

Summer 2013

Developing a library of display effects on pilot performance: Methods, meta-analyses, and performance estimates

Ellen Salud
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Salud, Ellen, "Developing a library of display effects on pilot performance: Methods, meta-analyses, and performance estimates" (2013). *Master's Theses*. 4363.

DOI: <https://doi.org/10.31979/etd.zdqa-cs43>

https://scholarworks.sjsu.edu/etd_theses/4363

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

DEVELOPING A LIBRARY OF
DISPLAY EFFECTS ON PILOT PERFORMANCE:
METHODS, META-ANALYSES, AND PERFORMANCE ESTIMATES

A Thesis

Presented to

The Faculty of the Department of Industrial & Systems Engineering
San José State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

by

Ellen Salud

August 2013

© 2013

Ellen Salud

ALL RIGHTS RESERVED

The Designated Committee Approves the Thesis Titled

DEVELOPING A LIBRARY OF
DISPLAY EFFECTS ON PILOT PERFORMANCE:
METHODS, META-ANALYSES, AND PERFORMANCE ESTIMATES

by

Ellen Salud

APPROVED FOR

THE DEPARTMENT OF INDUSTRIAL & SYSTEMS ENGINEERING

SAN JOSÉ STATE UNIVERSITY

August 2013

Dr. Kevin Jordan Department of Psychology

Dr. Sean Laraway Department of Psychology

Dr. Christopher Wickens Alion Science and Technology

ABSTRACT

DEVELOPING A LIBRARY OF DISPLAY EFFECTS ON PILOT PERFORMANCE: METHODS, META-ANALYSES, AND PERFORMANCE ESTIMATES

by Ellen Salud

The design of NextGen and current-day cockpit displays are critical for efficient pilot performance and situation awareness on the flight deck. Before deployment of a design into the cockpit the costs and benefits that a display design imposes on performance and situation awareness should be considered. In this thesis, a design tool was developed to support the design of NextGen displays for situation awareness and performance. This design tool is a library of pilot performance estimates. Through literature reviews and meta-analyses of empirical data, the library was developed to provide display designers 1) qualitative distinctions of display properties that either support or limit full situation awareness, and 2) quantitative performance time estimates until situation awareness as a function of various display formats. A systematic method was also developed for future augmentation of the library.

ACKNOWLEDGMENTS

This is dedicated to Kaya, who has been my faithful companion, sounding board, and teacher. Thank you to my family and friends, with special thanks to Edilberto Salud, Sr., and Robert and Koa Richardson, for their steadfast belief and support.

I would like to express my gratitude to my mentors Dr. Becky Hooey, Dr. Christopher Wickens, and Dr. Kevin Jordan-- for their generosity with their time, their perspective and teachings, their patience and positivity, and for seeing this thesis through to its end. Thank you to Audra Ruthruff and Jennifer Granada of San Jose State University Research Foundation, to Kathleen Simel and Alena Filip of the Office of Graduate Studies and Research at San Jose State University, and also to the Human Centered Systems (HCLS) Lab, Man-Machine Integration Design and Analysis (MIDAS) Lab, and the Vision Lab at NASA Ames Research Center who have provided support throughout this academic endeavor.

TABLE OF CONTENTS

LIST OF TABLES	xii
LIST OF FIGURES	xvi
Introduction.....	1
Designing for NextGen Displays	2
The Current Study: Addressing Cockpit Display Design with a Library of Performance Estimates	6
Gaps in the literature	6
Goals and objectives	7
Technical approach: Meta-analyses	7
The Library Framework and Theoretical Rationale.....	10
Information Accessibility library	12
Perception-to-Comprehension (PTC) time library.....	12
SA Limiter library.....	13
Summary	14
Method	16
Step 1: Identify Display Properties that Affect Pilot SA	16
Step 2: Identify Performance Parameter: IA, PTC, or SA Limiting Displays	17
Step 3: Gather Empirical Data	19
Step 4: Compute Performance Estimates.....	20
Step 4a: Computing accessibility time estimates.....	21

Step 4b: Computing PTC time estimates	22
Step 4c: Computing SL performance estimates	22
<i>A SA limiting variable (display property) with two or more levels</i>	<i>22</i>
<i>A SA limiting variable (display property) with quantitative levels</i>	<i>23</i>
Results: Library of Performance Estimates	25
Information Accessibility Library.....	26
Number of keystrokes	26
<i>Summary</i>	<i>26</i>
Perception-to-Comprehension Time Library	27
Search clutter	27
<i>Summary</i>	<i>27</i>
Background overlay clutter.....	28
<i>Summary</i>	<i>30</i>
Overlaid versus separate databases for focused tasks.....	31
<i>Summary</i>	<i>31</i>
Separate versus overlaid displays for information integration tasks	32
<i>Summary</i>	<i>33</i>
Wide display separation for information integration tasks	34
<i>Summary</i>	<i>35</i>
Text string length	36
<i>Summary</i>	<i>36</i>
Target highlighting for target identification tasks	36

<i>Summary</i>	37
SA Limiter Library	39
3D display angle of perspective viewing for spatial judgment tasks	39
<i>Summary</i>	41
Display legibility (height of text and symbols).....	41
<i>Summary</i>	42
2D tracking display size: Visual angle of maximum excursion	43
<i>Summary</i>	43
Auditory displays for spatial localization tasks	44
<i>Summary</i>	45
Auditory display string length	46
<i>Summary</i>	46
Symbol / abbreviation familiarity	47
<i>Summary</i>	48
Predictor format	48
<i>Summary</i>	49
Display conformality	50
<i>Summary</i>	55
Spatial displays for temporal judgments.....	55
<i>Summary</i>	55
Discussion	56
Overview	56

Implications.....	58
A tool for designing for compatibility	59
General SA guidelines	59
A compilation of display properties.....	60
Human performance modeling	60
Assumptions of Correlation Between Performance and SA.....	60
Gaps in the Current Library and Future Work.....	61
Background overlay clutter.....	61
Overlaid versus separate display for focused attention tasks.....	62
Display separation versus WIDE display separation for focused attention tasks.....	62
VAME.....	63
Digital display for tracking tasks	63
Predictor displays.....	63
Spatial displays for temporal estimation tasks.....	64
Location of non-conformal display elements	64
Limitations	65
Next Steps	66
Refining estimates.....	66
Augmenting the library	66
Affects of multiple display properties on SA	67
Projection of SA.....	67

Conclusion	68
References.....	69
Appendix.....	78
Information Accessibility Library: Literature Review, Data, and Meta-analysis.....	78
Number of keystrokes	78
PTC Library: Literature Reviews, Data, and Meta-analyses	79
Search clutter	79
Background clutter complexity.....	81
<i>The cost of mild background overlay clutter</i>	<i>81</i>
<i>The cost of moderate and severe background overlay clutter</i>	<i>82</i>
<i>Performance estimate notes for display design evaluation within a computational model of situation awareness (CSA; Hooey et al., 2010).....</i>	<i>87</i>
Overlaid versus separate databases for focused tasks.....	88
Separate versus overlaid displays for information integration tasks	94
Wide display separation for information integration tasks	97
Text string length	98
Target highlighting.....	100
<i>A benefit for single target identification</i>	<i>100</i>
<i>A benefit of highlighting a target set.....</i>	<i>102</i>
SA Limiter Library: Literatures Reviews, Data, and Meta-analyses	109
3D angle of perspective viewing.....	109
Display illegibility (height of text and symbols)	113

2D tracking display size: Visual angle of maximum excursion	114
Auditory displays for spatial localization tasks	123
Auditory display string length	127
Symbol / abbreviation familiarity	130
<i>The cost of information displayed with symbols or abbreviations</i>	130
<i>The cost of display unfamiliarity (training)</i>	131
<i>The cost of display infrequency</i>	132
Predictor format	134
Display conformality	135
<i>The cost of partially conformal (PC) display formats</i>	136
<i>The cost of non-conformal (NC) display formats</i>	141
Spatial displays for temporal judgments	148

LIST OF TABLES

Table		Page
1.	IF-THEN Decision Logic for Identifying Appropriate Sub-libraries for Display Properties	18
2.	Types of Data used for Developing Display Property Performance Estimates	21
3.	Number of Keystrokes Performance Estimate	26
4.	Search Clutter Performance Estimate.....	28
5.	Background Overlay Clutter (Mild, Moderate, and Severe) Performance Estimate	30
6.	Overlaid Display Performance Estimate for Focused Tasks (Relative to Separate Display Formats)	32
7.	Separate Display Performance Estimate for Information Integration Tasks (Relative to Overlaid Display Formats)	34
8.	Wide Display Performance Estimate for Information Integration Tasks	35
9.	Text String Length Performance Estimate	36
10.	Target Highlighting Performance Estimate.....	38
11.	Angle of Perspective Viewing Performance Estimate	41
12.	Display Legibility Performance Estimate	42
13.	VAME Performance Estimate	44
14.	Auditory Display Performance Estimate for Spatial Localization Tasks.....	45
15.	Auditory Display String Length Performance Estimate.....	46
16.	Symbol/Abbreviation Familiarity Performance Estimate	48

17.	Flight Path Predictor Performance Estimate	49
18.	Display Conformality Performance Estimate.....	55
19.	Spatial Display Performance Estimate for Temporal Estimation Tasks	56
A1.	Summary of Studies with Data for a Keypress Performance Estimate	79
A2.	Summary of Studies with Data for a Background Overlay Clutter Performance Estimate.....	87
A3.	Summary of Studies with Data for an Overlaid Display Performance Estimate (for Focused Attention Tasks).....	93
A4.	Summary of Studies with Data for a Separate Databases Performance Estimate (for Divided Attention Tasks)	96
A5.	Summary of Studies with Data for a Wide Display Separation Performance Estimate (for Divided Attention Tasks)	98
A6.	Summary of Studies with Data for a Text String Length Performance Estimate	99
A7.	Summary of Studies with Data for a Salient Highlighting Performance Estimate for a Single Target.....	102
A8.	Summary of Studies with Data for a Salient Highlighting Performance Estimate for a Target Set	108
A9.	Summary of Studies with Data for a 3D Angle of Perspective Viewing Performance Estimate.....	112
A10.	Summary of Legibility Recommendations for a Legibility Performance Estimate	114

A11. Parameters for Calculating VAME and Performance for a study by Alexander, Wickens, and Hardy (2003).....	116
A12. Parameters for calculating VAME and Performance for a study by Arthur, Prinzel, Kramer, Bailey, and Parrish (2003)	118
A13. Parameters for Calculating VAME and Performance for a study by Stelzer and Wickens (2006; Experiment 1).....	120
A14. VAMES and Performance from a study by Stelzer and Wickens (2006; Experiment 2).....	122
A15. Summary of Studies with Data for an Auditory Display Performance Estimate for Spatial Localization Tasks	127
A16. Summary of Studies with Data for an Auditory Display String Performance Estimate, with Various Auditory-string Lengths	129
A17. Summary of Data Comparing Performance with Abbreviated and Non-abbreviated Text.....	131
A18. Summary of Data Comparing Performance with Unfamiliar and Familiar Displays	132
A19. Summary of Data Comparing Performance with Infrequently and Frequently Encountered Displays.....	133
A20. Summary of Studies with Data for a Predictor Format Performance Estimate....	135
A21. Summary of Studies with Data for a Tracking Display Conformality Performance Estimate.....	147

A22. Summary of Studies with Data for a Spatial Display Performance Estimate
for TTA Estimation Tasks 153

LIST OF FIGURES

Figure	Page
1. An information processing model of situation awareness (Alion, 2011) overlaps Endsley's theoretical model of SA	11
2. Search clutter against a homogeneous (blank) background	27
3. Background overlay clutter	29
4. Overlaid displays	31
5. Separate displays	33
6. Display properties with separate information databases at varying degrees of separation	35
7. Highlighting	37
8a. Increasing angle of perspective view by manipulating the angle of elevation	40
8b. Increasing angle of perspective view by manipulating the angle of azimuth	40
9. Visual Angle of Maximum Excursion (VAME)	43
10. Abbreviated text on a FMS display	47
11. Predictor displays	49
12. Fully conformal tracking display symbols	52
13. Partially conformal ILS tracking display symbol	54
14. Non-conformal display of a 2-dimensional runway centerline superimposed on a 3-dimensional display of the far domain runway	54
15. Summary of display properties in the library	57

16. Non-conformal display elements (altitude information) in fixed text format located at varying distances from the center of the flight path	65
A1. None and relatively mild background clutter; displays used in an experiment by Yeh, Wickens, Merlo, and Brandenburg (2003)	82
A2. Relatively moderate background overlay clutter; displays used in an experiment by Beck, Lohrenz, and Trafton (2008)	84
A3. Relatively severe background overlay clutter; displays used in an experiment by Beck, Lohrenz, and Trafton (2010)	85
A4. Overlaid and separate display formats; displays used in an experiment by Kroft and Wickens (2001)	89
A5. Overlaid and separate display formats; displays used in an experiment by Lohrenz (2003)	90
A6. Overlaid and separate display formats; displays used in an experiment by Wickens et al. (2004)	92
A7. Separate and overlaid display formats; displays used in an experiment by Kroft and Wickens (2001)	95
A8. Separate and overlaid display formats; displays used in an experiment by Schons and Wickens (1993)	95
A9. Highlighting intensity; one of the displays used in an experiment by Beck, Lohrenz, and Trafton (2010)	100
A10. Highlighting intensity; display used in an experiment by Wickens, Ambinder, Alexander, and Martens (2004)	103

A11. Highlighting intensity; displays used in an experiment by Podczerwinski and Wickens (2002).....	105
A12. Background display against highlighted or non-highlighted targets used in an experiment by Nunes, Wickens, and Yin (2006, Experiment 3).....	106
A13. Comparison of displays without (A) and with (B) a perspective view	109
A14. 3D displays with 30- and 60-degree angle of perspective viewing (Panels A and B, respectively); displays used in an experiment by Merwin and Wickens (1996)	110
A15. Similar display with VAME from experiment by Alexander, Wickens, and Hardy (2003).....	115
A16. Display sizes "X" and "A" used in an experiment by Arthur et al. (2003).....	117
A17. 2D large, 2D small, 3D non-compressed horizontal, and 3D compressed longitudinal tracking displays used in an experiment by Stelzer and Wickens (2003, Experiment 1).....	119
A18. Tracking display (attitude direction indicator) within the greater cockpit display used in an experiment by Stelzer and Wickens (2003, Experiment 2)	121
A19. Fully (A) and partially (B) conformal display used in Martin-Emerson and Wickens (1997, Experiment 2).....	137
A20. Fully and partially conformal display used in Wickens and Long (1994).....	139
A21. Fully (A) and partially conformal (B) displays used in Foyle, Hooey, Wilson, and Johnson (2002)	142
A22. Hybrid display formats used in Foyle, Hooey, Wilsons, and Johnson (2002).....	143

Introduction

The Next Generation Air Transportation system (NextGen) is envisioned to implement long-term change in the management and operation of the national airspace system (NAS) in order to accommodate the forecast demands and expected increase in air traffic flow that has significantly grown over the past 30 years (JPDO, 2010). This transformation envisions new procedures such as four-dimensional (4D) trajectory-based operations (which will include timing constraints in NextGen in addition to three-dimensional (3D) spatial constraints of current-day operations), self-separation procedures (which will transfer some separation responsibility to the pilot in NextGen from air traffic control [ATC] in current-day operations), and Net-centric operations (which will use data-linked versus verbal communications as used in current-day).

Efficient and safe execution of NextGen operations to accommodate the increased traffic flow will require the development of supporting flight deck displays and technologies that meet performance time requirements, and that also support situation awareness (SA) in the cockpit. For instance, NextGen self-separation procedures may require that the pilot have SA of task-relevant information elements, such as distance from other aircraft, and perhaps also intended trajectories of other aircraft, so that separation minima can be maintained. Current-day flight deck technologies may not be effectively designed to support pilot SA of task-relevant information for such a procedure. As another example, NextGen 4DT operations may require pilots to have SA of their time trajectories, and perhaps those of other aircraft. However, current day

cockpit displays may not be designed such that maintaining, or even attaining, SA of time trajectories is possible.

Flight deck displays must then be designed in a way to support pilot performance and SA, where SA is defined as the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status into the near future (Endsley, 1995; 2000). Endsley's model of SA includes three levels, where level 1 involves the perception of task-relevant information, level 2 involves the comprehension of task-relevant information, and level 3 is the projection of comprehended information into future states. Any compromise of SA will impact performance in the cockpit, which can have cascading effects throughout the entire system (Endsley, 1994; 1995; 1999). Causal factors underlying aircraft accidents reported from the National Transportation Safety Board (NTSB) were analyzed by Endsley (1994; 1995; 1999) throughout a four year investigation, where, out of 71% of the accidents classified as resultant from human error, 88% involved problems with SA. The majority of accidents (72%) were attributed to a failure in correctly perceiving task-relevant information (level 1 SA), 22% involved a failure to correctly comprehend task-relevant information (level 2 SA), and 6% involved errors involving a failure to properly project near future states (level 3 SA) based on the aircrew's understanding of the situation.

Designing for NextGen Displays

A challenge imposed on designers, then, is to design flight deck displays so that they facilitate SA, which translates into displays that yield the least time delays until full

comprehension of display information. To meet this challenge, designers must determine how certain types of information should be displayed to the pilot in the cockpit, that is, which display properties are most compatible (or are a “good fit”) with the pilot’s perceptual and cognitive abilities, and also facilitate SA. The term “display” is used in this paper to describe display properties in the sense of “information characteristics” (e.g., visual versus auditory; highlighted versus non-highlighted; text versus pictorial displays) of any specific cockpit instrument (e.g., navigation display [ND]; primary flight display [PFD]) that the pilot must interact with on a perceptual and/or cognitive level.

The designer can address this challenge with a user-centered design (UCD) approach, where on the one hand the pilot’s perceptual and cognitive abilities and limitations are first taken into consideration BEFORE the display is developed. If the pilot’s perceptual and cognitive limitations are not first taken into consideration the full potential of the pilot-display interaction may not be realized, especially when re-design and re-evaluation efforts may be limited by time and cost constraints. In other words, the display design may be evaluated to be “good enough” for a given user and task, but perhaps not optimally compatible. On the other hand, if the design is tailored around what is known about the pilot’s perceptual and cognitive abilities and limitations, cockpit display designs can be adapted to exploit the compatibilities between the human operator and the information display rather than forcing the operator to adapt to a system that may be less compatible with his or her abilities. Such an approach would minimize the need for excessive and costly re-design and re-evaluation efforts after a display has already been developed.

Complimentary to a UCD approach, an ecological interface design (EID) approach otherwise refers to *compatibility* as the fit between constraints in the work domain and how they are represented on the display interface (Ellerbroek, Visser, van Dam, Mulder, & van Paassen, 2009; Vicente & Rasmussen, 1992), where the analysis of a design is approached with a method that uses abstraction hierarchies (AH) and the Skills, Rule, and Knowledge framework (SRK) (Rasmussen & Vicente, 1989). Abstraction hierarchies and the SRK framework are used to assess if an interface displays the constraints and relationships within the work environment in a way that the user's cognitive resources can be freed up for active problem solving (e.g., decision making), especially for managing unanticipated events. This type of compatibility analysis has been conducted for a number of aviation displays that include a 4D self-separation assistance display (Ellerbroek, Visser, van Dam, Mulder, & van Paassen, 2009), a terrain awareness display (Borst, Suijkerbuijk, Mulder, & van Paassen, 2006), and an energy-based flight path perspective display (Amelink, van Paassen, & Flach, 2005).

From a UCD approach, as that taken in this thesis, *compatibility* refers to the “fit” between a display property (that is, how the information is displayed, for instance, an auditory versus visual display) and the pilot's perceptual and cognitive limitations and abilities (for instance, thresholds for visual and auditory detection, or perhaps the limits of working memory). In contrast to the EID approach, analysis begins with the user's cognitive and perceptual abilities and limitations. At best, the selected display properties and configurations for a design are a means to exploit this compatibility.

There are a few tools available to assist in designing for compatibility, for instance, compatibility studies and design principles. We have learned from compatibility studies that how an operator processes displayed information, and the time and accuracy of an appropriate response, are influenced by their compatibility. Some compatibility studies have attempted to tease out optimal pairings, for instance, between a stimulus-type and response type (Brainard, Irby, Fitts, & Alluisi in Teichner & Krebs 1974; Fitts & Seeger, 1953), or, among stimulus-type, central processing modality, and response-type (Wickens, Sandry, & Vidulich, 1983), by examining performance effects as a function of various stimulus-response (S-R) pairings. Fitts and Seeger (1953) examined S-R compatibility by comparing performance with matching versus random spatial arrangements between visual stimuli and manual response inputs. They found that the spatial arrangement influenced both speed and accuracy of response times in favor of the matching spatial arrangements (Wickens, Sandry, & Vidulich, 1983). The idea of S-R compatibility was extended by Wickens et al. (1983) to include “central processing modalities” (stimulus-central processing-response, or SCR compatibility). In their study that examined spatial and verbal information encoding (1983, Experiment 2), they observed the results of various compatibility mappings. This included the effects of an auditory stimulus that required a speech response (that is, a “verbal task” that required the verbal information central processor). They also examined the effects of a visual stimulus that required a manual response (that is, a “spatial task” that required the spatial information processor). Response times were faster with these particular pairings when compared to alternative SCR pairings.

Various display design principles supported by empirical evidence may also summarize compatibility effects that are applicable for design and evaluation of cockpit displays. For instance, Wickens and Carswell (1995) formalized the Proximity Compatibility Principle (PCP) that relates to a number of psychological information–processing mechanisms and is based on a set of theoretical principles of human information processing (Wickens & Carswell, 1995). The PCP proposes that displays relevant to a common task or mental operation (close task or mental proximity) should be rendered close together in perceptual space (close display proximity), and has been applied to a number of design endeavors across various domains (Caroux, Bigot, & Vibert, 2011 study in the video gaming domain; Hopcroft, Burchat, & Vince, 2006 papers in the maritime domain; Lavie, Meyer, Bengler, & Coughlin, 2005 paper on in-vehicle displays; Marino & Mahan, 2005 paper in the medical domain).

The Current Study: Addressing Cockpit Display Design with a Library of Performance Estimates

Gaps in the literature. Although there are a variety of resources in the literature to assist in designing for compatibility and SA, there is a lack of quantitative estimates with which to make design decisions. Design guidelines and principles often state *how* a system or display should be designed for compatibility or to support SA, but the guidelines/principles are not always quantified. For example, Endsley, Bolstad, Jones, and Riley (2003) proposed SA-Oriented design guidelines for displaying *information requirements* for attaining SA, rather than guidelines for *how* the information requirements should be displayed. For example, among the SA-Oriented guidelines is

that, “support for parallel processing, such as multi-modal displays should be provided in data rich environments.” There are a number of display properties, however, that can be implemented by the designer into a multi-modal display that impede SA to different degrees. For instance, one display property may yield a longer time until full SA is attained or may even inhibit the attainment of full SA. In addition, although there are individual studies in the literature that compare various display properties for SA there is no consensus or generalized display design principles for designing for SA. Practical effects, or quantitative estimates of performance, and general guidelines for designing for SA would be useful for display design, for instance, when specific performance time or SA requirements must be met.

Goals and objectives. The objective of this thesis is to develop a tool that 1) assists in designing displays for compatibility with pilot’s perceptual and/or cognitive abilities and limitations; 2) provides a qualitative distinction of display properties that either support or limit full SA; and 3) provides estimated absolute and/or relative time costs until SA can be attained for a given display property.

Technical approach: Meta-analyses. Meta-analyses are central to this thesis for developing a tool that can be used by the designer towards a user-centered design approach. What is needed towards the objectives stated above are meta-analyses of existing empirical data to 1) support the inclusion of candidate display properties within the library, and 2) to calculate quantitative performance estimates as a function of various display properties relevant to the cockpit environment.

The aim of meta-analyses is to synthesize results across comparable studies for a given research focus, and to quantitatively combine the results to identify patterns,

relationships, and disagreements that may come to light in the context of multiple data points across studies. Traditional meta-analysis techniques estimate true effect size by synthesizing standardized effect sizes across studies (Glass, 1976), where various effect size indices can be applied. Quasi-meta-analysis techniques may also be used, where qualitative research is acceptable for inclusion and/or statistical techniques are not applied (Cooper, 1982). In this thesis, raw performance benefits/costs were calculated through meta-analyses of mean performance data rather than measures of effect sizes, as done in the conventional meta-analysis. Where possible, performance benefit/cost ratios were also calculated. What is of interest here, are the practical effects (raw and relative time and accuracy costs) of various display properties on performance.

Meta-analytic approaches similar to that taken in this thesis that did not use traditional statistical analyses are reported in the literature. For instance, Wickens, Hutchins, Carolan, & Cumming (2013) used two meta-analytic methods to examine the effects of various training strategies—increasing difficulty versus part-task training. Similar to the method used in this thesis, Wickens, Hutchins, Carolan, & Cumming used a non-statistical meta-analysis to calculate percentage costs and benefits (termed “transfer ratios”) of treatment conditions compared to control conditions. The authors compared the results of the non-statistical analysis with a Hedge’s *G* statistical analysis on aggregated transfer ratios, which reflected estimates of effect size in addition to the difference in means between treatment and control groups. A comparison of results from both methods showed a general agreement of results. Another meta-analysis that synthesized performance means from the literature was used to inform a human

performance model (HPM) with appropriate data for estimating response time and accuracy to off-nominal events in the Next Generation Airspace System (Hooey et al., 2010). Although statistical analyses were not performed on the data, data that were included in the analyses were reported as statistically significant.

Wickens (2005) also applied a non-statistical meta-analytic approach, although a different method of analysis than that taken in this thesis, to weigh performance costs and benefits of various display properties with the Display Formatting Situation Awareness Model (DFSAM; Wickens, 2005). Within DFSAM, practical and statistical significance for various display property effects found in the literature were first coded and aggregated into an overall figure of merit (FOM) that was then used to evaluate the overall cost and/or benefit of a display on performance. Many of the display formats identified for DFSAM along with references to data are relevant for this thesis in that they have been observed to impede SA, as summarized by Wickens (2005). The current work expands on DFSAM. In contrast to this thesis, however, quantitative estimates of performance as a function of these various display characteristics were not provided in DFSAM. DFSAM provided a method for calculating an overall figure of merit from which displays containing multiple information properties could then be ranked for their support for SA, rather than quantitative performance estimates for each individual display property. A unique contribution of this thesis will be the latter.

In summary, the meta-analyses will yield a compilation, or library, of pilot performance estimates that can 1) assist in designing displays for compatibility with pilot's perceptual and/or cognitive abilities and limitations; 2) provide a qualitative

distinction of display properties that either support or limit full SA; and 3) provide estimated absolute and relative time costs until SA can be attained for a given display property. The resulting library could be used to support the design of NextGen displays.

The Library Framework and Theoretical Rationale

The library characterizes three critical components that impede the attainment of SA (Alion, 2011):

1. Information accessibility (IA) time– a time delay required for the pilot to access display information before it can even be initially perceived.
2. Perception-to-comprehension (PTC) time– a time delay required for the pilot to fully comprehend display information after it has been accessed and initially perceived.
3. SA limiters (SLs)– factors that limit full comprehension of display information so that full SA is not attainable.

Within the library, display properties can be characterized by any of these three factors that impede SA.

On one hand, according to an information processing model of SA (Alion, 2011) these three critical components affect the speed of updating SA and/or the accuracy of subsequent SA about a particular variable (e.g., traffic awareness or flight path deviation awareness) (Figure 1). Endsley's model of SA, on the other hand, proposes levels of SA (level 1– perception; level 2– comprehension; level 3– projection). Information accessibility and perception-to-comprehension times can be mapped onto the first two levels of Endsley's model so that information accessibility time occurs before Endsley's

level 1 SA, and PTC time occurs between Endsley’s level 1 (perception) and level 2 SA (full comprehension). If Endsley’s model is viewed on a continuum, SA limiters can be thought of as display properties that inhibit full comprehension, or Endsley’s level 2 SA. That is, the pilot may get a sense of what display information means but he or she cannot fully comprehend it due to a cognitive or perceptual limitation. In this paper, attainment of full SA refers to attainment of full comprehension of display information.

