A COMPARISON OF THREE AIRCRAFT ATTITUDE DISPLAY SYMBOLOGY STRUCTURES

Thesis
Submitted to
The Graduate School of Arts and Sciences
UNIVERSITY OF DAYTON

By
Eric Edward Geiselman

In Partial Fulfillment of the Requirements for
The Degree of
Master of Arts in Psychology

UNIVERSITY OF DAYTON
Dayton, Ohio
December, 1991

APPROVED BY:

David W. Biers, Ph.D.
Chairperson, Thesis Committee

F. Thomas Eggemeier, Ph.D.
Thesis Committee Member

William F. Moroney, Ph.D.
Thesis Committee Member

Robert K. Osgood, Ph.D.
Thesis Committee Member

CONCURRENCE:

Kenneth J. Kuntz, M.A.
Chairperson, Department of Psychology

ACKNOWLEDGMENTS

The author dedicates this document to the memory of his grandfather, William V. Swihart, whose appreciation for quality and goodness continues to provide inspiration.

The author wishes to express his gratitude to a number of individuals who significantly influenced this work: Dr. David W. Biers, thesis advisor, for his constructive criticism and confidence throughout this long process; Dr. F. Thomas Eggemeier, thesis committee member, for his support and suggestions over the course of the author's graduate career; Dr. William F. Moroney, thesis committee member, whose knowledge of aviation systems proved to be invaluable; Dr. Robert K. Osgood, researcher at the Armstrong Laboratory, Human Engineering Division, Crew Systems Directorate, Visual Display Systems Branch, and thesis committee member, for his suggestions as well as providing the opportunity, facility, and resources that made this research possible; Dr. Michael Venturino, for providing the opportunity to become involved in his work and for the privilege of knowing his family; Dr. Lawrence Wolpert, for his editorial comments and expert guidance; Dr. Maxwell Wells, for his helpful comments; Mr. Ken Aldrich and Mr. Terry McClurg for their software expertise; Mr. Bob Schwartz, Mr. Mike Poole, and Mr. Doug Windle, for their hardware support during the development of this experiment; Mr. Jeff Agnew, Mrs. Beth Rogers-Adams, and Mr. Chris Calhoun for their general help throughout the experimental development and writing of this document; the author's family, for their continuing patience, understanding, and encouragement; and the author's fiancee, Kathryn Cook Bradley, for what is to come.

The research reported was sponsored by the United States Air Force Armstrong Laboratory, Human Engineering Division, Crew Systems Directorate, Visual Display Systems Branch. Contract number: F33615-89-C-0532.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

LIST OF ABBREVIATIONS

ADI	Head-down attitude/direction indicator.
ANOVA	Analysis of variance.
ARI	Attitude reference indicator.
CRT	Cathode ray tube.
EADI	Electronic attitude/direction indicator.
FOV	Field-of-view.
G-HUD	Global head-up attitude/direction indicator.
HUD	Head-up display portrayal of attitude information.
MCH	Modified Cooper-Harper.
OS	Ownship symbol.
OTW	Out-the-window.
RMS	Root mean squared error.
SL	Straight and level.

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (CONT.)

ABSTRACT

A COMPARISON OF THREE AIRCRAFT ATTITUDE DISPLAY SYMBOLOGY STRUCTURES

Geiselman, Eric Edward
University of Dayton, 1991

Advisor: Dr. David W. Biers

In modern tactical aircraft, primary flight attitude information appears in two basic display configurations: the head-down attitude/direction indicator (ADI) and the head-up display (HUD) portrayal of attitude information. Both formats contain the same information and both are present in the cockpit. Empirical evidence has suggested that separate utilization of ADI or HUD symbology in specific situations may result in task dependent performance differences. This indicates the existence of attitude display information structure differences such that, in order to optimize performance, the pilot must decide to view the correct display for the appropriate situation. Given the speed and maneuverability of modern tactical aircraft, display switching seems unacceptable.

The present study evaluated the attitude information conveyance of a new aircraft attitude display concept. The new symbology format, or global head-up attitude reference display (G-HUD), was developed by integrating the shape of the conventional ADI and the transparency of the HUD into a single format. The present research addressed whether or not there was a significant performance advantage or decrement incurred by the symbology integration. The G-HUD, ADI, and HUD were evaluated during an attitude maintenance task as well as an unusual attitude recovery task. Subjects were trained to a specific attitude maintenance task performance criterion on one of the three displays before they participated in a simulated attitude recovery task. The results of the study suggest that few differences exist between the G-HUD and ADI formats while performance and training time was better with both the G-HUD and the ADI than with the HUD format. The findings of the study lend support to the hypothesis that an attitude display formed of the integration of ADI and HUD type symbology will demonstrate a performance benefit over a pure HUD format.

CHAPTER 1

INTRODUCTION

Aircraft attitude reference indicators (ARIs) represent and display the relationship between an aircraft being flown and the earth's surface over which it flies. In the tactical environment, accurate spatial orientation knowledge is paramount to the pilot's overall situation awareness (Guttman, 1986).

In modern tactical aircraft, primary attitude information appears in two basic ARI visual display configurations: the head-down attitude/direction indicator (ADI) and the head-up display (HUD) portrayal of attitude information. The ADI is located in front of the pilot and is embedded in the instrument panel thus the pilot must look "into" the aircraft to read the display. The transparent HUD is located above the instrument panel and oriented centrally in the pilot's field-of-view so that the pilot may look through the display and see the outside world. Both ARI formats contain the same basic information (aircraft pitch, roll, and heading). Although, due to different display media and intended utility, the displays convey attitude information via different features and symbology structures.

Empirical evidence has suggested that separate utilization of ADI or HUD symbology in specific situations may result in distinct, task dependent, pilot performance differences (such as recovery from unusual attitudes). This indicates the existence of attitude display information structure differences such that, in order to optimize performance, the pilot must decide to view the correct display for the appropriate situation. For example, some pilots flying HUD equipped aircraft are trained to refer to their ADI in situations where spatial orientation is in doubt (Summers, 1985). With the speed and maneuverability of modern tactical aircraft, display switching seems unacceptable.

The present study evaluated a new attitude display symbology concept. The new symbology, or global head-up attitude reference display (G-HUD), was developed by combining features of the ADI and HUD. The integration involved combining the "best" features of the ADI and HUD formats by fusing the information structure differences onto a single display. In affect, this presents the different display features in such close proximity that the information structure difference existing between the present two displays may no longer exist in the G-HUD. The present research addressed whether or not there was a significant performance advantage or decrement incurred by the symbology integration. The G-HUD, as it appeared in the present study, was not intended to represent the display's final configuration. Instead, the intent was to demonstrate the performance costs and/or benefits associated with the integration of the two conventional information structures.

BACKGROUND

To provide the appropriate background to the present study, this section presents a brief description of ADI and HUD symbology followed by a review of previous ARI evaluation research. The ADI and HUD are then discussed in terms of their display features and principles as well as the theoretical basis on which the G-HUD was developed. Display features and principles include: pictorial realism, display objectness, and display compression. Theoretical issues include: object attention and global perception (associated with display objectness). The G-HUD is then described relative to the above features, principles and issues.

Display Symbology

The following paragraphs describe the basic formats in which head-down and head-up attitude reference indicators presently appear. It is important that the reader be somewhat familiar with the display symbologies so that different display attributes can be kept in mind throughout the following review.

ADI. The ADI (Figure 1) is a modification of the original Sperry artificial horizon developed around 1929 (Johnson and Roscoe, 1972). The electromechanical display is located in the instrument panel of modern tactical aircraft. The ADI is a three-dimensional opaque physical ball with symbology printed on the ball's surface. The ADI was developed to provide roll, pitch and directional information for pilot spatial orientation maintenance during instrument referenced flight throughout the aircraft maneuvering envelope.

The upper and lower hemispheres of the ADI are differentially shaded so that the contrast between the light (sky) upper-half and the dark (ground) lower-half defines an artificial horizon. In order to determine aircraft attitude, the pitch scale bars and longitudinal heading line position is compared to a fixed ownship symbol (OS) superimposed on the surface of the ball. The ADI is free to rotate about all three of its axes and is designed to represent translation relative to the aircraft as if the aircraft was stationary. This relationship forms a pure inside-out earth referenced display coordinate system (Johnson and Roscoe, 1972). The ADI also incorporates a bank scale

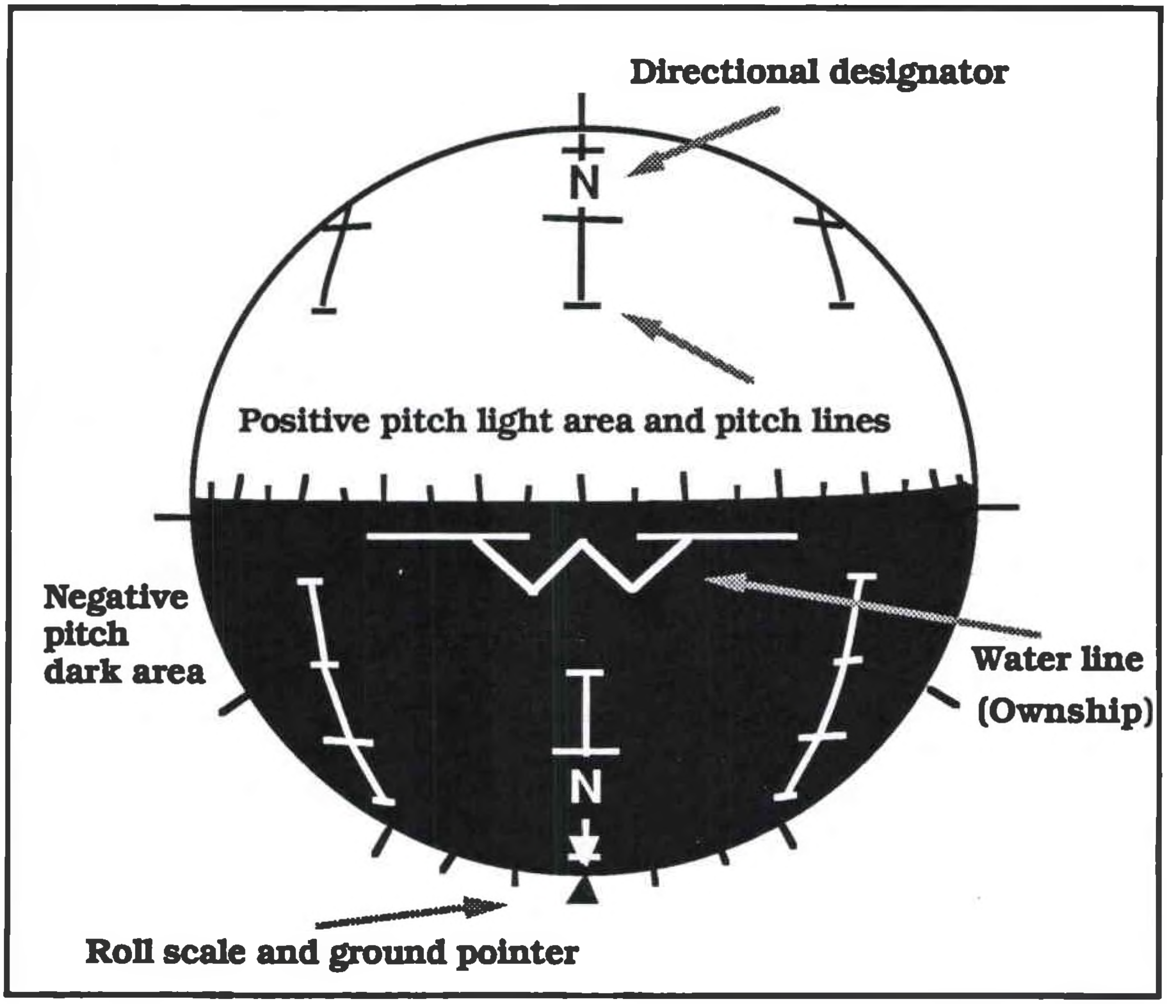


Figure 1. Computer Simulation of Conventional Attitude/Direction Indicator (ADI).

and ground-pointer along the bottom of the display. The bank scale or roll indicator gives precise roll information at bank angles equal to or less than 90 degrees.

HUD. In present aircraft, HUD symbology (Figure 2) is projected onto a transparent combiner glass in a collimated fashion so that the information represented is superimposed on the outside world. The intent of this approach is to form an integration of focal and ambient information so the pilot may constantly monitor available flight status information (in this case spatial orientation) without losing visual contact with the outside world. This is known as keeping the pilot's head out-the-window (OTW). HUD symbology is focused at visual infinity so, in theory, the pilot can look through the transparent display and view the distant world without a corresponding change in optical accommodation. It is important that the HUD be as uncluttered as possible so that the OTW scene is not excessively occluded.

