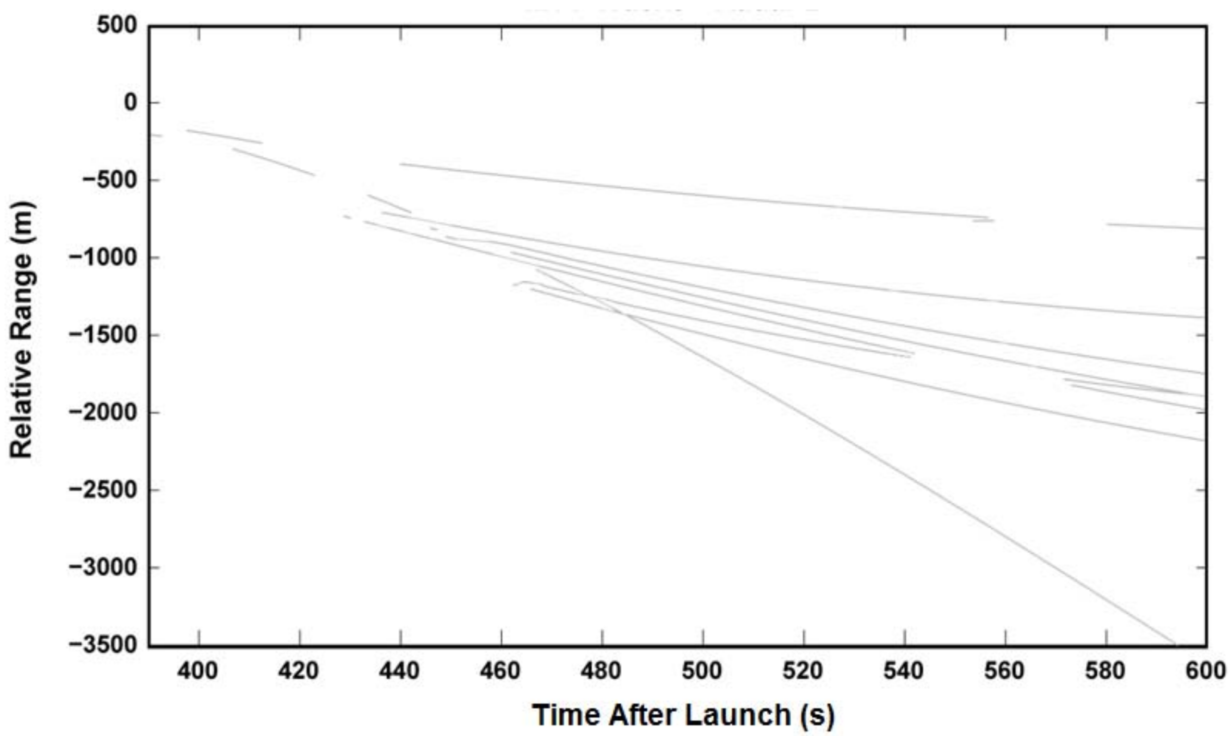


Figure H-5: Scenario 3 (Hard, 30%)

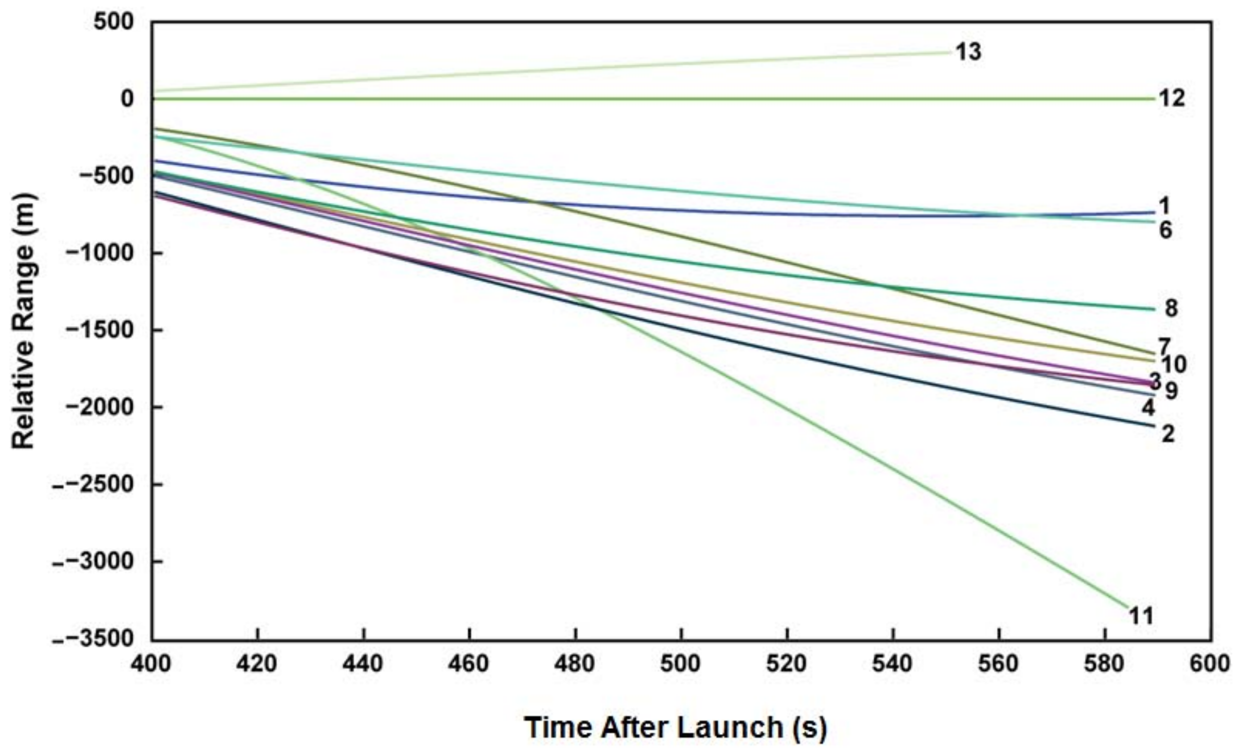


Figure H-6: Truth trajectories for Scenario 3 (Hard, 30%)