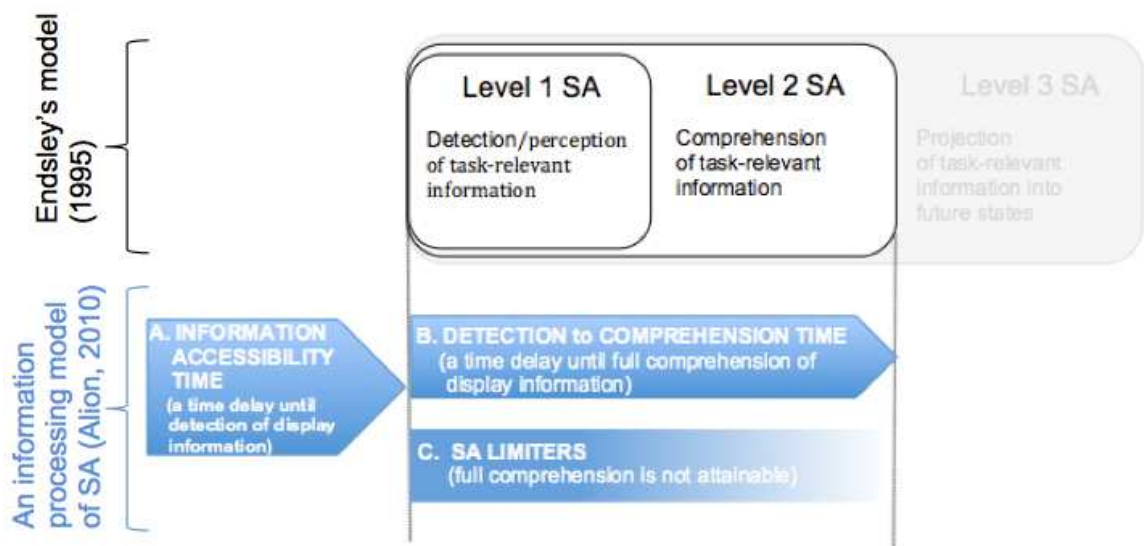


Figure 1. An information processing model of situation awareness (Alion, 2011) overlaps Endsley’s theoretical model of SA. (A) The time it takes to access information impedes detection (that is, initial perception). (B) The time it takes for information to go from perception to full comprehension impedes comprehension. (C) Limited comprehension of display information as a function of the display property impedes information comprehension. The information processing model does not yet address level 3 SA which is greyed-out in the figure.

The taxonomy of factors that impede the attainment of SA make up three sub-libraries within the greater library of performance estimates. They are referred to here as the Information Accessibility library, the Perception-to-Comprehension (PTC) time library, and the SA Limiter library.

Information Accessibility library. The Accessibility library includes performance estimates for display properties that impose a time cost before display information can be perceived (or detected; that is, before Endsley's level 1 SA is attained). Information accessibility time applies when display information necessary to attain SA is not visible and must be accessed by one or more requests (i.e., key strokes or visual scans). Hence the concept does not apply to auditory displays. For instance, if a keypress manipulation of a display is required by the pilot in order for the cockpit display of traffic information (CDTI) to become visually accessible, the manipulation will impose a time cost until the CDTI can even be initially perceived. Subsequently, this will also delay the speed until full comprehension (Endsley's level 2 SA) of information on the CDTI.

Perception-to-Comprehension (PTC) time library. The PTC library includes performance estimates for display properties that impose a time cost for information to go from initial perception (Endsley's level 1 SA) to full comprehension (Endsley's level 2 SA). PTC time costs apply only when display information can be fully comprehended, after it is accessed and initially perceived. A longer time requirement for information to go from perception to comprehension via any given display property suggests it will take more time for the pilot to attain SA. The time cost also reflects the compatibility of that

display property with the pilot's perceptual and cognitive abilities. For instance, the pilot can attain comprehension of display information on both a cluttered and non-cluttered CDTI; however, the PTC time for the less compatible cluttered display will be longer.

SA Limiter library. The idea of Data Limits was described by Norman and Bobrow (1975), who referred to the perceptual and cognitive limitations of the operator as resource limitations (e.g., accurate spatial resolution is limited by the auditory channel; Begault & Pittman, 1996; Perrott, Saberi, Brown, & Strybel, 1990). Within the library framework, these resource limitations can be thought of as comprehension limiters (or SA limiters). In this way, the term "SA Limiter" used in this paper has partial equivalence to the original "Data Limit" term used by Norman and Bobrow. When a pilot must gather information through the use of a data limiting (that is, SA limiting) display property, the result is that the information may not be accurately comprehended, and is reflected as a cost to performance accuracy. In essence, the data limiting characterization of a display property reflects its level of compatibility (or incompatibility) with the pilot's perceptual and cognitive abilities.

The SA limiter library includes display properties that inhibit full comprehension (that is, they limit attainment of full SA) for a given task. Hence, unlike the Accessibility and PTC libraries, performance time estimates until full SA can be attained are less relevant since full SA is not attainable. Instead, the SA limiter library includes display properties where empirical data show a significant degradation in performance accuracy (e.g., auditory displays significantly degrade performance accuracy for spatial localization tasks; Begault & Pittman, 1996; Perrott, Saberi, Brown, & Strybel, 1990).

Empirical data that show significant time or accuracy differences between LEVELS of a SA limiting display property (e.g., monaural versus binaural [levels] of auditory displays for spatial localization tasks) can be used to distinguish between relatively *mild* versus *severe* degraders of SA. This dichotomy indicates levels of a display variable that may serve as alternative design options for the designer. For instance, both binaural and monaural displays limit accurate spatial localization, but significant performance differences between the two types of displays show that performance degrades to a greater extent with the monaural display (Begault, 1990; Begault & Pittman, 1996; Perrott, Saberi, Brown, & Strybel, 1990). The monaural and binaural displays can then be characterized as severe and mild SA limiters, respectively. Mild and severe characterizations indicate to the designer which display properties further degrade the attainment of full SA; they also have implications for design evaluation within a human performance model (MIDAS; Hooey et al., 2010).

Summary

The aim of developing a library of performance estimates is to address the challenge of designing flight deck displays for compatibility and SA, and also, to support the design and evaluation of current-day and NextGen flight deck displays. Flight deck display properties that impede SA via accessibility and PTC time, or as SA limiters have not been identified in the existing body of literature. Also, empirical data have not been synthesized and compiled for some of the display properties identified in this thesis, in a way that supports development of performance estimates. Additionally, the development of new displays for NextGen procedures will bring new and creative ways for displaying

and interacting with information in the cockpit that may not be characterized by any of the display properties identified in this thesis. For instance, the need for displaying time-to-arrival information in the conceptual NextGen 4DT environment in a way that supports pilot SA may bring designers to develop a variety of ways for displaying this type of information, perhaps a text display of time-to-arrival information, a spatial display of geographical and velocity information, or a spatial display of temporal information. So, a systematic method for building onto the library in future efforts is needed. This thesis, then, aimed to:

- identify display properties that are appropriate for populating the IA, PTC, and SA limiter libraries through literature searches;
- search for already integrated data (existing meta-analyses or results of existing meta-analyses e.g., quantified design guidelines and principles) that can be included in the library;
- search for data and integrate those data using meta-analyses to develop quantitative performance estimates for the IA and PTC libraries;
- search for data and integrate those data using meta-analyses to identify SA-limiting display properties, or levels within a display property (and characterize them as either mild or severe where possible), for the SA limiter library; and
- develop a systematic method so that the library can be easily revised to incorporate new data and new display properties.

As the library is developed, it can serve as a more comprehensive design tool.

Method

The following methods were developed through iteration and used to build the library of quantitative performance time effects and qualitative estimates of display properties that impede SA. In summary, the approach used analyses of mean performance across studies to develop estimates.

Step 1: Identify Display Properties that Affect Pilot SA

To determine which display properties had potential to affect SA, a combination of top-down and bottom-up approaches were adopted.

- A top-down approach: A theoretically driven survey of the human factors, cognition, and aviation display research domains was used to identify display properties that had the potential to impact pilot SA via performance time and/or accuracy. This was done by querying subject matter experts for potential display properties that could affect SA, for instance, based on human factors principles (e.g., Proximity Compatibility Principle) or psychology or cognition theory (e.g., working memory load; resource limitations).
- A bottom-up approach: A survey was conducted to identify empirical studies that evaluated flight deck display properties. The scope of the literature search included both current-day and NextGen operations. Database searches were conducted using Google Scholar, the Human Factors Publications Database of the University of Illinois, and the NASA Technical Reports Server by using relevant keywords. In addition, a search through reference lists of papers in the human factors and aviation domains was conducted to find studies that examined display

effects on SA or performance. Additionally, cockpit display environments in Gore et al. (2010) were examined for display properties that could potentially affect SA.

Candidate display properties that had potential for developing pilot performance estimates for the library were marked as placeholders for the library.

Step 2: Identify Performance Parameter: IA, PTC, or SA Limiting Displays

The human performance parameter (either information accessibility time, perception-to-comprehension time, or a SA limitations) affected by each candidate display property was determined by using the IF-THEN decision logic presented in Table 1. Each display property was then assigned to one of three corresponding sub-libraries: the IA library, PTC library, or SA limiter library. Then, literature searches through related domains for empirical data related to these placeholders were conducted. If empirical data supported a candidate display property's effect on SA (in terms of significant performance time or accuracy effects) it was included in the library.

Table 1

IF-THEN Decision Logic for Identifying Appropriate Sub-libraries for Display Properties

IF...	THEN the display property is...
<p>the display requires manipulation of the display property before the relevant display information can be accessed...</p>	<p>an INFORMATION ACCESSIBILITY TIME factor and should populate the IA library</p>
<p>the display can be immediately detected without manipulation, AND full comprehension of the task relevant display is attainable given sufficient time...</p>	<p>a PERCEPTION-TO-COMREPEHSION TIME factor and should populate the PTC library</p>
<p>the information can be immediately detected without manipulation, but full comprehension is not attainable due to a perceptual or cognitive limitation regardless of how long the display is viewed...</p>	<p>a SA LIMITING factor and should populate the SA limiter library</p>

Step 3: Gather Empirical Data

A comprehensive literature search was conducted to identify studies for inclusion in the meta-analysis. Database searches (for example, through Google Scholar, the Human Factors Publications Database of the University of Illinois, or the NASA Technical Reports Server) served as useful resources for identifying studies with empirical data of display property performance effects. Keywords used for literature searches were developed through a brainstorm of candidate display properties for current-day and NextGen displays. They included display format, display property, display factor, cockpit, display, NextGen, keystroke, keypress, overlaid, clutter, search clutter, display separation, text height, tracking display, auditory display, spatial, binaural, digital, display highlighting, datalink, verbal, angle of perspective viewing, 3D ambiguity, predictor, color coding, conformality, temporal display, time display, time-to-contact, time-to-arrival.

The criteria used to select studies for inclusion into the meta-analyses were that the paper must have:

- been published in the public domain
- compared two or more relevant display properties
- provided sufficient detail to ascertain research method and display factors
- reported either performance-time (e.g., detection time, response time) or performance-accuracy effects (e.g., response accuracy, tracking accuracy)
- found a reliable display property effect, with statistical significance of at least $p < 0.1$

The most preferable were data from studies that applied directly to pilot performance in the aviation domain; however, for display properties where empirical data from the aviation domain were sparse, studies outside of the aviation domain were considered for the meta-analyses. The criterion of $p < 0.1$ was used to reduce type 2 error. Next, the data for display factor performance effects were synthesized.

Step 4: Compute Performance Estimates

Parameter meta-analyses techniques were used to average data across studies to estimate performance for a given display property. Data that explored the same display property were pooled across studies with different experimental conditions (e.g., flight test versus simulation), and varying subject populations (e.g., students versus pilots). In the present research, quantitative human performance parameters, such as target detection times and response times were averaged. The advantage of this parameter meta-analysis approach is that it produces estimates of human performance for each display property represented as performance “costs” or “benefits.” For some performance estimates, calculation of mean performance was not necessary if an equation of a display effect on performance was found in the literature.

On one hand, since IA and PTC factors are based on time delays until full comprehension, performance estimates for IA and PTC display properties were based on mean performance time delays as a function of the display property. On the other hand, since SA limiting factors are based on limited comprehension (inferred from a significant degradation in performance accuracy), SL display property performance estimates were based on performance accuracy data (see Table 2).

Table 2

Types of Data used for Developing Display Property Performance Estimates

Display property characterization that impedes SA	Sub-library	Type of data
IA factor	IA	Performance time
PTC factor	PTC	Performance time
SA limiter	SA limiter	Performance accuracy (mild vs. severe distinction based on performance time or accuracy)

Data from experimental conditions were included if they described a display property effect on the given task. For instance, if a tracking display was used for a tracking task, only tracking data would be used towards the mean performance estimate rather than target detection data that did not depend on the tracking display.

Additionally, studies used towards performance estimates fulfilled the criterion for inclusion previously described.

Step 4a: Computing accessibility time estimates. Accessibility performance estimates were calculated by averaging the mean time costs (response time, reaction time) across studies for a given accessibility display property.

Step 4b: Computing PTC time estimates. PTC performance estimates were calculated by averaging the mean time costs (response time, target detection time) across studies for a given PTC display property. There may be cases where different studies have very different absolute time costs, where the difference is a result from a difference in the baselines. For instance, two studies may reveal a mean difference of 5 s (an increase in performance time from 5 to 10 s) and 0.5 s (an increase in performance time from 0.5 to 1.0 s) caused by the change in display property. In such cases, a more consistent estimate of the performance time cost would be relative percentage cost, where there would be a 50% slowing in both cases. PTC performance estimates were calculated by averaging raw (or absolute) and relative performance time costs across studies where:

$$\text{Raw performance time cost} = |\text{Performance time} - \text{Baseline performance time}|,$$
$$\% \text{ performance cost} = \text{Raw performance time cost} / \text{Baseline performance time}.$$

Step 4c: Computing SL performance estimates. Unlike accessibility and PTC performance estimates, SA limiter performance estimates are not quantitative values to be calculated from mean performance times across studies. Rather, they are the identification of display properties that inhibit full SA, inferred from display properties that do not yield accurate performance (e.g., auditory displays for tracking tasks).

A SA limiting variable (display property) with two or more levels. Further, if a SA limiting display property could be distinguished by two or more levels, the levels were characterized as either relatively mild or severe. This is useful for the designer when there are different design options (different levels) within a single display property (variable). The mild/severe distinction was based on the relative severity of performance

degradation between levels of a given SA limiting display property, and was estimated on a case-by-case basis for each SA limiter. Thresholds of performance that can be systematically applied across SA limiting display properties to distinguish between mild and severe levels are not feasible, as different display properties may impose different ranges of performance error. So the levels of each SA limiting display property must be examined on a case-by-case basis. The aim is to estimate mild and severe levels, given the range of performance for a given display property. Where this distinction between mild and severe levels is not possible, or relevant, a single value can be averaged across studies to estimate the threshold where the display property acts as a SA limiter (in other words, the threshold where full comprehension of the display is not possible).

Significant performance differences were used to estimate mild versus severe levels of a single SA limiting display property characterized by qualitative properties (e.g., the SA limiting auditory display used for spatial judgment tasks can be qualitatively characterized as either binaural or monaural). If the SA limiting display property can be divided into two categories, the category that yields better performance is considered mild, and the other, severe (e.g., monaural displays are relatively severe SA degraders for spatial localization tasks).

A SA limiting variable (display property) with quantitative levels. Performance data were also used to estimate mild versus severe levels within a SA limiting display property characterized by a quantitative value or range of values (e.g., the angle of perspective viewing of a display may range from 30- to 90- degrees visual angle, with one category representing performance with 30- to 60-deg, and the second representing

61- to 90-deg). Estimating thresholds of mild and severe through meta-analyses was straightforward if the included studies used the same quantitative levels of the SA limiting variable. When they did not, thresholds for mild and severe were estimated with the following steps:

1. After identifying the SA limiting display property variable, studies were identified that manipulated the variable at two or more levels.
2. Error data (i.e., error rate, detection accuracy rate) were extracted. Where studies used different measures of error (e.g., error rate versus accuracy), they were converted to percent error rates for comparison. Where there was a statistically significant increase ($p < 0.10$) in error (or decrease in accuracy) across the levels of a SA limiter, the trend was examined to estimate three regions of severity (none, mild, and severe). For instance, from a graph of data points three regions of none, mild, and severe can be *estimated*.

Results: Library of Performance Estimates

Seventeen display properties were identified for the library that either summarize mean performance time effects and/or identify the level of SA supported by the display property. Data are presented in terms of performance costs (except where noted) for one of three performance parameters:

- 1) Information Accessibility; cost presented is the time delay (in seconds [s] or milliseconds [ms]) before initial perception of displayed information can begin
- 2) Perception-to-Comprehension; costs presented is the time (in s or ms) slowing to gain full comprehension given a display property (in some cases the time cost reported is relative to another display property)
- 3) Situation awareness limiter; display properties are presented that limit full comprehension, or SA. Where possible, SA limiters were dichotomized as either mild or severe. This was done when meta-analyses revealed significant differences between levels of a display variable that could be used as alternative design options by a designer.

For each performance estimate a description is given of the display property, followed by a table that indicates a performance time estimate for IA or DTC display properties, or the identification of SA limiters. Literature reviews and data points used to calculate performance estimates are in the Appendix.

Information Accessibility Library

Number of keystrokes. A keystroke display is one that requires a key-press manipulation before task relevant information can be detected by the pilot. The keystroke performance estimate does not apply to displays used for data entry, but rather, it applies when display information must be extracted by the pilot, but only after he or she executes a keystroke. For instance, a Multifunction Control Display Unit (MCDU) is a current-day text only device that displays messages to the pilot and accepts input through a keyboard. The MCDU may contain multiple systems (e.g., Communications Management Unit [CMU] or FMS [Flight Management System]) where a keystroke or keystrokes are required to interact with each system. Other examples of where keystroke is an information accessibility factor include conceptual NextGen selectable displays, such as terrain-selectable or weather-selectable display overlays on a CDTI.

Summary. A keystroke display delays information access, hence, detection of task-relevant information. The mean time delay is 1.4 s/keystroke as shown in Table 3.

Table 3

Number of Keystrokes Performance Estimate

Description	Performance time slowing	<i>N</i> = Number of studies: References
Performance slowing from a display that requires 1-2 keystroke manipulations before display information can be detected	1.4 s per keystroke	<i>N</i> = 2: Card, Moran, & Newell (1983); Olsen & Nilsen in Lane, Napier, Batsell, & Naman (1988)

Note: See Appendix for studies and data used to calculate this performance estimate.

Perception-to-Comprehension Time Library

Search clutter. Search clutter can be characterized by the number of potential search target items (N) displayed against a blank background (e.g., a continuous, homogeneous background) that the pilot must search through for a given task (Figure 2). An example of a flight deck display formatted with search clutter is a CDTI that displays a set of potential hazard aircraft against a blank (or continuous white) background.

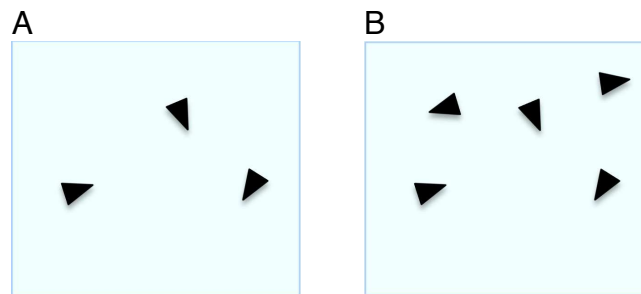


Figure 2. Search clutter against a homogeneous (blank) background. Arrays of search elements against blank backgrounds are illustrated, where the number of potential targets, N , is 3 in Panel A, and 5 in Panel B.

Summary. Search clutter results in delayed performance for identifying task-relevant information (or a target). Neisser's Serial Search Time (SST) model estimates the performance time delay as a function of search clutter when the target is present ($ST = a + (bN/2)$) or absent ($ST = a + bN$), as shown in Table 4.

Table 4

Search Clutter Performance Estimate

Description	Performance time slowing equation	N = Number of studies: References
Performance slowing caused by more elements to be searched through against a BLANK background	$ST = a + (bN/2)$ when the target is present $ST = a + bN$ when the target is absent	$N = 1$: Neisser as cited in Nunes, Wickens, & Yin (2006) and in Wickens, Nunes, Alexander, & Steelman (2005) Other references for clutter against a blank background: Beck, Lohrenz, & Trafton (2010); Wolfe, Oliva, Horowitz, Butcher, & Bompas (2002)

Notes: ST = search time per search element; N is the number of items in the field, a is the intercept parameter which characterizes the “read out” of the actual target; b is the time to inspect each of N items and decide it is not the target. When the target is present it takes a search through approximately $\frac{1}{2}$ the array to identify the target. When the target is not present, the entire array must be searched (Neisser as cited in Nunes, Wickens, & Yin [2006] and in Wickens, Nunes, Alexander, & Steelman [2005]). These equations only hold when searching through a display with no background clutter. See Appendix for studies and data that contribute to this performance estimate.

Background overlay clutter. Background clutter can be characterized by its complexity (Rosenholtz, Li, & Nakano, 2007; Wickens, Nunes, Alexander, & Steelman, 2005; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002). Increasing complexity of the background overlaid scene is referred to here as increasing background overlay clutter. It can be contrasted with search clutter, in which the effect is measured in search tasks against a homogeneous (or blank) background of uniform color. An example of displays with background overlay clutter are 1) heading or altitude overlaid against a continuous, heterogeneous, terrain background as illustrated in Figure 3A, and 2) displays of task-

relevant target information such as traffic icons displayed against terrain backgrounds as illustrated in Figures 3B and 3C. Upon qualitative inspection, background overlay clutter can be observed to increase from relatively mild in Figure 3A, to relatively moderate in Figure 3B, to relatively severe in Figure 3C.



Figure 3. Background overlay clutter. All panels present a display of task-relevant information overlaid on a non-homogenous background of terrain. Background clutter increases from relatively mild in Panel A, to relatively moderate in Panel B, to relatively severe in Panel C. Panel A Adapted from “Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling,” by M. Yeh, J. L. Merlo, C. D. Wickens, & D. L. Brandenburg, 2003, *Human Factors*, 45(3), 390-407. Panels B and C were adapted from “Measuring search efficiency in complex visual search tasks: Global and local clutter,” by M. R. Beck, M. C. Lohrenz, & J. G. Trafton, 2010, *Journal of Experimental Psychology: Applied*, 16(3), 238-250.

Summary. Increasing background overlay clutter prolongs target detection time compared to displays formatted without background overlay clutter. Performance time delay increases when background overlay clutter is increased from none to relatively mild background clutter by $M = 1.3$ s, by $M = 6.8$ s when increased from none to relatively moderate background clutter, and by $M = 20.2$ s when increased from none to relatively severe background clutter as summarized in Table 5. These performance time estimates have implications for a human performance model (Hooey et al., in preparation). As a design tool however, levels of clutter (e.g., severe, mild, moderate) will need to be better defined.

Table 5

Background Overlay Clutter (Mild, Moderate, and Severe) Performance Estimate

Description	Performance time slowing	Percent time slowing	N = Number of studies: References
Mild background overlay clutter- Performance slowing caused by mild clutter of an overlaid background	$M = 1.3$ s performance slowing relative to displays without background overlay clutter	16.3% performance slowing relative to displays without background overlay clutter	$N = 1$: Yeh, Merlo, Wickens, & Brandenburg (2003; Exp. 1)
Moderate background overlay clutter- Performance slowing caused by moderate clutter of an overlaid background	$M = 6.8$ s performance slowing relative to displays without background overlay clutter	283% performance slowing relative to displays without background overlay clutter	$N = 2$: Beck, Lohrenz, & Trafton (2010; Exps. 1-2)
Severe background overlay clutter- Performance slowing caused by severe clutter of an overlaid background	$M = 20.2$ s performance slowing relative to displays without background overlay clutter	842% performance slowing relative to displays without background overlay clutter	$N = 2$: Beck, Lohrenz, & Trafton (2010; Exps. 1-2)

Note: See Appendix for studies and data used to calculate this performance estimate.

Overlaid versus separate databases for focused tasks. Displays on the flight deck can be characterized as overlaid (Figure 4) when there are overlaid information databases that share a common frame of reference. An example of an overlaid display on the flight deck is an integrated display with overlaid traffic, weather and terrain domains. There may be advantages to overlaying databases into a single display. However when pilots must identify one piece of information separately for a focused task (i.e., identify a traffic hazard only), the overlaid databases may increase the time required for pilots to do so.

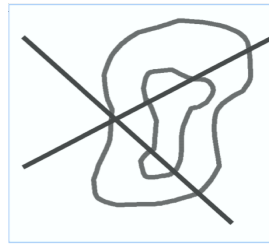


Figure 4. Overlaid displays. Two separate information databases that share a common reference frame can be overlaid (for instance, traffic, weather, and terrain databases). In the figure, the straight lines represent one information domain (e.g., traffic) and the curved lines represent another (e.g., weather). Adapted from *The Display of Multiple Geographical Data Bases: Implications of Visual Attention*, by P. Kroft & C. D. Wickens, 2001, Savoy: University of Illinois, Aviation Research Lab.

Summary. Searching for a target on a single information database (a focused task), such as imminent traffic hazards, is slowed when it is overlaid with another

database that shares a common reference frame (e.g., an overlay of terrain information) by $M = 2.8$ s as shown in Table 6. This value is highly clutter dependent and is expected to increase in a highly cluttered display.

Table 6

Overlaid Display Performance Estimate for Focused Tasks (Relative to Separate Display Formats)

Description	Performance time slowing	N = Number of studies: References
Performance slowing caused by overlaying multiple databases and searching for a target on one	$M = 2.8$ s relative to separate display formats	$N = 4$: Kroft & Wickens (2001); Lohrenz (2003); Wickens, Alexander, Thomas, Horrey, Nunes, Hardy, & Zheng (2004); Wickens, Kroft, & Yeh (2000)

Notes: One of the studies did not specify absolute baseline search times (Wickens, Kroft, & Yeh, 2000), so percentage slowing could not be calculated. See Appendix for studies and data used to calculate this performance estimate.

Separate versus overlaid displays for information integration tasks. When information databases share a common frame of reference, they can be either overlaid on a single display, or presented separately on individual displays (Figure 5). An example of a separate display format of multiple databases on the flight deck that must be used for a mental integration task would be three separate displays for each of traffic, terrain, and weather information domains that share a common geographical reference frame, and that require mental integration by the pilot for spatial integration tasks, such as for avoidance maneuver selection based on all types of hazards. For such a task the pilot must divide his attention among the three types of information, where after, the disparate sources of

information must be mentally integrated so that the relative spatial positions of each type of hazard can be comprehended with respect to each other and with respect to the pilot's aircraft, so that an appropriate maneuver can be selected— a time consuming task compared to performing the task with an overlaid display property that would otherwise illustrate the three information domains against a shared geographical reference frame.

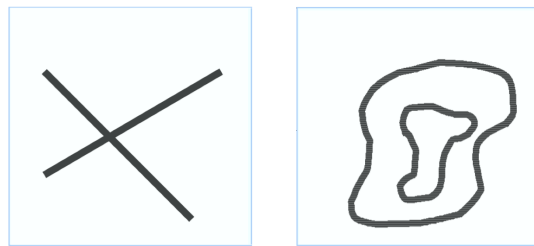


Figure 5. Separate displays. Two separate information databases that share a common geographical reference frame (for instance, traffic, weather, and terrain databases) are separately displayed, rather than overlaid. Adapted from *The Display of Multiple Geographical Data Bases: Implications of Visual Attention*, by P. Kroft & C. D. Wickens, 2001, Savoy: University of Illinois, Aviation Research Lab.

Summary. Displaying multiple databases on separate displays imposes a time cost for tasks that require mental integration across the multiple databases by $M = 1.2$ s, or 23.6%, relative to an overlaid display as shown in Table 7.

Table 7

Separate Display Performance Estimate for Information Integration Tasks (Relative to Overlaid Display Formats)

Description	Performance time slowing	Percent time slowing	N = Number of studies: References
Performance slowing caused by separating two spatially related databases, where mental integration is required	<i>M</i> = 1.2 s relative to overlaid display formats	23.6%	<i>N</i> = 2: Kroft & Wickens (2001); Schons & Wickens (1993)

Note: See Appendix for studies and data used to calculate this performance estimate.

Wide display separation for information integration tasks. For flight tasks that required mental integration across two or more data sources, a further cost may be assessed as a function of the degree of physical distance separating the display. A separation from 7.5 to 25.7-degrees visual angle is characterized here as "wide" display separation, for instance, widely separated traffic and weather displays (Figure 6) that require mental integration for a hazard-avoidance maneuver selection task.