The attitude information structure of the HUD can be easily described in terms of its components. The three basic components are: (1) the main attitude reference component--referred to as the pitch-ladder, (2) a heading indicator, and (3) a roll indicator with an associated ground-pointer. The attitude component is formed of a caging artificial horizon, pitch-bars with alphanumeric designators and a flight path marker. For the present study the OS is fixed with respect to the cockpit while the horizon and pitch-bars move about it. Like the ADI, this forms an inside-out oriented display (Johnson and Roscoe, 1972). The pitch-bar lines are solid above the horizon and segmented below. In addition, the pitch-bar lines are articulated: bent like chevrons pointing in the direction of the true horizon, giving the pilot a directional cue to follow when recovering from an extreme attitude back to horizon-level (Reising, 1988). The attitude "ladder" moves in both pitch and roll. Heading information is read off a tape along the top of the display. Whatever digital value appears above the OS's

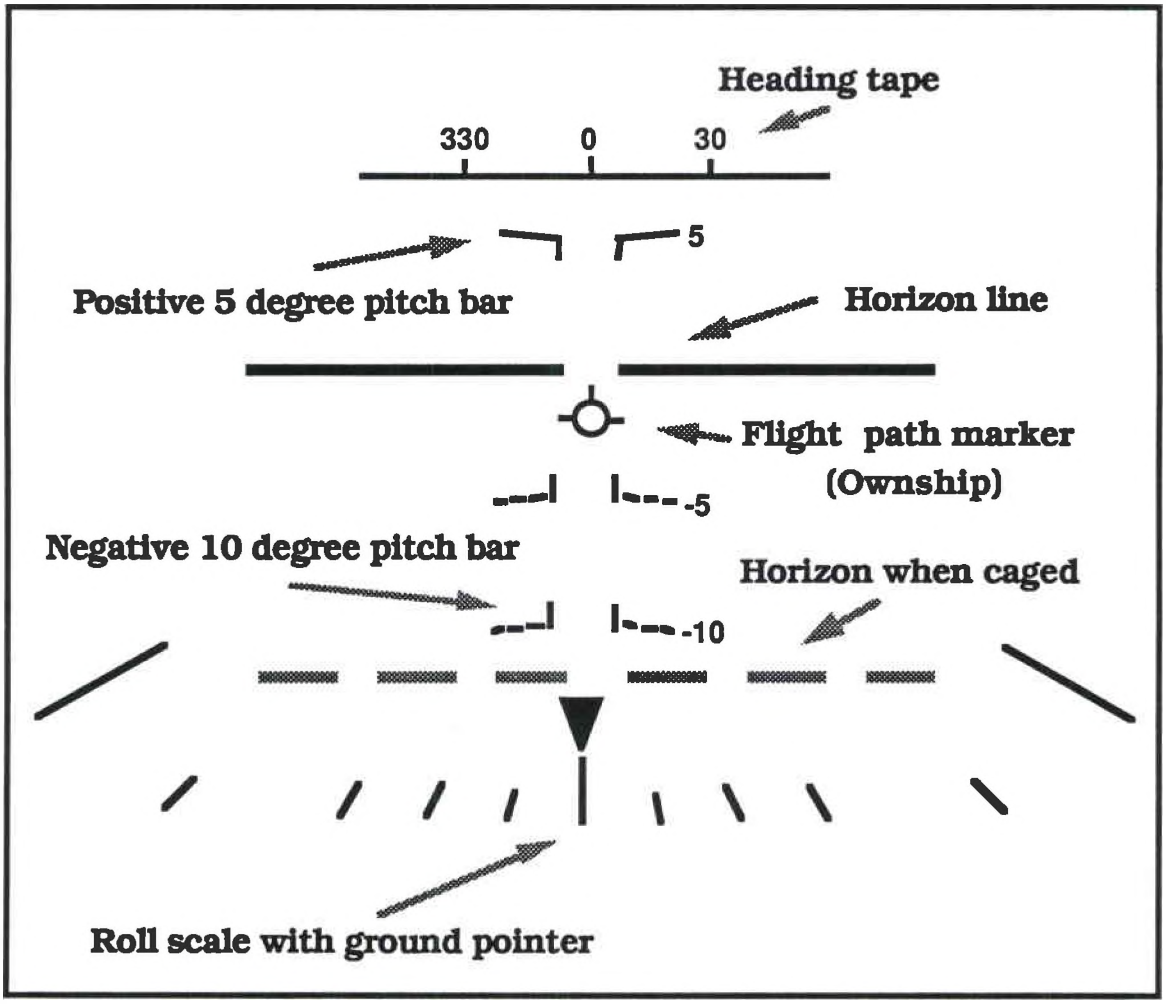


Figure 2. Modified Pitch Ladder HUD Portrayal of Attitude Information.

schematic tail is the aircraft's present heading. The roll indicator along the bottom of the display gives precise roll deviation when the ground-pointer symbol is compared to the bank angle scale (only at less then 45 degrees roll).

Previous Research

For conveyance of spatial orientation, it has been predicted that the ADI will out perform the standard HUD symbology (Roscoe, Corl, and Jensen, 1981). Some recent studies have suggested that this prediction is in fact accurate.

Kinsley, Warner, and Gleisner (1985) conducted two experiments to assess the relative merits of the current F/A-18 pitch-ladder ARI, a revised pitch-ladder HUD format, and a conventional ADI for recovery from unusual attitudes. The revised HUD was developed to enhance the global processing properties of the multi-dimensional HUD symbology by adding redundant orientation cues. The ADI was included in the investigation as a result of discussions with Naval Air Test Center pilots who indicated their preference for having an ADI present as an attitude indicator. The first experiment compared static presentations of the three displays. The ability of these formats to aid the subject in deciding how to recover from unusual attitudes was assessed. The metrics utilized were decision time (defined as the duration of time before a control input was initiated) and control error (control reversals). Three pitch values (0, +55 and -55 degrees) were crossed with six roll values (0, 60, 120, 180, -120 and -60 degrees) to form the stimuli set. Twelve non-pilot subjects viewed the stimuli as static events projected onto a screen and were asked to make a determination as to the proper recovery action by making a control stick input.

Data analysis of Experiment 1 indicated that the ADI resulted in significantly faster decision times than either the conventional or enhanced HUD format. No error differences for the static presentation reached statistical significance.

Experiment 2 assessed dynamic ADI and current F/A-18 HUD pitch-ladder ARI formats. Eight subjects viewed an actual ADI and HUD in an aircraft simulator. Display presentation was manipulated within-subjects and was counterbalanced. Pitch and roll values were identical to those used in experiment one. Non-pilot subjects flew the simulator with only one of the experimental displays activated per experimental trial. After flying straight and level for a period of time, subjects were forced to look away from the ARI to make a data input. Upon completion of the data input task, the display was reoriented so that it represented one of the unusual attitude conditions. The subject's task was to recover the aircraft back to straight and level flight (SL) as quickly as possible. Decision and recovery times were recorded.

Analysis of Experiment 2 data indicated that the display condition main effect for decision time failed to reach statistical significance. On the other hand, the recovery time metric reveled display condition performance differences in which mean time for recovery was significantly shorter for the ADI than the HUD pitch-ladder (the revised HUD format was not tested). Kinsley, Warner, and Gleisner (1985) recommended that an optimally located electromechanical ADI be included within the fighter's display suite as the primary or secondary spatial orientation instrument.

A study closely related to Kinsley et al. (1985) was completed by Guttman (1986) in order to evaluate whether the addition of a computer generated two-dimensional ADI located directly below the existing F/A-18 HUD would aid pilots in recovery from unusual attitudes. The study compared three display formats including a graphic representation of an ADI, an F/A-18 HUD, and the concurrent use of the ADI and HUD. It was hypothesized that the concurrent use of the HUD and ADI would result in faster decision and recovery times than the use of the HUD alone. It was also believed that the ADI would demonstrate better performance than the HUD in the single display conditions. Guttman used 10 naval aviators with and without previous HUD experience as subjects. Data were collected via a generic fixed-base simulator, actual HUD combiner equipment, a simulated three inch cathode ray tube (CRT), and a coupled control stick. The display images were dynamic and modeled the performance characteristics of the F/A-18. Again, three values of pitch (0, 55, and -55 degrees) were crossed with six values of roll (0, 60, 120, 180, -60, and -120 degrees) to form 18 unique pitch and roll orientations for testing. Roll, pitch, and display format were manipulated within-subjects while pilot HUD experience formed the sole between-subjects variable. The experimental procedure was identical to the dynamic display presentation experiment performed by Kinsley et al. (1985). Recorded dependent variables were decision and recovery time. At the end of trial data collection, subjects participated in a structured interview in which the pilots were asked to indicate the display format they preferred overall and why.

The recovery time metric was sensitive to the display format manipulation. Statistical analysis showed that recovery times were significantly faster using the ADI compared to the HUD. There was a significant format effect in which, at the 55 and -55 degree pitch conditions, the ADI resulted in the fastest recovery times, concurrent use of the HUD and ADI was next fastest while the HUD demonstrated the worst recovery performance. During the post experiment interviews, a majority (60%) of the subjects stated that they most preferred the concurrent use of the ADI and HUD because the displays each contained features that were helpful during different stages of attitude recovery. The remaining subjects (40%) preferred the use of the ADI alone. When asked which display was least preferred, a 70 percent majority indicated the HUD format while none of the subjects least preferred the ADI/HUD combination. Guttman (1986) speculated that, during recovery from extreme pitch attitudes, the strengths of each format acted to compensate for the weakness of the other. This statement was qualified by the comments of several of the naval aviators (the ADI alone showed the best objective performance). The pilots pointed out that the ADI display did not contain airspeed and altitude information that would be necessary for actual extreme attitude recovery. No differences were found between the HUD experience groups.

Osgood and Venturino (1990) investigated the effects of various attitude displays and their associated visual features on performance of a simulated unusual attitude recovery task. Among the evaluated attitude display formats was a simulated ADI, modified inside-out HUD, an outside-in HUD, and an egocentric HUD format. Subjects performed a "warm-up" tracking task phase followed by an unusual attitude recovery phase. For presentation of the unusual attitude position, the helmet-mounted display blanked for one second after the tracking phase and reappeared in an unusual attitude. There were four roll and three pitch angle groupings utilized in order to form the portrayed unusual attitudes (roll groups: +/-31-60, +/-61-90, +/-91-120 and +/-121-150 degrees--pitch groups: +/-1-15, +/-16-30 and +/-31-45 degrees). Subjects were instructed to recover their simulated aircraft back to straight and level. A successful recovery was designated as a return to +/-4 degrees roll and pitch without exceeding +/-15 degrees roll or pitch for a duration of two seconds. Six dependent variables were recorded: time to first control input in roll and pitch, input accuracy in roll and pitch, and the number of control reversals in the first second of input--in roll and pitch.

Among the findings for the displays, the data suggested that error and initial control reversals with the ADI format were consistently fewer or no different than that of the inside-out HUD. Performance for the ADI was equal to or better than that of the HUD but at no time was performance via the HUD better than the ADI.

Display Features and Principles

The previously reviewed studies demonstrate a clear performance advantage for the ADI display format in an experimental setting. Possible reasons for the performance differences may be explained in terms of the differences between the display features.

Pictorial realism. Researchers have theorized that, for a spatial orientation representation, pictorial realism as a display feature is important (Burns and Lovering, 1988; Previc, 1989; Roscoe, Corl, and Jensen, 1981; Taylor, 1982, 1984, 1988).

Roscoe, Corl and Jensen (1981) explain that pictorial realism for orientation representation presents a spatial analog of the real world in which the position of an object (the OS) is convincingly seen in depth as well as up-down and left-right (Roscoe, 1968). It was theorized that pictorially realistic displays permit the pilot to use highly learned perceptual rules about the world to reduce the amount of information that must be processed before making control responses. Roscoe, Corl, and Jensen (1981) write: "No such direct comparison (between the display and the earth) can be made if the information is presented on single-parameter symbolic displays". This statement suggests that some amount of pictorial realism seems necessary in order to cue the observer's natural orientation.

The presently utilized ADI (Figure 1) contains some degree of pictorial realism. The ADI presents attitude information as a contact-analogue of the real world such that the display behaves in a manner that is analogous to viewing a pictorial representation of the world through a port hole at the front of the aircraft (Roscoe, 1968). Comparing the position of the OS to the display's artificial horizon gives general aircraft roll, pitch and heading (when scaled) information (the ball rotates about its vertical axis as heading changes occur). Roll, pitch, and heading scales afford precise status reference as well as motion trend cues. The contrast between the display hemispheres (light on top and dark on the bottom) simulates the contrast between the sky and the earth when viewing the actual OTW scene. This contrast also acts as an inversion cue.

For superimposed displays such as the HUD, there is an obvious need to minimize obscuration of the outside world. Apparently it is for this reason that HUDs (Figure 2) display abstract, symbolic codes that rely on lines, shapes and numerals for roll, pitch and heading information (Taylor, 1984, 1988). Taylor (1982) concluded that none of the HUD symbology features he tested (segmented pitch-bars, pitch-bar tags, and etc.) appeared to have a natural orientation or to provide inherently meaningful cues to aircraft orientation. The HUD symbology structure results in a display that is divisible into separate information components (separable) such as heading information which is not included on the pitch-ladder.

For the ADI, pictorial realism results in a display that is a unitary whole formed of integrated continuous information similar to that information found in the real world. The symbolic HUD is formed of separable redundant discrete information. The following sections discuss the processing consequence of the information structure differences existing between the present displays.

Display objectness. Display objectness refers to the representation of one or several information variables as dimensions of a single object (Wickens, 1986). In the case of the ADI, the dimensions of the object represent variables of aircraft movement. As the aircraft changes in pitch, the ADI tilts, as a roll change occurs the ADI rolls, and as heading translates the ADI spins.

According to Wickens (1986), the most critical defining attribute of display objectness is the presence of contours (real or subjective). An object is that which is defined by closed contours. Because the ADI is enclosed within a contour, and simulates a three dimensional object, the display maintains the feature of objectness. The objectness of the symbolic HUD may not be as intuitive as the ADI. Also, the contours of the ADI may encompass more meaning than those of the HUD. The line and alphanumeric symbology on the ADI is included within the overall contour of the ball while the lines of the HUD, by definition, form independent objects.

Using the taxonomy of object display research, an object is more integral than its symbolic counterpart. The contours of the object are joined so that their combined features create a new or "emergent feature" (Wickens, 1986). The lines and differentially shaded hemispheres of the ADI are joined to form the overall contour of a "ball" so that the newly formed angles and contrast are perceived as roll, tilt, and spin. The lines on the HUD are more separable, less easily seen as integral, and seemingly do not form an emergent feature.

The objectness of the ADI may allow it to be processed as a single unitary object while the HUD, formed of many objects within close proximity, may require more time consuming and complicated processing. Theoretically this is analogous to processing a pictogram instead of mentally integrating actual text. Wickens and Andre (1990) point out that performance differences depend on the information integration requirement associated with a specific task.

Object attention. For attitude display design, the potential difference in attention requirements for unitary integral displays (objects) versus separable information (HUD symbology) may be critical, as may the relative benefit of global perception. Object attention refers to the manner in which humans distinguish objects and object features from their background, other objects, and other features (Previc, 1989). Object attention research has suggested that symbology structures formed of non-separable (integrated) features, such as the ADI, may be discriminated with less use of focal attention and processing effort (Treisman and Gelade, 1980).

Redundancy is often added to symbolic attitude displays under the logic that presenting similar information in more and different ways should aid the information conveyance efficiency of the display in the form of a redundancy gain (Taylor, 1984). An example of redundancy in the HUD of the present study are the indicators for the area below the horizon (negative pitch): (1) pitch-bar lines are segmented, (2) the alphanumerics have negative signs and, (3) the chevrons point up as do the pitch-bar tags--when not inverted. On the HUD, addition of information in the form of redundancy is not integrated into an existing display object. Redundancy creates conjunctions of separable features which may in actuality work against the display by pulling upon unnecessary focal attention as well as information processing time required for feature integration. According to Previc (1989) the optimal attitude display should be processed as a unitary percept.

Treisman and Gelade (1980) tested the feature-integration theory of attention which holds that attention must be directed serially to each stimulus in a display whenever conjunctions of more then one separable feature are needed to characterize or distinguish the symbol meaning. Integral features are conjoined automatically, while separable features require attention for their integration. Treisman and Gelade's (1980) position included: for visual search and discrimination, if two physical properties are integral, they should function as a single feature (under the experimental paradigm) allowing parallel search. If the stimuli are separable, their conjunctions will require focused attention. Conjunctions were expected to require serial search and should have no effect on performance unless focally attended.