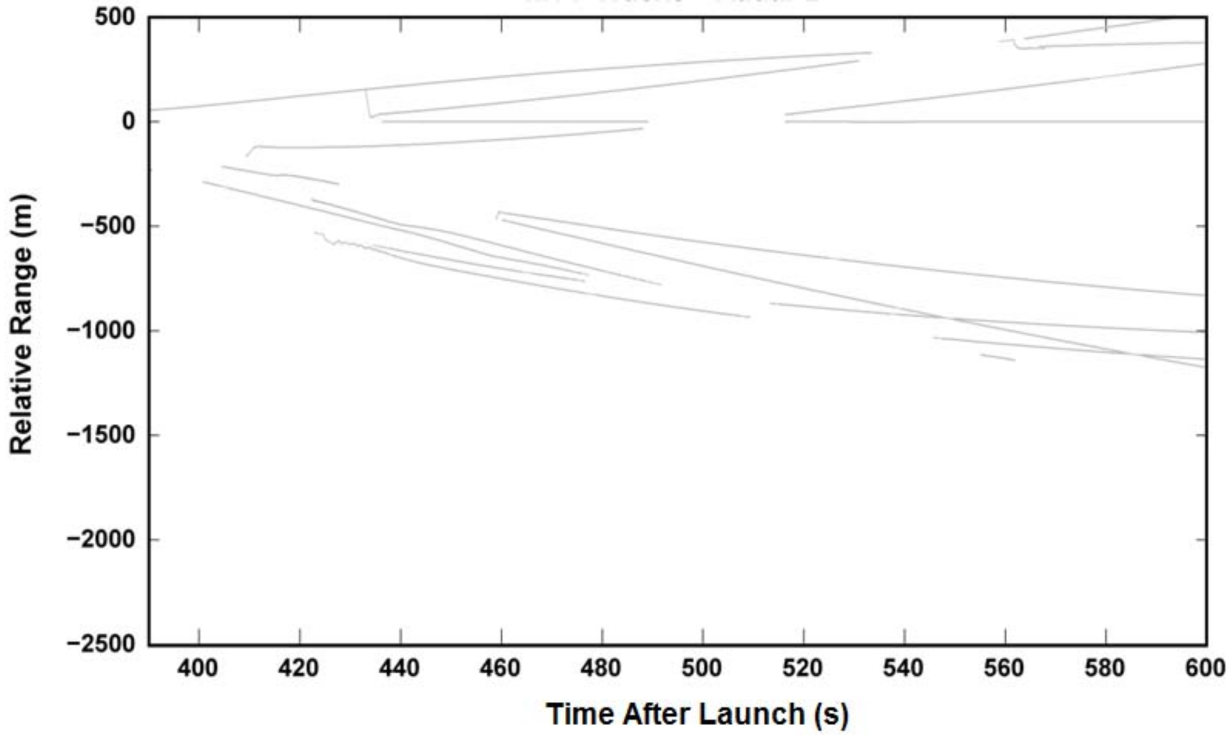


Figure H-7: Scenario 4 (Hard, 30%)