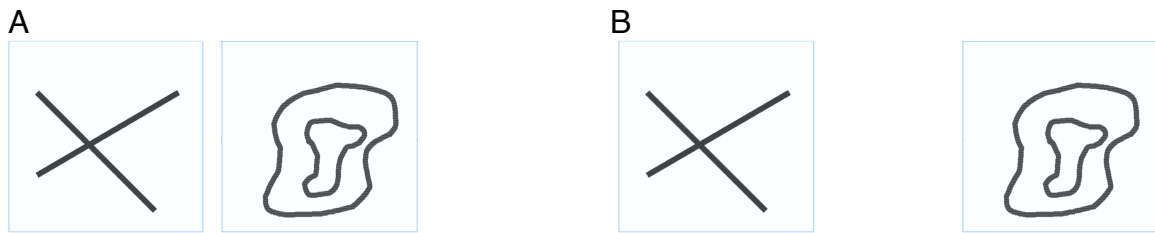


Figure 6. Display properties with separate information databases at varying degrees of separation. Close separation is depicted in Panel A and wide separation in Panel B

Adapted from *The Display of Multiple Geographical Data Bases: Implications of Visual Attention*, by P. Kroft & C. D. Wickens, 2001, Savoy: University of Illinois, Aviation Research Lab.

Summary. Performance time increases by $M = 1.6$ s, or 23%, for display separations from 7.5- to 25.7- degrees visual angle, compared to overlaid display properties with a 0-degree visual angle separation (Schons & Wickens, 1993), as shown in Table 8.

Table 8

Wide Display Performance Estimate for Information Integration Tasks

Description	Performance time slowing	Percent time slowing	N = Number of studies: References
Performance slowing when separation of two displays that must be integrated is 7.5 degrees visual angle or greater	$M = 1.6$ s relative to a closely spaced (<7.5-deg) or overlaid display format	23% averaged across display separations from 7.5- to 25.7- degrees visual angle	$N = 1$: Schons & Wickens (1993)

Note: See Appendix for studies and data used to calculate this performance estimate.

Text string length. Text displays can be characterized by the length of a text string. An example of text strings displayed on the flight deck is an ATC command displayed in text strings of various lengths on a Datalink display. The time required per character to access the information depends on the type of character (numerical digit vs. letter).

Summary. Comprehension time increases for a text string by 33.4 ms per numerical digit, 40.2 ms per letter character, and/or 47 ms per word (Cavanagh, 1972 as cited in Card, Moran, & Newell, 1983), as summarized in Table 9.

Table 9

Text String Length Performance Estimate

Description	Performance time slowing	N = Number of studies: References
Slowing from the time it takes to comprehend text strings	33.4 ms per numerical digit	N = 8 Cavanagh (1972)
	40.2 ms per letter	N = 13 Cavanagh (1972)
	47 ms per word	N = 4 Cavanagh (1972)

Note: See Appendix for studies and data used to calculate this performance estimate.

Target highlighting for target identification tasks. Displays can manipulate the salience of targets relative to a background (or distractor information domain) with a single unique color or intensity. This results in high-lighted or non-lighted targets; that is, a target can be highlighted against distractors or, it can be of uniform intensity with

distractors (Figure 7). An example of a display formatted with a highlighted target on the flight deck is an overlaid display of traffic, where only the most imminent, single traffic hazard is made salient by displaying it at the highest relative intensity, compared to non-highlighted less hazardous aircraft. Alike, a set of potential targets could be highlighted.

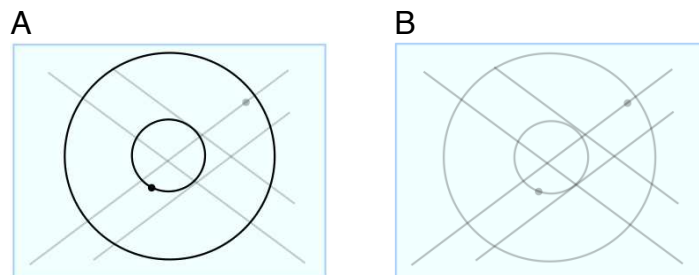


Figure 7. Highlighting. Two domains are displayed, circles versus lines, where the circle domain is highlighted at a greater intensity in Panel A, and is of uniform intensity with the other information domain in Panel B.

Summary. A benefit of highlighting, or intensity coding, is to enable visual segregation of displayed information in a cluttered display, for instance, an overlaid display (Wickens, Ambinder, Alexander, & Martens, 2004). When a single target is made salient through highlighting, detection performance time decreases by $M = 18.5$ s, or 84%, relative to when highlighting is not employed. When a target set is salient or, relatively highlighted, target detection performance time decreases by $M = 2.1$ s, or $M = 13.9\%$, compared to when it is non-salient (that is either of uniform intensity or relatively low-lighted against the background). Based on empirical data, salient highlighting of a **set** of targets showed the smaller benefit where as salient highlighting of a **single** target

showed the larger benefit: a true "pop out" effect (Beck, Lohrenz, & Trafton, 2010). This is summarized in Table 10.

Table 10

Target Highlighting Performance Estimate

Description	Performance time gain (benefit)	Percent time gain (benefit)	<i>N</i> = Number of studies: References
Faster target discrimination due to highlighting of a single target , against a background with a single unique intensity	18.5 s relative to a non-salient single target	84% relative to a non-salient single target	<i>N</i> = 1: Beck, Lohrenz, & Trafton (2010, Exp. 3)
Faster target discrimination due to highlighting of a target set against a background with a single unique intensity	<i>M</i> = 2.2 s relative to a non-salient target set	17.9% relative to a non-salient target set	<i>N</i> = 3: Nunes, Wickens, & Yin (2006); Podscerwinski & Wickens (2002); Wickens, Ambinder, Alexander, & Martens (2004)

Note: Salient highlighting of a single target showed a larger benefit compared to highlighting of a target set: a true "pop out" effect (Beck, Lohrenz, & Trafton, 2001).

SA Limiter Library

3D display angle of perspective viewing for spatial judgment tasks. 3D

perspective display formats can be characterized by their angle of perspective view (or, the displayed vantage point; Figures 8A and B), which can be created by manipulating the displayed viewpoint along the vertical axis (angle of elevation) and/or the viewpoint along the lateral axis (angle of azimuth), for example, with a 3D CDTI. In the figures, perspective viewing is manipulated by changing the angle of elevation (Figure 8A) or the angle of azimuth (Figure 8B). The camera viewpoints labeled A, B, C, and D in the top panels correspond with the displays A, B, C, and D illustrated in the bottom panels. Viewing vectors are illustrated with black dashed lines, while x-, y-, and z-axes are labeled and illustrated with blue solid lines. Ownship is illustrated with a magenta triangular symbol, and objects in space are illustrated with solid black circles. A progression of increasing angles of elevation or azimuth are illustrated in the bottom panels from Panel A, to Panel B, then C. Panel D is an extreme view where angle of elevation or azimuth is maximized at the expense of no resolution along the y- or x- axes (respectively), otherwise creating a 2-dimensional view.

A benefit of perspective view is that all three spatial axes can be viewed in an integrated format. This comes at the cost, however, of accurate spatial judgments (Merwin & Wickens, 1998). This is sometimes referred to as “compression” (McGreevy & Ellis, 1986).

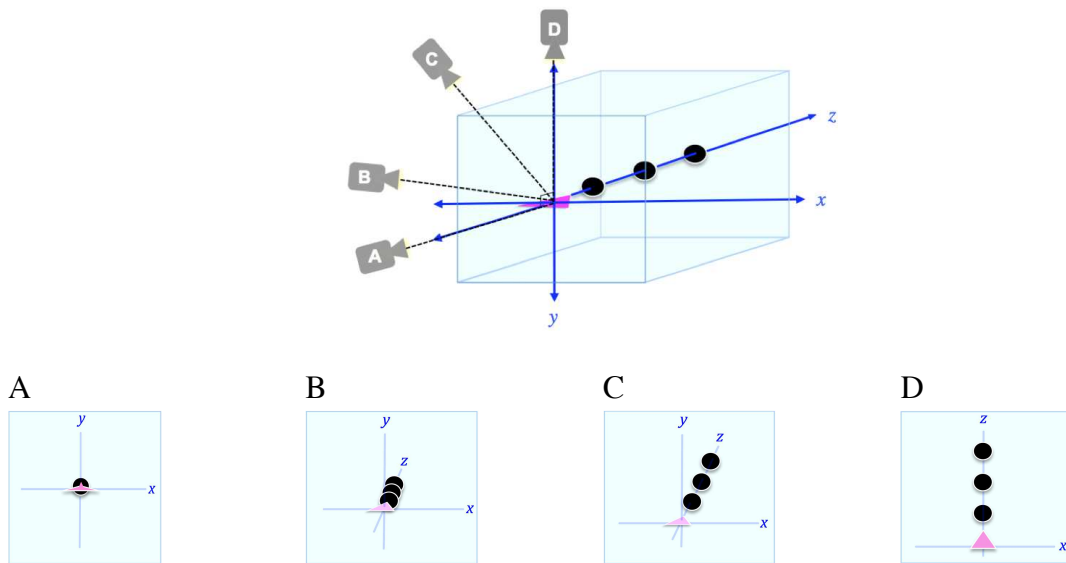


Figure 8a. Increasing angle of perspective view by manipulating the angle of elevation.

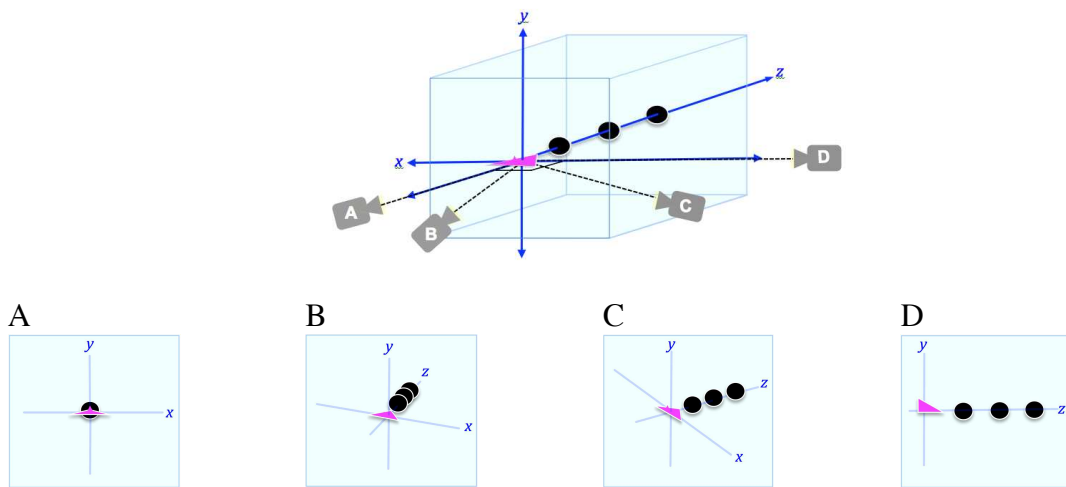


Figure 8b. Increasing angle of perspective view by manipulating the angle of azimuth.

Summary. Spatial judgment accuracy is limited along the axis that perspective viewing is increased (e.g., vertical perspective viewing angle degrades accurate vertical axis judgment) as a function of 3D display ambiguity that comes with increasing angles of perspective viewing (Boeckman, 1998; Merwin & Wickens, 1998). There is a non-linear increase in error with increasing angles of perspective viewing. Performance accuracy degradation is mild when the perspective view is 30-60 degrees of visual angle, and severe for viewing angles >60-90 degrees of visual angle viewing (Boeckman, 1998; Merwin & Wickens, 1998) as shown in Table 11.

Table 11

Angle of Perspective Viewing Performance Estimate

Description	SA limiter		N = Number of studies: References
	Mild	Severe	
Spatial judgment accuracy is limited along the axis that perspective viewing is increased (e.g., vertical perspective viewing angle degrades spatial judgment accuracy along the vertical axis)	30-60 degrees angle of perspective viewing	>60-90 degrees angle of perspective viewing	N = 2: Boeckman (1996); Merwin & Wickens (1998)

Note: See Appendix for studies and data related to this performance estimate.

Display legibility (height of text and symbols). Alphanumeric text and symbols can be characterized by the visual angle that they subtend. An example on the flight deck where display legibility is a performance factor includes detection and comprehension of

aircraft data tag information on a CDTI, or airport taxiway labels on a taxi navigation display. Alphanumeric text and symbols that subtend a larger visual angle will be comprehended with greater accuracy. The cost of displayed text and symbols that are too small is that more time and effort is required for comprehension.

Summary. Recommendation from MIL-STD 1472F (1999) for displayed text and symbols indicate that character height should not subtend a height of less than 0.3 degrees visual angle, with a 0.5 degree visual angle preferred. Comparable recommendations are made in the FAA Human Factors Handbook (Ahlstrom & Kudrick, 2007). This can be applied to both alphanumeric text and symbols. A performance accuracy penalty will be imposed when text and symbols are too small; this cost is expected to be mild for heights between 0.3 and 0.5 degrees visual angle, and severe when less than 0.3 degrees visual angle as summarized in Table 12.

Table 12

Display Legibility Performance Estimate

Description	SA limiter		N = Number of studies: References
	Mild	Severe	
Visual angle height of text/symbols	0.3 ° - 0.5° visual angle height	<0.3° visual angle height	N = 2: DoD MIL STD 1472F (1999); FAA Human Factors Handbook DOT/FAA/TC-07/11 (Ahlstrom & Kudrick, 2007)

Notes: See Appendix for studies and data related to this performance estimate. The appendix also provides estimates for black and white text, colored text, warning and caution symbols, and head-up display (HUD) alphanumeric and symbols.

2D tracking display size: Visual angle of maximum excursion. Visual angle of maximum excursion (VAME) is the visual angle spanned by the maximum possible tracking error on a tracking display, as depicted in Figure 9. It is an absolute measuring scale for tracking display-size (C. D. Wickens, personal communication, April 2011) that takes into account varying geometric fields of view and/or physical size (e.g., the subtended visual angle) of the tracking display. That is, display compression and expansion are accounted for.

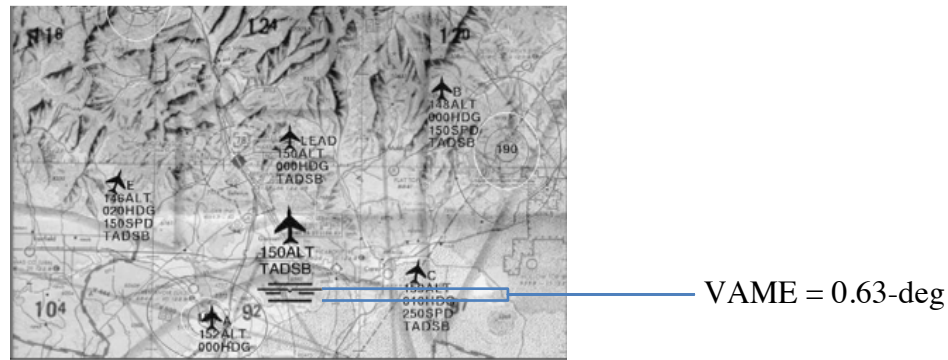


Figure 9. Visual Angle of Maximum Excursion (VAME). The tracking display (attitude direction indicator) within the greater cockpit display shows a maximum altitude tracking error that spans 0.63 degrees of visual angle. For this display VAME = 0.63. Adapted from “Pilots strategically compensate for display enlargements in surveillance and flight control tasks,” by E. M. Stelzer & C. D. Wickens, 2006, *Human Factors*, 48, 166-181.

Summary. Decreasing VAME from 0.63 to .30 degrees visual angle increases error by double (Stelzer & Wickens, 2006). As summarized in Table 13, 0.63 degrees is estimated to be the VAME threshold where error begins to increase.

Table 13

VAME Performance Estimate

Description	SA limiter	N = Number of studies: References
The visual angle subtended by displayed tracking error when it reaches its maximum likely value	<0.63-deg VAME	N = 1: Stelzer & Wickens (2006, Exp. 3)

Note: See Appendix for studies and data related to this performance estimate.

Auditory displays for spatial localization tasks. Displays that are formatted to convey spatial information, such as the location of traffic aircraft, can be characterized by the sensory modality required to perceive the display. For instance, a visual text format such as a Datalink display, a visual spatial analog format such as a graphical CDTI, or, an auditory format such as a speech or tone display. These display formats can each be used to identify the presence of traffic, and also, for making spatial judgments of traffic.

Spatial localization accuracy, however, is limited by auditory displays- either monaural or binaural. Generally, the term “monaural” means: relating to, or designating sound reception by one ear. In the present context, “monaural” refers to an auditory stimuli that does not have directional quality. For instance, a monaural display on the flight deck may consist of an auditory tone that sounds the same regardless of whether the traffic is located in either direction of ownship, right or left. In both of these examples, the non-spatial monaural display does not provide any spatial information of where the traffic is located. A binaural display format does, however, have a directional quality that corresponds with the spatial location of the displayed objects. For instance,

with a binaural display format, an auditory warning of traffic aircraft to the left of ownship could be delivered to the pilot’s left ear via headset, and traffic to the right of ownship could be represented by an auditory stimulus to the pilot’s right ear, providing the pilot with directional information of where the traffic is located.

Summary. Spatial localization accuracy is limited by auditory displays of spatial information, either monaural or binaural. Performance accuracy degradation however, is mild with binaural displays (which do provide some directional cueing) and severe with monaural displays (which do not provide any directional cueing) as summarized in Table 14.

Table 14

Auditory Display Performance Estimate for Spatial Localization Tasks

Description	SA limiter		N = Number of studies; References
	Mild	Severe	
Auditory display of sounds (e.g., a tone) to identify spatial information limits performance accuracy	Binaural displays of spatial information	Monaural display (e.g., a non-spatial tone or spoken word)	N = 3: Begault (1993); Begault & Pittman (1996); Perrott, Saberi, Brown, & Strybel (1990; Exp. 1)

Note: See Appendix for studies and data related to this performance estimate.

Auditory display string length. Auditory displays that convey information can be characterized by the length of the auditory string, according to the number of digits included. For instance, the length of an auditory string increases from an auditory command of altitude with heading, for example, “altitude 250, heading 315” which consists of 6 digits, to a command of auditory + heading + airspeed, "altitude 250, heading 315, airspeed 170 knots,” which consists of a 9 digits.

Summary. Auditory displays can benefit performance by off-loading visual demands in the cockpit. A cost is imposed, however, if the auditory information imposes a load on working memory, where it cannot be accurately comprehended. The number of digits in an auditory command affects performance accuracy where 5-7 digits mildly degrades performance, and 8+ digits relatively severely degrades performance, as summarized in Table 15.

Table 15

Auditory Display String Length Performance Estimate

Description	SA limiter		N = Number of studies: References
	Mild	Severe	
Increased number of digits in an auditory string degrades performance accuracy due to working memory load	5-7 digit auditory string	≥8 digit auditory string	N = 3: Helleberg & Wickens (2003); Loftus (1979); Wickens, Goh, Helleberg, & Talleur (2002)

Note: See Appendix for studies and data related to this performance estimate.

Symbol / abbreviation familiarity. Flight deck displays frequently use symbol or abbreviated text to convey information due to limitations in available display real estate and clutter consequences. An example of flight deck display that uses abbreviation is a FMS display with abbreviated menu text for minimizing clutter (Figure 10). An example of a symbol or abbreviation that is trained or infrequently encountered by the pilot is text in a different language or, a symbol that is required for a task that is rarely encountered (off-nominal or non-normal indicators).

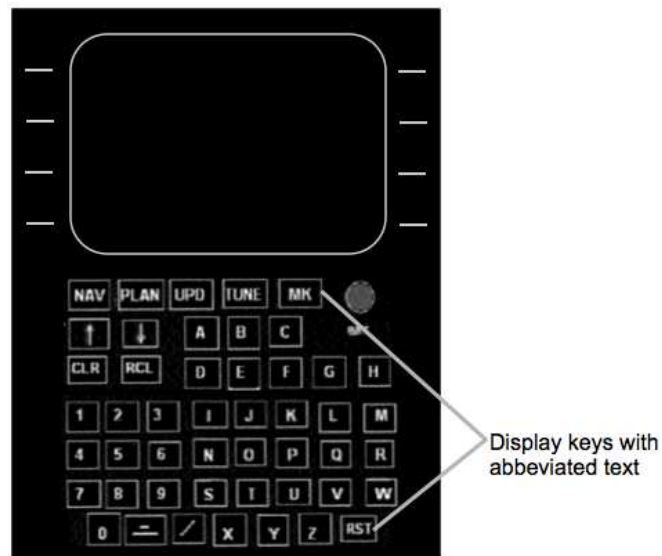


Figure 10. Abbreviated text on a FMS display. Abbreviated text and symbols require transformation by the pilot into something meaningful. The task is a greater challenge if the display is infrequently encountered, or if the pilot is untrained on the symbol/abbreviation meaning.

Summary. The benefit of symbols and abbreviations is to minimize display clutter. However, they require that the pilot transform the information into something that is meaningful. Although the pilot may have had formal training of the symbology or abbreviation meaning, if it is infrequently encountered (e.g., rarely seen or encountered on a cockpit display during the pilot's flying or simulator experience), performance accuracy will be mildly degraded. A completely untrained symbol/abbreviation, however, is a relatively severe degrader of performance as, summarized in Table 16.

Table 16

Symbol/Abbreviation Familiarity Performance Estimate

Description	SA limiter		N = Number of studies: References
	Mild	Severe	
Untrained or infrequently encountered symbology or abbreviations	Trained, but infrequently encountered	Untrained	N = 3: Fennell, Sherry, Roberts, & Feary (2006); Rehman, Reynolds & Neumeier (1995)

Note: See Appendix for studies and data related to this performance estimate.

Predictor format. A flight path predictor is a graphical depiction of the pilot's flight path trajectory in space and/or time. It can be characterized by the absence or presence of a predictor, the shape of the displayed predictor vector when it is present, and the shape of the corresponding path that it represents. For instance, a linear or curved predictor may be absent as in Figure 11A, or when present it may represent either a curved (Figure 11B) or curved (Figure 11B or C) flight path trajectory.

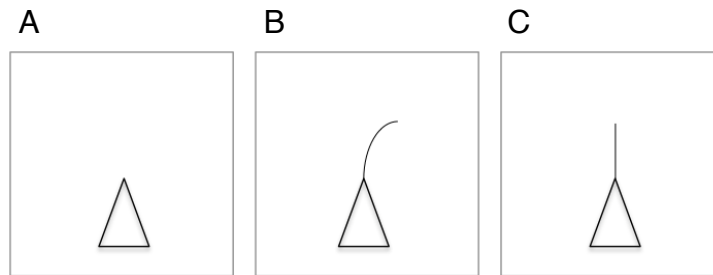


Figure 11. Predictor displays. The triangular ownship symbols are depicted with no predictor in Panel A, curved predictor in Panel B, and a straight predictor in Panel C. The straight predictor of ownship in Panel C may represent either linear or curved flight paths.

Summary. The benefit of a predictor display is that the pilot can spatially extrapolate their position in time and space with respect to other aircraft. Accurate flight path projection, however, suffers when no predictor is present, as summarized in Table 17. Performance accuracy degrades when there is no predictor display at all. Effects of linear predictors for curved encounters is another display property that has potential for inclusion within the library, given that there is supporting data in the literature.

Table 17

Flight Path Predictor Performance Estimate

Description	SA limiter	N = Number of studies: References
Displayed projection of an aircraft's future position, or flight path	No predictor	N = 2: Hart & Loomis (1980); Jago & Palmer (1982)

Note: See Appendix for studies and data related to this performance estimate.

Display conformality. Conformality refers to the preservation of angles within the real world onto a flight path tracking display, where the level of display conformality can lie on a continuum from fully conformal to non-conformal. Tracking display conformality has been characterized in the literature as

1. Display symbol that overlays the spatial position of its far domain counterpart in the real world, as the real world image moves across the display surface (e.g., if the aircraft rolls right the HUD image of a symbolic runway moves towards the left across the display) (Naish cited in Martin-Emerson & Wickens, 1997),
2. Display symbol that is analogous in shape with its far domain counterpart in the real world (or, forms an “object” with its far domain counterpart [Martin-Emerson & Wickens, 1997] (e.g., the symbolic runway in the near domain [or on the cockpit display] follows the shape of the real runway in the far domain even as its image across the display changes perspective),
3. Display symbol that is analogous with the directional motion of its far domain counterpart (e.g., the symbolic runway in the near domain moves in all directions, not just laterally or vertically on the display) (Martin-Emerson & Wickens, 1997), and
4. The display symbol is displayed with a 1:1 ratio with its far domain counterpart in the real world (e.g., the symbolic runway is not compressed in size to allow for more information to be presented on the display panel); that is, display

elements move with the same angular scaling as do their far domain counterparts (Bray cited in Martin-Emerson & Wickens, 1997).

For the purpose of estimating performance, three levels of display conformality are characterized here as 1) fully conformal (FC), 2) partially-conformal (PC), and 3) non-conformal (NC). A FC display has all four of the characteristics described above. For example, the FC runway symbol in Figure 12 overlays the spatial position of its far domain counterpart in the real world, is analogous in shape with the runway in the real world, is analogous with the directional motion of the far domain runway as its image moves across the display, and is displayed with a 1:1 ratio with the real world).

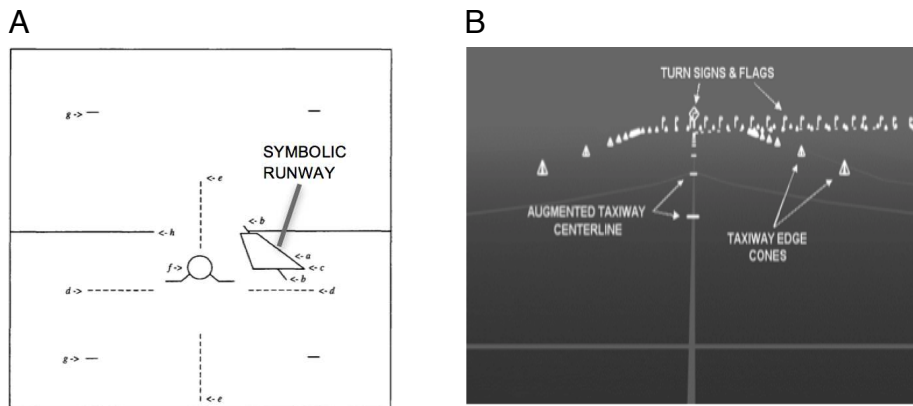


Figure 12. Fully conformal tracking display symbols. A) A symbolic runway used in Martin-Emerson and Wickens (1997) consists of the four characteristics listed above for a fully conformal display. Adapted from “Superimposition, symbology, visual attention, and the head-up display,” by R. Martin-Emerson and C. D. Wickens, 1997, *Human Factors*, 39(4), 581-601. B) The virtual runway centerline (called scene-linked symbology) used in Foyle, Hooey, Wilson, and Johnson (2002) also consists of the four characteristics listed above for a fully conformal display (2002). Adapted from “HUD symbology for surface operations: Command-guidance vs. situation-guidance formats,” by D. C. Foyle, B. L. Hooey, J. R. Wilson, and W. A. Johnson, 2002, *SAE Transactions: Journal of Aerospace*, 111, 647-658.

A PC display has some but not all of the four characteristics described above (a display is more conformal to the extent that more requirements are fulfilled). Martin-Emerson and Wickens (1997) describe that a symbol is characterized as partially conformal in that it conforms only to a specified parameter of the far domain analogue (e.g., motion OR shape). For example, the PC ILS (Instrument Landing System) display

crosshairs in Figure 13 consist of *some* but not all of the four characteristics listed above for a FC display, where only two of the four characteristics are displayed: 1) The ILS crosshairs *are* positioned on the pilot's display that correspond with localizer and glideslope frequencies emitted from near the runway so they **DO overlay the spatial position** of the localizer and glideslope frequencies as their images, if visible, would reach the aircraft; and 2) they **ARE analogous in directional motion** to its far domain counterpart (i.e., the ILS display crosshairs move across the display panel as the aircraft deviates from glideslope and localizer). However, the ILS crosshairs **are NOT fully analogous in shape** to its far domain counterpart (i.e., only lateral and vertical representations of glideslope and localizer ranges are displayed; there is no longitudinal [or perspective] representation displayed], nor are they displayed **with a 1:1 ratio** with their far domain counterpart in the real world (i.e., the ILS crosshair display is usually compressed along one or both dimensions). This display can be considered as partially conformal (Martin-Emerson & Wickens, 1997).

A NC display does not display any of the four characteristics described above. That is, it does not overlay the spatial position of its far domain counterpart, it is not analogous in shape with its far domain counterpart, it is not analogous in directional motion with its far domain counterpart, and it is not displayed with a 1:1 ratio with its far domain counterpart. A T-NASA display used in an experiment by Foyle, Hooey, Wilson, & Johnson (2002) shows a 2-dimensional runway centerline overlaid on a 3-D perspective display (Figure 14), and can be considered non-conformal.