A number of experiments were performed by Treisman and Gelade (1980) in order to investigate the above hypotheses. The experiments utilized the visual search paradigm to compare color-shape conjunctions with disjunctive color and shape features as targets and distractors. The experimental results suggested that search was serial when separable features were required to characterize targets and parallel when integral features were utilized.

Taylor (1984) performed several studies in order to examine the effects of various HUD symbology features intended to produce a redundancy gain in attitude recovery performance. The experimental findings indicated that a redundancy gain occurs only when combinations of integral stimuli dimensions were formed. Addition of integral stimuli form emergent features that are processed as a single unified whole. Taylor (1984) hypothesized that the presence of separate non-integral stimuli (the addition of horizon-pointing pitch line tags and asymmetric numerals) may attract unnecessary attention, increase response latencies, they may reduce saliency and discriminability of the most effective cues, and they will cause clutter and obscuration of the outside world. Taylor (1984) uses a clever speech analogy: Integral structures shout clearer and louder and separate redundant structures merely repeat the message. Taylor's findings for HUD symbology can not be directly compared to the ADI because, to date, the ADI has not been examined at the feature level. The performance gain afforded by the ADI pitch-bars and longitudinal lines is not known.

Global form perception. Global form perception (GFP) refers the perception of a level of visual structure that is made up of constituents at a lower or "local" level (Hoffman, 1980). GFP is the perception of a whole object which is made up of one or more features. The importance of global form perception for head-up attitude display designs lies with its ability to create vivid percepts which are natural and minimally dependent on actual contour information (Previc, 1989). Incorporating globally perceived object features into attitude display symbology design could result in reduced display clutter and enhanced "see-through" visibility while allowing interpretation under degraded visual conditions.

Navon (1977) explored the principle of global precedence which suggests that the global structuring of a scene precedes analysis of local features in human visual perception. Navon (1977) performed a number of experiments reminiscent of the Stroop (1935) studies in order to investigate the questions: Do we perceive a visual scene feature-by-feature? Or is the process instantaneous and simultaneous? According to Navon (1977), the globality of a visual feature corresponds to the place it occupies in the scene structure hierarchy: At the top of the hierarchy the nodes are more global then the nodes at the bottom or local level. It is claimed that processing of a scene proceeds from global-to-local where, as if focusing, gross features are processed followed by fine-grained analysis. For display design, in may be true that the less an observer has to focus on the image, the less fine-gain analysis required, the less processing time and effort will be required for information extraction.

Navon (1977) equated the properties of global and local features by using stimuli in which possible global features were identical to that of the set of possible local ones. This was accomplished by using large letters formed of smaller letters. Navon (1977) collected data suggesting that the global pattern of the letter stimuli was responded to faster than the elements of which it was formed. Results indicated that people can voluntarily attend to the global pattern without being affected by the local features but, they are not able to process the local features without being affected by the whole. A further study, utilizing global and local features of spatial patterns, found that global differences are more frequently detected than local differences and that the global configuration is more likely to be perceived on brief exposures than the local pattern. This finding suggests, according to Navon (1977), that local analysis must occur relatively late in the process.

Hoffman (1980) performed two experiments designed to investigate questions similar to those asked by Navon (1977). Does the human perceptual system process the existence of local structures and a subsequent global form is predicted (bottom-up processing) or is the global form perceived thus inferring the existence of appropriate local features (top-down processing)? Hoffman (1980) found that global level of form is not invariably processed prior to local levels. The data suggested that the quality of local and global features may predict the direction of processing. If the global features are degraded, the probability of local feature processing preceding global processing increases.

The HUD and the ADI both contain fine grain local features but, much like the objectness of the displays, the ADI appears to be more global. This can be easily demonstrated by degrading the focus of the displays and determining which becomes useless first. The "roundness" and sky/ground contrast of the ADI is responsible for this phenomenon. Under degraded conditions, such as vibration, high luminance levels, and high rates of translation, the user has both global and local features available within the ADI while the HUD may require more fine grained analysis. This may result in the ADI requiring less focused attention and processing effort. The ADI may be processed more "ambiently" than the HUD.

An example of a HUD global feature is the shape relationship formed by the asymmetric pitch-bar numerals as well as pitch-bar articulation. Experimental evidence has suggested that features within the HUD which are thought to be globally processed result in better performance than that of its local counterparts. Taylor (1984) found that global configural characteristics were superior to local features for indicating roll orientation. Particularly, as mentioned above, for the use of asymmetric positioning of pitch-bar numerals. In a separate pitch task experiment, Taylor (1984) found that sloped pitch-bars (pitch-bar inclination did not vary with pitch angle) produced better performance than pitch-bars without an inclination. Zenyuh, Reising, and McClain (1987) helped to confirm the articulated (inclination varying with pitch angle) performance gain.

Display compression. Display compression is the relationship between a display and the real world in terms of rate of motion, resolution, and field-of-view. The rate of motion relationship between the HUD--of the present study--and the world is 1:1 in both pitch and roll axes. This relationship translates into high resolution pitch information at the cost of a reduced field-of-view (FOV) and the potential for high rates of motion during high energy aircraft maneuvering. According to Burns and Lovering (1988), the use of a 1:1 display such as the HUD enhances precise control in the pitch axis but may degrade the pilot's ability to acquire and maintain spatial awareness due to the reduced FOV and high rates of display translation. The HUD has a limited FOV in pitch as well as heading although the HUD's transparency enables OTW contact when the world is not obscured. This HUD format displays a maximum of 15 degrees pitch and 90 degrees of azimuth area.

The ADI is compressed so that its display/world ratio is greater then one. The ADI artificial horizon does not have to move as far as the natural horizon to represent the same deflection. The ADI world-to-display ratio results in a display translation "slowing". The large FOV of the ADI maintains some portion of the horizon in view at all times. Only at extreme attitudes such as +/-90 degrees pitch is there not some portion of both the upside and the down side of the display in view during translation.

Burns and Lovering (1988) performed a study comparing three different electronic attitude/direction indicators (EADIs). The evaluation tested two modified EADI formats (a two-dimensional EADI with the addition of sky-pointing arrow graphics, and a three-dimensional EADI with sky-pointer arrows) and a conventional EADI for their relative merit enhancing attitude awareness. The two-dimensional displays had a 60 deg FOV while the three-dimensional EADI had a 180 deg FOV. Military pilot subjects performed unusual attitude recoveries via each of the three displays at 24 different pitch/roll positions (0, 60, and -60 degrees pitch crossed with 0, +/-60, +/-90, +/-120, and 180 degrees roll). Subjects were instructed to perform recoveries as if they were flying an actual fighter aircraft. Subjects were not to race the clock but were instructed to consider energy management and 'g' loading. The collected objective data included decision time, recovery time, correct response, altitude gain and altitude loss. Subjective data were also collected via a verbal debriefing, Modified Cooper-Harper, and written questionnaire completed after each experimental session and at the end of the evaluation.

Analysis of the objective data indicated significantly faster decision times for the conventional EADI and modified two-dimensional EADI than the three-dimensional EADI. This difference was present only in the nose low pitch position while no difference was detected between the display at the nose high or nose level positions. For recovery time, no significant differences were found between the displays. Response accuracy for the three-dimensional EADI did not differ significantly from either of the other displays. Altitude gain and loss data failed to reach any statistically significant differences.

Burns and Lovering (1988) submitted modified Cooper-Harper (MCH) rating data to an analysis of variance (ANOVA) and found the three-dimensional EADI to be significantly more desirable than the other displays. A preference for the three-dimensional display continued to be significant when the data were separated by attitude condition and analyzed. The subjective questionnaires indicated that a majority of the subjects preferred the three-dimensional EADI format for recovery from unusual attitudes. The pilots commented on the large FOV EADI format's ability to convey attitude information as the "big picture" both instantly and intuitively.

The incongruity between the objective and subjective data indicates that the large FOV EADI may indeed be superior to the other displays but the effect was not detected experimentally for whatever reason. Overall the pilots felt that the larger FOV EADI was the most useful display. The fact that the pilots reacted more favorably to the large FOV EADI may have significant impact on the success of the display in training as well as overall usability.

Global HUD

The G-HUD was developed by the author by integrating the symbology structures of the standard ADI and HUD into a single display format. The G-HUD (Figure 3) is a simulated three-dimensional transparent "half-ball" with an enhanced horizon line and directional alphanumeric designators. The simulated ball has a shaded outer surface which forms the outer edge of the display. The transparent imaginary inner surface of the virtual half-ball is formed by arched line symbology. The vertical lines that give the ball its apparent roundness indicate azimuth position in 30 degree increments. The G-HUD, like the ADI, is free to translate in all three of its axes. Like the HUD, upper and lower hemispheres of the display are indicated by continuous and segmented longitudinal line symbology. Aircraft attitude is read from the G-HUD by comparing a HUD-like OS--with elongated wings--to the artificial horizon, pitch-bars, and longitudinal heading lines.

The following sections discuss the G-HUD format in terms of the display features and principles on which it is based.

Pictorial realism. The G-HUD was designed to mimic the pictorial realism of the ADI. The G-HUD represents a spatial analog of the real world but, because of its transparency, the G-HUD can not include a sky/ground contrast to the same extent as the ADI. For the G-HUD, a sky/ground contrast is formed between continuous and segmented longitudinal azimuth lines. Due to the incomplete contrast between the two degrees of line integrity, it was necessary to draw an artificial horizon line. It is possible that lack of sky/ground contrast in the G-HUD may remove some of the natural orientation thought to be inherent in the ADI (Taylor, 1982).

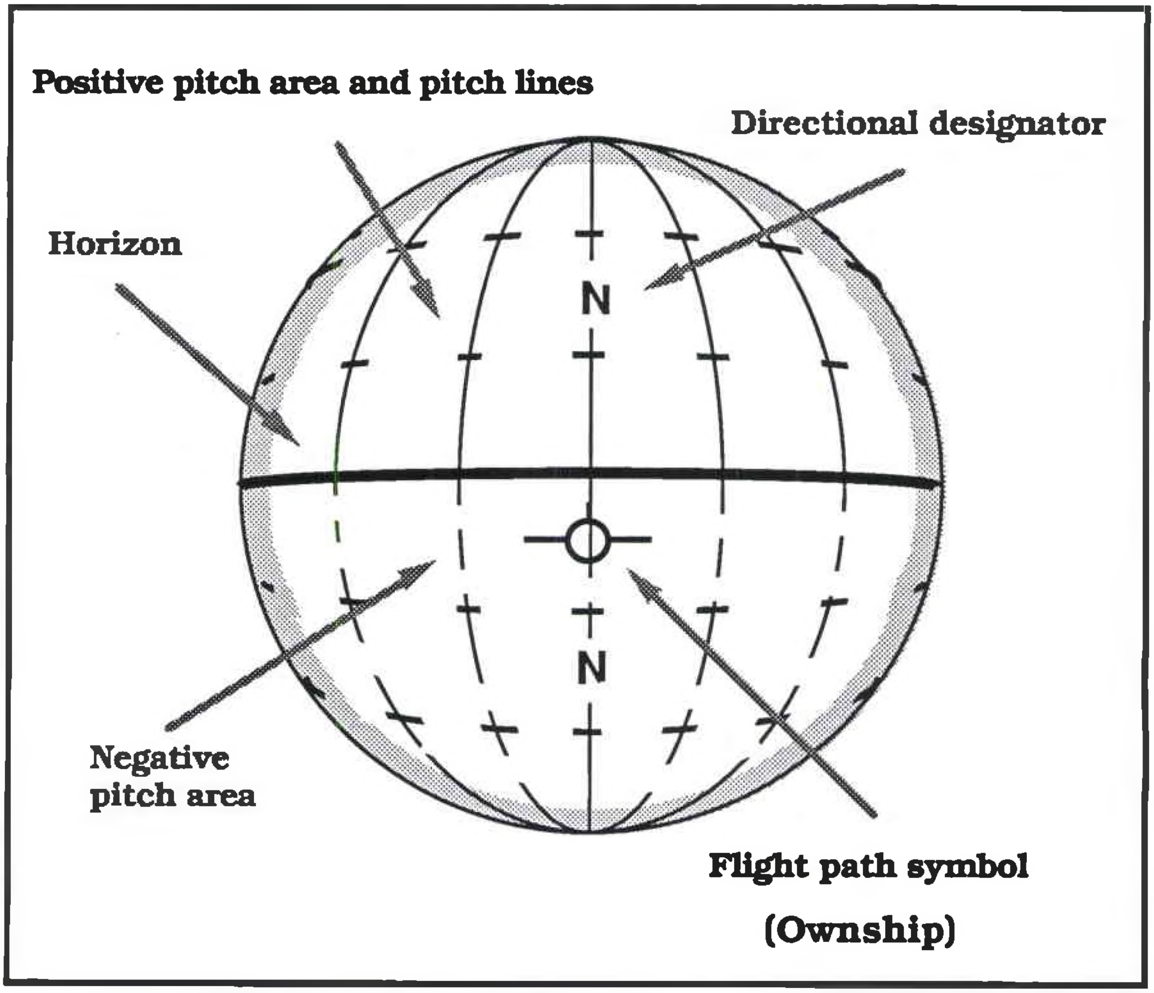


Figure 3. New Global Attitude/Direction Indicator Head-Up Format (G-HUD).

The G-HUD, like the ADI, is free to move about all three of its axes. The OS compared to the G-HUD artificial horizon and azimuth lines represents aircraft roll, pitch, heading status, and motion trend cues as a miniature pictorial representation of the OTW earth as if viewed through a port hole.

Objectness. The G-HUD represents aircraft attitude variables in much the same way as the ADI. Aircraft roll, pitch, and azimuth translation is integrated into a simulated ball which represents movement via a corresponding display roll, tilt, and/or spin. The most pronounced difference between the the ADI and the G-HUD is that the ADI simulates an opaque solid ball while the G-HUD simulates a transparent "wire frame" half-ball. Configuring line symbology in such a way that individual contours form the outline of various segments of a ball results in an objectness emergent feature. Because only the minimally necessary contours are used to form the objectness of the structure, the display continues to be see-through.

Like the ADI, the G-HUD incorporates strong contours which encompass the display symbology. A strong "edge" contour is formed by the visible shaded surface of the G-HUD. The objectness of the G-HUD may allow it to be processed as a single unitary object affording some of the ADI performance advantages. The G-HUD relies on as little redundancy as possible and separable features are not created. The G-HUD does not utilize a sky or ground-pointer and azimuth information is integrated into the display. The integral features of the display form a global percept that may require little actual contour information for interpretation.