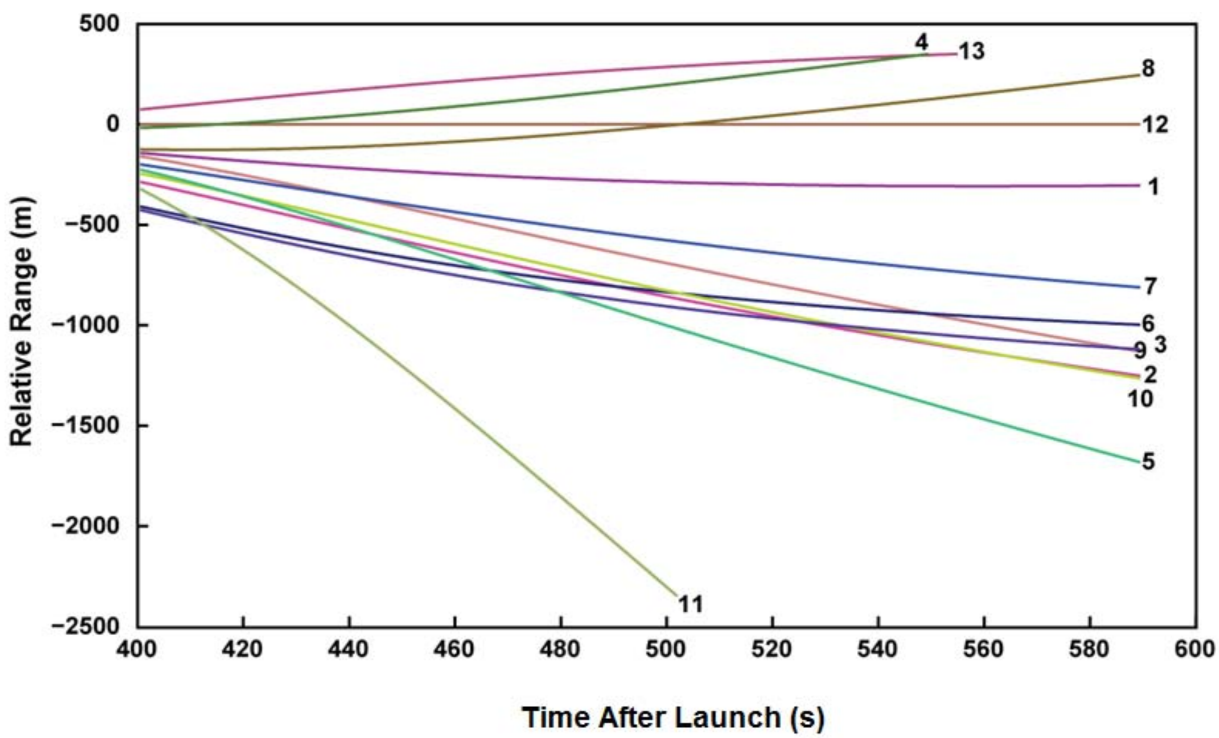


Figure H-8: Truth trajectories for Scenario 4 (Hard, 30%)

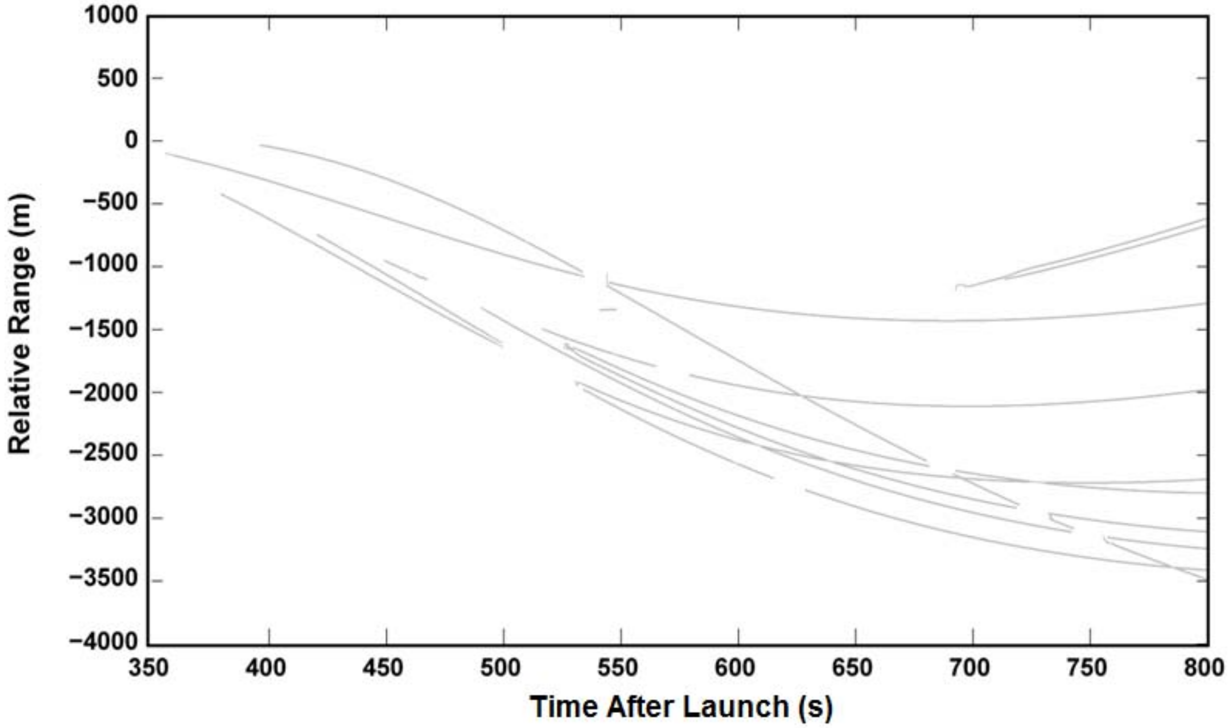


Figure H-9: Scenario 5 (Hard, 60%)