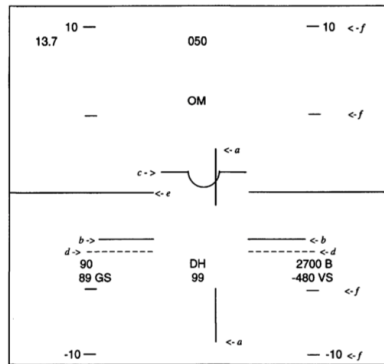


Figure 13. Partially conformal ILS tracking display symbol. Adapted from “Superimposition, symbology, visual attention, and the head-up display,” by R. Martin-Emerson and C.D. Wickens, 1997, *Human Factors*, 39(4), 581-601.

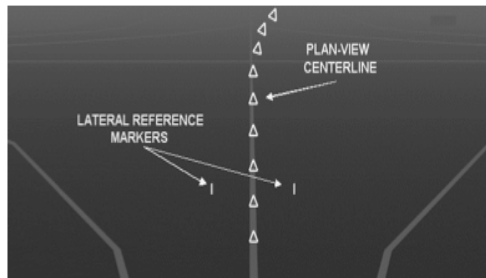


Figure 14. Non-conformal display of a 2-dimensional runway centerline superimposed on a 3-dimensional display of the far domain runway. Adapted from “HUD symbology for surface operations: Command-guidance vs. situation-guidance formats,” by D.C. Foyle, B. L. Hooey, J. R. Wilson, and W. A. Johnson, 2002, *SAE Transactions: Journal of Aerospace*, 111, 647-658.

Summary. Tracking performance accuracy degrades with partially conformal displays, and to a greater extent with non-conformal displays, when compared to fully conformal displays. PC and NC displays are characterized as mild and severe performance degraders, respectively as summarized in Table 18.

Table 18

Display Conformality Performance Estimate

Description	SA limiter		N = Number of studies: References
	Mild	Severe	
Conformality (the preservation of real world angles displayed on a tracking displays) affect tracking accuracy	Partially conformal tracking displays	Non-conformal tracking displays	N = 4 Foyle, Hooey, Wilson, & Johnson (2002); Martin-Emerson & Wickens (1997); Wickens & Long (1994); Wilson, Hooey, and Foyle, (2005)

Note: See Appendix for studies and data related to this performance estimate.

Spatial displays for temporal judgments. Temporal information, such as time-to-arrive (or TTA) information can be displayed in a variety of ways, for instance, by text that explicitly tells the pilot how much time remains until contact with a hazard or destination or perhaps with a spatial display that consists of objects moving at various velocities in space.

Summary. Significant degradation in TTA judgment accuracy has been observed with spatial display formats as a function of resource limitations in deciphering relative velocities of moving objects in space. This may result from a distance over speed bias (Law, Pellegrino, Mitchell, Fischer, McDonald, & Hunt, 1993; Xu, Wickens, &

Rantenen, 2004). Although temporal estimates can be made from objects in motion (speed, or relative velocity information), performance will not be accurate. Spatial displays for TTA judgments then, act as SA limiters as summarized in Table 19.

Table 19

Spatial Display Performance Estimate for Temporal Estimation Tasks

Definition	SA Limiter	N = Number of studies: References
Spatial displays that include both distance and motion (i.e., velocity) information used for making time-to-arrive (TTA) estimates	Spatial display formats	N = 3: Law, Pellegrino, Mitchell, Fischer, McDonald, & Hunt (1993) Experiment 1; Law, Pellegrino, Mitchell, Fischer, McDonald, & Hunt (1993) Experiment 2; Xu, Wickens, & Rantenen (2004)

Note: See Appendix for studies and data related to this performance estimate.

Discussion

Overview

In this thesis a library of performance estimates relevant to the NextGen and current-day cockpit display environments was developed. Display properties that impede full SA were identified for the library: 1) within the Information Accessibility library are display properties that impose time delays until display information can be accessed; 2) within the Perception-to-Comprehension time library are display properties that impose time delays until full comprehension of display information after it has been perceived; and 3) the Situation Awareness limiter library was populated with display properties that inhibit full SA. In addition to the identification of display properties that impede SA,

quantitative time estimates until full SA (raw and/or percentage cost estimates) were calculated based on literature searches and meta-analyses of empirical data, design guidelines, and design principles. A summary of the library is provided in Figure 15.

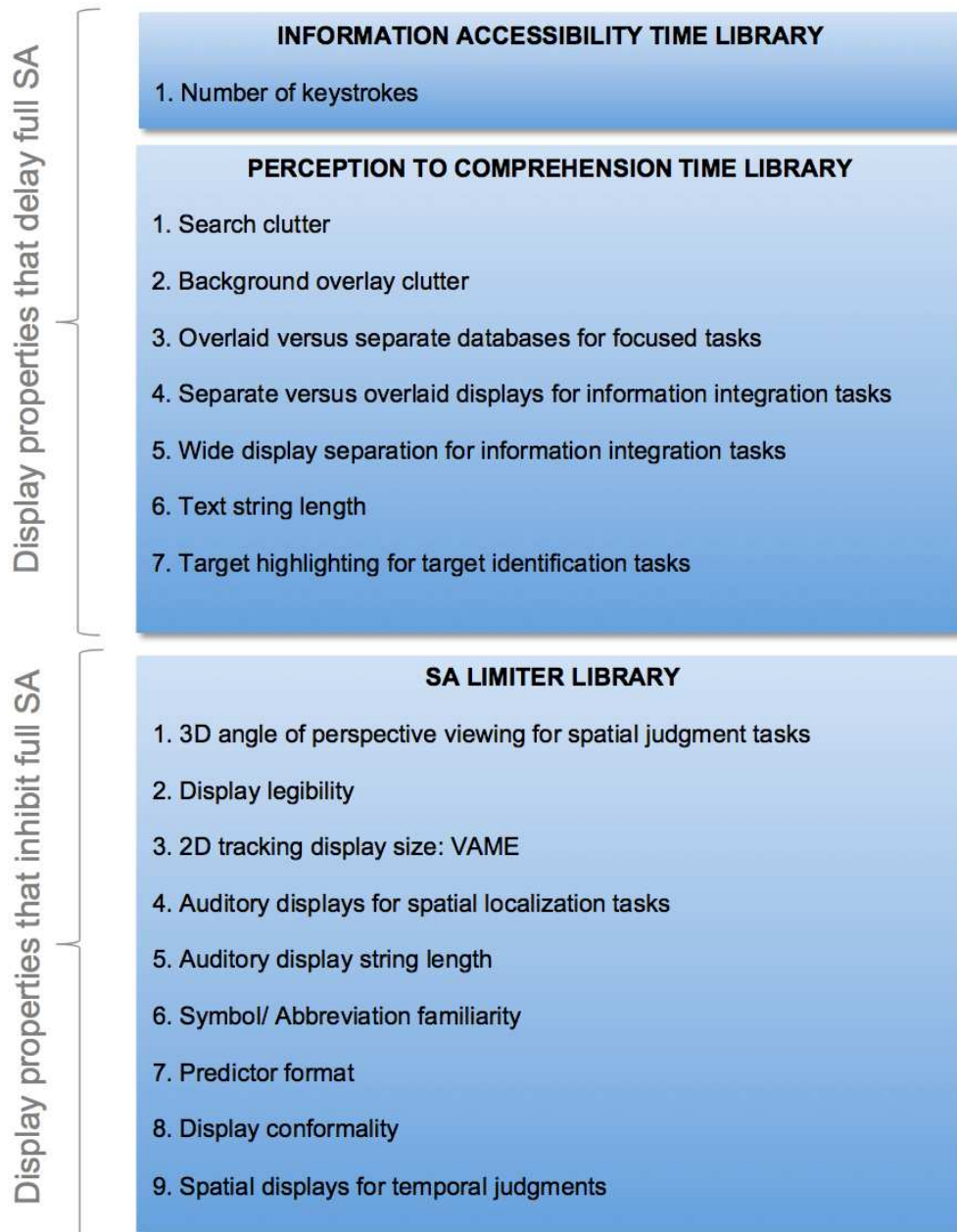


Figure 15. Summary of display properties in the library.

Many of the display properties and data that populate the library stem from display formats and references identified in Wickens' Display Formatting and Situation Awareness Model (2005). In particular, DFSAM guided the selection of the following display properties as those that affect SA, and also provided data references for calculating performance estimates: background overlay clutter; overlaid versus separate databases for focused tasks; separate versus overlaid displays for information integration tasks; display separation for information integration tasks; wide display separation for information integration tasks; target highlighting for target identification tasks; 3D angle of perspective viewing for spatial judgment tasks; and predictor format. DFSAM generated amalgamated performance units rather than quantitative and qualitative performance estimates, as developed in this thesis.

The qualitative distinction of display properties for which the only resource limitation is time until full SA can be attained versus those for which full SA cannot be attained are unique contributions to the SA, human factors, and aviation literature. Additionally, the quantitative performance estimates developed in this thesis are also unique contributions. Also, a systematic method that used meta-analyses to calculate mean performance estimates was developed. This method can be used to further refine and/or expand the library as data become available in the literature.

Implications

Recall that the objectives of this thesis were to develop a tool that 1) assists in designing displays for compatibility with pilot's perceptual and/or cognitive abilities and limitations; 2) provides a qualitative distinction of display properties that either support

or limit full SA; and 3) provides estimated absolute and relative time costs until SA can be attained for a given display property.

A tool for designing for compatibility. In developing the library within the framework of the SA information processing model (Alion, 2011), which is based on performance times until accurate comprehension and the identification of display properties that limit comprehension, a compilation of display properties has been created that reflects the level of compatibility for the given display formats. This addresses the first objective of this thesis. In essence, the IA and PTC libraries consist of display properties that can be described as “compatible” with the cognitive and perceptual limitations of the human user, whereas the SA Limiter library consists of display properties that are less compatible. For the latter, the designer must take heed that if full comprehension of display information is desired, then perhaps, supplemental means of presenting the information should be considered. Additionally, SA limiting display properties that are either continuous (e.g., angle of perspective view) or discrete (e.g., auditory displays) have been further broken down (e.g., 30-60 versus 60-90 degrees angle of perspective view; monaural versus binaural auditory displays) to help the designer understand the magnitude of the effect that an SA limiter may have.

General SA guidelines. As a SA design guideline, where SA limiters are distinguished from display properties that do not inhibit SA (but rather impose time costs until SA), the library fulfills the second objective of this thesis by providing a qualitative distinction of display properties that either support or limit full SA, and further, as mild versus severe SA degraders. As a general SA guideline, the library can also be used by

the designer to compare a design against SA requirements. In addition, the third objective is fulfilled, as the estimated absolute and relative time costs until SA can be attained for a given display property were calculated through literature searches for data and meta-analyses.

A compilation of display properties. A convenience for the designer is that performance time effects and support for SA of various display properties that are relevant to the cockpit environment are compiled in one place. Also, the mean quantitative estimates derived from meta-analyses are in a format that can be directly interpreted by the designer (as a raw and/or percentage cost/benefit), whereas a traditional meta-analysis that analyses statistical effect sizes may be less useful for the design decision-making process.

Human performance modeling. An added utility of the library is that both the qualitative and quantitative estimates can be used for human performance modeling. The results of this research are being directly integrated into a human performance model (Man-machine Integrated Design and Analysis System or MIDAS; Hooey et al., 2010) for predicting SA of NextGen flight deck display configurations (Hooey et al., in preparation).

Assumptions of Correlation Between Performance and SA

The framework of the library is based on an information processing model of SA (Alion, 2011) in which the attainment of full comprehension and limited SA is inferred from performance. It is clear that, in the literature, time and accuracy measures are relevant to SA as seen with the Situation Present Assessment Method (SPAM; Durso & Dattel, 2006) and the Situation Awareness Global Assessment Technique (SAGAT;

Endsley, 1998) measures of SA. Rather than using data from such query methods, the information processing model employed as a framework for the library uses performance time and accuracy measures as a function of a given display to assess if SA is attainable, and if so, how long it would take. A challenge for SA measures is that SA is dynamic; it changes with time in that it can be continually updated or degraded from one moment to the next. Also, it is possible for good performance to occur even without SA.

Gaps in the Current Library and Future Work

Gaps have been identified in the empirical literature related to some of the display properties in the existing library. In some cases a display property was identified to affect SA as either an IA, PTC, or SA limiting display property but the appropriate data could not be found (performance time or accuracy). The following areas have potential for future human-in-the-loop research.

Background overlay clutter. Recall that background overlay clutter estimates a PTC time cost associated with increased clutter on a display. However, the data used to generate these PTC time costs were limited because 1) only one study was available for the “mildly” cluttered display, and 2) the characterizations of “mild,” “moderate,” and “severe” in addition to the calculated mean estimates are specific to the displays used in the referenced studies. More clutter research using a variety of flight-relevant displays is needed for a generalizable estimate. Also, there is a need for objective clutter metrics to characterize (or quantify) different levels (e.g., mild, moderate, and severe) of clutter. A literature review of clutter metrics is provided in the Appendix.

Overlaid versus separate display for focused attention tasks. A PTC time cost was estimated for the overlaid versus separate display property for focused tasks. A percentage cost estimate was not calculated because one of the four studies did not specify absolute baseline search time. This is an area for future work towards refining the library. More data from studies in the aviation domain that compared separate versus overlaid displays would contribute to refining the PTC percentage time cost, while providing greater power. Also, the 2.8 s PTC cost is expected to increase in a highly cluttered overlaid display. This emphasizes the need for objective clutter metrics as previously mentioned.

Display separation versus WIDE display separation for focused attention tasks. Two performance estimates were calculated regarding display separation. They are referred to in the Results as “Separate versus overlaid displays for information integration tasks” and “WIDE display separation for information integration tasks.” The first performance estimate does not specify the degree of separation; one of the two studies for calculating the performance estimate (Kroft & Wickens, 2001) did not specify these parameters. The second estimate, however, did specify that data from a single study were for “widely” separated displays that ranged from 7.5-degrees to 25.7-degrees visual angle. The raw time costs for each of the estimates are relatively close (a 1.2 second estimate when degree of display separation was not specified versus 1.6 seconds for displays ranging from 7.5- to 25.7-degrees visual angle). Also, the percentage time costs are similar (23.6% cost estimate when the degree of display separation was not

specified versus 23.0% cost estimate for displays ranging from 7.5- to 25.7-degrees visual angle). The data need a closer look to determine if they can be synthesized across the two performance estimates into a single estimate.

VAME. Recall that VAME refers to the visual angle spanned by the maximum possible error on a tracking display, or visual angle of maximum excursion. The current estimates for mild and severe VAMEs were based on data from a single study that did not directly manipulate VAME as an experimental objective. To estimate a more accurate value of VAME with greater power, experimental studies that manipulate VAME around the estimated values noted in the library could be conducted. None were found in the literature.

Digital display for tracking tasks. This display property had been identified as a SA limiter but supporting empirical data could not be found in the literature. A literature search for performance data as a function of this display property could support its inclusion within the library.

Predictor displays. In the current library, the absence of a predictor display was identified as a SA limiter. There are, however, a number of ways to present predictor information in the cockpit that might, perhaps, act as SA limiters. For instance, predictor information could be presented for ownship only, or for both ownship and intruder aircrafts, or also at varying levels of prediction (e.g., varied prediction times or space). Literature searches to identify these predictor formats and also for supporting data could be conducted to augment the library.

Spatial displays for temporal estimation tasks. With the transition to NextGen, designing for the display of time constraints in the 4DT environment is an important area for future work. Four-dimensional trajectory operations are anticipated to include time constraints in addition to 3-dimensional space constraints (JPDO, 2010). Only one of the three studies used for this performance estimate was found from the aviation domain. More studies that look at various ways of displaying temporal information and their effects on pilot performance would benefit NextGen display design for 4DT operations.

Location of non-conformal display elements. This display property can be described as fixed text (e.g., non-conformal / non scene-linked) that can be overlaid against an out-the window scene (e.g., HUD) at various locations from the center field view (or from the primary tracking task symbology) to support pilot SA (Dowell, Foyle, Hooey, & Williams, 2002; Foyle, Dowell, & Hooey, 2001). For instance, near-domain display text indicating the altitude of ownship can be displayed at varying distances away from the center of the displayed flight path (Figure 16). This display property can be characterized as an information accessibility factor, as it will take time for the viewer to visually scan, or access, the text that is located in the periphery from the line of sight.

Two studies observed performance accuracy degradation when the non-conformal text was within 8-degrees visual angle from the center of the display (Dowell, Foyle, Hooey, & Williams, 2002; Foyle, Dowell, & Hooey, 2001); however, it is not clear what time cost was imposed to access the information. A performance estimate for this property could be augmented into the Information Accessibility library if performance time data for this display parameter is found in the literature.

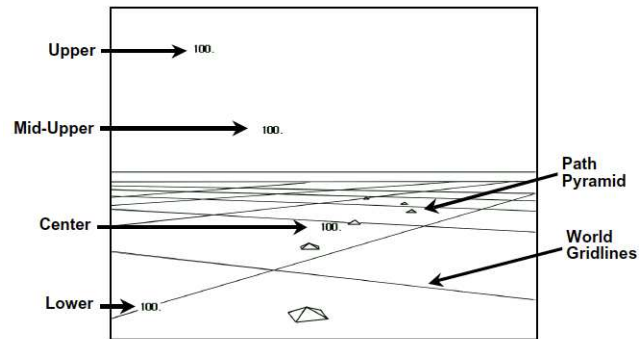


Figure 16. Non-conformal display elements (altitude information) in fixed text format located at varying distances from the center of the flight path. The “Lower” text is located 15.43 degrees from the center; “Center” text is located 0 degrees from the center, “Mid-upper” text is located 7.71 degrees from the center, and “Upper” is located 15.43 degrees from the center. Adapted from “Cognitive tunneling in head-up display (HUD) superimposed symbology: Effects of information location,” by D. C. Foyle, S. R. Dowell, and B. L. Hooey, 2001, *Proceedings of the 11th International Symposium on Aviation Psychology*, Columbus, Ohio: Ohio State University.

Limitations

The method conducted for this thesis did not employ statistical effect sizes to calculate weighted means. Rather, straight averages across multiple data points from different studies were used to calculate quantitative performance estimates. A limitation is that weighted means could have been calculated using a measure such as Hedge’s *g* or Cohen’s *d* to give weight to studies that tested more participants. A benefit of such methods is that they have a structured way for characterizing the statistical power within an individual study from which inferences regarding an effect size can be drawn

(Wickens, Hutchins, Carolan, & Cumming, 2012). Also, an examination of moderator variables may have been possible with these methods. Many studies, however, do not include the necessary statistical data in order to use Hedge's *g* or Cohen's *d* measure.

In this thesis, some of the studies that contributed to a performance estimate may have included enough information to employ one of the techniques for calculating weighted means. But in some cases, only one of a few (or no studies at all) contained enough information. A benefit of the transfer ratio method (or percentage cost/benefit method) used in this thesis allowed the meta-analysis to be more inclusive. It also produced values (in terms of time costs) that are directly interpretable by the designer.

Next Steps

Refining estimates. In this thesis, the term “estimates” is emphasized, partially because the values calculated for the time estimates may be refined as more data become available in the aviation domain.

Augmenting the library. The display formats identified in this thesis may not be the only ones that affect performance and SA in the cockpit. The evolution of technology continues to bring different means of displaying and interacting with information; as examples, there are displays with tactile feedback, swipe functions, display resizing functions with tactile manipulation, cube display concepts, 3D text displays, projection displays, or temporal displays. These display formats have not yet been addressed in the library presented in this thesis. This is in part due to the limited availability of data in the aviation literature, which was uncovered during the literature search process. With the method developed in this thesis, the library can be expanded as more empirical data

become available, particularly for display properties that are relevant to the NextGen environment.

Affects of multiple display properties on SA. The library of estimates aids the designer in estimating the impact of a single display property on performance time and SA. However, it is certainly possible for any given display design to be characterized by multiple display properties; for instance, an integrated hazard display of traffic, weather, and terrain may be characterized as both three dimensional and also as an overlaid display of the three separate information databases. According to the library, this display would then be an SA limiter for spatial location tasks as a function of the angle of 3D perspective viewing, but would also be subject to a time delay for focused tasks as a function of the overlaid nature of the display. With the current library of estimates, it would not be possible to estimate the overall (rather than task-specific) effect on SA. It would also not be possible to rank displays that feature multiple display properties, as does DFSAM (Wickens, 2005), in terms of time costs and/or support for overall SA. This is because many of the performance estimates in the current library are task specific, and also because the interacting effects across multiple display properties may not be understood. This is something that can be considered for future work, and that also has potential to contribute to human performance modeling.

Projection of SA. Finally, the information processing model, which the library framework is based on (Alion, 2011), addresses Endsley's first two levels of SA- perception and comprehension. It is not clear how to incorporate the third level, projection, into the library framework. This is a topic for future work.

Conclusion

A library of performance estimates was developed through a systematic method of literature searches and meta-analyses that summarized display properties relevant to the flight deck environment and that also impede SA. The library includes time costs/benefits in terms of raw time and relative percentage time slowing/gain until SA for a given display property (specifically, for IA and PTC display properties). It also includes display properties that inhibit full SA (specifically, SA limiters). These estimates have implication on design decisions for NextGen cockpit displays.

References

- Ahlstrom, V., & Kudrick, B. (2007). *Human factors criteria for displays: A human factors design standard update of chapter 5* (DOT/FAA/TC-07/11). Washington, DC: FAA.
- Alexander, A. L., Wickens, C. D., & Hardy, T. J. (2003). Synthetic vision systems: The effects of guidance symbology, display size, and field of view. *Human Factors*, *47*, 693.
- Alion Science and Technology (2011). *CSA- Situation accessibility heuristics* (Subcontract #1670-ALI-001). Alion Science and Technology: Boulder, Colorado.
- Amelink, N. H. J., Mulder, M., van Paassen, M. M., & Flach, J. (2005). Theoretical foundations for a total energy-based perspective flight-path display. *International Journal of Aviation Psychology*, *15*, 205-231.
- Arthur, J. J., Prinzel, P. L., Kramer, L. J., Bailey, R. E., & Parrish, R. V. (2003). *CFIT prevention using synthetic vision* (Technical Report 20040086539). Hampton, VA: NASA Langley Research Center.
- Begault, D. (1993). Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigations. *Human Factors*, *35*, 707-717.
- Begault, D. & Pittman, M. T. (1995). 3-D Audio versus head down TCAS displays. *Proceedings of the Eighth International Symposium on Aviation Psychology*, *1*, 116-121.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cavanagh, J. P. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review*, *79*(6) 525-520.
- Beck, M. R., Lohrenz, M. C., & Trafton, J. G. (2010). Measuring search efficiency in complex visual search tasks: Global and local clutter. *Journal of Experimental Psychology: Applied*, *16*, 238-253.
- Boeckman, K. J. (1998). *The resolution and performance effects of three-dimensional display rotation on local guidance and spatial awareness measures*. (Master's thesis). University of Illinois, Urbana Champaign.

- Borst, C., Suijkerbuijk, H. C. H., Mulder, M. & Van Paassen, M. M. V. (2006). Ecological interface design for terrain awareness. *International Journal of Aviation Psychology*, *16*, 375–400.
- Brainard, R. W., Irby, T. S., Fitts, P. M. & Alluisi, E. A. (1962). Some variables influencing the rate of gain of information. *Journal of Experimental Psychology*, *63*, 105-110.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Caroux, L., Bigot, L., & Vibert, N. (2011). Maximizing players' anticipation by applying the proximity-compatibility principle to the design of video games. *Human Factors*, *53*, 103-117.
- Cavanagh, J. P. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review*, *79*(6) 525-520.
- Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of educational research*, *52*, 291-302.
- DoD (Department of Defense) (1999). *MIL-STD-1472F: Human engineering design criteria for military systems, equipment and facilities*. Washington, D.C.: DoD.
- Dowell, S. R., Foyle, D. C., Hooey, B. L. & Williams, J. L. (2002). The effect of visual locating on cognitive tunneling with superimposed HUD symbology. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomic Society*. Santa Monica, CA: HFES.
- Durso, F. T., & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury and S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (137-154). Aldershot, UK: Ashgate Publishing.
- Ellerbroek, J., Visser, M., van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2009). *Towards an ecological four-dimensional self-separation assistance display*. Paper presented at AIAA Guidance, Navigation, and Control Conference, Chicago, Illinois, August 10-13.
- Endsley, M. R. (1994). Situation awareness in dynamic human decision making: Theory. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), *Situational Awareness in Complex Systems* (pp. 27-58). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.
- Endsley, M. R. (1998). *A comparative analysis of SAGAT and SART for evaluation of situation awareness*. Paper presented at the 42nd Annual Meeting of the Human Factors & Ergonomic Society. Chicago, IL.
- Endsley, M. R. (1999). Situation awareness in aviation systems. In: D. J. Garland, J. A. Wise, and V. D. Hopkin. (Eds.), *Handbook of Aviation Human Factors*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley and D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates.
- Endsley, M. R., Bolstad, C. A., Jones, D. G., & Riley, J. M. (2003). *Situation awareness oriented design: From user's cognitive requirements to creating effective supporting technologies*. Paper presented at the Human Factors and Ergonomics 47th Annual Meeting, October 2003, Denver, Colorado.
- Endsley, M. R., Farley, T. C., Jones, W. M., Midkiff, A. H., & Hansman, R. J. (1998). *Situation awareness information requirements for commercial airline pilots (ICAT-98-1)*. Cambridge, MA: Massachusetts Institute of Technology International Center for Air Transportation.
- Fadden, S., Ververs, P. M., & Wickens, C. D. (1998). Costs and benefits of head-up display use: A meta-analytic approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 1998*, 42, 16-2.
- Fennel, K., Sherry, L., Roberts, R. J. Jr., & Feary, M. (2006). Difficult access: The impact of recall steps on Flight Management System errors. *International Journal of Aviation Psychology*, 16, 175-196.
- Fitts, P. M. & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 46(3), 199-210.
- Foyle, D. C., Dowell, S. R. & Hooey, B. L. (2001). *The effect of visual location on cognitive tunneling with superimposed HUD symbology*. In R. S. Jensen, L. Chang, & K. Singleton (Eds.), *Proceedings of the Eleventh International Symposium on Aviation Psychology*, 143, 1-6, Columbus, Ohio: Ohio State University.

- Foyle, D. C., Hooley, B. L., Wilson, J. R., & Johnson, W. A. (2002). HUD symbology for surface operations: Command-guidance vs. situation-guidance formats. *SAE Transactions: Journal of Aerospace*, *111*, 647-658.
- Gempler, K. S., & Wickens, C. D. (1998). *Display of predictor reliability on a cockpit display of traffic information* (ARL-98-6/ROCKWELL-98-1). Savoy, Illinois: University of Illinois Aviation Research Lab.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(19), 3-8.
- Gore, B. F., Hooley, B. L., Wickens, C. D., Sebok, A., Hutchins, S., Salud, E., Small, R., Koenecke, C., & Bzostek, J. (2010). *Identification of pilot performance parameters for human performance models of off-nominal events in the NextGen environment* (NASA/CR—2010-216411). NASA Ames Research Center: Moffett Field, CA.
- Hart, S. G., & Loomis, L. L. (1980). Evaluation of the potential format and content of a cockpit display of traffic information. *Human Factors*, *22*(5), 591-604.
- Helleberg, J., & Wickens, C. D. (2003). Effects of Data-link modality and display redundancy on pilot performance: An attentional perspective. *International Journal of Aviation Psychology*, *13*, 189-210.
- Holland, S. K. (1998). *Conflict detection and maneuver assessment with combined traffic information* (Master's thesis). Retrieved from http://scholarworks.sjsu.edu/etd_theses/1750.
- Hooley et al. (in preparation). A library of performance estimates implemented into MIDAS.
- Hooley, B. L., Gore, B. F., Wickens, C. D., Scott-Nash, S., Socash, C., Salud, E., & Foyle, D. C. (2010). Modeling Pilot Situation Awareness. *Proceeding of the Human Modeling in Assisted Transportation Conference, June 30-July 2, 2010*. Belgirate, Italy: Springer-Verlag Berlin Heidelberg.
- Hooley, B. L., Wickens, C. D., Salud, E., Sebok, A., Hutchins, S., & Gore, B. (2009). Predicting the unpredictable: Estimating human performance parameters for off-nominal events. *Proceedings of the 15th International Symposium on Aviation Psychology*. Dayton, OH: Wright State University.
- Hopcroft, R., Burchat, E., & Vince, J. (2006). *Unmanned aerial vehicles for maritime patrol: Human factors issues* (DSTO-GD-0463). Victoria, Australia: DSTO Defense Science and Technology Organization.