Display compression. The G-HUD is a compressed display format that may benefit from a large FOV and decreased apparent motion. The G-HUD FOV includes 180 degrees of azimuth by 180 degrees in elevation. Like the ADI, the G-HUD's three dimensionality may be effective in presenting "big picture" spatial awareness. The enhanced horizon line is constantly in view as are both the up and down reference hemispheres (with the exception of +/-90 degrees pitch positions).

THE PRESENT STUDY

The present study involved the evaluation of a new ARI symbology structure concept. The G-HUD was developed by integrating the successful attributes of two existing display formats onto a single new display format. The G-HUD maintains characteristics in common with both the ADI and the HUD but forms a completely new display. Because the ADI and HUD have consistently demonstrated task dependent performance differences, such as significantly fastest recovery performance via ADI-like formats, it was expected that the G-HUD would result in performance consistent with that of the symbology features of the conventional displays. The ADI has proven to be successful as an aid to attitude awareness and the HUD affords awareness while the pilot maintains OTW contact with the world. The G-HUD should afford ADI-like attitude awareness while being HUD-like transparent. The present study addressed the question of whether there would be a significant performance increase or decrement incurred by the G-HUD symbology structure compared to the simulated conventional ADI and HUD ARI formats.

The present study evaluated the G-HUD by comparing the new format to the conventional ARIs in much the same manner as the the ADI and HUD have been compared to one another in the past (Kinsley, Warner, and Gleisner, 1985; Guttman, 1986; Burns and Lovering, 1988). Subjects performed experimental trials consisting of both a practice attitude maintenance task and an attitude recovery task. Subjects learned their specific display by performing an attitude maintenance task. The attitude maintenance task required that the subject maintain a SL display representation while being buffeted by a roll and pitch disturbance function. Subjects' root mean squared deflection error from SL comprised the attitude maintenance performance data. After training and performing to criterion on three different disturbance function amplitudes, subjects moved onto an unusual attitude recovery task. The unusual attitude recovery Experimental Session borrowed the method incorporated by Kinsley, Warner, and Gleisner 1985 and Guttman, 1986. Subjects were required to recover the simulated aircraft display representation to a status of SL from a predetermined unusual attitude. Attitude recovery performance data included recovery accuracy, decision time, and total recovery time.

It was hypothesized that performance for the G-HUD on the attitude maintenance and recovery tasks would be at the same level if not better than the conventional ADI ball and performance for both the G-HUD and the ADI would be superior to that of the HUD pitch ladder format.

CHAPTER 2

METHOD

Subjects

Subjects were drawn from a Logicon Technical Services Incorporated managed subject pool at Wright-Patterson AFB, OH. Subjects (N=45) were paid five dollars per hour for their participation. Out of the total of 45 subjects that attempted the initial Training Session, 15 subjects failed to reach performance criterion (one under the ADI format and 14 under the HUD format) and thus were not eligible to return for the Experimental Session. Of the 14 HUD subjects who failed to meet performance criterion, nine were returned to laboratory in order to rerun the Training Session under a different display from that which they originally participated. No experienced pilots or subjects, including those with any flight simulator experience, were permitted to take part in the study. Subjects were male, right-hand-dominant (self-reported), college students between the ages of 18 and 30 with normal or corrected to normal vision (20/20).

Subjects were given complete information concerning their participation and were free to withdraw from the study at any time. No deception was employed and subjects were thoroughly debriefed.

Design

Three independent groups each received a different display. Subjects participated in a familiarization training session followed by a data collection training session. The data collection training session can be described as a 3 by 3 mixed factorial with display format manipulated between-subjects and three levels of a disturbance function manipulated within-subjects. Following the two training sessions, each subject participated in 54 experimental trials, each of which was divided into two phases: an attitude maintenance phase and an attitude recovery phase.

During the attitude maintenance phase, subjects attempted to maintain a SL display representation while experiencing a roll and pitch disturbance function. For analysis purposes, the attitude maintenance trial phase can be represented as a one-way design in which display format is manipulated between-subjects. Root mean squared error in degrees deviation from straight and level (RMS) for both pitch and roll was recorded and analyzed separately. A 10 trial running RMS average or moving means window was also recorded for both roll and pitch.

The attitude recovery phase is represented by a 3 by 3 by 6 mixed factorial design formed by the three display formats crossed with three initial pitch positions (0, 55 and -55 degrees) and six initial roll positions (0, 60, 120, 180, -120 and -60 degrees). Roll and pitch were manipulated as within-subjects variables while display format remained the sole between-subjects variable. Recovery accuracy, decision time, and total recovery time were recorded as recovery phase dependent measures. Correct recovery accuracy was defined as an initial stick input that resulted in display deflection in the direction of the shortest route toward the horizon. In the case of an initial roll and/or pitch position of zero, the display must have been deflected beyond a +/-3.7 degree dead-band in order to have been registered as a purposeful, although incorrect input. The +/-3.7 degree area around the zero degree positions was empirically determined by pilot study to be unintentional noise. When subjects attempted to move the display in only one axis, deflection on the other axis averaged +/-3.7 degrees. Decision time was defined as the period between display presentation initiation and the first stick input. Recovery time was defined as the total time required to recover from the predetermined unusual attitude to a straight and level status. Straight and level for the present experiment meant that subjects had recovered the OS to within positive or negative five degrees in both pitch and roll for a duration of at least three seconds.

Apparatus

Trial stimuli were presented on a high resolution Silicon Graphics Iris 3130 graphics generator monitor. The monitor screen in terms of total size measured 29.5 cm by 39.5 cm and contains a matrix of 768 by 1024 pixels. All images were generated by the Iris at an update rate of 15 Hz. Trials were initiated by the subject and responses made on a Measurement Systems Inc. right hand two-axis force stick equipped with an index finger trigger. Trials and trial phases were initiated by the subject, by pressing the force stick trigger. Data were recorded by the Iris with a resolution of 15Hz. Data were transferred to a Digital Corporation Vax 11/785 to undergo reduction and analysis. Subjects were seated at a student type desk/chair that incorporated an attached right-arm half desktop. The force stick was mounted in a hole cut in the center of the desktop. This chair configuration afforded the subject full arm support while his right hand was on the force stick. The front of the desktop was clamped to the tabletop on which Iris video monitor and keyboard were set. The clamping insured that the desktop did not shift due to stick inputs. Subjects were seated in front of the Iris video monitor so that the display viewing distance was approximately 40cm.

Disturbance Function

The roll and pitch disturbance function file was developed so that no apparent pattern existed. The disturbance function was created for a previous study (Osgood and Venturino, 1990) and modified accordingly for use in the present study. The disturbance time history spectrum was gaussian random with a sample length of 1024 at 60Hz. The spectrum was Butterworth filtered with a cut off frequency of 0.5Hz (1 pass-12dB per octave rolloff). The filtered time history was then fast fourier transformed (forward full range), converted from real and imaginary to modulus and phase, and then converted to ASCII for the simulation to use as the disturbance. The amplitude of the forcing function was scaled in order to produce the desired RMS values. Both roll and pitch were driven by the same basic function loop but utilized different starting points within the loop. Five different starting points were designated within the 17 second disturbance function. A different starting point was randomly selected for each trial with the stipulation that roll and pitch could not share the same starting point on any particular trial. The disturbance function starting points were selected so that, at trial initiation, roll and pitch deflection was zero degrees (Figure 4). The disturbance function acted as a direct display driver which was designed to produce a constant predictable RMS in both roll and pitch for each of three disturbance amplitude levels (low=10, medium=15 and high=20 degrees mean display deflection).

Displays

ADI. The ADI format was a computer generated simulation of an inside-out referenced electromechanical ADI found in the cockpits of modern fighter aircraft (Fig. 1). The ADI was shaped like a three dimensional ball (23.3 deg) and has roll, pitch and directional symbology printed on the ball surface. The OS on the ADI was the classic

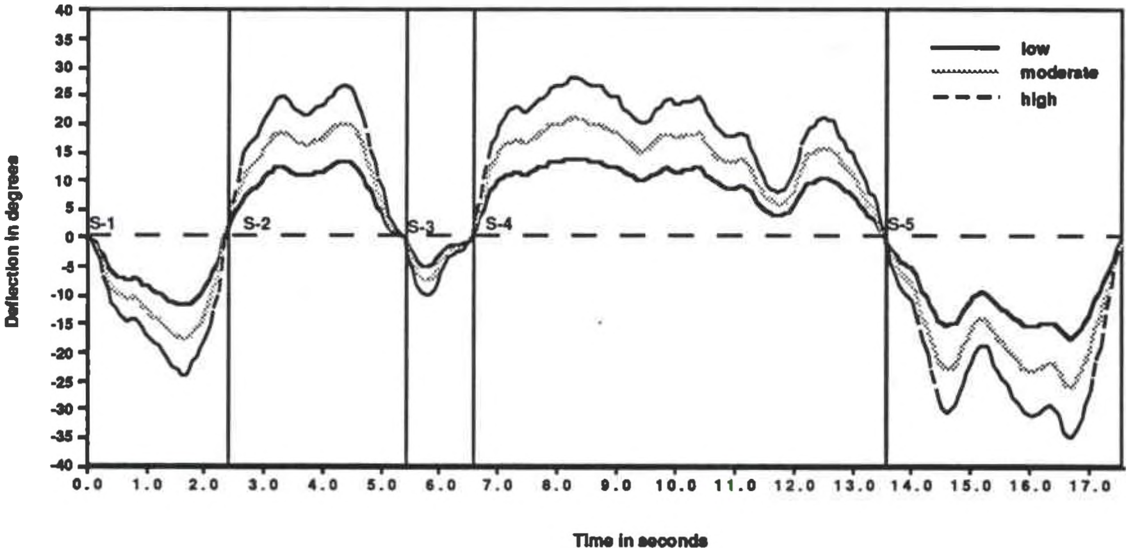


Figure 4. Disturbance Function Amplitude Levels and Maintenance Task Starting Points.

waterline schematic "W" (1.6 by 12.8 deg). The waterline symbol was fixed with respect to the instrument panel and the display ball rotated about it in roll. The ball was free to rotate about all three of its axes. At any one time, approximately 90 degrees of aircraft attitude was represented by the display (+/-45 degrees from the waterline symbol) in the vertical component, and approximately 90 degrees in the horizontal component (45 degrees right and left of ownship heading). Pitch scaling was ticked with horizontal pitch bars in 10 degree increments from +/-10 degrees to +/-60 degrees pitch. A roll and bank scale along the bottom of the display represented aircraft bank via ground pointer and scale depicting from zero degrees bank to right or left 90 degrees. The upper half or hemisphere of the ADI ball was light in shade (green on the monitor) depicting the sky and positive pitch angles. The bottom hemisphere was dark (black on the monitor) to represent the earth and negative pitch angles. An artificial horizon line was printed where the light and dark half of the display contrasted. The horizon line was ticked at five degree increments depicting aircraft heading.

Aircraft attitude information was obtained from the ADI by comparing the OS location to the ball orientation. For example, when the subject saw the waterline symbol superimposed over the horizon line and the light or sky portion of the ball was positioned on the upside of the OS, then the display represented straight and level flight. If only the light sky portion of the ball was seen, then the display represented an extreme nose-high attitude or vertical flight. Likewise, a nose-low attitude was represented by the dark half of the ball. At extreme attitudes, longitudinal lines on which the pitch scale bars were fixed, appeared to converge and meet at both zenith and nadir positions on the ADI ball. At an attitude of +/-90 degrees the ownship symbol was superimposed on the point where these lines came together and formed what appeared to be spokes emanating from a hub. The longitudinal lines gave heading information and cardinal headings were tagged with an appropriate (2.6 deg) heading symbol (i.e. N, E, S, or W).

HUD. The HUD format was made up of a culmination of symbology that is currently being utilized in operational displays as well as some symbols that have been modified and are being evaluated for use in future cockpits (Reising, Zenyuh, and Barthelemy, 1988).

The pitch-ladder HUD (Fig. 2) included a bold centralized caging horizon line and "near" parallel articulated lines (one half the angle of incidence) depicting pitch angle in five degree increments up to +90 degrees (zenith) and -90 degrees (nadir). The ladder subtended 6.3 by 15 degrees of visual angle. Pitch line bars formed the rungs of the ladder and, relative to a fixed OS (2.5 deg), appeared to move down as the aircraft increased pitch and up as the aircraft decreased pitch. The aircraft symbol seemed to "climb" the pitch ladder as an increase in aircraft pitch continued and vise versa for a pitch decrease. Positive and negative pitch angles were distinguished by continuous or broken pitch-bar lines. The broken pitch lines designated a negative pitch angle while the continuous lines are necessary above the horizon line to maintain a high level of contrast against the sky.

Pitch-bar tags were also included into the design and were oriented toward the horizon at the inner pitch line extremities on both positive and negative pitch lines. Further roll information was available from a wide angle bank scale (41 deg) along the bottom of the display and its associated ground-pointer. Aircraft heading appeared as tape and scale along the top of the display. Aircraft heading was whatever digital value appeared closest to the center of the heading scale.

The OS in the pitch-ladder HUD format was a schematic representation of ownship's tail. Included was a single vertical stabilizer and a conventional pair of horizontal stabilizers which indicated the ownship's respective axes. The observer was oriented in relation to the display so he is was looking straight down the aircraft's longitudinal axis. For the proposes of the present study, the aircraft symbol was fixed in the center of the display while, during maneuvering, the pitch ladder representation of the outside world moved about it. This formed a pure aircraft coordinate referenced or inside-out display/motion relationship (Johnson and Roscoe, 1972).

G-HUD. The G-HUD format was a combination of ADI and HUD symbology and structure. The G-HUD was a computer generated virtual three-dimensional half ball with symbology printed on the inside as well as the outside of the ball (Fig. 3). The overall display subtended 22.6 degrees visual angle. The half ball was oriented so that an observer viewed the concave portion or inner ball as if the ball was cut open at just less than half (the clipping plane is located at +/-85 degrees). The outside surface of the ball was shaded while the inside of the ball was transparent. As the display translated, the clipping plane remained fixed so that the subject continued to look directly into the ball. Like the ADI, the G-HUD incorporated longitudinal lines to represent heading and a single highlighted (twice as thick as the other lines) artificial horizon line. A vertical heading line appeared every 30 degrees of azimuth. Cardinal headings were labeled with a (1.9 deg) heading symbol (i.e. N, E, S, or W) centered +/-55 degrees from the horizon. Each heading line was tagged with pitch-bars at 30 and 60 degrees. The longitudinal heading lines were continuous above the artificial horizon (depicting high pitch angles) and segmented below the horizon (low pitch angles).