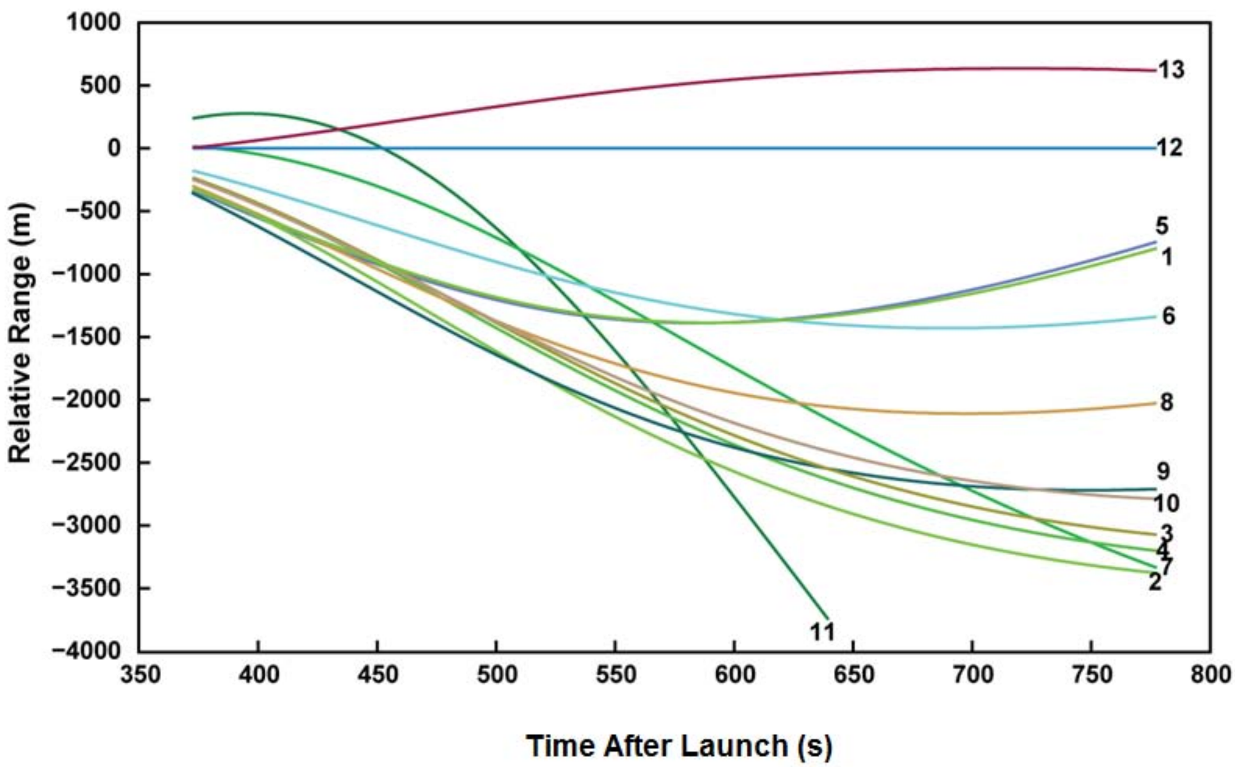


Figure H-10: Truth trajectories for Scenario 5 (Hard, 60%)

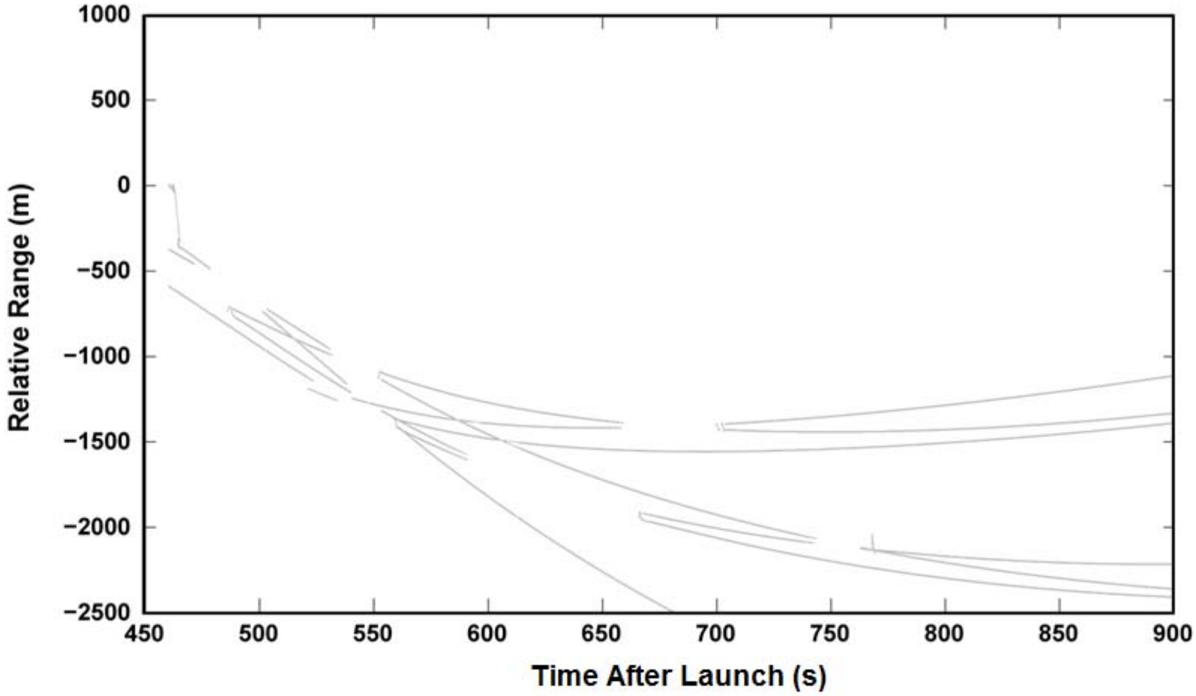


Figure H-11: Scenario 6 (Hard, 60%)