- Jago, S., & Palmer, E. (1982). *Separation monitoring with four types of predictors on a cockpit display of traffic information* (19820026187). Moffett Field, CA.: NASA Ames Research Center.
- Joint Planning and Development Office (JPDO). (2010). *Concept of operations for the next generation air transportation system*. Version 3.2, September 30, 2010. Washington, D.C: FAA.
- Johnson, W. J., Batiste, V., Delzell, S., Holland, S., Belcher, S., & Jordan, K. (2003). *Development and demonstration of a prototype free flight cockpit display of traffic information* (NASA/TM-2003-212273). Moffett Field, CA.: NASA Ames Research Center.
- Kroft, P., & Wickens, C. D. (2001). The display of multiple geographical data base implications of visual attention (ARL-01-2/NASA-01-2). Savoy, Illinois: University of Illinois Aviation Research Lab.
- Lane, D. M., Napier, A., Batsell, R. R., & Naman, J. L. (1993). Predicting the skilled use of hierarchical menus with the Keystroke-Level-Model. *Human Computer Interaction*, 8, 185-192.
- Lavie, T., Meyer, J., Bengler, K. J., & Coughlin, J. F. (2005). Visual displays of route properties in route guidance systems: Effects on driving performance and route selection. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting September 2005*, 49, 1535-1539.
- Law, D. J., Pellegrino, J. W., Mitchell, S. R., Fischer, S. C., McDonald, T. P., & Hunt, E. B. (1993). Perceptual and cognitive factors governing performance in comparative arrival-time judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 19, II8J-II99.
- Loftus, G. R., Dark, V. J., & Williams, D. (1979). Short-term memory factors in ground controller/pilot communication. *Human Factors*, 21, 169-181.
- Lohrenz, M. C. (2003). *Cognitive issues related to advanced cockpit displays: Supporting the transition between internal and external guidance* (Master's thesis). Cambridge, Massachusetts: MIT.
- Marino, C. J., & Mahan, R. R. (2005). Configural displays can improve nutrition-related decisions: an application of the proximity compatibility principle. *Human Factors*, 47, 121-30.

- Martin-Emerson, R., & Wickens, C., D. (1997). Superimposition, symbology, visual attention, and the head-up display. *Human Factors*, 39(4), 581-601.
- McGreevy, M. W., & Ellis, S. R. (1986). The effect of perspective geometry on judged direction in spatial information instruments. *Human Factors*, 28, 439-456
- Merwin, D. H., & Wickens, C. D. (1996). *Evaluation of perspective and coplanar cockpit displays of traffic information to support hazard awareness in free flight* (Technical report: ARL-96-5/NASA-96-1). Savoy, IL: University of Illinois, Aviation Research Lab.
- Morphew, M. E., & Wickens, C. D. (1998). Pilot performance and workload using traffic displays to support free flight. *Proceedings of the Human Factors 42nd Annual Meeting*, pp. 52-56
- Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision*, 11(14), 1-16.
- Neisser, U., Novick, R., & Lazar, R. (1964). Searching for novel targets. *Perceptual and Motor Skills*, 19, 427-432.
- Norman, D. A., & Bobrow, D. J. (1975). On data-limited and resource limited processes. *Cognitive Psychology*, 7, 44-64.
- Nunes, A., Wickens, C., & Yin, S. (2006). Examining the viability of the Neisser Search Model in the flight domain and the benefits of highlighting in visual search. *Proceedings of the 50th Annual Meeting of the Human Factors & Ergonomics Society* (pp. 35-39). San Francisco, CA: Human Factors & Ergonomic Society.
- Olsen, J. R., & Nilsen, E. (1988). Analysis of the cognition involved in spreadsheet software interaction. *Human-computer Interaction*, 3, 309-349.
- Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. S. (1990). Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics*, 48, 214-226.
- Podczerwinski, E. S., Wickens, C. D., & Alexander, A. (2002). Exploring the “Out-of-Sight-Out-of-Mind” phenomenon in dynamic settings across electronic map displays (ARL-01-8/NASA-01-4). Savoy, IL: University of Illinois, Aviation Research Lab.
- Rasmussen, J. & Vicente, K. J. (1989). Coping with human errors through system design: Implications for ecological interface design. *International Journal of Man-Machine Studies*, 31, 517-534.

- Remington, R. W., & Williams, D. (1986). On the selection and evaluation of display symbology: Factors influencing search and identification times. *Human Factors*, 28, 407-420.
- Rehman, A. J., Reynolds, M. C., & Neumeier, M. E. (1995). *A pilot evaluation of text display formats for weather information in the cockpit* (DOT/FAA/CT-TN95/42). Atlantic City, NJ: DOT.
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7(2), 1-22.
- Schons, V., & Wickens, C. D. (1993). *Visual separation and information access in aircraft display layout* (ARL-93-7/NASA-A³I-93-1). Savoy, Illinois: University of Illinois, Aviation Research Lab.
- Stelzer, E. M., Wickens, C. D. (2006). Pilots strategically compensate for display enlargements in surveillance and flight control tasks. *Human Factors*, 48, 166-181.
- Teichner, W. H., & Krebs, M. J. (1974). Laws of visual choice reaction time. *Psychological Review*, 81, 75-98.
- Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions of Systems, Man, and Cybernetics*, 22, 589-606.
- Wickens, C. D. (2005). *Display Formatting and Situation Awareness Model (DFSAM): An Approach to Aviation Display Design* (Technical Report AHFD-05-14/NASA-05-5). Savoy, IL: University of Illinois, Aviation Research Lab.
- Wickens, C. D., Alexander, A., Thomas, L. C., Horrey, W. J., Nunes, A., Hardy, T. J., & Zheng, X. S. (2004). *Traffic and flight guidance depiction on a Synthetic Vision System Display: The effects of clutter on performance and visual attention allocation* (AHFD-04-10/NASA(HPM)-04-1). Savoy, Illinois: University of Illinois, Aviation Research Lab.
- Wickens, C. D., Ambinder, M. S., Alexander, A. L., & Martens, M. (2004). The role of highlighting in visual search through maps. *Proceedings of the 48th Annual Meeting of the Human Factors & Ergonomics Society*. Santa Monica, CA: Human Factors & Ergonomic Society.
- Wickens, C. D., & Carswell, M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37, 473-494.

- Wickens, C. D., Goh, J., Helleberg, J., & Talleur, D. A. (2002). *Modality differences in advanced cockpit displays: Comparing auditory vision and redundancy for navigational communications and traffic awareness* (Technical Report ARL-02-8/NASA-02-6). Savoy, Illinois: Aviation Research Lab Institute of Aviation.
- Wickens, C. D., Hooey, B. L., Gore, B. F., Sebok, A., Koenecke, C., & Salud, E. (2009). Predicting pilot performance in off-nominal conditions: A meta-analysis and model validation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting October 2009* 53: 86-90.
- Wickens, C. D., Hutchins, S., Carolan, T., & Cummings, J. (2013). Effectiveness of part-task training and increasing-difficulty training strategies: A meta-analysis approach. *Human Factors*, 55, 461-470.
- Wickens, C. D., Kroft, P., & Yeh, M. (2000). Database overlay in electronic map design. *Proceedings of the IEA 2000/HFES 2000 Congress* (pp. 3-451-3- 454). Santa Monica, CA: Human Factors & Ergonomics Society.
- Wickens, C. D., & Long, J. (1994). Conformal symbology, attention shifts, and the head-up display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1994*, 38, 6-10
- Wickens C. D., Nunes, A., Alexander, A. L., & Steelman, K. (2005). *3D Navigation and integrated hazard display in advanced avionics: Workload, performance, and situation awareness display clutter issues in SVS design* (AHFD-05-19/NASA-05-7). Savoy, IL: University of Illinois, Aviation Human Factors Division.
- Wickens, C. D., Sandry, D. L. & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25, 227-248.
- Wickens, C. D., Vidulich, M., & Sandry-Garza, D. (1984). Principles of S-C-R compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. *Human Factors*, 26, 533-543
- Wilson, J. R., Hooey, B. L., & Foyle, D. C. (2005). Head-up display symbology for surface operations: Eye tracking analysis of command-guidance vs. situation-guidance formats. *Proceedings of the 13th International Symposium of Aviation Psychology: Oklahoma City, OK*.
- Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, 42, 2985-3004.

- Xu, X., & Wickens, C. D., & Rantanen (2004). Effects of air traffic geometry on pilots' conflict detection with cockpit display of traffic information. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, Santa Monica, CA*.
- Yeh, M., Merlo, J. L., Wickens, C. D., & Brandenburg, D. L. (2003). Head-up versus head-down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors*, 45, 390-407.

Appendix

Reviews of the studies used to calculate performance estimates, relevant experimental conditions, and data are reported in this Appendix.

Information Accessibility Library: Literature Review, Data, and Meta-analysis

Number of keystrokes. An analysis of data from two studies in the HCI domain (Card, Moran, & Newell in Lane, Napier, Batsell, & Naman, 1991; Olsen & Nilsen in Lane, Napier, Batsell, & Naman, 1991) yield a mean keypress time of 1.4 s per keystroke. Card, Moran, and Newell suggested the Keystroke-Level Model (or KLM) as a system designer's quantitative analysis tool that is simple, accurate, and flexible enough for designers to estimate the time for expert users to perform a given task that required keystrokes. KLM estimated a total keystroke time of 1.43 s per keystroke that consisted of 0.08 s for the actual key-press and 1.35 s for mental preparation time. The mental preparation time for a keystroke is not the same as decision-making time, but rather, it is the time to prepare to issue a command once it has been decided it should be executed (Lane, Napier, Batsell, & Naman, 1991).

Olsen & Nilsen (cited in Lane, Napier, Batsell, & Naman, 1991) tested Card, Moran, and Newell's KLM in an experiment that examined performance time of experienced users with a spreadsheet task that required keystroke interaction. They reported a comparable time cost to Card, Moran, and Newell's keystroke estimate, of 1.28 s per keystroke, that consisted of 0.2 s per keystroke and a mental preparation time of 1.08 seconds. These data are summarized in Table A1.

Table A1

Summary of Studies with Data for a Keypress Performance Estimate

Reference	Measure	Performance time	Notes
Card, Moran, & Newell (1980)	Time per keystroke, for 1-2 keystrokes	1.4 s per keystroke	1.43 s total = Actual key-press of 0.08 s + mental preparation time of 1.35 s
Olsen & Nilsen (1988) in Lane, Napier, Batsell, & Naman		1.3 s per keystroke	1.28 s total = Actual key-press of 0.2 s + mental preparation time of 1.08 s
		Mean information accessibility time = 1.4 s	

PTC Library: Literature Reviews, Data, and Meta-analyses

Search clutter. Clutter can be described as “the state in which excess items, or their representations or organizations, lead to degradation of task performance,” (Rosenholtz, Li, & Nakano, 2007), and according to Neider & Zelinsky (2011) clutter has been used as a standard for characterizing visual search efficiency. The literature reveals that an “excess” item, or items, may refer to a number of things, for instance, the set size of possible targets, where reaction time versus set-size functions as a measure of search difficulty has been a focus of visual search research (Beck, Lohrenz, & Trafton, 2010; Nunes, Wickens, & Yin, 2006; Rosenholtz, Li, & Nakano, 2007; Wickens, Nunes, Alexander, & Steelman, 2005; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002). An excess item may also characterize a background image that overlays a set of possible targets (Beck, Lohrenz, & Trafton, 2010; Nunes, Wickens, & Yin, 2006; Rosenholtz, Li,

& Nakano, 2007; Wickens, Nunes, Alexander, & Steelman, 2005; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002). Furthermore, a number of modeling efforts reveal a number of display features that have been proposed to constitute “clutter,” for instance: subjective clutter, edge count, feature congestion (Neider & Zelinsky, 2011), color density, color saliency, distractor similarity, background complexity, color density, (Beck, Lohrenz, & Trafton, 2010); edge density (Mack & Oliva, as cited in Rosenholtz, Li, & Nakano, 2007), local clutter, global clutter (Beck, Lohrenz, & Trafton, 2010), orientation variation, feature density, number of vectors in a display, and ink per unit area (Rosenholtz, Li, & Nakano, 2007 provide a review of clutter measures and models). For the library, performance heuristics for *search clutter* and *background overlay clutter* were examined.

Neisser’s Serial Self Terminating (SST) Search Model offers a model of search time cost as a function of search clutter, where clutter is a function of the number of potential targets displayed in a search task. The SST model proposes that search time (ST) for a target item, included among N potential target items increases linearly as N increases. When the target is present, $ST = a + (bN/2)$, where a is the intercept term that “characterizes the readout of the actual target” (Wickens & Nunes, 2005) (Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002, explain that the intercept for reaction time by set size functions “represent the fixed costs of processes such as those involved in the motor response”); b is the time required to examine each potential target item, and to determine that the item is not the target (Nunes, Wickens, & Yin, 2006) (Wolfe et al., 2002, note that this is the slope that represents the added cost of each additional item); and the

division of N by 2 reflects that when a target is present, it is located on average after searching half the array of items (Nunes, Wickens, & Yin, 2006). When the target is not present, the entire array of search elements must be searched through until it is realized that the target is, in fact, absent (Wickens, Nunes, Alexander, & Steelman, 2005). In this case $ST = a + bN$.

In sum, Neisser's SST model can be used as a performance estimate for the library that estimates the performance time delay as a function of search clutter (that is, potential targets) when the target is present versus when it is absent.

Background clutter complexity.

The cost of mild background overlay clutter. An analysis of data from the clutter literature in the aviation and military domains reveal that performance time increases when displays are formatted with background overlay clutter, compared to displays that are without (Beck, Lohrenz, & Trafton, 2008 [Experiments 1 and 2]; Yeh, Merlo, Wickens, & Brandenburg, 2003 [Experiment 1]). For a target detection task, Yeh, Merlo, Wickens, & Brandenburg compared target detection performance with a "sparse" hand-held display consisting of only a target detection-aid display against a blank background (Figure A1A), compared to an overlaid display format (Figure A1B) with both a target detection-aid display in the near domain and terrain background in the far HUD domain.

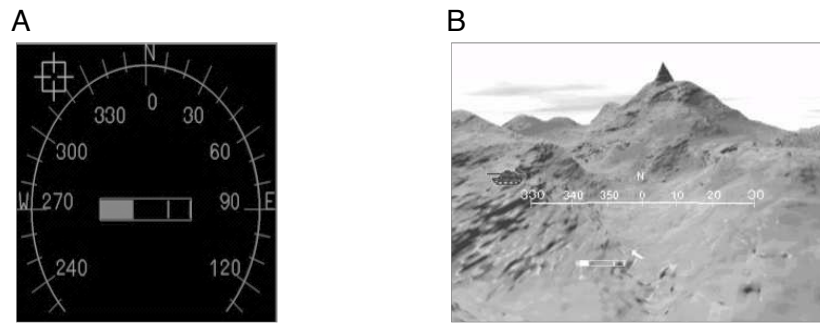


Figure A1. None and relatively mild background clutter; displays used in an experiment by Yeh, Wickens, Merlo, and Brandenburg (2003). A target detection-aid display formatted without background overlay clutter is illustrated in Panel A, and with background overlay clutter in Panel B. Adapted from “Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling,” by M. Yeh, J. L. Merlo, C. D. Wickens, & D. L. Brandenburg, 2003, *Human Factors*, 45(3), 390-407.

Yeh, Merlo, Wickens, & Brandenburg observed that target detection performance was slower when background overlay clutter was present ($M = 9.3$ s) compared to when it was not ($M = 8.0$ s; $p < 0.01$) by ~ 1.3 s (means are estimated from data plotted for the “low salience, un-cued condition” in Yeh, Merlo, Wickens, & Brandenburg, 2003, Figure 2, pp. 397). This 1.3-s raw performance time cost yields a 16.3% relative time cost for the display formatted with background overlay clutter.

The cost of moderate and severe background overlay clutter. Also for a target detection task, Beck, Lohrenz, & Trafton (2010; Experiment 1) measured search times for single-peaked targets among double-peaked distractors, overlaid against six levels of

display clutter consisting of background maps. The 6 levels were made with a combination of three levels of global clutter (low, medium, or high) and two levels of local clutter (low or high). The six combinations of global and local clutter variables (Figure A2 and Figure A3) were rated by the C3 (color-cluster clutter) algorithm (Lohrenz & Dendron, 2008), which computes a quantitative clutter value from 0 to 12 based on a display's color density (or cluster of similar colors) and color saliency. According to the C3 ratings reported in this study, the six displays increased in clutter from the least to the greatest as follows: low global/low local clutter display had a C3 rating of 1.8; low global/high local clutter had a C3 rating of 4.4; medium global/low local clutter display had a C3 rating of 4.9; medium global/high local clutter display had a C3 rating of 7.2; high global/low local clutter display had a C3 rating of 7.3; and finally, the display formatted with high global/high local clutter was rated with the highest degree of clutter with a C3 rating of 9.6. Based on these clutter ratings, the six displays can be characterized here into two categories where the three displays with the lowest C3 ratings are characterized here as relatively mild background overlay clutter (Figure A2), and the three displays with the highest ratings are characterized here as relatively severe background overlay clutter (Figure A3).

A Low global / low local clutter ;
C3 rating= 1.8



B Low global/ high local clutter;
C3 rating= 4.4



C Medium global/Low local clutter;
C3 rating= 4.9



Figure A2. Relatively moderate background overlay clutter; displays used in an experiment by Beck, Lohrenz, and Trafton (2008). Based on the C3 algorithm ratings, and also upon visual observation, background overlay clutter increases from left to right. The outlined icons represent a blow-out of single-peaked targets for a target detection task. Adapted from “Measuring search efficiency in complex visual search tasks: Global and local clutter,” by M. R. Beck, M. C. Lohrenz, & J.G. Trafton, 2010, *Journal of Experimental Psychology: Applied*, 16(3), 238-250.

A Medium global/ high local clutter;
C3 rating= 7.2



B High global / high local clutter;
C3 rating= 7.3



C High global /High local clutter;
C3 rating= 9.6



Figure A3. Relatively severe background overlay clutter; displays used in an experiment by Beck, Lohrenz, and Trafton (2010). Based on the C3 algorithm ratings, and also upon visual observation, background overlay clutter for Panels A, B, and C is greater than the displays in Figure A2. The outlined icons represent a blow-out of single-peaked targets for a target detection task. Adapted from “Measuring search efficiency in complex visual search tasks: Global and local clutter,” by M. R. Beck, M. C. Lohrenz, & J.G. Trafton, 2010, *Journal of Experimental Psychology: Applied*, 16(3), 238-250.

The mean target detection time for the displays formatted with moderate background clutter (Figure A2) is estimated as 9,200 ms (this is an average across data points plotted in Beck, Lohrenz, & Trafton, 2008, Figure 4, pp. 243; target detection times are estimated as ~6,000 ms for low global/low local clutter, ~9,000 ms for low global/high local clutter, and ~12,500 ms for medium global/high local clutter). A mean detection time of 20,200 ms for displays formatted with severe background overlay clutter was considerably higher (this is also an average across data points plotted in Beck, Lohrenz, & Trafton, 2008, Figure 4, pp. 243; target detection are estimated as ~23,000 ms for medium global/high local clutter, ~15,000 ms for high local/medium local clutter, and ~30,000 ms for high global/high local clutter).

Beck, Lohrenz, and Trafton conducted a second experiment for a baseline measure of target detection times in the absence of background overlay clutter using the same single-peak target as in Experiment 1, where a shorter mean target detection time of 2,451 ms was observed compared to performance times with moderately and severely cluttered displays (they did not include an image of the “blank” display in their paper). A comparison among data between Experiments 1 and 2 reveals that, for what was characterized here as moderate background overlay clutter, there was a 6,800 ms delay in target detection performance compared to performance when background overlay clutter was absent-- a relative percentage time cost of 283%. The time cost was exaggerated for what is characterized here as severe background overlay clutter, where there is an estimated 22,600 ms time delay compared to performance when background overlay clutter was absent, resulting in an 842% relative time delay. A summary of these data are below (Table A2).

Table A2

Summary of Studies with Data for a Background Overlay Clutter Performance Estimate

Reference	Measure	Performance time		Raw time cost for overlay clutter	% time cost for overlay clutter	Notes
		Display <i>with</i> overlay clutter	Display <i>without</i> overlay clutter (baseline)			
Yeh, Merlo, Wickens, & Brandenburg (2003; Exp. 1)	Target search time (for relatively mild display clutter)	$M = 9.3$ s	$M = 8.0$ s	1.3 s	16.3%	$p < 0.01$; data estimated from Fig. 2, pp. 397
Beck, Lohrenz, & Trafton (2008; Exps. 1 & 2)	Target detection time (for relatively moderate display clutter)	$M = 9.2$ s (Exp. 1)	$M = 2.4$ s (Exp. 2)	6.8 s	283%	Data for displays <i>with</i> clutter (Exp. 1) were estimated from Fig. 4, pp. 243
Beck, Lohrenz, & Trafton (2008; Exps. 1 & 2)	Target detection time for relatively severe display clutter	$M = 22.6$ s (Exp. 1)	$M = 2.4$ s (Exp. 2)	20.2 s	842%	

Performance estimate notes for display design evaluation within a computational model of situation awareness (CSA; Hooey et al., 2010). In summary, a performance estimate for background overlay clutter can be included in the library, where the presence of an overlaid background prolongs target detection time compared to displays formatted without background overlay clutter. Display formats can be evaluated by the CSA model, along a dichotomy, where the background overlay clutter variable is either present or absent. From the meta-analysis, it is clear that increasing background clutter prolongs performance time. However, as far as quantitative evaluation of time costs for different levels of background overlay clutter for flight deck display formats other than those presented here, it is not clear how the mild, moderate, and severe categories would apply since the performance estimates calculated here are specific to the particular displays used in the three experiments (Beck, Lohrenz, & Trafton, 2010 Exps. 1-2; Yeh, Merlo, Wickens, & Brandenburg, 2003; Exp. 1). It is possible to apply something like the C3 algorithm to quantitatively evaluate background overlay clutter across multiple flight deck displays, so that there is a standard, or common baseline against which multiple displays can be evaluated. Without a common baseline it is possible, however, for the CSA model to make a single evaluation for a selection of competing displays, for their relative support for SA against one another.

Overlaid versus separate databases for focused tasks. An analysis of data from the aviation and military domains reveals that performance time increases with overlaid display formats when used for focused attention tasks (Kroft & Wickens, 2001; Lohrenz, 2003; Wickens et al., 2004; Wickens, Kroft, & Yeh, 2000). Kroft and Wickens

(2001) measured reaction times of student pilots to questions related to one of two overlaid databases on a “small integrated display”; one domain consisted of navigational ground features (e.g., roads, power lines, bridges), and the other consisted of air hazards (a composite display of weather and air traffic) (Figure A4).

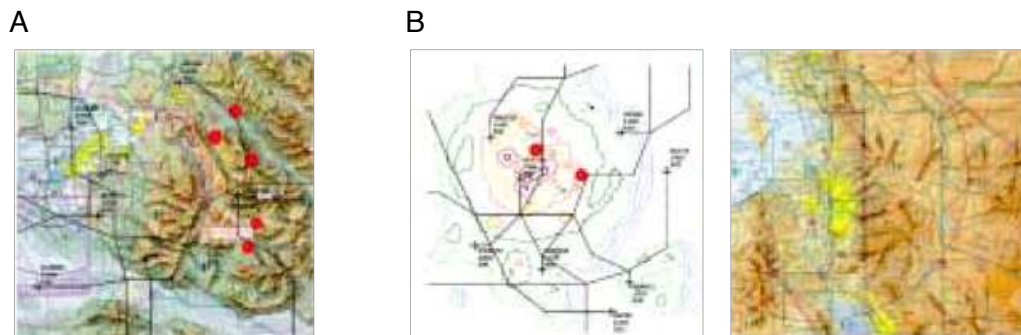


Figure A4. Overlaid and separate display formats; displays used in an experiment by Kroft and Wickens (2001). An overlaid display format of ground features and air hazard databases is illustrated in Panel A, and a separate display format in Panel B. Adapted from *The Display of Multiple Geographical Data Bases: Implications of Visual Attention*, by P. Kroft & C. D. Wickens, 2001, Savoy: University of Illinois, Aviation Research Lab.

Mean reaction performance time for correctly answered focused-attention questions was slower with the overlaid display ($M = 13.3$ s for what Kroft & Wickens referred to as a “small” integrated display) compared to the *separate* display format ($M = 11.3$ seconds; $p < 0.01$) by 2.0 s, yielding a relative percentage time cost of 17.7%. The

authors report that the cost is related to the presence of task-irrelevant information, or clutter.

In a low-fidelity simulation, Lohrenz (2003) also examined the effects of display overlay by measuring target acquisition times by participants with either a pilot's license or flight simulator experience, for targets in a military aviation environment.

Topographic map and flight path information databases were either overlaid or displayed separately (Figure A5).

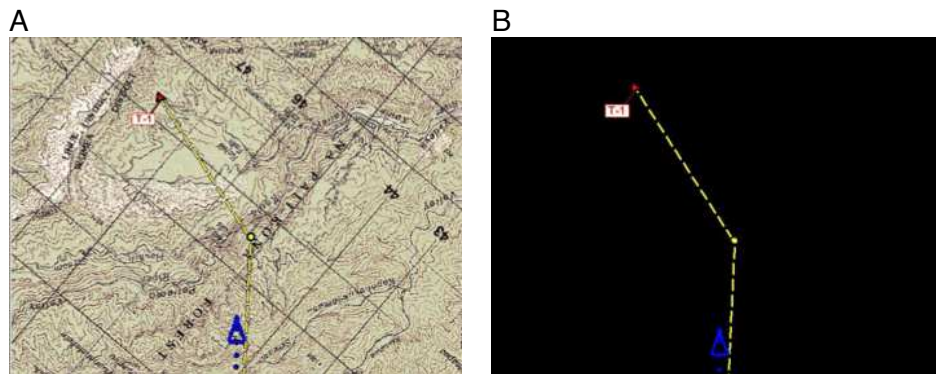


Figure A5. Overlaid and separate display formats; displays used in an experiment by Lohrenz (2003). Ground vehicle targets (not illustrated here) were displayed against an overlaid topographic map with a flight path database illustrated in Panel A, or, targets were displayed against a separate display of flight path information alone as illustrated in Panel B. Adapted from *Cognitive issues related to advanced cockpit displays: Supporting the transition between internal and external guidance*, by M. C. Lohrenz, 2003, MIT, Cambridge, Massachusetts.

Target acquisition times were slower for the display formatted with more overlaid databases (map, flight path, and targets) ($M = 2.1$ s), compared to when targets were displayed against the flight path database information alone ($M = 1.2$ seconds; $p < 0.04$) by ~ 0.9 s, equal to a 75% relative percentage time cost for the display with more overlaid databases. The author reported that the “results underscore the “less is more” philosophy: identify and present only the information required to accomplish the task at hand.”

Also for a focused attention task, Wickens et al. (2004) compared traffic detection performance for 14 instrument-rated pilots who flew a high fidelity simulation with a Synthetic Vision System (SVS) display suite that displayed traffic either overlaid with, or separate from, the instrument panel. The SVS display also varied, with or without the presence of an overlaid tunnel against the traffic and/or instrument panel (Figure A6).

Traffic detection time was slower with the overlaid instrument panel display ($M = 18.8$ s) compared to when it was separate ($M = 12.4$ s; $p < 0.01$) by ~ 6.4 s-- a 62.2% relative percentage time cost of the overlaid versus separate display format. Interestingly, the overlay of the tunnel actually improved target detection times, which the authors attributed to the possibility that the tunnel was easier to integrate with the scene and/or because the cognitive load of flight path tracking was alleviated, hence, lowering overall workload. The instrument panel overlay, however, imposed a time delay on traffic detection compared to when it was separately displayed, whether or not the tunnel was present. A summary of these data are below (Table A3).