The OS for the G-HUD was identical to that utilized in the conventional HUD with the exception of elongated "wings" (twice the length of the wings on the conventional ADI (4.6 deg)). As for both the ADI and HUD, the G-HUD OS was fixed with respect to the aircraft while the display translated about it. The G-HUD OS was oriented on the inside of a virtual ball so that all three of the ball's axes run through the OS. As the aircraft maneuvers, an observer compared the position of the OS to the inside of the G-HUD ball in order to obtain attitude information.

Procedure

Upon entering the research facility, subjects read and signed a standard consent form. The consent form included a brief description of the study procedure. The consent form was written to inform subjects of the risk, responsibility, and liability involved in the study. Subjects were asked to read and sign the consent form before participation in the study was authorized.

After reading and signing the consent form, subjects were asked to read a set of written instructions which familiarized them with their particular display format as well as the basic experimental procedure. A different instruction set was used for each of the three display formats. The three instruction sets differed only in what was minimally necessary between the three display types. Subjects were permitted to ask questions at any time during the instruction set as well as the remainder of the practice and experimental sessions.

The study was divided in to three distinct sessions of which the first two were run on the first day (Day 1) and the third on a return visit (Day 2).

Day 1. Session 1 of Day 1 consisted of a "free-flight" period intended to familiarize subjects with the interaction between the force stick and the movement of their particular display. Session 1 was performed in absence of any disturbance. The general features of the display were again explained as well as the experimental procedures. When the subject expressed satisfaction with the display familiarization and all questions were answered, Session 1 was complete. Session 1 lasted approximately five minutes. After Session 1, subjects read instructions on how to think about and complete the Modified Cooper-Harper (MCH) subjective workload rating (Wierwille and Casali, 1983). See Appendix C.

Session 2 of Day 1 was comprised of an attitude maintenance task intended to train subjects to accurately interpret and react to their particular attitude display. Subjects were instructed--both in the pre-experiment instructions and prior to each trial on the monitor itself--to press the force stick trigger in order to initiate each trial. Subjects were instructed to view the computer generated ARI and use the force stick to dampen any disturbance perturbation and maintain as close to SL as possible. Each trial lasted for 17 seconds.

Each subject began Session 2 by performing first the maintenance task for the low disturbance amplitude followed by the moderate amplitude and finally the high disturbance amplitude. Subjects were required to obtain a 10 trial RMS average (moving means window) of 14 degrees or less in roll and 13.5 degrees or less in pitch before moving onto the next disturbance amplitude and ultimately Day 2. This procedure insured that the subjects were all trained to a minimum level of performance as well as guaranteed that all subjects experienced at least 30 maintenance task trials (10 of each disturbance amplitude). The moving means performance criteria for Session 2 were empirically derived from pilot study data. Any subject who failed to reach criterion within 120 trials of a single disturbance amplitude was dropped and replaced. Subjects were informed of the criterion stipulation and were given continuous performance feedback including trial RMS and the moving means scores. The number of trials to reach criterion, individual trial roll and pitch RMS, the moving means windows, and MCH ratings--performed after each of the three disturbance levels--were recorded as Session 2 dependent measures. Upon reaching criterion for the low, moderate and high disturbance levels respectively, subjects were dismissed and asked to return for Day 2 (Session 3) data collection.

Day 2. The Day 2 Session consisted of 12 practice trials followed by 54 experimental trials each of which included an attitude maintenance phase followed by an attitude recovery phase. For the attitude maintenance phase, subjects were instructed--prior to Session 3--to press the force stick trigger in order to initiate each trial. Similar to Session 2, subjects were instructed to view the computer-generated ARI and use the force stick to dampen any disturbance perturbation and maintain as close to SL as possible. After the 17 second maintenance phase, subjects initiated the attitude recovery phase. Subjects viewed a static presentation of the ARI which represented one of the predesignated roll/pitch position combinations. Subjects were to dynamically return the display to a status of SL as quickly as possible in the direction of the closest horizon. Subjects were instructed that a successful recovery requires that the OS remain within positive and negative five degrees in both pitch and roll for a duration of at least three seconds. Subjects were informed of the relative meaning of the performance feedback they received at the completion of each trial.

The 12 practice trials of Session 3 were intended to familiarize subjects with the experimental procedure as well as warm them up for the experimental trials. The practice trials were comprised of the high disturbance amplitude for the attitude maintenance phase. For each trial, subjects saw a "get ready" message as well as instructions to press the trigger on the force stick to begin a trial. The subject then performed the attitude maintenance task. After 17 seconds the display went blank denoting the end of the first trial phase. A message on the monitor instructed subjects to press the stick trigger in order to move onto the recovery phase. At this point the computer read the force stick input to insure that it was in a neutral position before the display was presented. If the stick was not neutral, subjects saw a "ZERO FORCE STICK" command on the monitor. This acted as a reminder for subjects to lighten their grip on the stick so that the attitude recovery phase would begin with zero stick input. Subjects were not permitted to take their hand off the force stick during a trial. The force stick had to be within its neutral dead band for two consecutive seconds before the ARI was presented. The attitude recovery phase, for the practice trials, was comprised of six unique recovery conditions formed by crossing two pitch positions (30 and -30 degrees) and three roll positions (45, 180, and -45 degrees) with two replications.

The 54 experimental trials proceeded in much the same manner as the practice trials with the exception that the trials were formed by crossing three pitch positions (0, 55 , and -55 degrees) and six roll positions (0, 60, 120, 180, -60 , and -120 degrees) with three replications. As in the practice trials, subjects experienced only the high disturbance amplitude for the attitude maintenance phase of each trial. After the experimental trials, subjects were thoroughly debriefed and released.

Balancing. There were 10 subjects for each display condition. Each subject participated in one experimental session of 54 trials formed by crossing the three pitch positions with the six roll positions (18 trials) and three replications of each unique unusual attitude condition. For the attitude maintenance phase, each amplitude level (low, medium and high) was repeated 18 times. The conditions were randomly presented within three blocks of 18 trials which contained equal proportions of each unique condition. Replications were collapsed and not treated as an experimental variable.

CHAPTER 3

RESULTS

The following experimental results include three main sections that essentially represent the results of three individual experiments. The sections include the Training Session, the Experimental Session and the Training Rerun Session. The Training Session represents experimental results of 30 subjects that successfully met the performance criterion on an attitude maintenance task. These subjects returned to the laboratory to participate in the Experimental Session and its associated attitude maintenance and unusual attitude recovery tasks. Out of the total of 45 subjects that attempted the initial Training Session, 15 subjects failed to reach performance criterion and thus were not eligible to return for the Experimental Session. Of the 15 subjects who failed, 14 failed to reach the performance criterion on the HUD format. The remaining subject failed to reach criterion under the ADI format. Of the 14 HUD subjects who failed to meet performance criterion, nine were returned to laboratory in order to rerun the Training Session under a different display from that which they originally participated. Four of the return subjects participated under the G-HUD format and the remaining five subjects performed the rerun trials using the ADI. These results are reported as the Training Rerun Session.

Training Session

Three variables were of interest for the subject Training Session. These variables included the number of trials each subject required to reach performance criterion, mean performance in terms of RMS error in roll and pitch, and the MCH ratings that subjects assigned to the task after completing each of the three disturbance amplitude levels. Each of the variables were analyzed for any statistical differences existing between the display formats, the disturbance amplitude levels, and their associated interactions.

Trials to reach criterion. A 3 by 3 mixed factorial analysis of variance (ANOVA) procedure was performed on the trials to reach performance criterion data in order to test the effects of display format and the disturbance amplitude levels. A significant main effect for display format was indicated (F (2, 27) = 4.58, p = 0.0193). Display format accounted for 8.32% of the variance. A Tukey HSD Post Hoc comparison (Tukey, 1977) found significant differences between the display formats such that the G-HUD subjects, on average, took fewer trials (mean = 17.40 trials) to meet the training performance criterion than HUD subjects (mean = 36.80 trials). The mean number of trials for ADI subjects (mean = 27.73 trials) to meet performance criterion was not significantly different from that of either the G-HUD or HUD subjects. Figure 5 illustrates this effect. See Appendix A for the ANOVA summary tables.

A significant main effect was found for the disturbance amplitude independent variable (F (2, 54) = 32.82, p = 0.0001). disturbance amplitude accounted for 31.29% of the variance. A Tukey HSD test found that the high disturbance amplitude condition (mean = 48.27 trials) required significantly more trials than both the medium and low amplitudes. The medium (mean = 21.83 trials) and low (mean = 11.83 trials) amplitudes

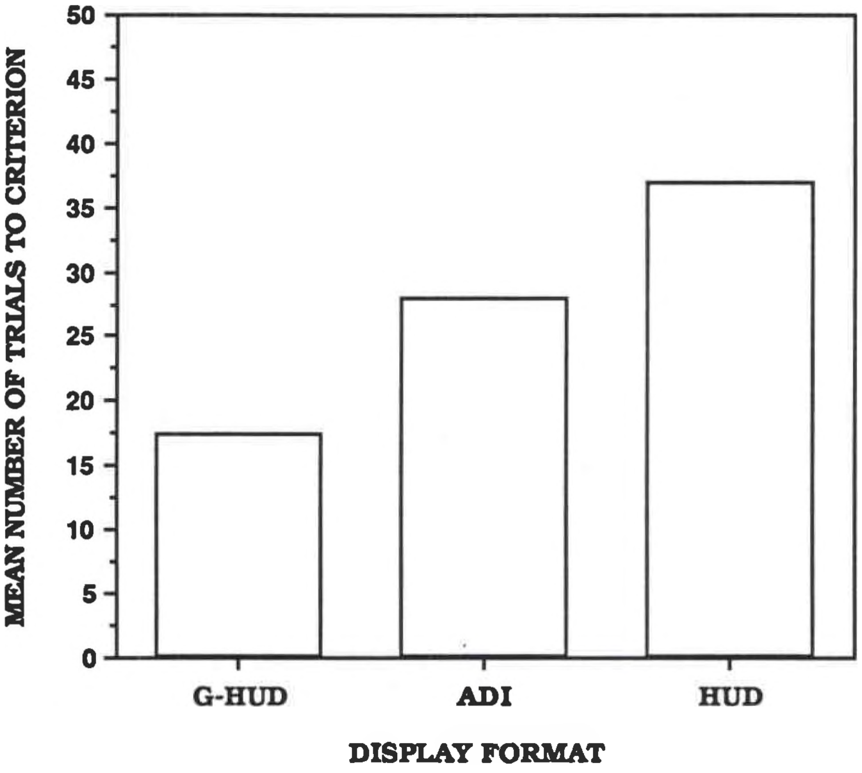


Figure 5. Mean Number of Trials to Reach Criterion as a Function of Display Format During the Training Session.

did not significantly differ from one another. This effect can be seen in Figure 6.

A significant interaction among display formats and disturbance amplitude was found (*F* (4, 54) = 5.32, *p* = 0.0011). The display format by disturbance amplitude interaction accounted for 10.14% of the variance. An analysis of simple effects revealed that the only significant difference between display format was located at the high disturbance amplitude level (*F* (2, 27) = 7.94, *p* = 0.0019). Within the high disturbance amplitude, the G-HUD (mean = 25.70 trials) took significantly fewer trials to reach performance criterion than the HUD format (mean = 73.80 trials). Figure 7 illustrates the nature of the interaction. The ADI format (mean =45.3 trials) did not significantly differ from either the G-HUD or the HUD format at the high disturbance level. See Appendix A for the ANOVA summary tables.

RMS error performance. A 3 by 3 ANOVA procedure was performed on the per trial roll and pitch RMS error data. The data were averaged across trials.

Two statistically significant effects were found with roll RMS as a dependent measure. The first was a main effect of display format (*F* (2, 27) = 3.94, *p* = 0.0314). A Tukey HSD test found that performance in roll was significantly better with the G-HUD (mean = 12.12 deg) than with the HUD (mean = 14.81 deg). The ADI (mean = 13.37 deg) format did not differ significantly from either the G-HUD or the HUD. This effect (Figure 8) accounted for 12.10% of the variance. The second effect was that of a disturbance amplitude main effect (*F* (2, 54) = 27.23, *p* = 0.0001). A Tukey HSD test found a significant difference between all three of the disturbance amplitude levels. Low disturbance amplitude resulted in the lowest roll RMS (mean = 11.62 deg), the medium disturbance amplitude level resulted in the next highest roll RMS (mean = 13.34 deg), and the high disturbance amplitude level resulted in the highest roll RMS (mean = 15.35

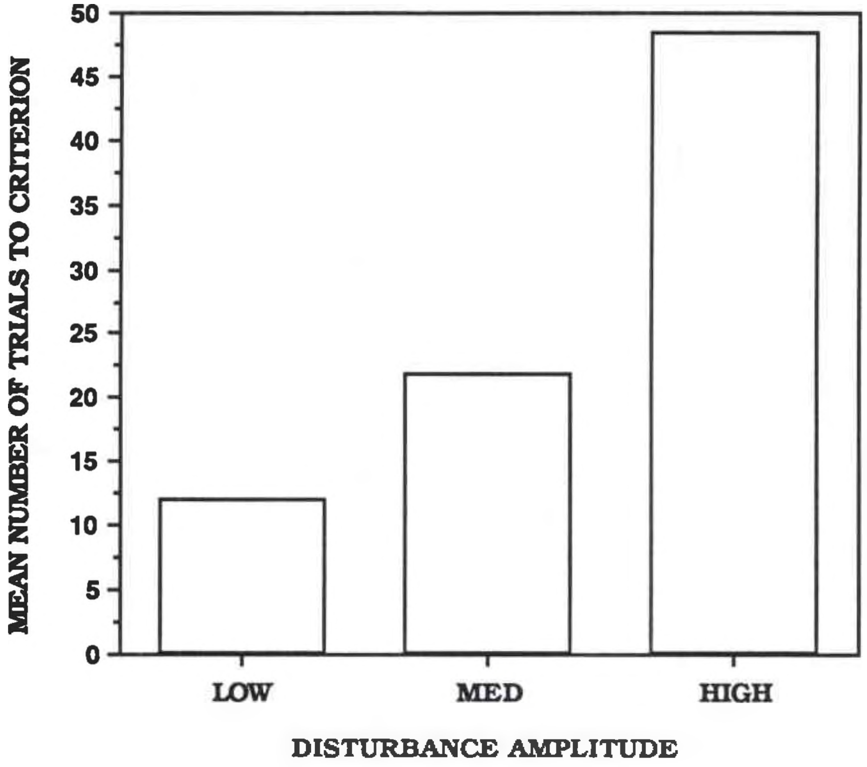


Figure 6. Mean Number of Trials to Reach Criterion as a Function of Disturbance Amplitude During the Training Session.