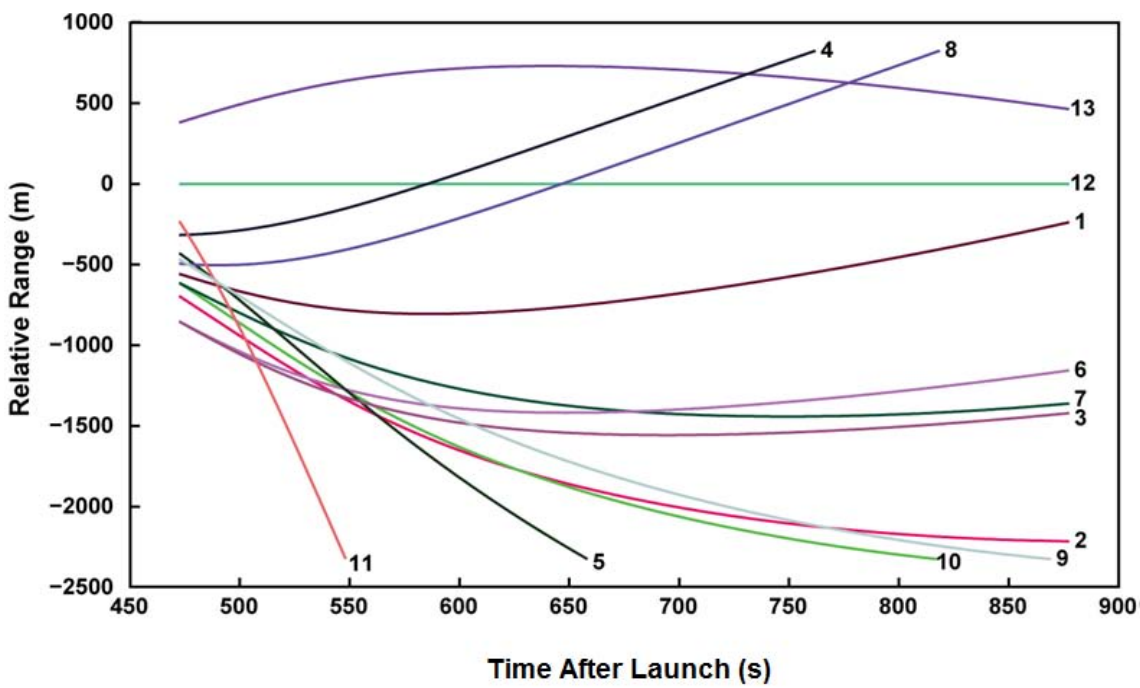


Figure H-12: Truth trajectories for Scenario 6 (Hard, 60%)

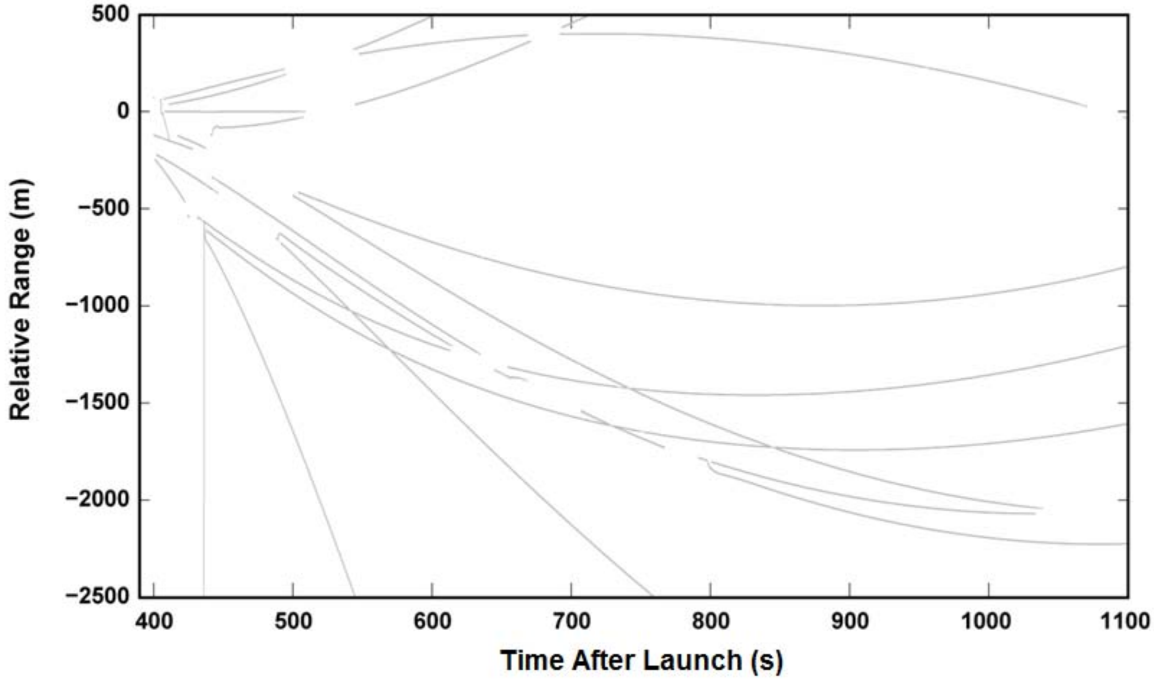


Figure H-13: Scenario 7 (Hard, 100%)