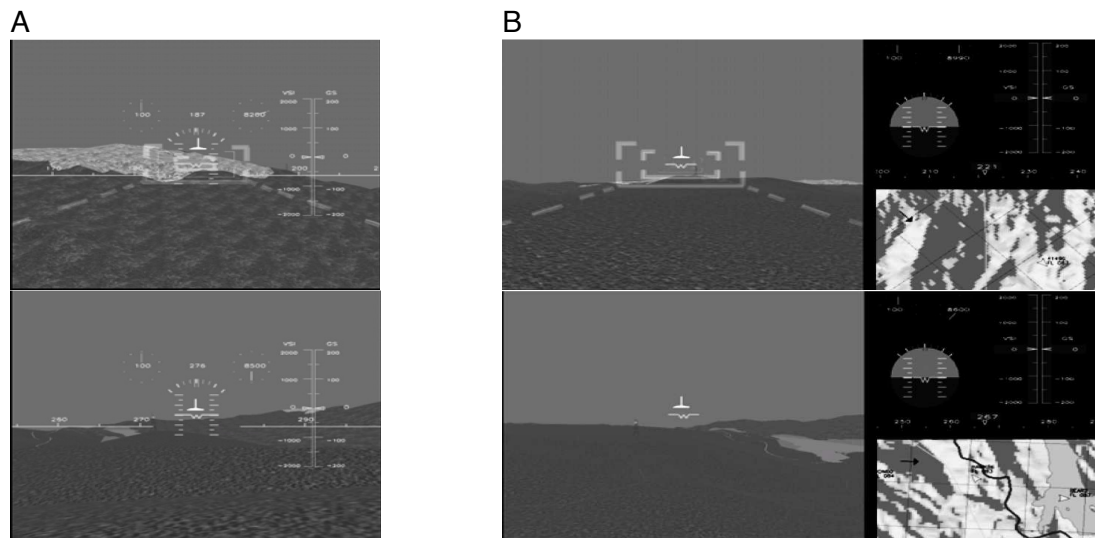


Figure A6. Overlaid and separate display formats; displays used in an experiment by Wickens et al. (2004). The instrument panel is overlaid to the right of the SVS displays in Column A, whereas it is displayed separately in the upper right corners of the displays in Column B. Adapted from *Traffic and flight guidance depiction on a Synthetic Vision System Display: The effects of clutter on performance and visual attention allocation*, by C.D. Wickens et al., 2003, Savoy: University of Illinois, Aviation Research Lab.

Table A3

Summary of Studies with Data for an Overlaid Display Performance Estimate (for Focused Attention Tasks)

Reference	Measure	Performance time		Raw time cost for overlaid databases	% time cost for overlaid databases	Notes
		Overlaid displays	Separate displays			
Kroft & Wickens (2001)	Reaction time to focused attention tasks	$M = 13.3$ s	$M = 11.3$ s	2.0 s	17.7%	$p < 0.01$; data estimated from Fig. 3.3, pp. 25.
Lohrenz (2003; Exp. B)	Time to answer focused attention question	$M = 2.1$ s	$M = 1.2$ s	0.9 s	75%	$p < 0.04$; data estimated from Fig. 4-18, pp. 53.
Wickens et al. (2004)	Detection time for a focused target detection task	$M = 18.8$ s	$M = 12.4$ s	6.4 s	51.6%	$p < 0.01$
Wickens, Kroft, & Yeh (2000); data from 1/5 studies	Time to answer focused attention question	NA (raw data not reported)	NA (raw data not reported)	2.0 s	NA (raw data not reported)	$p < 0.01$
				Mean time delay =2.8 s		

Separate versus overlaid displays for information integration tasks. An analysis of data from the aviation and military domains reveal that performance time is delayed when separate display formats of multiple information databases are used for divided attention tasks (Kroft & Wickens, 2001; Schons & Wickens, 2003). For a mental integration task Kroft and Wickens (2001) measured reaction times of student pilots to questions related to two databases, that were either formatted on a “large” integrated display, or, on separate displays; one domain consisted of navigational ground features (e.g., roads, power lines, bridges), and the other consisted of air hazards (a composite display of weather and air traffic) (Figure A7). Reaction time performance to questions that required mental integration across the two information domains was slower with the separate display format ($M = 11.8$ s) compared to performance with a “large integrated,” or overlaid, display format ($M = 9.8$ s; $p < 0.01$) by ~ 2.0 seconds, yielding a relative percentage time cost of 20.4%.

Also for a mental integration task, Schons and Wickens (1993) measured turn reaction times when a microwave landing system (MLS) indicator was either overlaid against the HUD or separated at varying degrees, where the flight control task required the two information sources to be mentally integrated (Figure A8). Mean turn reaction time was slower when the MLS and HUD were formatted as separate displays ($M = 1.9$ seconds) compared to when the MLS and HUD were “integrated” (overlaid) ($M = 1.5$ sec; $p < 0.04$) by ~ 0.4 seconds, yielding a relative percentage time cost of 26.7%. A summary of these data are in Table A4.



Figure A7. Separate and overlaid display formats; displays used in an experiment by Kroft and Wickens (2001). A separate display format of ground features and air hazard databases is illustrated in Panel A, and an overlaid display format in Panel B. Adapted from P. Kroft & C. D. Wickens, 2001, *Technical Report ARL-01-2/NASA-01-2*, Savoy: University of Illinois, Aviation Research Lab.

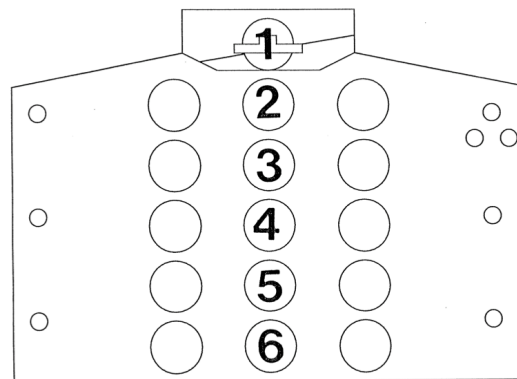


Figure A8. Separate and overlaid display formats; displays used in an experiment by Schons and Wickens (1993). A HUD was either overlaid with a microwave landing system (MLS) indicator (as illustrated in position 1) or, the MLS was separately displayed at varying distances from the HUD (as illustrated in positions 2-6). Adapted from *Visual separation and information access in aircraft display layout*, by V. Schons & C. D. Wickens, 1993, Savoy: University of Illinois, Aviation Research Lab.

Table A4

Summary of Studies with Data for a Separate Databases Performance Estimate (for Divided Attention Tasks)

Reference	Measure	Performance time		Raw time cost for separate databases	% time cost for separate databases	Notes
		Display formatted with separate databases	Display formatted with overlaid databases			
Kroft & Wickens (2001)	Reaction time to divided attention questions	$M = 11.8$ s	$M = 9.8$ s	2.0 s	20.4%	data estimated from Fig. 3.1, pp. 23
Schons & Wickens (1993)	Turn reaction time (a task requiring divided attention between 2 displays)	$M = 1.9$ s	$M = 1.5$ s	0.4 s	26.7%	data estimated from Fig. 16, pp. 40
				Mean time delay = 1.2 s	Mean % time delay = 23.6%	

Wide display separation for information integration tasks. Separate display formats can be characterized by their degree of physical separation. A number of studies have compared performance with overlaid (superimposed) versus separate displays that require mental integration, where significant performance differences were not observed for display formats with relatively narrow separation. For instance, as reported by Schons and Wickens (1993), Martin-Emerson and Wickens did not find performance differences for an integration task between an overlaid display of 0-degrees separation versus a 3.2- and a 6.4-degree display separation. Also, Andre and Cashion (in Schons & Wickens, 1993) found equivalent performance across separation angles from 0- to 8-degrees. Schons and Wickens, however, examined varying degrees of display separation that were greater than those examined by Martin-Emerson and Wickens as well as by Andre and Cashion. Examining display separation angles of 7.5-, 12.3-, 17-, 21.5-, and 25.7-degrees visual angle separation, Schons and Wickens did find performance differences for mental integration tasks. This range of display separations can be characterized as “wide” display separation.

For a mental integration task, Schons and Wickens (1993) had participants perform a dual task of maintaining their commanded course and airspeed with both an overlaid HUD with airspeed display, or, the airspeed display was separated at varying positions from 7.5-, 12.3-, 17-, 21.5-, and 25.7-degrees visual angle separation. Airspeed-change initiation times were measured when the airspeed display was overlaid versus separately displayed. Schons and Wickens observed that the airspeed change initiation performance degraded when separation of the airspeed display increased from

the HUD ($p < 0.074$), where the mean airspeed initiation time across the various degrees of display separation was 8.48 s. Performance across these separate display conditions were significantly higher, compared to when the airspeed display and HUD were overlaid (without extraneous clutter between the displays; $M = 6.84$ s; $p < 0.001$) by ~ 1.6 s. This is a relative percentage time cost of 23%. Schons and Wickens showed that increasing display separation greater than 7.5-degrees to 25.7-degrees visual angle degraded performance compared to performance with an overlaid display format. A summary of these data are below (Table A5).

Table A5

Summary of Studies with Data for a Wide Display Separation Performance Estimate (for Divided Attention Tasks)

Reference	Measure	Performance time		Raw time cost for wide display separation	% time cost for wide display separation	Notes
		Widely separated displays	Overlaid display			
Schons & Wickens (1993)	Reaction time for varying degrees of display separation	$M = 8.48$ s	$M = 6.84$ s	1.6 s	23%	data estimated from Figs. 20-21, pp. 45-46

Text string length. Cavanagh (1979) examined the memory span and memory search literature for memory span and item recognition data across several classes of stimuli. Among a number of stimuli for which data were collected, his findings included a mean processing rate of 33.4 ms per digit-character across eight studies and a mean

processing rate of 40.2 ms per letter-character across thirteen studies. Selected data were from studies that used adult participants and visual stimuli. Items were presented either successively or simultaneously. These data are summarized in Table A6.

Table A6

Summary of Studies with Data for a Text String Length Performance Estimate

Reference	Measure	Processing time*
Cavanagh (1972) (also summarized in Card, Moran, & Newell, 1983)	Mean processing for time per digit across eight studies	$M = 33.4$ ms / digit ----- 34 ms (Bracey, 1969) 27 ms (Burrows & Okada, 1971) 30 ms (Cruse & Clifton, 1971) 38 ms (Sternberg, 1966) 39 ms (Sternberg, 1967) 36 ms (Sternberg, 1969) 28 ms (Theios, Smith, Haviland, & Traupmann, 197) 34 ms (Yio & Santa, 1970)
	Mean processing time per letter across thirteen studies	$M = 40.2$ ms / letter ----- 29 ms (Cavanagh & Chase, 1971) 44 ms (Chase & Calfee, 1969) 53 ms (Chase & Posner, 1965) 43 ms (Cruse & Clifton, 1971) 41 ms (Ellis & Chase, 1971) 24 ms (Egeth & Smith, 1967) 42 ms (Forrin & Morin, 1969) 33 ms (Klatzky & Atkinson, 1971) 44 ms (Klatzky, Juola, & Atkinson, 1971) 65 ms (Nickerson, 1966) 26 ms (Williams, 1971) 38 ms (Wimberly, 1968) 39 ms (Yio & Santa, 1970)
	Mean Processing time per word across four studies	$M = 47.0$ ms / word ----- 52 ms (Burrows & Okada, 1972) 50 ms (Goldring, 1968) 36 ms (Juola & Atkinson, 1971) 50 ms (Smith, 1967)

Note: *Individual and mean processing times were taken from a meta-analysis in Cavanagh (1972).

Target highlighting

A benefit for single target identification. The benefit of salient highlighting for a single target. Data from the following study showed a benefit for salient highlighting for a single target. Also for a target detection task, Beck, Lohrenz, and Trafton (2010; Experiment 3) measured target detection times of 24 undergraduate students for “salient” and “non-salient” targets among salient and non-salient distractors displayed against a map background. However, there was only a single unique target on the display that required identification (a single-peak terrain icon), as opposed to a set of potential targets (Figure A9).



Figure A9. Highlighting intensity; one of the displays used in an experiment by Beck, Lohrenz, and Trafton (2010). A blowout of the single target is depicted in the upper right hand corner. The display is an example of those used for comparing salient versus non-salient single targets; illustrations of salient versus non-salient targets were not included in the paper. Adapted from “Measuring search efficiency in complex visual search tasks: Global and local clutter,” by M. R. Beck, M. C. Lohrenz, & J. G. Trafton, 2010, *Journal of Experimental Psychology: Applied*, 16(3), 238-250.

The authors reported that:

Saliency was calculated as the difference between the color of the target symbol and the background (predominant) color of the chart, according to several color difference formulas, including dLab (the Euclidean distance between colors in CIE L*a*b* space), de2000 (the CIE de2000 color difference formula: CIE, 2000), and dHSV (difference in Hue, Saturation, and Value). On average, a non-salient target's color was different from the background color by 15 (SD = .04) dLab units, .22 (SD = .05) de2000 units, and .50 (SD = .08) dHSV units (all normalized from 0–1). A salient target's color was different from the background by .61 (SD = .07) dLab, 0.55 (SD = 0.08) de2000, and 0.88 (SD = .08) dHSV units, normalized.

Beck, Lohrenz, & Trafton found that when salient targets (relatively highlighted targets) were displayed with non-salient distractors, the salient, single, targets were detected faster ($M = 3,500$ ms) compared to when non-salient, single, targets (relatively low-lighted targets) were displayed with salient distractors ($M = 22,000$ ms; $p < 0.01$). This is a detection time benefit of 18,500 ms for the salient targets-- a relative percentage cost of 84%. This is summarized below in Table A7. The data from this study shows a greater benefit when a single target is made salient through highlighting, compared to data where entire target sets were made salient through highlighting (Podczerwinski & Wickens, 2002; Wickens, Ambinder, Alexander, & Martens, 2004) as described in the following section.

Table A7

Summary of Studies with Data for a Salient Highlighting Performance Estimate for a Single Target

Reference	Measure	Performance time		Raw time benefit for a salient single target	% time benefit for a salient single target	Notes
		Salient single target	Non-salient single target			
Beck, Lohrenz, & Trafton (2010; Exp. 3)	Target detection time for a single, salient target against a map background, and similar distractors	$M = 3.5$ s	$M = 22.0$ s	18.5 s	84%	$p < 0.01$; estimated from Fig. 8, pp. 247

A benefit of highlighting a target set. Wickens, Ambinder, Alexander, & Martens (2004) measured response times of 24 undergraduate students to focused attention questions regarding targets in a destination-information domain. The destination domain was presented at a constant intensity (“level 3”) while the vehicle domain varied from intensity “level 1” to “level 4” (the four intensity levels were reported as: 3.28×10^{-2} fL for intensity level 1, 8.09×10^{-1} fL for intensity level 2, 1.74 fL for intensity level 3, and 3.89 fL for intensity level 4). The intensity variation of the distractor vehicle-information domain made it so the target destination-information domain (which remained at level 3 intensity) was displayed either at uniform intensity, was relatively highlighted (when the vehicle domain intensity was at levels 1 and 2), or, was relatively low-lighted (when the vehicle domain was displayed with intensity level 4), compared to the non-target vehicle-information domain (see Figure A10).

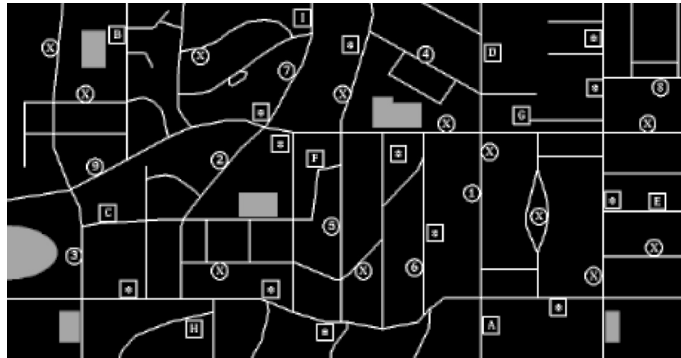


Figure A10. Highlighting intensity; display used in an experiment by Wickens, Ambinder, Alexander, and Martens (2004). A map of the two information domains is illustrated where the vehicle domain is represented by numbers in circles, the destination domain is represented by letters in squares, and distractors are represented by an “X” or “*” in squares. This illustration is a negative image of what was presented to participants, which displays both vehicles and destinations at uniform intensity (level 3). Adapted from “The role of highlighting in visual search through maps,” by C. D. Wickens, M. S. Ambinder, A. L. Alexander, & M. Martens 2004, *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, Santa Monica: HFES.

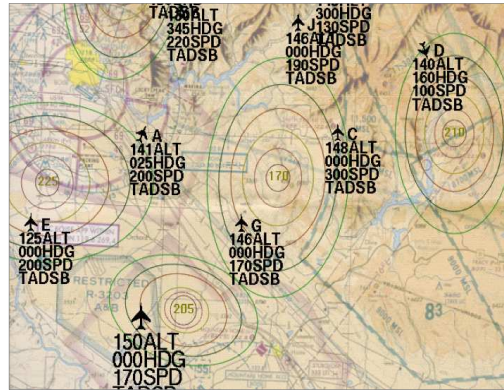
Response times for answering questions related to the destination targets were slower when they were non-salient ($M = 18.3$ s for non-salient targets; $M = 18.0$ s for targets of uniform intensity), compared to when they were salient ($M = 16.25$ s for targets that were relatively highlighted; $p = 0.03$). A comparison between the salient and both non-salient target domains reveal a performance time benefit of 1.9 s (18.2–16.3) when the target domain was highlighted-- a relative benefit of 10.4%. The authors noted that

discriminability rather than salience may be a factor of target detection against the non-target domain.

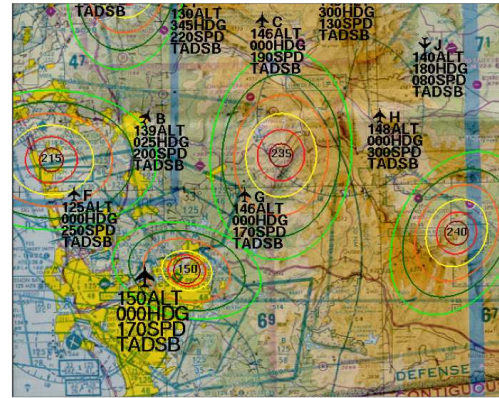
Also for a target detection task among a set of potential targets, Podczerwinski and Wickens (2002) measured change detection times of student pilots to traffic and weather changes; the two databases were overlaid on a map display where the target and non-target domains were of uniform intensity, or, the target domain was relatively low-lighted, or, the target domain was relatively highlighted (Figure A11).

The physical units of intensity were not reported, that is, the distinction between "highlighted," "uniform intensity," and "low-lighted" targets in this study is qualitative. Podczerwinski and Wickens found that change detection times for both weather and traffic were faster when the target information domains were salient ($M = 12.5$ s across highlighted traffic and weather targets) than when they were non-salient ($M = 14.0$ s for uniform intensity targets, and $M = 18.5$ s for low-lighted targets $p < 0.01$). A comparison of detection times for the salient ($M = 12.5$ s) and non-salient targets ($M = 16.3$ s) reveals a 3.8 second benefit for the salient, highlighted targets-- a relative benefit of 23.3%. It should be noted that the term "salience" used in this meta-analysis refers to the attention capturing properties of a relatively highlighted information domain, whereas, in the study by Podczerwinski and Wickens, the term "salience" refers to the spatial quality of an information domain (versus a non-salient non-spatial digital tag).

A



B



C

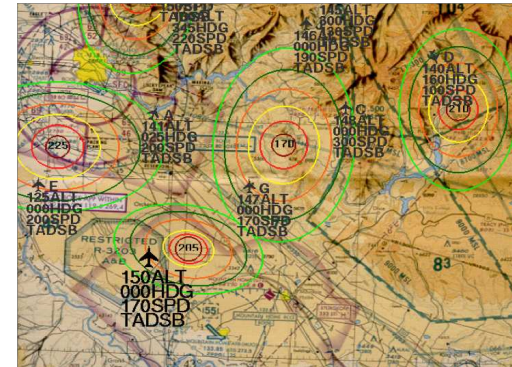


Figure A11. Highlighting intensity; displays used in an experiment by Podczerwinski and Wickens (2002). The traffic domain is highlighted relative to the weather domain and background in Panel A, is of uniform intensity in Panel B, and is relatively low-lighted in Panel C. Adapted from *Exploring the “Out-of-Sight-Out-of-Mind” phenomenon in dynamic settings across electronic map displays*, by E. S. Podczerwinski & C. D. Wickens, 2002, Savoy: University of Illinois, Aviation Research Lab.

In another study Nunes, Wickens, and Yin (2006; Experiment 3) measured target detection times in a search task using a terrain display that highlighted either four of eight potential targets versus a display without highlighted targets. The targets were either an aircraft icon or altitude text from which participants were asked to detect altitude information. The terrain display used in the study is illustrated below (Figure A12), but without the aircraft or text targets (an illustration of these were not included in the paper).



Figure A12. Background display against highlighted or non-highlighted targets used in an experiment by Nunes, Wickens, and Yin (2006, Experiment 3). Highlighted targets for the text and aircraft icon conditions are not pictured. Adapted from “Examining the viability of the Neisser Search Model in the flight domain and the benefits of highlighting in visual search,” by A. Nunes, C. D. Wickens, & S. Yin, 2006, *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomic Society Meeting*, San Francisco, CA: HFES.

Nunes, Wickens, and Yin (2006) found that target detection times for both text and aircraft icon were faster when the target information domains were made salient by highlighting ($M = 3.6$ s across highlighted text and aircraft icon targets) than when they were non-salient ($M = 4.4$ s for uniform intensity targets across both text and aircraft icon conditions). A comparison of detection times for the salient and non-salient targets reveals a 0.9 second benefit for the salient, highlighted targets-- a relative benefit of 20.1%. Collectively, the data from these studies showed a mean raw time benefit of 2.2 s and a mean percentage benefit of 17.9% for a target set made salient through highlighting compared to a non-salient target set, as summarized in Table A8.

Table A8

Summary of Studies with Data for a Salient Highlighting Performance Estimate for a Target Set

Reference	Measure	Performance time			Raw time benefit for a salient target sets	% time benefit for a salient target set	Notes
		Salient target set (high-lighted)	Non-salient target set (uniform intensity)	Non-salient target set (low-lighted)			
Wickens, Ambinder, Alexander, & Martens (2004)	Target detection time within a target domain (destination information domain)	$M = 16.3$ s	$M = 18.0$ s	$M = 18.3$ s	1.9 s	10.4%	$p = 0.03$; data estimated from Fig. 3a, pp. 4
			$M = 18.2$ s				
Podczerwinski & Wickens (2002)	Target detection time within a target information domain (traffic information)	$M = 12.5$ s	$M = 14.0$ s	$M = 18.6$ s	3.8 s	23.3%	$p < 0.01$; data estimated from Fig. 14, pp. 19
			$M = 16.2$ s				
Nunes, Wickens, and Yin (2006)	Target detection times for both text and aircraft icon	$M = 3.6$ s	$M = 4.4$ s across both highlighted conditions (either highlighted text or aircraft icon targets)		0.9 s	20.1%	$p < 0.01$; data estimated from Fig. 5, pp. 38
					Mean time benefit = 2.2 s	Mean % time benefit = 17.9 s	

SA Limiter Library: Literatures Reviews, Data, and Meta-analyses

3D angle of perspective viewing. A benefit of perspective viewing is that all three spatial axes can be viewed in an integrated format; however, this comes at the cost of accurate spatial judgments. An example where this is a performance factor on the flight deck is when the pilot must make spatial judgments of hazard traffic on a 3D CDTI. These cost and benefit tradeoffs can be displayed with a comparison between a 2D CDTI void of a perspective view, and a 3D CDTI with perspective viewing, both displayed in Figure A13.

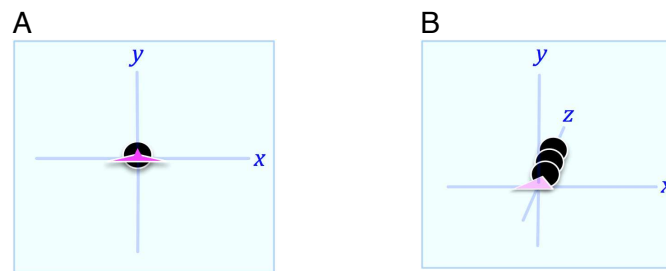


Figure A13. Comparison of displays without (A) and with (B) a perspective view.

In the 2D display illustrated in Panel A, there are three separate targets aligned through the z -axis, at the same altitude as ownship. However, the targets appear as a single dot, since the nearest target occludes the visibility of those directly behind it. When perspective viewing is allowed through increasing the angle of elevation and/or azimuth, the three targets are revealed as in Panel B, where there is some resolution, or decompression, of the z -axis relative to Panel A. Of course, the resolution depends on how much of a vantage point is displayed. However, as resolution increases along the z -axis, there is a tradeoff of decreased resolution along the axis along which perspective

viewing is increased. That is, altitude spatial judgments along the y -axis will degrade with increasing angles of elevation and lateral spatial judgments along the x -axis will degrade with increasing angles of azimuth (as illustrated in Figure A14). This is referred to as 3D ambiguity (Merwin & Wickens, 1998; Wickens & Alexander 2004).

Data from two studies (Boeckman, 1996; Merwin & Wickens, 1996) show empirical evidence for characterizing increasing angles of perspective viewing from 15-75 degrees as data limiters, where the range of data have been shown to limit accurate spatial judgments. Merwin & Wickens (1996) compared conflict detection performance with 3D perspective displays of 60-degree angle of elevation, a 3D display of 30-degree angle of perspective viewing, and a 2D coplanar traffic display void of a perspective view. The displays with perspective view are illustrated in Figure A14.

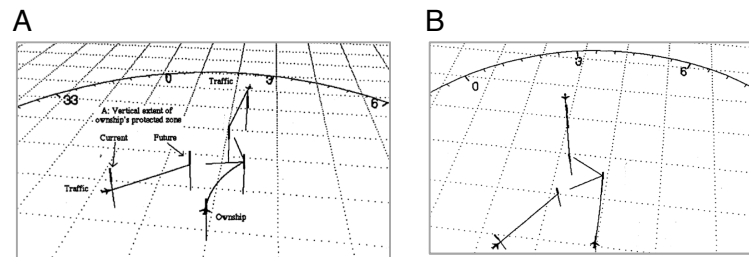


Figure A14. 3D displays with 30- and 60-degree angle of perspective viewing (Panels A and B, respectively); displays used in an experiment by Merwin and Wickens (1996).

Adapted from *Evaluation of perspective and coplanar cockpit displays of traffic information to support hazard awareness in free fight* by D. H. Merwin and C. D.

Wickens, 1996, Savoy: University of Illinois, Aviation Research Lab.

Mean conflict detection rates for the two perspective displays showed lower mean performance accuracy for conflict detection with the 60-degree display angle of perspective viewing ($M = 92.5\%$) compared to the 30-degree display angle of perspective viewing ($M = 93.0\%$); the coplanar display had the highest accuracy rate ($p = 0.034$). Decreased conflict detection accuracy with an increasing angle of perspective view was reported as a likely result of 3D display ambiguity, where the authors reported that, “the ambiguity of the perspective displays likely impaired the accurate judgment of traffic with respect to ownship.”

Boeckman (1996) found a similar trend in results where performance decreased as display angle of elevation increased across five elevation angles for a CDTI: 15, 35, 45, 55, and 75 degree angles of elevation (displays used were not illustrated in this study). Participants were asked to perform spatial judgment tasks that included estimating the angle of elevation, the angle of azimuth, and the distance of intruder aircraft. Reduced vertical resolution caused by increasing the displayed elevation angles reduced vertical judgment performance; mean vertical judgment error (MVJE) increased from 11.6% with a 15-degree angle, to 11.8% with a 35-degree angle, 12.5% with a 45-degree angle, 13.5% with a 55-degree angle, to 17% with a 75-degree angle of elevation ($p < 0.001$; these values were estimated from plotted data in Boeckman, 1996, Figure 3.5, page 76). Boeckman concluded that the reduced resolution with increasing angles of elevation reduced performance for vertical judgments in a non-linear fashion. Boeckman also observed a non-linear effect of increasing vertical RMSE as the angle of perspective viewing increased ($p < 0.001$); RMSE increased from 275 ft for a 15-degree angle of

perspective viewing, 400 ft for 35-degrees, 475 ft for 45-degrees, 600 ft for 55-degrees, to 800 ft for 75-degrees vertical angle of perspective viewing. These data are summarized in Table A9 below.

Table A9

Summary of Studies with Data for a 3D Angle of Perspective Viewing Performance

Estimate

Reference	Measure	Performance accuracy						
		15	30	35	45	55	60	75
Merwin & Wickens (1996)	Conflict detection accuracy	-	96.0%*	-	-	-	92.5%*	-
Boeckman (1998)	Vertical judgment error	11.6 %	-	11.8%	12.5%	13.5%	-	17.0%
	Vertical RMSE	275 ft	-	400 ft	475 ft	600 ft	-	800 ft

Note: * $p = 0.034$; estimated from Merwin & Wickens (1996), Fig. 19 pp. 36.

From the collective data from both studies, a challenge is to estimate at what point or points performance severely degrades. This would be straightforward if the same variable levels were used across studies. Given the available data, however, and the non-linear effect observed by Boeckman, three categories of none, mild, and severe can be estimated as <30, 30-60, and >60-90 degrees, respectively. If a linear effect were observed, the midpoint of the range of angles (45-degrees) could instead be estimated as a threshold between mild and severe categories for the library.

Display illegibility (height of text and symbols). Military Standards (MIL STD) 1472F (DoD, 1999), offers the following recommendation for the visual angle heights at which alphanumeric and non-alphanumeric symbols should be displayed:

Visual displays.