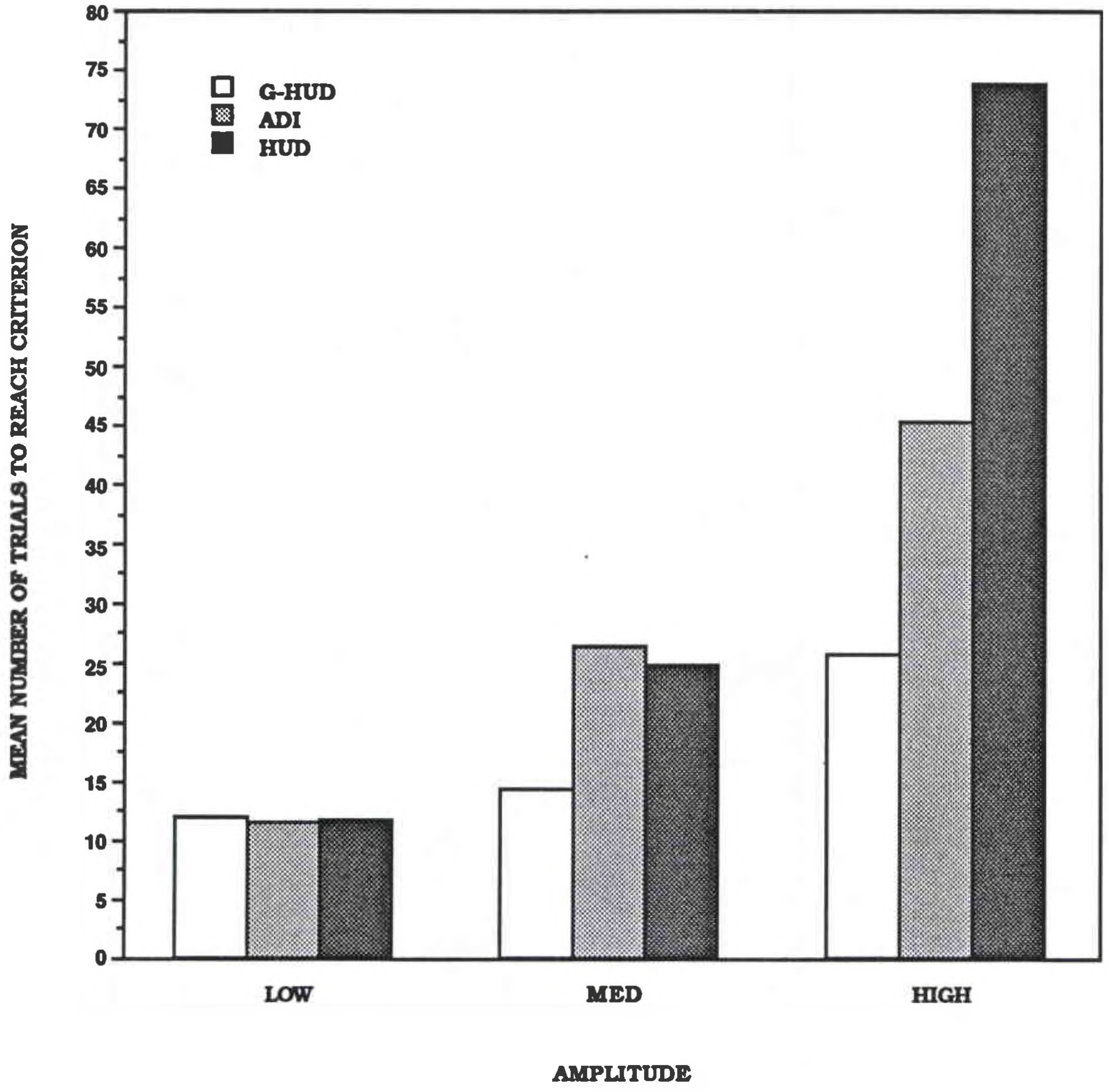


Figure 7. Mean Number of Trials to Reach Criterion as a Function of Disturbance Amplitude and Display Format During the Training Session.

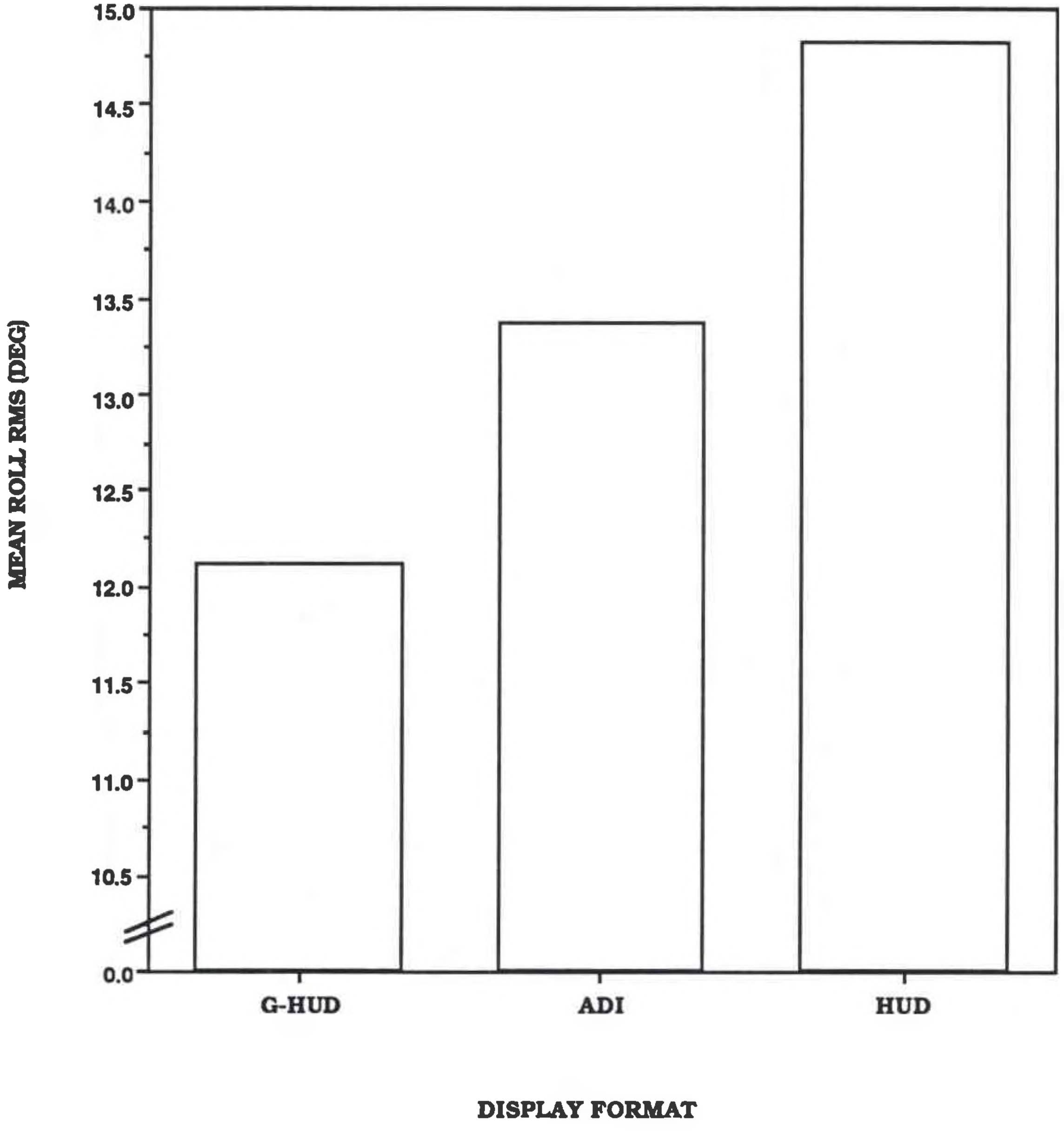


Figure 8. Mean Roll Root Mean Squared Error as a Function of Display Format During the Training Session.

deg). The disturbance amplitude effect accounted for 23.19% of the variance. No interaction between display format and disturbance amplitude was indicated. See Appendix A for the ANOVA summary tables.

As with the roll RMS dependent measure, two statistically significant effects were found with pitch RMS as a dependent measure. The first was a main effect of display format (F (2, 27) = 3.70, p = 0.0380). A Tukey HSD test found that performance in pitch was significantly better with the G-HUD (mean = 10.60 deg) than with the HUD (mean = 12.89 deg). The ADI (mean = 11.56 deg) format did not differ significantly from either the G-HUD or the HUD (Figure 9). This effect accounted for 9.37% of the variance. The second effect was that of a disturbance amplitude main effect (F (2, 54) = 86.86, p = 0.0001). A Tukey HSD test indicated significant differences such that the low disturbance amplitude resulted in the lowest roll RMS (mean = 9.25 deg), the medium disturbance amplitude level resulted in the next highest roll RMS (mean = 11.71 deg), and the high disturbance amplitude level resulted in the highest roll RMS (mean = 14.09 deg). The disturbance amplitude effect accounted for 41.83% of the variance. No interaction between display format and disturbance amplitude was found. See Appendix A for the ANOVA summary tables.

Modified Cooper-Harper. A 3 by 3 mixed factorial ANOVA procedure was performed on the MCH ratings. The MCH ratings were recorded for the maintenance task at the end of each of the disturbance amplitude levels. A main effect for disturbance amplitude was the only statistically significant difference indicated for the MCH ratings (F (2, 54) = 24.53, p = .0001). The disturbance amplitude effect accounted for 23.16% of the variance. A Tukey HSD Post Hoc comparison found that each of the three amplitudes were significantly different from one another. Low amplitude resulted in the lowest mean MCH rating (mean = 3.17 MCH). Medium amplitude was rated next

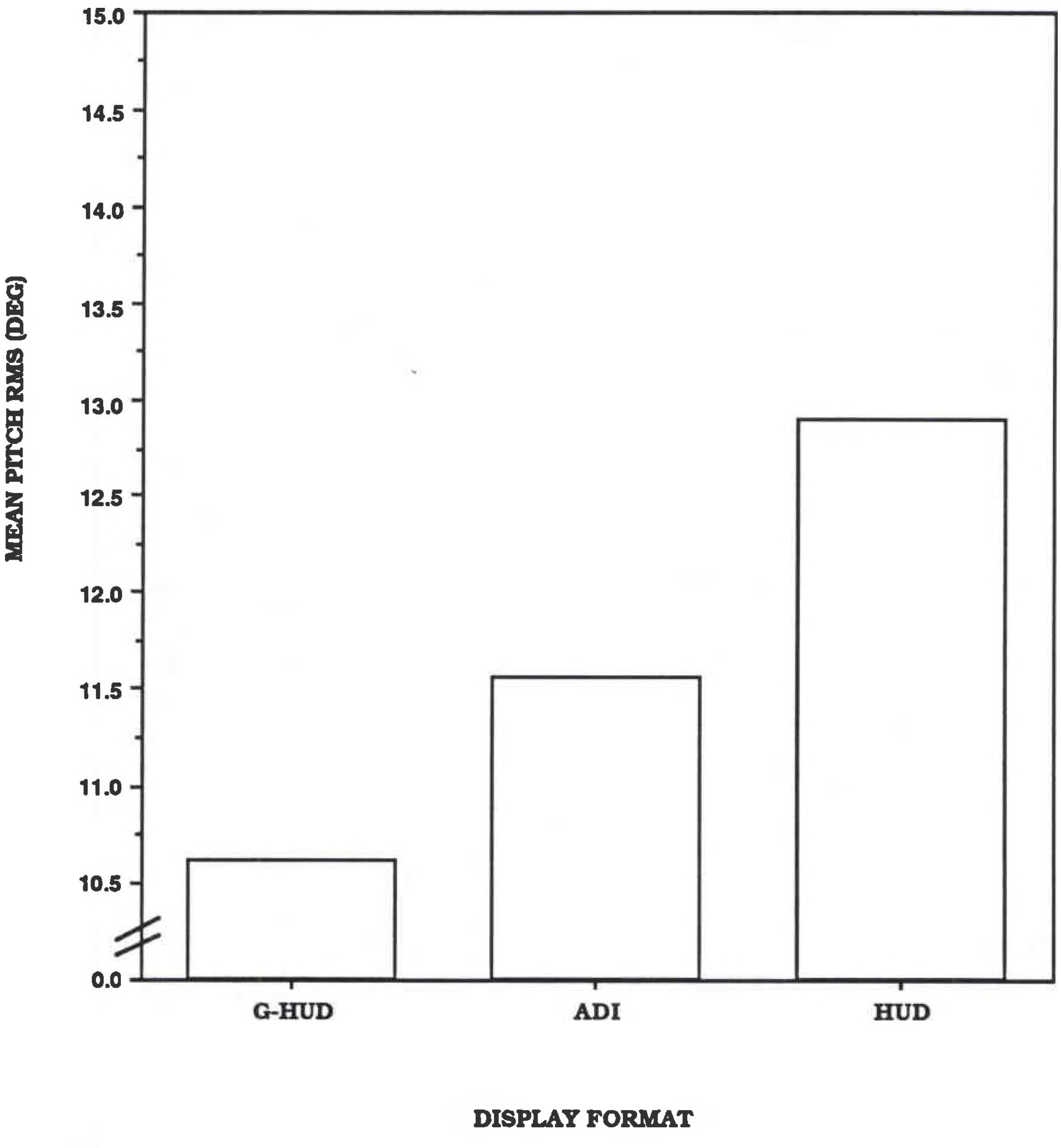


Figure 9. Mean Pitch Root Mean Squared Error as a Function of Display Format During the Training Session.

highest (mean = 4.30 MCH) and logically the highest amplitude condition was rated with the highest overall mean MCH score (mean = 5.63 MCH). Figure 10 represents the disturbance amplitude effect.

The MCH ratings for display format failed to reach statistical significance as did the interaction between disturbance amplitude and display format. The mean MCH rating for the G-HUD format was 4.27 while the ADI and HUD format resulted in mean ratings of 3.93 and 4.90 respectively.

Experimental Session

Trials in the Experimental Session consisted of an attitude maintenance task phase followed by an unusual attitude recovery task phase. The attitude maintenance task consisted of only the high disturbance amplitude which was identical to that used during the Training Session. The effect of display format on RMS error in roll and pitch was of interest for the attitude maintenance task. Decision time, recovery time, and accuracy were of interest for the unusual attitude recovery task phase.