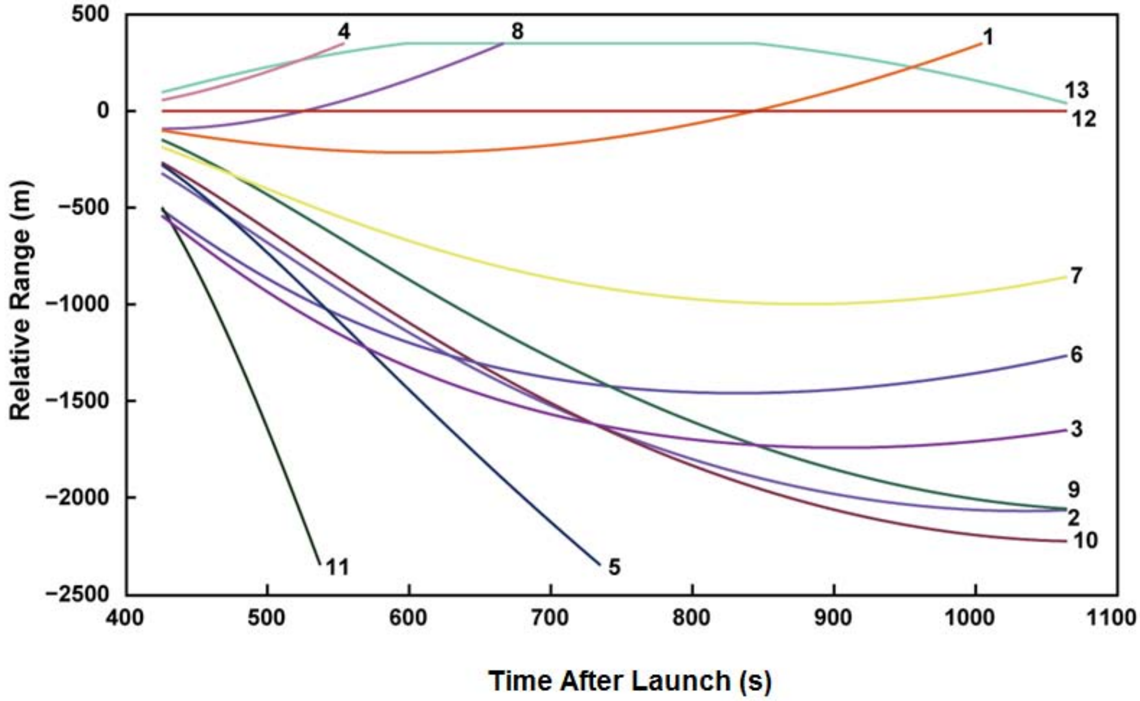


Figure H-14: Truth trajectories for Scenario 7 (Hard, 100%)

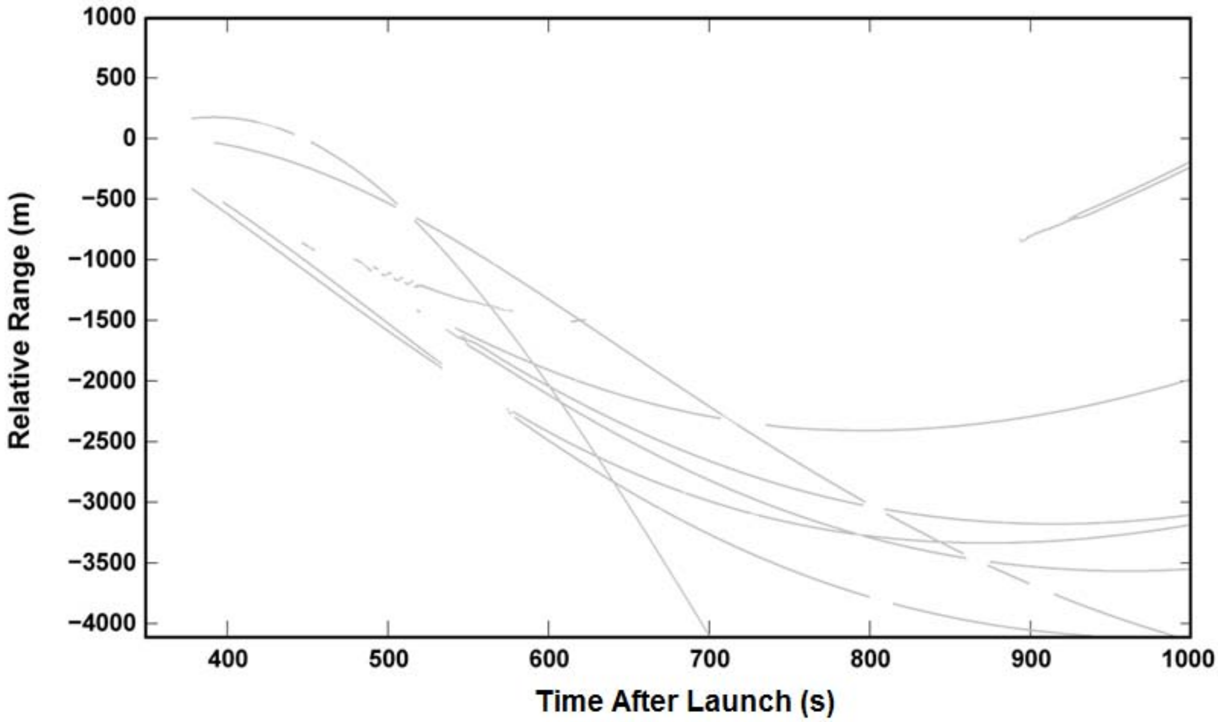


Figure H-15: Scenario 8 (Hard, 100%)