5.2.1.6.4.1 Character height. As measured from the greatest anticipated viewing distance, the visual angle subtended by height of black-and-white characters should be not less than 4.6 mrad (16 min) with 5.8 mrad (20 min) preferred; the visual angle subtended by height of colored characters should be not less than 6.1 mrad (21 min) with 8.7 mrad (30 min) preferred.

Similarly, the FAA Human Factors Handbook DOT/FAA/TC-07/11 (Ahlstrom & Kudrick, 2007) recommends a similar recommendation of at least 16 min of arc for information critical to a task or when readability is important. These legibility standards (character height) are summarized in Table A10 below (values were translated into degrees of visual angle).

Table A10

Summary of Legibility Recommendations for a Legibility Performance Estimate

References	Recommended least height	Recommended preferred height
MIL-STD 1472F, Paragraph 5.2.1.6.4.1: Visual display character height	16 min. of arc = 0.3° visual angle	30 min. of arc = 0.5° visual angle
FAA Human Factors Handbook DOT/FAA/TC-07/11 (2007), Paragraph 5.1.8.10: Alphanumeric character and symbol size	16 min. of arc = 0.3° visual angle	

2D tracking display size: Visual angle of maximum excursion. Recall that VAME is an absolute measuring scale for tracking display-size (C. D. Wickens, personal communication, April 2011) that takes into account varying geometric fields of view and physical size (e.g., that is, the maximum tracking error subtended in degrees of visual angle-- MaxTE) of a tracking display. A goal in developing this performance estimate was to find the visual angle of maximum excursion (VAME) threshold where performance significantly degrades. To estimate a display size performance estimate, VAMEs were calculated from three studies in the aviation literature. In sum, for two of the three studies, substantial effects of increased tracking error with decreasing display size (that is, decreasing VAMEs) were not observed (Alexander, Wickens, & Hardy, 2005; Arthur, Prinzl, Kramer, Bailey, & Parrish, 2003). However, in two experiments (Stelzer & Wickens, 2006) effects of increased tracking error with decreasing display size (VAME) were observed. It is suspected that the particular display sizes that did not yield performance decrements with decreasing display size were not within the display size

threshold where tracking performance begins to degrade. All three studies are further detailed and synthesized below.

Alexander, Wickens, and Hardy (2003; Experiment 1) crossed two display sizes ('small' and 'large') with 2 geometric field of views (GFOV) (30 and 60-degrees visual angle of the world, or, VAW), where the tracking error displays appeared to span the entire visual angle of the display (VAD), or width of the displays. A display similar to the four displays used in the experiment is illustrated below in Figure A15.

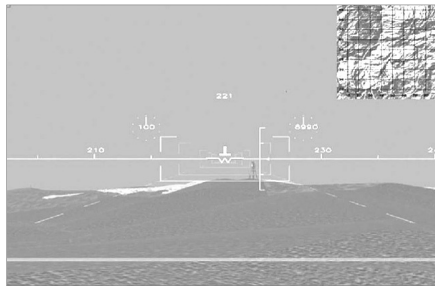


Figure A15. Similar display with VAME from experiment by Alexander, Wickens, and Hardy (2003).

The VAD:VAW ratio for all four displays were reported, from which the VADs , MaxTEs, and VAMEs can be calculated (calculated VAMEs are listed in bold in Figure A11). VAMEs can be calculated by multiplying VAD/VAW by the absolute value of the maximum possible tracking error. That is:

$$\text{VAME} = \text{VAD/VAW} \times |\text{maximum possible tracking error from 0}| .$$

For instance, for the large display with 30-degrees GFOV, VAD/VAW was reported as 1, from which the VAME is easily estimated as half the GFOV (VAW), or 15-degrees. This 15-degrees represents the absolute value of the tracking error display, also taking into consideration GFOV. For the small display with 30-degrees GFOV, VAD/VAW was reported as 0.77, so, VAD is estimated as 23.1-degrees (0.77 x 30-degrees), MaxTE is 11.5-degrees (23.1 /2), and VAME calculated as 8.9-degrees visual angle (.77 x 11.5-deg). Data for the four displays are summarized in the Table A11. Lateral and vertical tracking errors are also reported in Table A11, where for both measures, the data are contrary to what would be expected. That is, error actually decreased as display size (or, VAME) decreased. Alexander and Wickens did not report results as a function of VAME, but they did report that the differences due to display size and GFOV were so small that they may not be of practical significance.

Table A11

Parameters for Calculating VAME and Performance for a study by Alexander, Wickens, and Hardy (2003)

Display	VAD:VAW (reported)	MaxTE	VAME	Lateral RMSE	Vertical RMSE
Large, 30-deg GFOV	1-deg	15-deg	15-deg	10m	5m
Small, 30-deg GFOV	0.77-deg	11.5-deg	8.9-deg	9m	4.7m
Large, 60-deg GFOV	0.5-deg	15-deg	7.5-deg	7m	4.6m
Small, 60-deg GFOV	0.4-deg	12-deg	4.8-deg	7m	4.4m

In another study, Arthur, Prinzel, Kramer, Bailey, & Parrish (2003) examined the effects of various SVS terrain displays, while also comparing tracking error as a function of four PFD display sizes: a baseline EFIS 757 display, a size “X” 8”x10” display, size “A” 5”x5.25” display, and a HUD. Size X and A displays are depicted in Figure A16. Display parameters from which VAMEs were calculated were reported in the study for size “X” and “A” displays, and are listed below in Table A12 along with images of the displays used in the study.

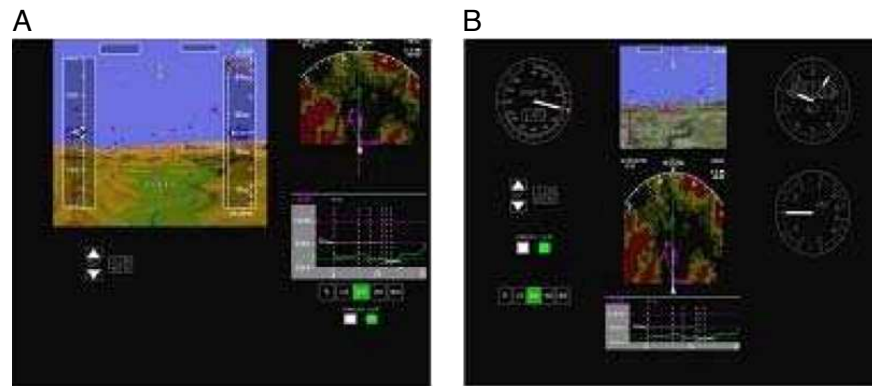


Figure A16. Display sizes “X” and “A” used in an experiment by Arthur et al. (2003). Display size “X” is depicted in the top left screen of Panel A. Display size “A” is depicted in the top right middle screen in Panel B. Adapted from *CFIT Prevention Using Synthetic Vision*, by J. J. Arthur, P. L. Prinzel, L. J. Kramer, R.E. Bailey, & R.V. Parrish, 2003, Langley Research Center: Hampton, VA.

Table A12

Parameters for calculating VAME and Performance for a study by Arthur, Prinzel, Kramer, Bailey, and Parrish (2003)

Display	VAD:VAW	MaxTE	VAME	Lateral RMSE	Vertical RMSE
Display X, 60-deg GFOV	23:60= 0.38	11.5-deg	4.37-deg	80ft	63ft
Display A, 60-deg GFOV	12:60= 0.2	6-deg	1.2-deg	82ft	69ft

Again, VAMEs can be calculated by multiplying VAD/VAW by the absolute value of the maximum possible tracking error. For the size “X” 8-inch display, a minification factor of 2.6 was reported. Dividing the reported 60-degree VAW by the reported minification factor reveals a 23-deg VAD. Then, the VAD:VAW ratio of 23:60 can be calculated as 0.38, and MaxTE is estimated as 23-deg/2, or 11.5-deg. VAME is then calculated as $(VAD/VAW) \times MaxTE = 0.38 \times 11.5 = 4.4$. The same procedure for the size “A” display reveals a VAME of 1.2, which was calculated from the reported 5.0 minification factor for a 60-degree FOV (or 60-deg VAW). These calculated VAMEs assume that the tracking error display spans the entire width of the PFD. Lateral and vertical tracking error are also reported in Table A12, where for both measures, the trend of the raw data differs from the previous study. Here, error actually increases as display size, or VAME, decreases. In addition, the VAMEs used in this study of 4.4 and 1.2 are much smaller than the VAMEs calculated for the displays used by Alexander, Wickens, and Hardy. The data trends indicate that perhaps the smaller VAMEs used in the study by Arthur et al. are closer to the “size,” or VAME, threshold where tracking performance begins to suffer (error increases). The next two studies provide more context.

In another study that examined flight control, surveillance, and target search as a function of display size, Stelzer and Wickens (2006; Experiment 1) examined the effects of display size with a 2D tracking display, and compression within a 3D perspective display (Figure A17). First, the 2D display consisted of a “large” vertical tracking display of 4-degrees visual angle, and another 2D display consisted of a “small” horizontal tracking display of 2-degrees visual angle, so that the small display was half the physical size of the large display. Next, a 3D perspective display consisted of a 4-degree horizontal axis, and a 2-degree relatively compressed longitudinal axis so that the degree of compression between the horizontal and longitudinal axis within the 3D perspective display was the same amount of compression between the large and small 2D displays.

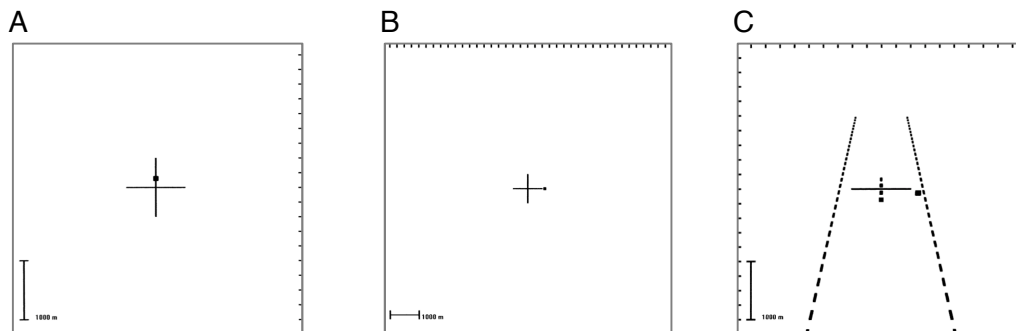


Figure A17. 2D large, 2D small, 3D non-compressed horizontal, and 3D compressed longitudinal tracking displays used in an experiment by Stelzer and Wickens (2003, Experiment 1). The large vertical tracking display in Panel A is the same size as the 3D horizontal display in Panel C. The small horizontal tracking display in Panel B is the same size as the 3D longitudinal compressed display in Panel C. Adapted from “Pilots strategically compensate for display enlargements in surveillance and flight control tasks,” by E. M. Stelzer and C. D. Wickens, 2006, *Human Factors*, 48, 166.

In essence, the 2D large tracking display had the same VAME as the 3D horizontal tracking display (assuming a 1:1 VAD to VAW ratio, VAME = 1-degree for both displays), and the 2D small display had an equivalent VAME as the 3D compressed display (again, assuming a 1:1 VAD to VAW ratio, VAME = 2-deg for both displays). Display parameters for calculating VAMEs (VAD:VAW and MaxTE) are listed in the bold in Table A13. In essence, the 2D large tracking display had the same VAME as the 3D horizontal tracking display (assuming a 1:1 VAD to VAW ratio, VAME = 1-degree for both displays), and the 2D small display had an equivalent VAME as the 3D compressed display (again, assuming a 1:1 VAD to VAW ratio, VAME = 2-deg for both displays). Display parameters for calculating VAMEs (VAD:VAW and MaxTE) are listed in Table A13.

Table A13

Parameters for Calculating VAME and Performance for a study by Stelzer and Wickens (2006; Experiment 1)

Display	VAD:VAW	MaxTE	VAME	Tracking RMSE
Large 2D vertical tracking display	1-deg	2-deg	2-deg	<i>M</i> = 235
3D perspective (large) horizontal tracking display	1-deg	2-deg	2-deg	
Small 2D horizontal tracking display	1-deg	1-deg	1-deg	<i>M</i> = 274
3D perspective (small) longitudinal tracking display	1-deg	1-deg	1-deg	

In another study that also examined flight control, surveillance, and target search as a function of display size, Stelzer and Wickens (2006; Experiment 2) examined the effects of display size with a 2D attitude direction indicator (ADI) display (Figure A18). The VAME for the ADI within a greater integrated hazard display was varied as either .63- or .30-degrees visual angle (Wickens, personal communication, May 17, 2013).

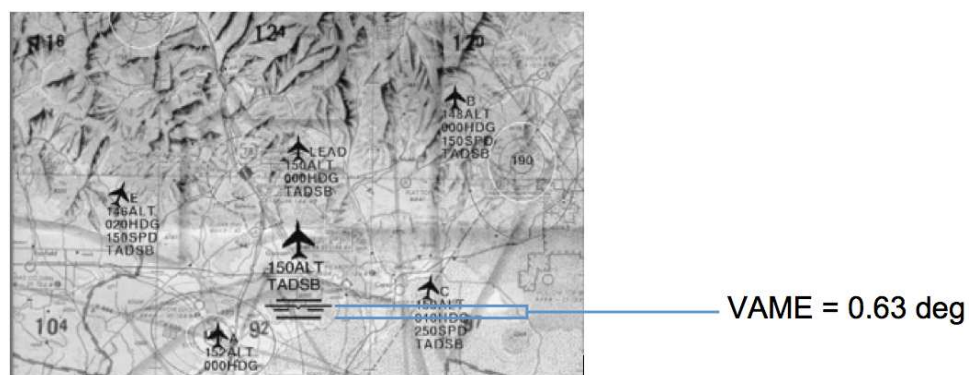


Figure A18. Tracking display (attitude direction indicator) within the greater cockpit display used in an experiment by Stelzer and Wickens (2003, Experiment 2). VAME for the ADI was varied at .63 versus .30. Adapted from “Pilots strategically compensate for display enlargements in surveillance and flight control tasks,” by E. M. Stelzer & C. D. Wickens, 2006, *Human Factors*, 48, 166-181.

Tracking error was observed to double from 315 ft to 615 ft. This large increase in error depicts that there is certainly an increase in error with the decrease in VAME from .63- to .30 VAME as summarized in Table A14.

Table A14

VAMES and Performance from a study by Stelzer and Wickens (2006; Experiment 2)

Display	VAME	Tracking RMSE
Attitude direction indicator within a greater integrated hazard	.63-deg	<i>M</i> = 315 RMSE
	.30-deg	<i>M</i> = 615 RMSE

Data analysis. Looking at the raw data collectively across the four studies (Alexander, Wickens, & Hardy, 2003, Experiment 1; Arthur, Prinzel, Kramer, Bailey, & Parrish, 2003; Stelzer & Wickens, 2006), VAMEs of the largest tracking display used in the Alexander and Wickens study of 15-, 8.9-, 7.5-, and 4.8- degrees showed that tracking error actually decreased with decreasing VAME, however, the authors suggested that the error differences were so small that they were not of practical significance. The smaller display VAMEs from the Arthur et al. study of 4.4- and 1.2-degrees showed a trend of results that were opposite of the Alexander and Wickens' study. Tracking error actually increased as VAME decreased. Similar to the Alexander and Wickens study, Arthur et al. reported that the differences in tracking error for the displays used in this experiment were not operationally significant. Stelzer and Wickens looked at displays with even smaller VAMEs, 2- and 1-degree VAME, and their results showed a significant increase in mean error from 235 ft to 274 ft RMSE. This difference in error however, is relatively small to that found in Experiment 2, where a decrease in VAME from .63 to .30 showed a larger increase in mean tracking error (double the error) from 315 to 615 ft RMSE.

From the collective raw data across the three studies, it can be *estimated* that the VAMEs from the latter 3 studies are within the threshold where tracking performance

begins to increase with decreasing display VAME. Not enough data in the literature however, has been found to examine VAMEs within a more detailed range so the exact threshold where performance begins to degrade cannot be identified here. However, as an estimate, we can take the VAME of 0.63 used by Stelzer and Wickens in Experiment 2 as the threshold from which tracking error begins to increase.

Auditory displays for spatial localization tasks. Wickens, Sandry, and Vidulich (1983) examined the optimal assignment of modalities for both a spatial target acquisition and verbal memory task when performed concurrently with a manual flight-path tracking task. For the verbal memory task, either an auditory speech or visual text command was displayed that required either a manual data entry or a verbal speech response from the pilot. For the target localization task, a stimulus command that designated the identity of one of three targets to be localized was conveyed with either an auditory-speech or a visual-text display format. Wickens, Sandry, and Vidulich observed that spatial location task latencies were significantly slower with the auditory-speech display when compared to the visual-text display of spatial location ($p < 0.0001$; pp. 241). The reverse effect was observed for the verbal task where the auditory display was observed as more compatible than the visual display where error rates correlated positively with latencies.

Wickens, Vidulich, and Sandry-Garza (1984; Experiment 2) also observed incompatibility effects of an auditory display for a spatial location task. For a spatial threat evaluation task, performance was compared between a visual display format of an *X-Y* horizontal situation display, and an auditory display that consisted of either a low or

high tone stimulus indicating that traffic was either behind or ahead of ownship. The auditory tones were spatially correlated with traffic, oriented either to the left, right, or in mid-plane with respect to ownship. Lower performance accuracy was observed when the spatial threats were aurally displayed ($p < 0.001$; Vidulich & Wickens, 1985); error rates were reported as greater in magnitude compared to the incompatibility observed when the visual display was used for the verbal task.

These studies both report that response time and performance accuracy suffer with auditory displays compared to both spatial and non-spatial (text) visual displays, and in general, it is more difficult for humans to make spatial judgments with aural versus visual stimuli. Auditory displays can then be categorized as data-limiters (or, SA limiters) for spatial localization tasks. Furthermore, as described next, non-spatial (or monaural) auditory displays are relatively greater data-limiters compared to spatial auditory (or binaural) display formats.

Spatial versus non-spatial auditory displays. The data limiting characteristic of auditory displays for spatial tasks would suggest that they not be used for spatial location tasks such as for spatial judgment of hazard traffic. Despite display format incompatibility, it may be beneficial in some cases, to implement less than optimal display-compatibility mappings in a multi-task environment. For instance, auditory displays can be used to offload visual demands, or to even act as a redundant source of information when the information is critical. For instance an aural traffic alert could cue the pilot to the visual display if she was not already looking at it. Such auditory displays used for spatial location of traffic can be characterized as either monaural or binaural.

The cost of spatial and non-spatial auditory displays. Data from the literature show evidence that monaural displays significantly slow down performance when compared to binaural display formats. Begault (1993) found a 2.25 s cost for target acquisition times for a monaural auditory TCAS display when compared to a binaural TCAS auditory display (4.74 versus 2.5 s; $p = 0.002$), although significant accuracy effects were not observed. Begault and Pittman (1996) also found slower mean performance times for a traffic target acquisition task of 0.5 s (2.63 versus 2.13 s; $p < 0.001$) for a head-up spatial auditory TCAS display, compared to a standard visual-audio TCAS display consisting of a head down visual display with a monaural traffic alerts. No accuracy effects were observed.

Data from basic research mimic the temporal effect between spatial and non-spatial auditory displays. For instance, for a visual search task, Perrott, Saberi, Brown, and Strybel (1990, Experiment 1) varied visual target locations within a 260 degree region along the horizontal plane at fixed elevation, accompanied by either a collocated (spatial) 10hz auditory click train, or, with a non-collocated click train displayed from a source in front of participants. They observed slower performance times for the non-spatially correlated display across a range of visual target locations from 0 +/-130 degrees along a horizontal plane ($p < 0.001$). The cost for the non-spatial displays increased for targets located in the rear hemifield more than 90 degrees from initial line of gaze; performance time costs from 500 to 775 ms for the non-spatial auditory display were observed when compared to the spatial click-train for visual targets at 90-175 degrees in either the left or right direction. Accuracy effects were not reported. In a second

experiment, the procedure and tasks were repeated with the added variation of target locations in the vertical field +/- 46 degrees with a collocated click train in both the lateral and vertical direction. Again, greater performance costs were observed for the non-spatial auditory versus the spatial auditory display ($p < 0.001$). Accuracy was not reported.

Although both types of auditory displays are data limiters, the collective data from the three studies point to monaural displays as relatively severe degraders of performance compared to binaural displays, for spatial location tasks. Data are summarized in Table A15.

Table A15

Summary of Studies with Data for an Auditory Display Performance Estimate for Spatial Localization Tasks

Reference	Measure	Performance time		Notes
		Monaural (non-spatial) display	Binaural (spatial) display	
Begault (1990)	target acquisition times	4.74 s	2.5 s	$p = 0.002$
Begault & Pittman (1996)	a traffic target acquisition	2.63 s	2.13 s	$p < 0.001$
Perrott, Saberi, Brown, and Strybel (1990; Exp. 1)	visual search task	500 to 775 ms time cost for the non-spatial auditory display compared to the spatial auditory display, for visual targets at 90- to 175-degrees in either the left or right direction.		$p < 0.001$

Auditory display string length. Three studies are described where lengthier auditory strings resulted in performance decrements as a function of working memory load. That is, too much information was displayed for accurate comprehension of the information. Loftus, Dark, and Williams (1979) showed that increasing auditory string length imposes an increased load on the pilot's limited working memory capacity. Readback error rates were measured for either a 4-digit auditory string (e.g., a radio frequency) or an 8-digit auditory string (e.g., a 4-digit radio frequency followed by 4-digit transponder number) with varying retention times from 0 to 15 s. Loftus observed effects of information load, where increasing the number of digits to be recalled resulted

in greater readback errors. There was roughly a mean difference in error rates of 14% between high (8-digit) and low (4-digit) memory load conditions ($M = 65\%$ response probability for the high memory load condition versus $M = 45\%$ for the low memory load condition for a 0 s retention interval).

Helleberg and Wickens (2003) found a similar pattern of results when participants were asked to read back auditory commands between two and six parameters (among heading, altitude, airspeed, communication radio frequency, transponder radio frequency, and an altimeter calibration setting) displayed with a synthesized voice. Two-parameter commands were on average 7 digits, and 3 or more parameter-commands consisted of 8 or more digits. The greatest number of errors occurred with the auditory display format when compared to a redundant (auditory+visual) display, and visual display format (10%, 3%, and 4% error rates respectively, $p < 0.01$). Also, as the length of ATC instructions increased, so did the proportion of communication errors ($p = 0.08$; Helleberg and Wickens, 2003). For 2- and 3-auditory parameter commands, the readback error rate averaged 4%, and increased to an average of 6.75% across 4, 5, and 6-parameter commands. Again, lengthier auditory strings resulted in greater read back errors.

In another study by Wickens, Goh, Helleberg, and Talleur (2002), pilots flew a simulated flight using 3 display formats: auditory, visual, and redundant displays for ATC clearance and traffic location information. The information varied between a single and a 3-parameter command of heading, altitude, and/or airspeed. Single-parameter commands averaged 3-4 digits whereas 3-parameter commands were longer than 8 digits.

When the communication load was short, there was no penalty for the auditory delivery of information (mean error rate = 2.3% for single parameter commands), but lengthy auditory commands that exceeded the capacity of working memory significantly increasing readback error rate ($M = 11%$ for 3-parameter commands; $p < 0.05$). In sum, readback errors were highest for the auditory display; as auditory string length increased from the single-parameter to 3-parameter commands so did the readback error rates.

Data are summarized in Table A16.

Table A16

Summary of Studies with Data for an Auditory Display String Performance Estimate, with Various Auditory-string Lengths

Reference	Measure	Accuracy rate per # of digits						Notes
		3	4	5	6	7	8+	
Loftus, Dark, & Williams (1979)	Readback error-rate	-	45%	-	-	-	65%	44.4% relative cost; estimated from Fig. 2, pp. 176
Helleberg & Wickens (2003)	Readback error rate	-	-	-	-	4.5%	6%	33.3% relative cost
Wickens, Goh, Helleberg, & Talleur (2002)	Readback error rate	2.3%		-	-	-	11%	378.3% relative cost; Fig. R8, pp. 29).

From these data it can be estimated that, when an auditory string contains 8 or more digits to be remembered, performance accuracy severely degrades relative to when less than 8 digits must be remembered. Based on the finding by Wickens, Goh, Helleberg, and Talleur (2002), that the 2.3% error rate for 3- to 4-digit auditory strings was not significant, it can be estimated that 5- to 7-digit auditory strings are relatively mild performance degraders.

Symbol / abbreviation familiarity. The benefit of symbols and abbreviations is to minimize display clutter; however, they require that the pilot transform the information into something that is meaningful. Remington and Williams (1986) reported that familiarity is a factor that affects the time to identify symbols. At an extreme, for an unfamiliar or untrained display, full comprehension by the pilot will be limited, and 100% accurate performance is at best, by chance.

The cost of information displayed with symbols or abbreviations. Rehman, Reynolds and Neumeier (1995) examined the presentation of weather information in the cockpit for 4 display formats as well as for four methods of data entry. The 4 formats for weather presentation included either plain English or highly coded teletype (TTY) abbreviations that have been used with Datalink displays, were either vertically or horizontally oriented. Participants included GA, private pilots, and commercial pilots where all were highly trained and familiar with English. Pilots were asked to retrieve information regarding weather. Rehman, Reynolds and Neumeier (1995) found that Datalink messages containing TTY weather information required longer response times ($M = 14.3$ s) by 34% compared to those in plain English ($M = 10.7$ s; $p < 0.05$). Also,

error rate was higher for TTY abbreviations ($M = 55\%$) by 53% compared to English ($M = 36\%$; $p < 0.05$). Data are summarized in Table A17.

Table A17

Summary of Data Comparing Performance with Abbreviated and Non-abbreviated Text

Reference	Measure	Performance time	
		Abbreviated text	Non-abbreviated text
Rehman, Reynolds & Neumeier (1995)	Error rate for weather information retrieval	55%	36%
	Performance time	14.3 s	10.7 s

The cost of display unfamiliarity (training). Fennell, Sherry, Roberts, and Feary (2006) also found that experience (or training), defined as flight time with an FMS system ranging from 25 (least experienced) to 200 hours (the most experience), affected performance. Using the RAFIV model, the authors examined errors as a function of the number of recall steps required to reformulate, access, format, insert, and verify FMS display information for a number of tasks carried out with the FMS, which includes abbreviated text. The RAFIV model contains 3 basic stages: comprehension (which involves r reformulating the task into FMS functionality), communication (which involves a accessing the correct feature, f formatting, and i nserting information into the FMS), and confirmation (which involves v erifying the information is correct). They found that as experience with the FMS display decreased, error rate (measured by the number of incorrect display inputs) increased. Fennell, Sherry, Roberts, and Feary reported a mean

of 18 errors with the FMS display for the least experienced pilots, and a mean of 10 errors for the most experienced pilots ($p = 0.025$). “Experience” with the display system can be viewed as a form of training, where less training is prone to increased performance error. Data are summarized in Table A18.

Table A18

Summary of Data Comparing Performance with Unfamiliar and Familiar Displays

Reference	Measure	Performance time	
		Unfamiliar	Familiar
Fennell, Sherry, Roberts, & Feary (2006)	Number of errors	18	10

The cost of display infrequency. Fennell, Sherry, Roberts, and Feary also found performance costs as a function of the frequency that tasks were called for (hence, the frequency that task-relevant FMS displays were encountered). In their study, each of 20 tasks were given a frequency rating, where a task was considered “frequent” if it was estimated to occur in greater than 1/20 missions. A 68% error rate was reported for tasks that were both infrequent and that also contained recall steps in the reformulation stage. This finding “highlight[s] the importance of analyzing frequency in both design and training... [where] a design should support the reformulate and access stage directly with salient labels and easy access. Frequent tasks might not need as much support as infrequent tasks and could be designed for ease and quickness of task execution, while infrequent tasks require direct and clear support.” Sherry, Fennell, Feary, & Polson

(2006) describe that unfamiliarity of messages often leads to persistent interaction where the pilot seeks an appropriate response by exploring the user interface, but the exact interaction is not quite known. Although infrequently encountered displays may not be needed as often as other displays used for nominal tasks, when it is critical that infrequently used displays must be comprehended, any unfamiliarity with the system will delay appropriate action. In sum, infrequently encountered displays require direct and clear support through salient and familiar symbology and text. Data are summarized in Table A19.