Maintenance task phase. Two 3-way ANOVAs were performed on the maintenance task data. The first ANOVA tested roll performance and the second tested pitch performance. A main effect of display format in roll performance failed to reach statistical significance. Mean roll performance on the G-HUD was 13.04 degrees while roll performance for the ADI was 13.71 degrees. Mean roll performance with the HUD was 13.41 degrees. For pitch performance, there was a significant display format main

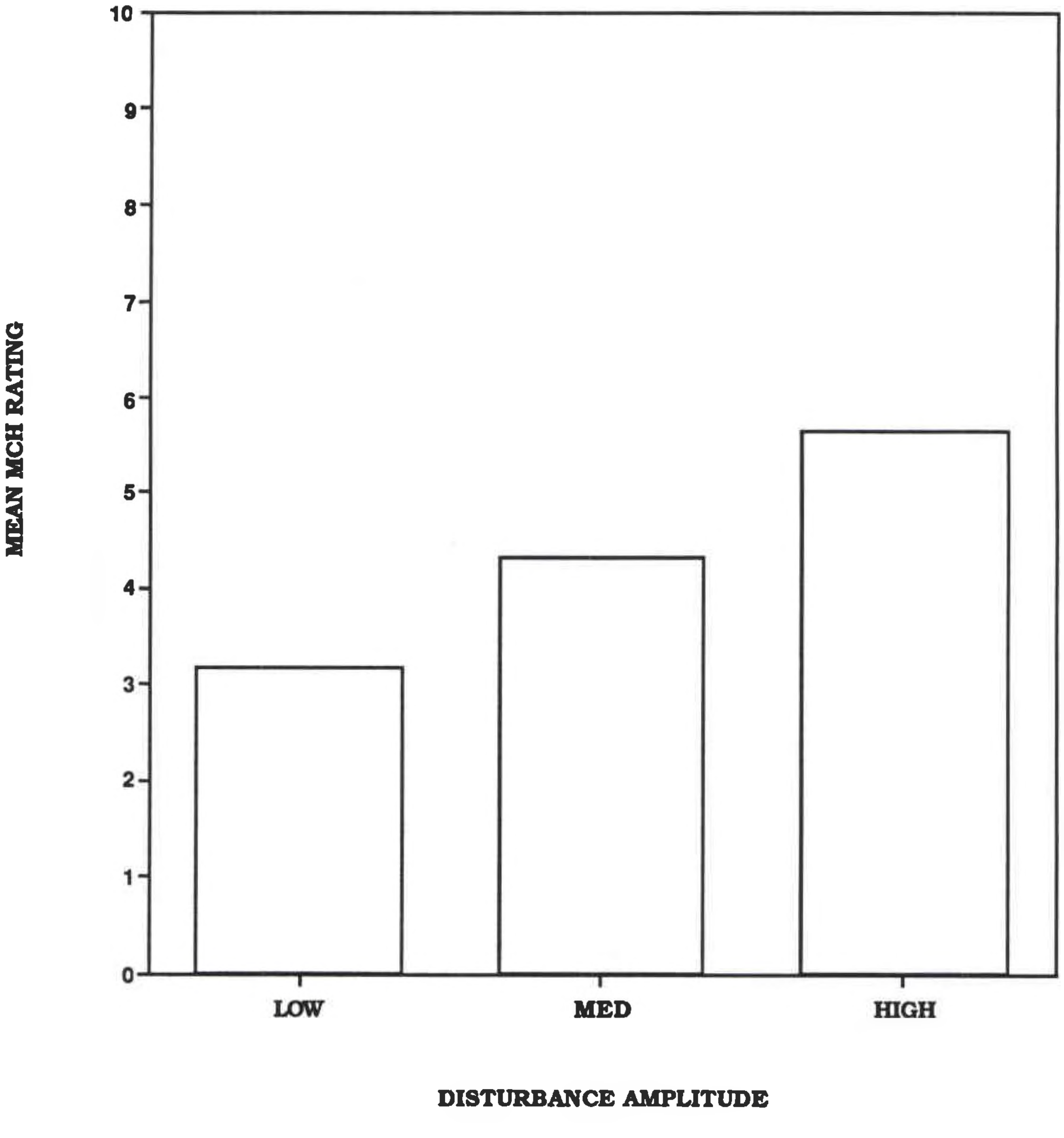


Figure 10. Mean Modified Cooper-Harper Rating as a Function of Disturbance Amplitude During the Training Session.

effect (F (2, 27) = 5.15, p = .0128). The display format effect accounted for 30.41% of the variance. A Tukey HSD comparison indicated a significant difference between the G-HUD and the HUD format. Pitch performance with the G-HUD format (mean = 12.31 deg) and the ADI format (mean = 12.34 deg) was significantly better than performance with the HUD format (mean = 13.81). There was no significant performance difference between the G-HUD and the ADI. This effect can be seen in Figure 11.

Recovery task phase. The recovery task phase was formed of three independent variables. Included were: display format, initial roll position (-120, -60, 0, 60, 120, and 180 degrees), and initial pitch position (-55, 0, 55 degrees). A first pass analysis of variance was performed to determine if there were any statistical differences between positive and negative direction--right/left roll and up/down pitch dimensions--for decision time and recovery time. It was found that no significant differences existed between positive and negative direction within a dimension. For reasons of simplification it was decided to perform and interpret the remainder of the analyses using the absolute values of direction within a dimension. Positive and negative direction data were collapsed and averaged.

Four 3 by 2 by 4 ANOVAs were performed on the recovery task data. The first analysis tested the decision time dependent variable for differences attributable to display format, initial roll, initial pitch, and their associated interactions. The recovery time variable was analyzed similarly to decision time. Two final ANOVA procedures tested roll and pitch initial input accuracy differences due to display format.

The decision time variable consisted of the time between the static presentation of the display--in it's unusual attitude--and the first stick input made by the subject. The analysis indicated main effects for both initial roll (F (3, 81) = 76.26, p = .0001) and

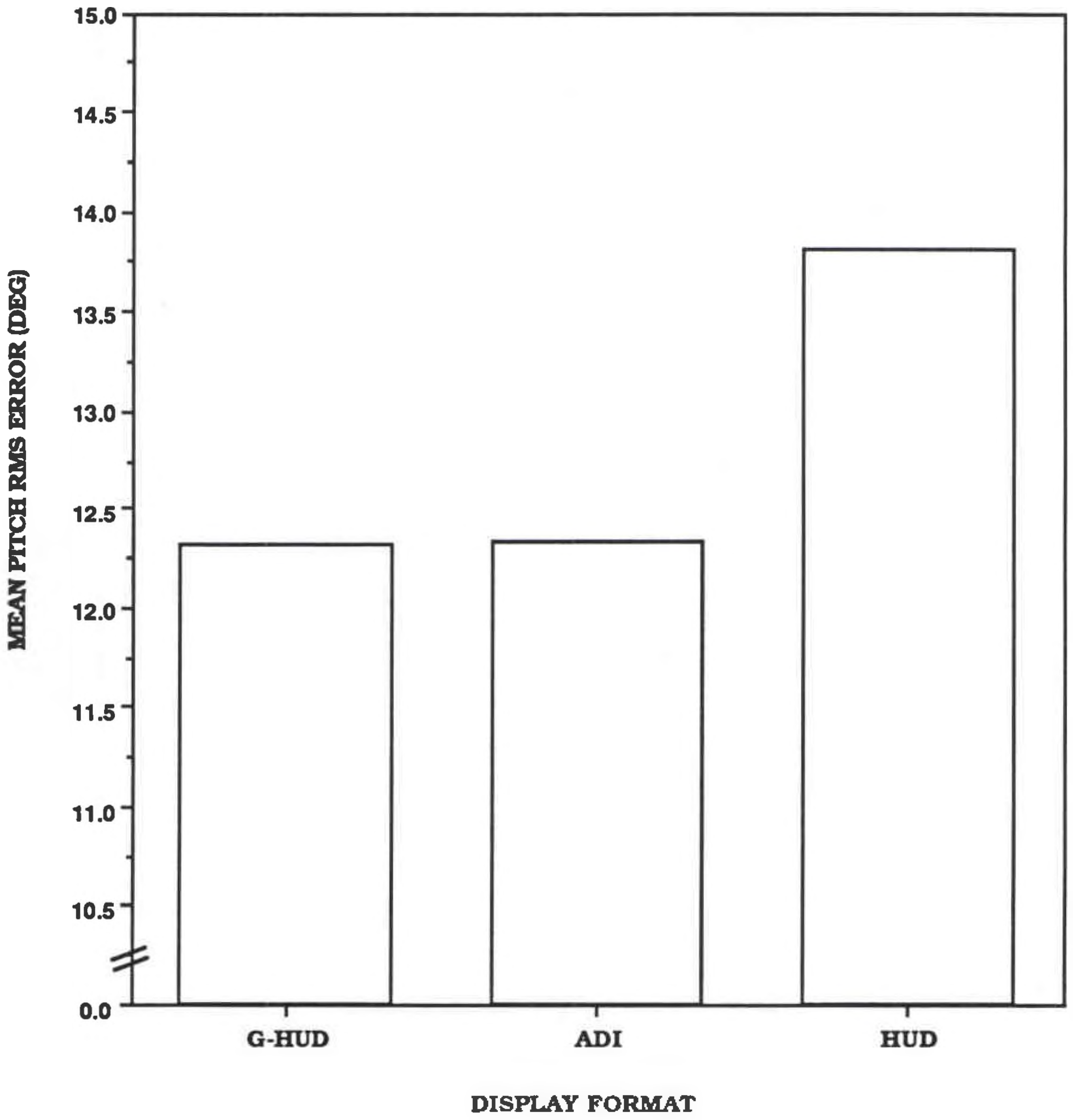


Figure 11. Mean Pitch Root Mean Squared Error as a Function of Display Format During the Experimental Session.

initial pitch (*F* (1, 27) = 141.74, *p* = .0001) but not for display format. Display format resulted in the following mean decision times: G-HUD mean decision time was .780 seconds, the ADI resulted in a mean decision time of .744 seconds, and the HUD resulted in a mean decision time of .825 seconds. A Tukey HSD comparison found significant differences between the initial roll means such that the zero degree roll level resulted in the fastest decision time. Decision time for the 60 degree roll level was significantly faster than the 120 degree and 180 degree levels but was significantly slower than the zero degree initial roll level. There was no statistically significant difference between the 120 and 180 degree initial roll levels. Figure 12a illustrates the effects. For initial pitch, a Tukey HSD comparison found that the zero degree initial pitch level resulted in a significantly faster mean decision time than the 55 degree initial pitch level. Initial roll and initial pitch accounted for 23.01% and 11.74% of the variance respectively. Figure 12b illustrates the effects. A significant interaction was found between initial roll and initial pitch (*F* (3, 81) = 111.58, *p* = .0001). The interaction accounted for 16.41% of the model variance. Figure 13 illustrates the interaction effect. An analysis of simple effects indicated a significant decision time difference for roll only at the initial pitch level of zero degrees (*F* (3, 116) = 69.11, *p* = 0.0001). Zero degrees roll resulted in the fastest decision time (mean = 0.03 sec) while the 60 degree roll level took significantly longer (mean = 0.73 sec). The 60 degree initial roll level resulted in a significantly shorter decision time than 180 degree initial roll (mean = 0.93 sec). The 120 degree initial roll level (mean = 0.87 sec) did not significantly differ from either the 60 or 180 degree levels. See Appendix A for the ANOVA summary tables.

A three-way interaction was found between initial roll, initial pitch, and display format (*F* (6, 81) = 4.39, *p* = .0007). The interaction accounted for only 1.28% of the variance. An analysis of simple interactions for display format by initial pitch at each level of initial roll indicated only one significant two-way interaction within the

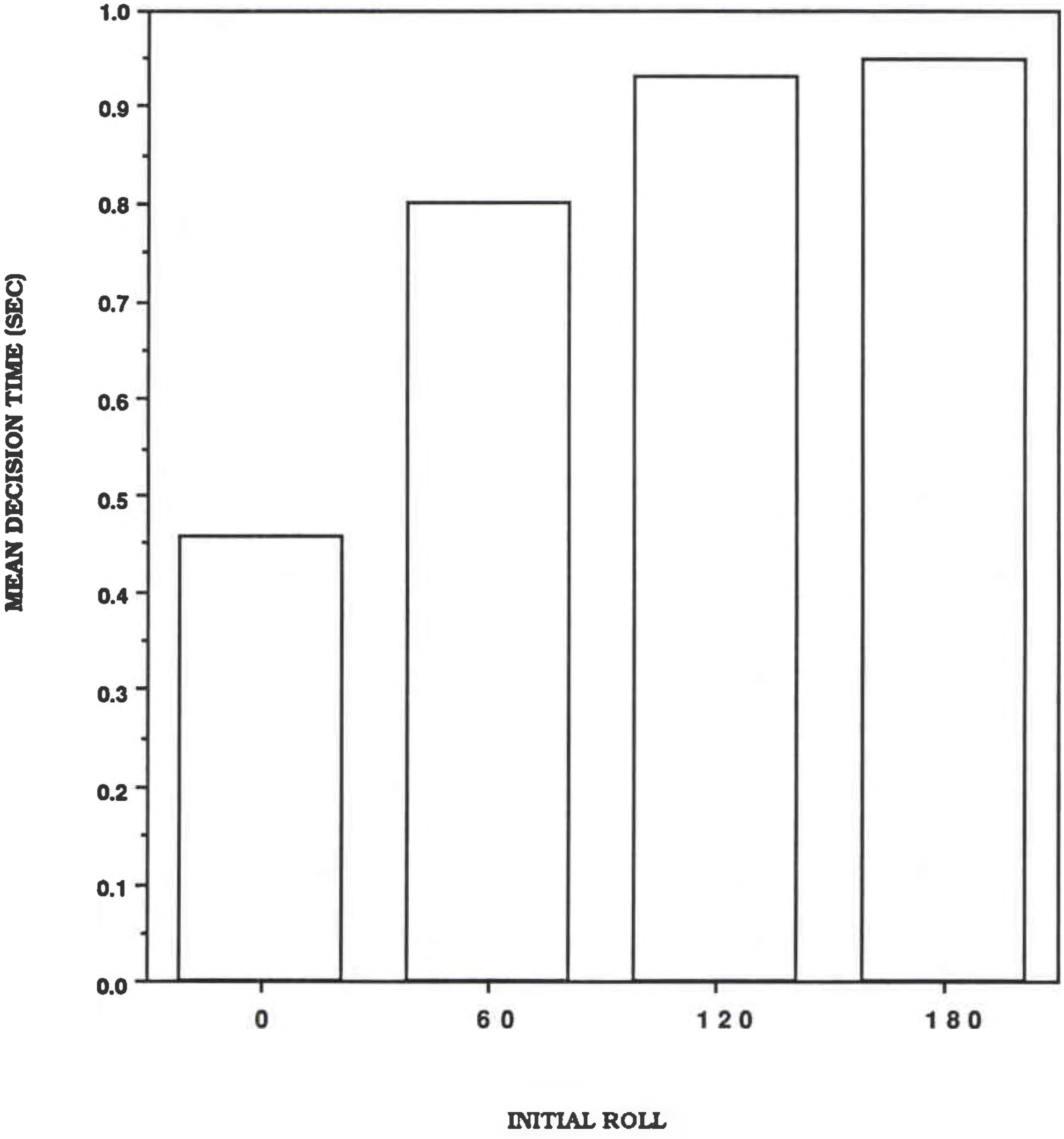


Figure 12a. Mean Decision Time as a Function of Initial Roll During the Experimental Session.