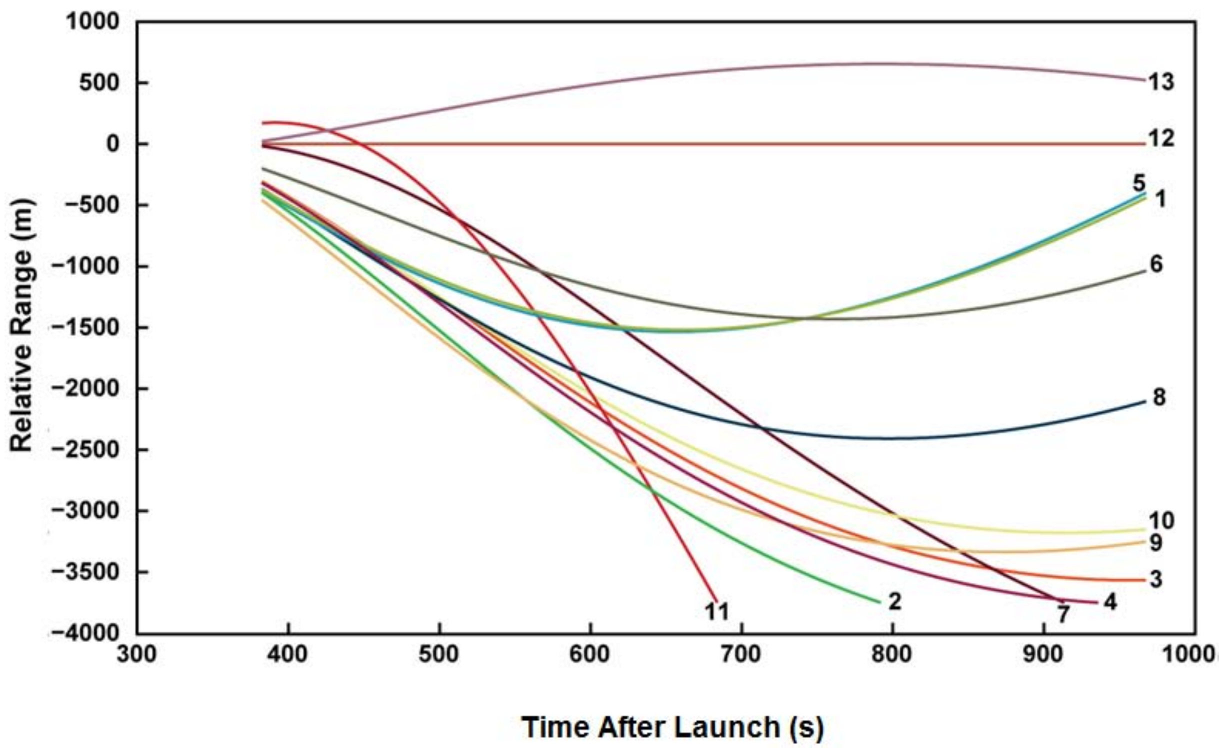


Figure H-16: Truth trajectories for Scenario 8 (Hard, 100%)

Appendix I: Experiment Two Statistics

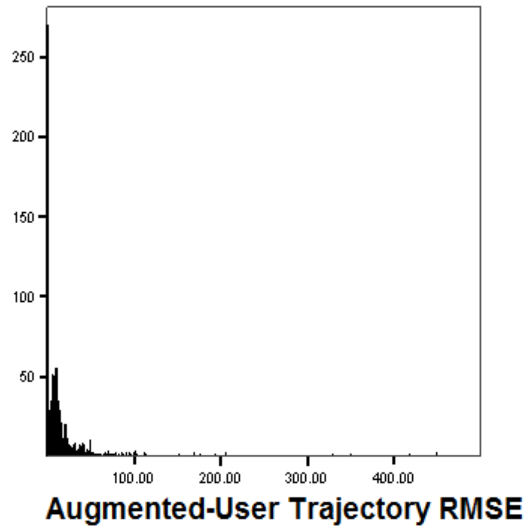


Figure I-1: Histogram of the RMSE for both decision sources. It's heavily weighted towards a "0" score which represents a perfect match

Missed and False Trajectories

Table I-1:
The Number Participants Who Missed Each of the Trajectories

Scenario/Trajectory	1	2	3	4	5	6
1	3	0	0	0	10	0
2	0	0	0	0	0	0
3	10	0	0	0	11	0
4	0	1	0	0	0	0
5	2	0	0	0	1	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	1	1

Scenario/Trajectory	7	8	9	10	11	12	13
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0

Table I-2:
The Percentage of Missed Trajectories Per Scenario, Per Track.

(Scenario, Track)	(1,3)	(1,5)	(3,1)	(3,5)	(4,2)	(5,1)	(5,5)	(7,13)	(8,5)	(8,6)
Augmented Missed Trajectories	3	10	10	11	1	2	1	1	1	1
Missed Trajectory Percentage	0.25	0.83	0.83	0.92	0.08	0.17	0.08	0.08	0.09	0.09

Please refer to Appendix H for the Actual Trajectories

Table I-3: Number of More Accurate Trajectories and Number of Standard Deviations From Mean For All Participants
Overall Mean Number = 13.75, Standard Deviation = 2.987

Participant	1	2	3	4	5	6
# of More Accurate Trajectories	16	16	15	8	12	10
Std Dev from Mean	0.7529	0.7529	0.4183	-1.924	-0.586	-1.255

Participant	7	8	9	10	11	12
# of More Accurate Trajectories	15	16	10	16	17	14
Std Dev from Mean	0.4183	0.7529	-1.255	0.7529	1.0875	0.0837

Table I-4:
The Number of False Trajectories per Participant

Participant	1	2	3	4	5	6	7	8	9	10	11	12
# False Trajectories	2	0	1	3	9	3	2	1	3	2	4	1

Mean Comparisons

Mean comparison t tests for all scenarios for between decision source comparisons, in which the significant differences are depicted with an asterisk.