Table A19

Summary of Data Comparing Performance with Infrequently and Frequently Encountered Displays

Reference	Measure	Performance time	
		Infrequent ($< 1/20$ occurrences)	Frequent
Fennell, Sherry, Roberts, & Feary (2006)	Error rate on infrequent and frequent tasks	68% for infrequent tasks	-

Unfamiliar (e.g., untrained) symbols and abbreviations can be included in the library as a SA limiting display factor, where infrequently encountered displays for which pilots have some familiarity (perhaps through training) are mild SA limiters, and those for which pilots have not had experience/training are relatively severe data limiters. It can be argued that unfamiliarity (either due to lack of training or infrequent encounters) is actually a function of either pilot experience or the infrequent nature of off-nominal

events for which displayed information is rarely encountered, rather than a function of the display format itself. Hence, training can be considered more of a “human” factor, and infrequency a “task” factor, rather than “display” factors. For the current purpose, however, display unfamiliarity will be characterized as a SA Limiting display factor in the library which has implications for human performance modelling (Hooey et al., in preparation).

Predictor format. A number of studies have reported effects of various predictor variables, for instance, for selectable predictor displays (Johnson, Battiste, Delzell, Holland, & Belcher, 2003), predictive threat vectors (Morphew & Wickens, 1998), and predictor frames of reference (Holland, 1998) (for a review see Gempler & Wickens, 1998). Two studies are summarized here that depict performance accuracy effects for the absence of a predictor when a pilot must project a flight path trajectory, and also, the cost of linear predictors for projecting curved trajectories.

Jago and Palmer (1982) had pilots monitor a CDTI to make perceptual judgments of future positions of an intruder aircraft while varying predictor type by reference frame (either ground referenced or ownship referenced) and whether the predictor was linear (absent of turn rate information) or curved (which did include turn rate information). In a pre- and post-test, participants carried out the same task of predicting whether an intruder would pass in front or behind ownship without any predictor information at all. This required the pilot to project the flight trajectory of ownship and/or other traffic. There was no significant difference between the no-predictor conditions between the pre- and post- test; however, Jago and Palmer found that mean error rates were much higher in the

no-predictor conditions of the pre- and post-tests compared to the conditions when a predictor was present. That is, the presence of predictor displays significantly improved performance. For the pre and post tests together, there was a mean error rate of 30.4 % when no predictor was present, which is three times the error rate of 9.9% when a predictor was present.

Hart and Loomis (1980) duplicated the experimental conditions described above, and the trend of results from Jago and Palmer was replicated. Performance accuracy was reduced with the addition of a predictor, either linear or curved, for straight encounters (8.8%). The mean error rate was worse across straight and curved encounters when no predictor was displayed= 19.5%. In sum, across both studies, the absence of a significantly increased the error rate for spatial judgment tasks.

Table A20

Summary of Studies with Data for a Predictor Format Performance Estimate

Reference	Measure	Condition		Relative Cost of no predictor vs. predictor
		No predictor	Predictor	
Jago & Palmer (1982)	Error rate for spatial judgment task	30.4 %	9.9%	20.5%
Hart & Loomis (1980)	Error rate spatial judgment task	19.5%	8.8% (for straight encounters)	10.7%
Average error rate		24.5%	9.4%	15.6%

Display conformality. Earlier it was described that tracking display conformality lies on a continuum, and three levels of conformality have been defined for the library

(fully conformal [FC], partially conformal [PC], and non-conformal [NC]). Based on the data from Fadden, Ververs, and Wickens (1998), Martin-Emerson and Wickens (1997; Experiment 2), Foyle, Hooey, Wilson, and Johnson (2002), and Wilson, Hooey, and Foyle (2005), where degradations in performance accuracy were observed with both the PC and NC display formats (compared to FC display formats).

Fadden, Ververs, and Wickens (1998) evaluated the costs and benefits associated with the HUD by an analysis of data obtained from eighteen studies in the aviation domain. Five of the eighteen studies evaluated conformality and showed benefits for either increased flight path tracking accuracy or faster detection responses to changes in symbology and presentation of traffic when compared to non-conformal displays. Studies contributing to the analysis were significantly heterogeneous, indicating that confounding variables may be influencing conformality effects which were possibly the result of noise. However, empirical data from the literature support the results of Fadden, Ververs, and Wickens' meta-analysis that PC and NC display formats fare worse for tracking performance accuracy compared to FC displays. In the following section four of these studies will be reviewed.

The cost of partially conformal (PC) display formats. Martin-Emerson and Wickens (1997; Experiment 2) measured tracking accuracy performance and event detection (both near and far) across eight levels of visibility (ranging from none, to some, to full visibility) with both fully conformal and partially conformal navigation displays. The fully conformal symbolic runway on the navigation display relevant for the task of tracking consisted of all four characteristics previously listed in the Results section for a

FC display. Near domain display symbols also formed meaningful objects as they aligned with one another on the display (e.g., a velocity vector symbol aligned with the symbolic runway threshold and reference lines to signify the aircraft was on course with the runway in the real world). This FC display is illustrated in Figure A19.

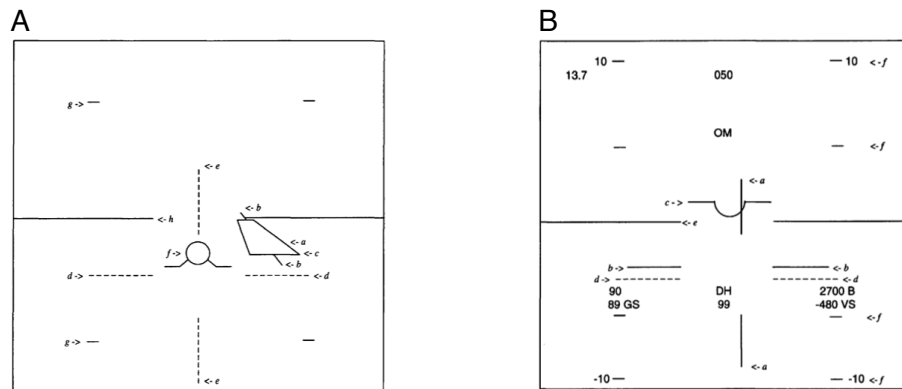


Figure A19. Fully (A) and partially (B) conformal display used in Martin-Emerson and Wickens (1997, Experiment 2). Adapted from “Superimposition, symbology, visual attention, and the head-up display,” R. Martin-Emerson and C. D. Wickens, 1997, *Human Factors*, 39(4), 581-601.

Tracking performance with the FC display (Figure 19A) was compared to a partially conformal navigation display that consisted of only two of the characteristics for a FC display (the ILS crosshairs were NOT fully analogous in shape to its far domain counterpart nor were they displayed with a 1:1 ratio with their far domain counterpart in the real world). Martin-Emerson & Wickens found significantly greater lateral tracking error for the partially-conformal display ($M = 35$ ft) across all levels of visibility

($p = 0.08$) compared to the FC display ($M = 15$ ft), which reflects a relative performance accuracy cost of 133% for the PC display. As visibility cycled through zero visibility to maximum and again to zero, greater performance differences were observed for the partially conformal display ($p = 0.06$), while performance with the fully conformal display stayed relatively consistent. This is an indication of discrete shifts of attention between the near and far domain when pilots used the PC display, as visibility varied from none to full visibility. For the condition where full visibility (rather than fluctuating visibility) was displayed, lateral tracking error was still greater for the PC display (RMSE = 37 ft) compared to the FC display (RMSE = 15 ft), by 147%.

Martin-Emerson & Wickens suggested that the fully conformal symbology was consistent with the far domain even in poor visibility, which provided for better division of attention between the two domains, as performance stayed relatively consistent even in changing visibility conditions. With the partially conformal display, however, discrete shifts in focus between the near and far domain were attributable to the perceptual discrepancy between the ILS display's abstract symbology of far domain elements. Based on the large performance accuracy cost of 133% for the partially-conformal display across varying visibility (and the comparable 147% cost under consistently full visibility) it can be estimated that the partially conformal display format used in this study limited comprehension of task-relevant information that was required to continuously maintain accurate tracking.

Similar to Martin-Emerson and Wickens, Wickens and Long (1994) also found a cost for a PC display format when compared to a FC format. Wickens and Long varied

display location (HUD and HDD), visibility (pre- and post-breakout conditions), and conformality (fully-conformal or partially-conformal displays) and found that, overall, tracking performance accuracy was best with the FC display format. Wickens and Long note that what they had initially considered as non-conformal may otherwise be considered partially conformal, and will be characterized here as so. The fully conformal navigation display (Figure A20A) relevant for the task of tracking consisted of a fully conformal trapezoid runway in the near domain that fulfilled the four listed requirements for a FC display format. Near domain display symbols also formed meaningful objects as they aligned with one another on the display (i.e., a velocity vector symbol aligned with the symbolic runway threshold to signify the aircraft was aligned in the real-world. Additional symbology was displayed: a fully conformal horizon, a flight path velocity vector that formed an object with the FC symbolic runway, a stationary aircraft symbol, and alphanumeric aircraft parameters.

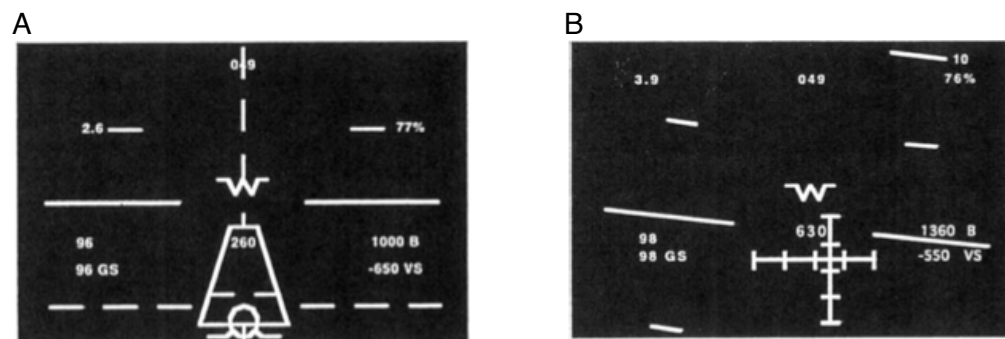


Figure A20. Fully and partially conformal display used in Wickens and Long (1994). Adapted from “Conformal symbology, attention shifts, and the head-up display,” by C. D. Wickens and J. Long, 1994, *Human Factors Proceedings*, 38(1), 6-10.

A partially conformal ILS navigation display relevant to the tracking task (Figure A20B) consisted of ILS crosshairs representing glideslope and localizer that consisted of only two of the four characteristics previously listed for FC displays (the crosshairs were not fully analogous in shape with localizer and glideslope trajectories, nor were they displayed with a 1:1 ratio with the real world since the traditional ILS is usually scaled up or down along either the localizer or glideslope dimension). Additional symbology was displayed: a horizon symbol, a stationary aircraft symbol, and alphanumeric aircraft parameters.

For the HUD pre-breakout condition, where visibility was essentially zero, tracking error was greater with the partially conformal display ($M = 88$ ft) when compared to performance with the fully conformal display ($M = 65$ ft; $p < 0.01$) indicating a performance accuracy cost for the partially conformal display of 35.4%. In the post breakout condition where there was some visibility, Wickens and Long found that the partially conformal HUD symbology yielded greater tracking error performance ($M = 78$ ft) compared to the fully conformal HUD ($M = 68$ ft)-- a performance cost of 14.75% for the partially conformal display. Across pre- and post- breakout HUD conditions, the data reflect a mean performance accuracy cost for the partially conformal display of 21.85% when compared to the fully conformal format. The authors suggest that during full visibility tracking performance benefits in the HUD conditions can be attributed to perceptual fusion of the near and far domain, whereas performance with the partially conformal display suffers from clutter. It is possible that the conformal display

contributed to perceptual fusion of near and far domains to a greater extent than the partially conformal display in post break out conditions, and the abstract representation of the partially conformal display contributed to performance decrement during pre-breakout conditions.

The cost of non-conformal (NC) display formats. Hooey, Wilson, and Johnson (2002) found a cost for a NC display formats. They measured tracking performance and event detection during a taxi navigation task with what they referred to as a' fully conformal T-NASA situation guidance display, a non-conformal command guidance display, and a hybrid display with redundant navigation symbology in both non-conformal and fully conformal formats. The fully conformal navigation display relevant for the task of taxiway tracking (Figure A21A) consisted of fully conformal scene-linked T-NASA HUD symbology that included a virtual taxi centerline and runway edges with turn and flag symbols that consisted of the four characteristics for full conformality, previously listed in the Results section.

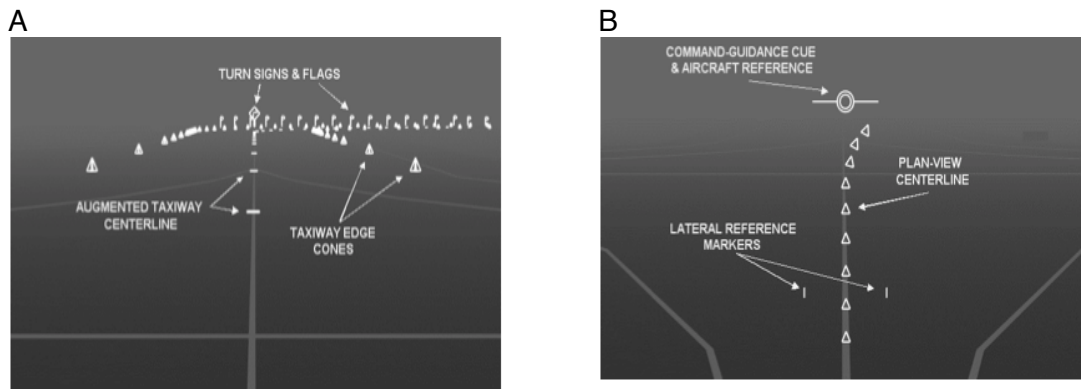


Figure A21. Fully (A) and partially conformal (B) displays used in Foyle, Hooey, Wilson, and Johnson (2002). Adapted from “HUD symbology for surface operations: Command-guidance vs. situation-guidance formats,” by D.C. Foyle, B. L. Hooey, J. R. Wilson, and W. A. Johnson, 2002, *SAE Transactions: Journal of Aerospace*, 111, 647-658.

A non-conformal navigation display relevant to the tracking task consisted of a Non-conformal taxi way centerline. The centerline was a two-dimensional overlay onto a three dimensional perspective view of the far domain centerline. Foyle, Hooey, Wilson, & Johnson (2002) found greater tracking error for turn segments with the NC command display ($M = 9\text{ft}$) compared to the FC situation guidance display ($M = 6.23\text{ ft}$; $p = 0.013$), which reflects a relative performance accuracy cost of 44.5% for the NC display format. Performance for a hybrid display, which can be considered as a fully conformal display format with redundant partially conformal display symbology for taxi navigation, was also measured. The hybrid format (Figure A22) consisted of a Fully Conformal runway centerline.

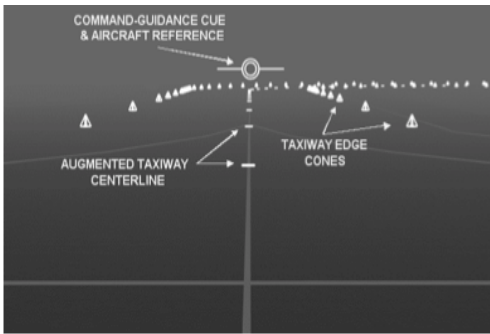


Figure A22. Hybrid display formats used in Foyle, Hooey, Wilson, and Johnson (2002). Adapted from “HUD symbology for surface operations: Command-guidance vs. situation-guidance formats,” by D.C. Foyle, B. L. Hooey, J. R. Wilson, and W. A. Johnson, 2002, *SAE Transactions: Journal of Aerospace*, 111, 647-658.

No significant difference between the redundant hybrid display and the fully conformal situation guidance display were observed. The authors suggest that since both the fully conformal and the hybrid display (with redundant non- and fully-conformal navigation symbols) yielded better performance than the partially conformal display, the benefits of full conformality outweigh the costs of the non-conformal display. They note that decreased performance on turn segments with the non-conformal display may result from cognitive tunneling; the attention capture of the partially conformal display requires the pilot to focus on a small portion of the display. This focus of attention comes at the cost of awareness of the bigger picture. Essentially, the pilot may be flying the aircraft with respect to the display symbology rather than with respect to the real world, where the conformal symbology serves as a closer proxy of the real world. The mean performance cost of 44.5% for the NC display format during turn segments indicates that

it limits full comprehension of the information required for continuous tracking tasks. It can be imagined that this effect would be exacerbated in zero visibility conditions, as was shown with the 145% performance accuracy decrement in Martin-Emerson and Wickens (1997) that did include trials with zero visibility.

In a follow-up study to measure eye tracking, Wilson, Hooey, and Foyle (2005) duplicated the study above using the same displays and condition with the exception of added non-conformal symbology to the non-conformal display; a non-conformal preview turn symbol indicated if and in what direction an upcoming turn would be oriented but it did not overlay any far domain counterpart, nor follow in shape, spatial position nor directional motion. The results were similar in trend to those of Martin-Emerson and Wickens (1997), Wickens and Long (1994), and Foyle, Hooey, Wilson, and Johnson (2002). Overall, the non-conformal command guidance display led to greater tracking error ($M = 5.75$ ft) compared to the FC situation guidance display ($M = 4.25$ ft RMSE; $p < 0.05$) yielding a 35.3% performance accuracy cost imposed by the non-conformal display format. The non-conformal display also yielded greater tracking error compared to a hybrid display ($M = 3.0$ ft RMSE; $p < 0.001$). The eye tracking data revealed the pilot spent more time looking at the forward scene with the fully conformal and hybrid displays when compared with the non-conformal displays. The authors suggest that less route information was available with the non-conformal display so pilots spent more time looking elsewhere to gain awareness of route information. The authors suggest that the benefit of the conformal route information provided with the situation guidance and hybrid HUD formats provided a common reference with the environment, which may

have supported better distribution of attention between near and far domain, leading to better tracking performance.

In summary, varying levels of display conformality have been examined for tracking tasks in zero through full visibility. The role of the pilot is to comprehend what the tracking display symbols represent (for navigation tracking, this may include a representation of the runway centerline and boundaries), how the symbols relate to one another (for instance, the spatial relationship between the runway centerline and runway boundary symbols), and to comprehend how the symbols relate to their representations in the real world (that is, the relationship between the spatial positioning of symbols on the display and spatial position of their counterparts in the real world). When the symbols are abstract and not to scale with the real world (or, non-conformal), the transformation of displayed information into an accurate assessment of the situation is a greater challenge than if the pilot were dealing with symbols that are purely analogous with the real world (fully conformal). Of course, through training the pilot will be equipped to carry out the task of navigation where some level of comprehension of the display symbols and their relationships are attainable. However, full comprehension of the dynamic and changing situation is a greater challenge with the abstract non- or less-conformal display formats (Martin-Emerson & Wickens, 1997, Wickens & Long, 1994; Foyle, Hooey, Wilson, & Johnson, 2002; Wilson, Hooey, & Foyle, 2005). The pilot is left to translate the abstract symbols and relationships that are not to scale with the real world, and performance accuracy for the continuous tracking task is limited (Foyle, Hooey, Wilson, & Johnson, 2002; Martin-Emerson & Wickens, 1997, Wickens & Long,

1994; Wilson, Hooey, & Foyle, 2005). The pilot may then be flying the aircraft with precision according to the display symbology, but may not be flying accurately with respect to the real world.

With non-conformal symbology, these effects may be exacerbated in zero visibility. In visible conditions, there is the cost of cognitive tunneling on display symbology that does not allow for divided attention with the far domain. If on the other hand, the level of display conformality did not affect performance, then we would see equally accurate navigation performance with both displays, which however, is not the case, as shown with the data across the four studies reviewed here. A summary of these data, where the pilot had some visibility, is summarized in Table A21.

Table A21

Summary of Studies with Data for a Tracking Display Conformality Performance Estimate

Reference	Measure	Tracking error			Notes
		Non-conformal	Partially conformal	Fully conformal	
Martin-Emerson & Wickens (1997)	Tracking error (visible condition)	-	$M = 37$ ft	$M = 15$ ft	147% cost; $p = 0.013$ for the PC display
Wickens & Long (1994)	Tracking error (visible post-breakout condition)	-	$M = 78$ ft	$M = 68$ ft	14.8% cost for the PC display
Foyle, Hooey, Wilson, & Johnson (2002)	Tracking error during taxi	$M = 9.0$ ft	-	$M = 6.23$ ft	44.5% cost; $p = 0.013$
Wilson, Hooey, & Foyle (2005)	Tracking error during taxi	$M = 5.75$ ft	-	$M = 4.25$ ft	35.3% cost; $p < 0.05$

From a comparison between the 14.8% cost for the PC display format in Wickens and Long (1994) and the relatively higher 44.5% and 35.3% costs for the non-conformal displays in Foyle, Hooey, Wilson, and Johnson (2002) and Wilson, Hooey, and Foyle (2005), it can be estimated that the NC display formats are relatively severe performance degraders compared to the PC formats. The high cost of 147% for the PC display in Martin, Emerson, and Wickens (1997) may be attributed to experimental differences, or other display factors used in the study.

Spatial displays for temporal judgments. Four empirical studies from the basic research and one from the aviation domain were identified that measured time-to-arrival information with spatial display formats (Law, Pellegrino, Mitchell, Fischer, McDonald, & Hunt, 1993 Experiments 1-4; Xu, Wickens, & Rantenen, 2004). All five studies illustrate a distance over speed bias in TTA judgments tasks that impede accurate TTA judgments.

Law, Pellegrino, Mitchell, Fischer, McDonald, and Hunt (1993) examined factors relative to arrival time judgments in the transverse plane across 4 experiments. The experiments used a spatial 2D display of two moving objects ('0' and '1' symbols) approaching at varying speeds, configurations, and distances to common targets. A consistent finding across the 4 experiments revealed an overreliance on relative distance information for making relative arrival time judgments. In Experiment 1, after a freezing point subjects were asked to estimate which symbol would arrive at a target first. Trials were designed so that the slower object was closer to its target throughout each trial. However, trials were balanced so that the closer object would have arrived first in only

50%, and in the other 50% the faster object would have overcome the closer object, arriving first. A comparison of results between trial types revealed higher performance accuracy for trials where the closer object would arrive first. Overall, closer objects were correctly estimated to arrive first with a 93.89% accuracy rate, whereas farther objects were correctly estimated to arrive first with only a 32.34% accuracy rate ($p < 0.0001$). Across both types of trials this is a mean TTA estimation accuracy rate of 63.1%. These results suggest that viewers used a distance rule to estimate time-to-arrival.

To further tease out the effects of velocity and distance, in Experiment 2, in addition to an arrival time judgment task, Law et al. asked subjects to perform distance judgment where subjects indicated which object was closer after a freeze, and a velocity judgment task where subjects were required to indicate whether objects were moving at the same velocity, and if not, to identify which was closer to a target point after a freeze. Arrival time judgments followed the same trend as Experiment 1 where accuracy was higher for closer objects than for farther objects (90.63% vs. 35.16%; $p < 0.0001$). Across both types of trials this is a mean accuracy rate of 62.9%. These results suggest again, that viewers used a distance rule to estimate time to arrival. In the velocity judgment task, subjects demonstrated sensitivity to relative velocity that increased as the velocity ratio of the two objects increased. This suggests that the relative velocity information in the arrival-time task display is accessible and the differences can be discriminated. However, although velocity information is accessible, the authors note that it is systematically underrepresented in relative arrival-time decisions as the effect of distance seems to outweigh the effect of velocity for TTA judgment tasks. Since

Experiment 2 ruled out that the distance over velocity rule was a result of insensitivity to velocity information, the authors suggest resource limitations may play a role. Law et al. state that, “This could occur because velocity information, being the rate of change of distance information, is more complex than distance information. Furthermore, when the conjoint processing demands of two or more sources of information exceed processing capacity, the graceful degradation of the more complex information occurs first (Norman & Bobrow, 1975).” This is exactly what is referred to as a data limiter, or, SA limiter in library of performance estimates.

Law et al. conducted two additional experiments. In Experiment 3, they attempted to evaluate velocity estimation under pre-cue and post cue conditions. In the pre-cue condition, attention was directed to which information should be extracted from the display; in the post-cue condition attention was directed to which information should be extracted *after* the display of information; in essence, viewers had to pay attention to both distance and velocity information in the post-cue condition. The authors concluded that “the substantial decrement in relative velocity judgments under post-cueing suggests that the distance bias observed in Experiments 1 and 2 is caused by resource limitations.” Otherwise, performance in the pre- and post- cue conditions would not be different. That is, viewers would be able to attend to all sources of information and answer accurately in the post cue trial, but this was not the case. The authors note that “possible candidates for resources that could limit the concurrent assessment and integration of relative velocity and distance information are working memory capacity (Baddeley in Law, et al., 1993; Just & Carpenter in Law, et al., 1993) and limits on individuals' ability to coordinate

information from multiple sources (Yee, Hunt, & Pellegrino in Law, et al., 1993).” In Experiment 4 they attempted to tease out the effects of distance with velocity on participants’ ability to separate, versus integrate, the information for TTA judgments. It was observed that the “relative arrival-time judgments of the type used in these experiments depend as much on the ability to coordinate information from multiple sources as it does on the separate abilities to judge relative velocity and distance.” In addition, the distance over velocity bias was replicated.

The results of the four experiments by Law et al. apply to the library of performance estimates in that they provide evidence that TTA estimations using a spatially formatted display, which displays both distance and velocity, is a data limiting (or SA limiting) display format for TTA estimation tasks. In addition, the four experiments consistently show a perceptual bias where subjects use a “distance over speed” heuristic at the expense of accurate TTA judgment when using spatially formatted display. Looking to the aviation domain, Xu, Wickens, and Rantenen (2004) found similar results where pilots were observed to use a distance over speed bias at the expense of accurate time estimates of the closest point of approach (CPA) between two aircraft.

Xu, Wickens, and Rantenen (2004) found the same trend using a similar measure of TTA estimates as Law et al. Law et al. asked subjects to select which one of two objects would arrive at a target point first, whereas, Xu, Wickens, and Rantenen asked pilots to mentally extrapolate display information and press a button at the time when they thought an object (ownship) would arrive at its closest point of approach (CPA) with

a converging aircraft. Xu, Rantenen, and Wickens measured time to closest point of approach (TCPA) using a spatially formatted top-down 2D CDTI. They varied the intruder's distance to the closest point of approach at freezing (1.33, 2.67, or 4.0 nm), intruder's speed relative to ownship in an ownship-centered frame of reference (160, 240, or 480 knots), and miss distance (0.67, 2.67, and 4.67 nm). A comparison of absolute TCPA estimation errors to true TCPAs reveal that for true TCPAs of 10 seconds (across varying distances) the mean TCPA accuracy cost equals 15%; for true TCPAs of 30 seconds, the mean TCPA accuracy cost equals 26%; and for true TCPAs of 60 seconds, the mean TCPA accuracy cost equals 25%. On average, this is an accuracy cost of 22%. Xu, Wickens, and Rantenen also observed that absolute TCPA estimation error increased with increasing distance to closest point of approach (mean absolute errors = 8 sec at 1.33 miles, 14 sec at 2.67 miles, and 13 seconds at 4 miles distance to closest approach; $p < 0.05$). Similar to the Law et al. study, results revealed a distance over speed bias for TCPA estimates, where the authors observed that "estimated TCPA was always shorter for that point with the shorter distance and slower speed" for distance-speed combinations that had the same true TCPA ($p < 0.001$).

The empirical results from the Law et al. and Xu, Wickens, and Rantenen studies are summarized in Table A20. In summary, an average of data across 3 empirical studies reveals a low mean TTA estimate accuracy rate of 49.3%. This accuracy rate seems low enough to estimate that performance accuracy for TTA estimates suffers with spatial display formats that display both distance and relative velocity.

Table A22

*Summary of Studies with Data for a Spatial Display Performance Estimate for TTA**Estimation Tasks*

Reference	Measure	Accuracy rate	Notes
Law, Pellegrino, Mitchell, Fischer, McDonald, & Hunt (1993) Exp. 1	Time to arrive estimates	63.1%	Closer objects were correctly estimated to arrive first with a 93.89% accuracy rate, whereas farther objects were correctly estimated to arrive first with only a 32.34% accuracy rate. Averaging across these conditions yields a mean 63.1% accuracy rate ($p < 0.0001$).
Law, Pellegrino, Mitchell, Fischer, McDonald, & Hunt (1993) Exp. 2	Time to arrive estimates	62.9%	Closer objects were correctly estimated to arrive first with a 90.63% accuracy rate, whereas farther objects were correctly estimated to arrive first with only 35.16% accuracy rate. Averaging across these conditions yields a 62.9% accuracy rate ($p < 0.0001$).
Xu, Wickens, & Rantenen (2004)	Time to closest point of approach estimates	22%	Estimated from Fig. 6
		Average TTA estimate accuracy rate = 49.3%	