Table I-5: Pairwise Comparisons (Wilcoxon) Based on RMSE Decision Source Scores

Scenario #, Factor Level	*Two-Tailed Significance (α = .005)*

Combined	t= -1.427, p=0.154
Scenario 1	t = 0.502, p = 0.616
Scenario 2	t = 4.482, p = 0.000*
Scenario 3	t = -0.672, p = 0.502
Scenario 4	t = -1.802, p =0.073
Scenario 5	t = -2.844, p =0.005*
Scenario 6	t = -1.334, p =0.183
Scenario 7	t = -0.712, p =0.477
Scenario 8	t = -0.554, p =0.580

References

[1] H. A. Bethe, J. Boutwell, and R. L. Garwin, "BMD Technologies and Concepts in the 1980s," *Daedalus*, vol. 144, pp. 53-71, 1985.

[2] J. Rasmussen, "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models," in *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, pp. 257 - 266, 1983.
[3] T. B. Sheridan, "Humans and Automation: System Design and Research Issues," John Wiley & Sons, Ed. Santa Monica, CA, 2002.
[4] P. M. Fitts, "Human Engineering for an Effective Air Navigation and Traffic Control system," Washington, DC: National Research Council, 1951.
[5] J. M. Fulvio, M. Singh, and L. T. Maloney, "Precision and consistency of contour interpolation," *Vision Research,* vol. 48, pp. 831-849, 2007.
[6] C. Ware, *Information Visualization*, Second ed. San Francisco: Morgan Kaufmann Publisher, 2004.
[7] H. Lambert, Private Communication, MIT Lincoln Laboratory
[8] W. C. Hoffman and P. C. Dodwell, "Geometric psychology generates the visual gestalt," *Canadian Journal of Psychology,* vol. 39, pp. 491-528, 1985.
[9] D. J. Field, A. Hayes, and R. F. Hess, "Contour integration by the human visual system: evidence for a local "association field"," *Vision Research,* vol. 33, pp. 173-193, 1992.
[10] P. J. Kellman and T. F. Shipley, "A theory of visual interpolation in object perception," *Cognitive Psychology,* vol. 23, pp. 141-221, 1991.
[11] P. Parent and S. W. Zucker, "Trace inference, curvature consistency, and curve detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 11, pp. 823-839, 1989.
[12] M. Singh and J. M. Fulvio, "Visual extrapolation of contour geometry," *Proceedings of the National Academy of Sciences USA,* vol. 102, pp. 783-798, 2005.
[13] P. E. Warren, L. T. Maloney, and M. S. Landy, "Interpolating sampled contours in 3D: Analyses of variability and bias," *Vision Research,* vol. 42, pp. 2431-2446, 2002.
[14] A. Saidpour, M. L. Braunstein, and D. D. Hoffmann, "Interpolation across surface discontinuities in structure from motion," *Perception & Psychophysics*, vol. 55, pp. 611-622, 1994.
[15] T. F. Shipley and P. J. Kellman, "Boundary completion in illusory contours: Interpolation or extrapolation," *Perception*, vol. 32, pp. 985-999, 2003.
[16] D. P. Andrews, A. K. Butcher, and B. R. Buckley, "Acuities for spatial arrangement in line figures: human and ideal observers compared," *Vision Research,* vol. 13, pp. 599-620, 1972.
[17] D. H. Foster, "Visual discrimination, categorical identification, and categorical rating in brief displays of curved lines: implications for discrete encoding processes, " *Journal of Experimental Psychology-Human Perception and Performance,* vol. 9, pp. 785-796, 1983.
[18] D. H. Foster and C. J. Savage, "Uniformity and asymmetry of rapid curved-line selection explained by parallel categorical coding of contour curvature," *Visual Research,* vol. 42, pp. 2163-2175, 17 May 2002.
[19] D. J. Marr and E. Hildreth, "Theory of edge detection" *Proceedings of the Royal Society of London* vol. B207, pp. 187-217, 1980.
[20] D. P. Andrews, "Perception of contour orientation in the central fovea part 1: short lines," *Vision Research,* vol. 7, pp. 975-997, 1967.
[21] J. C. Oglive and M. M. Taylor, "Effect of length on the visibility of a fine line," *Journal of the Optical Society of America,* vol. 45, 1959.
[22] A. Matsubayashi, "Forschung uber die Tiefenwahrnehmung, II," *Acta Societatis Ophthalmologicae Japonicae,* vol. 41, 1937.

[23] J. W. French, "Height and width of vernier lines and separation," *Transactions of the Optical Society*, vol. 21, 1920.
[24] C. Fantoni, J. D. Hilger, W. Gerbino, and P. J. Kellman, "Surface interpolation and 3D relatability," *Journal of Vision*, vol. 8, pp. 1-19, 2008.
[25] C. Layton, P. J. Smith, and C. E. McCoy, "Design of a Cooperative Problem-Solving System for En-Route Flight Planning: An Empirical Evaluation," *Human Factors*, vol. 36, p. 94, 1994.
[26] G. W. Klau, N. Lesh, J. Marks, M. Mitzemacher, and G. T. Schafer, "The HuGS platform: a toolkit for interactive optimization," in *Advanced Visual Interfaces* Trento, Italy, 2002.
[27] J. J. Marquez, "Human-Automation Collaboration: Decision Support for Lunar and Planetary Exploration," in *Aeronautics and Astronautics* Cambridge, Massachusetts: Massachusetts Institute of Technology, 2007.
[28] Fitts, P. M., *Human Engineering for an Effective Air Navigation and Traffic Control system*. National Research Council: Washington, DC, 1951.
[29] J. Nielsen, *Usability engineering*. Cambridge, MA: Academic Press 1993.
[30] M. H. Kutner and M. H. Kutner, *Applied linear statistical models*, 5th ed. Boston: McGraw-Hill Irwin, 2005.