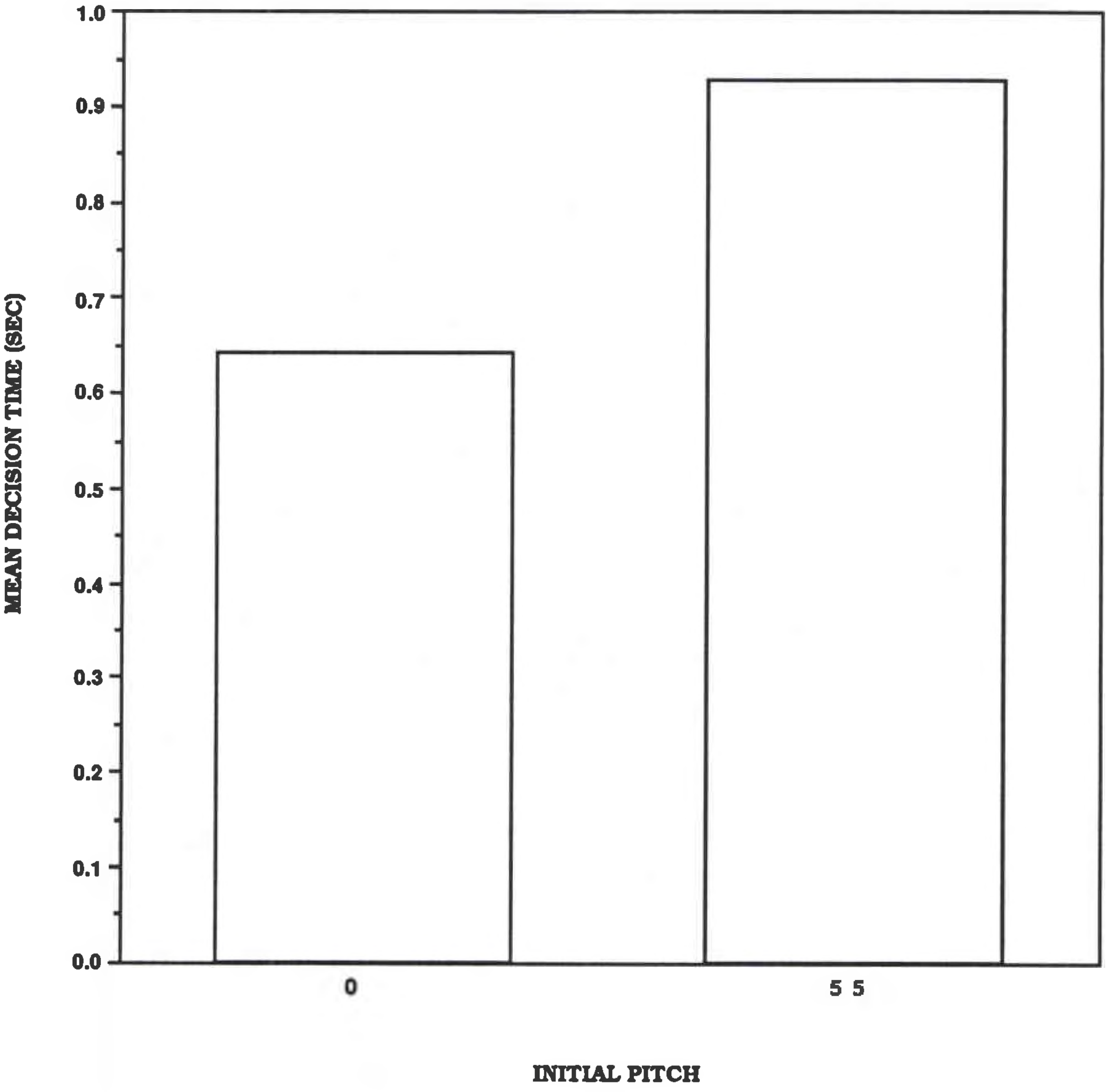


Figure 12b. Mean Decision Time as a Function of Initial Pitch During the Experimental Session.

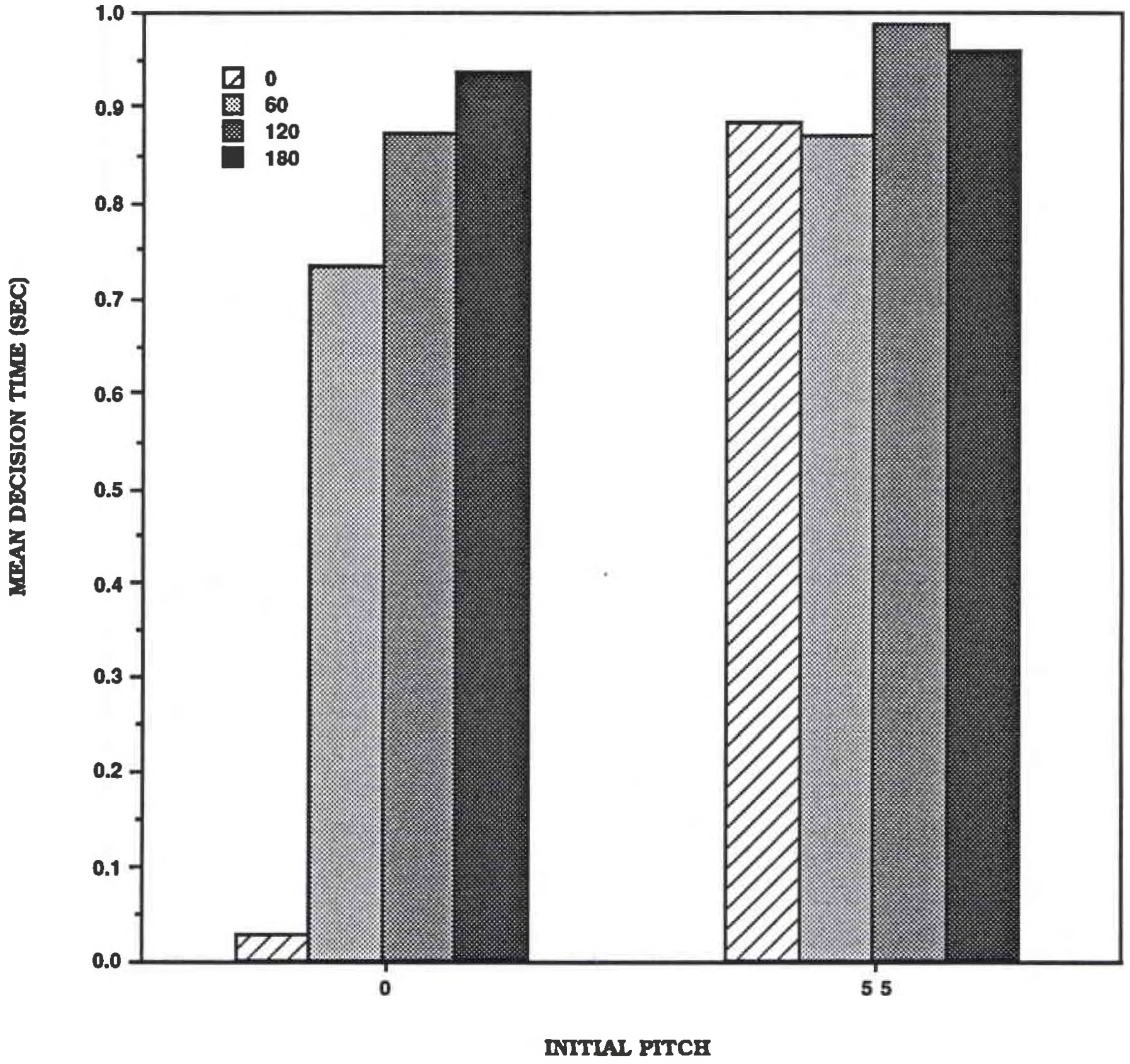


Figure 13. Mean Decision Time as a Function of Initial Roll and Initial Pitch During the Experimental Session.

three-way interaction. The two-way interaction was present among display formats and initial pitch at the zero degree initial roll level. No other significant two-way interactions were found at any of the other initial roll levels. The effect at the zero degree initial roll level is illustrated in Figure 14. An analysis of simple effects found significant differences between the display format decision time means at only the 55 degree initial pitch level. Mean decision time for G-HUD (mean = 0.80 seconds) and ADI (mean = 0.78 seconds) subjects was significantly faster than HUD subjects' mean decision time (mean = 1.06 seconds). Decision time means for the G-HUD and the ADI were not significantly different from one another.

The recovery time variable consisted of the time between the static presentation of the display--in it's unusual attitude--and the point at which the display was recovered to within +/- five degrees in both roll and pitch. The analysis indicated significant main effects for both initial roll (F (3, 81) = 100.20, p = .0001) and initial pitch (F (1, 27) = 170.31, p = .0001) but not for display format. Display format resulted in the following mean recovery times: G-HUD mean recovery time was 3.01 seconds, the ADI resulted in a mean recovery time of 3.68 seconds, and the HUD format resulted in a mean recovery time of 3.60 seconds. Initial roll and initial pitch accounted for 29.61% and 27.98% of the variance respectively. A Tukey HSD comparison found significant differences between the initial roll recovery time means such that the zero degree initial roll level resulted in the fastest mean recovery time. The 60 degree initial roll level resulted in a significantly faster recovery time than either the 120 degree or 180 degree levels. There was no statistically significant difference found between the 120 and 180 degree initial roll levels. Figure 15a illustrates the effects. For initial pitch, a Tukey HSD comparison found that the zero degree initial pitch level resulted in a significantly faster recovery time than that of the 55 degree initial pitch level. Figure 15b illustrates the effects. A significant interaction was found between initial roll

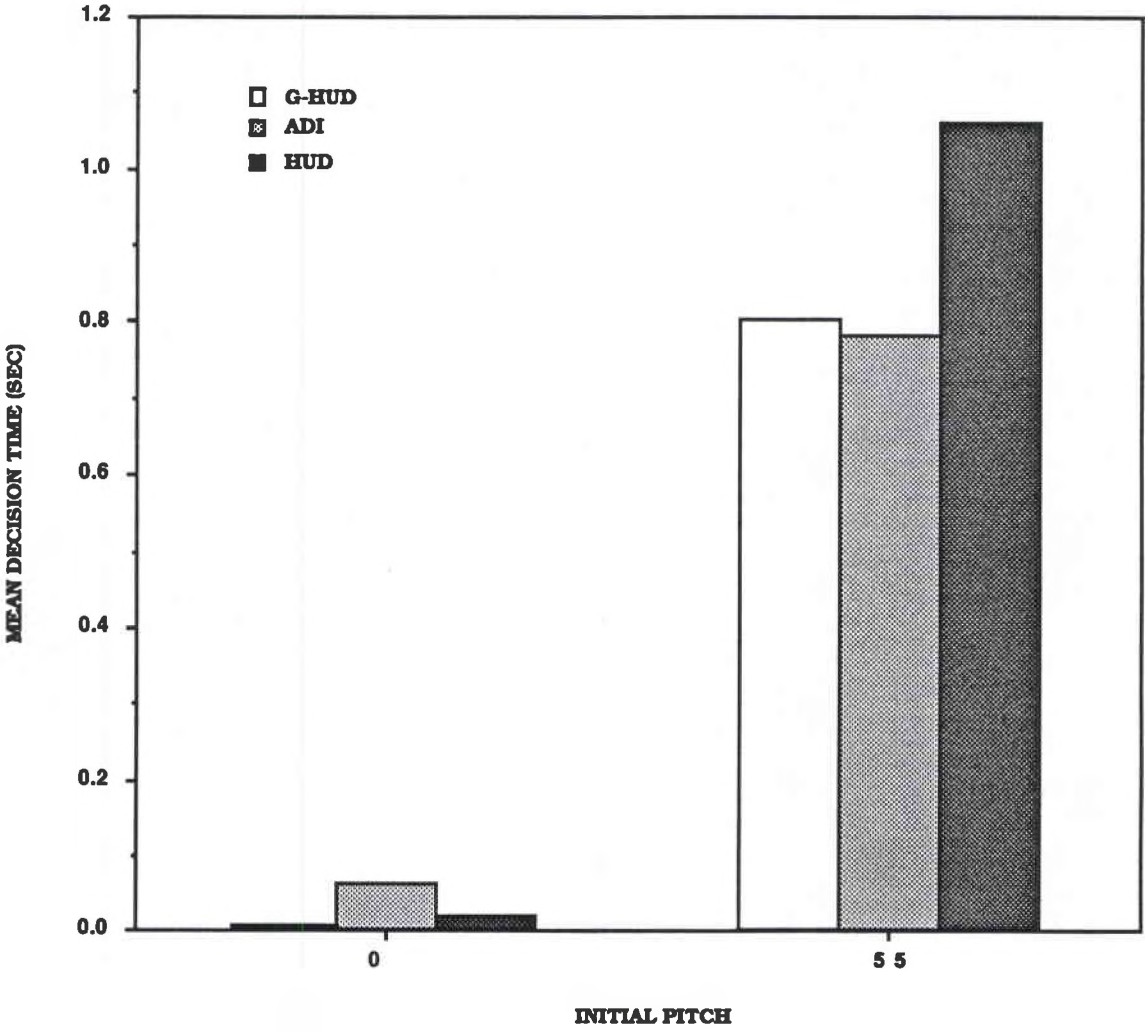


Figure 14. Mean Decision Time as a Function of Display Format and Initial Pitch at Zero Degrees Initial Roll.

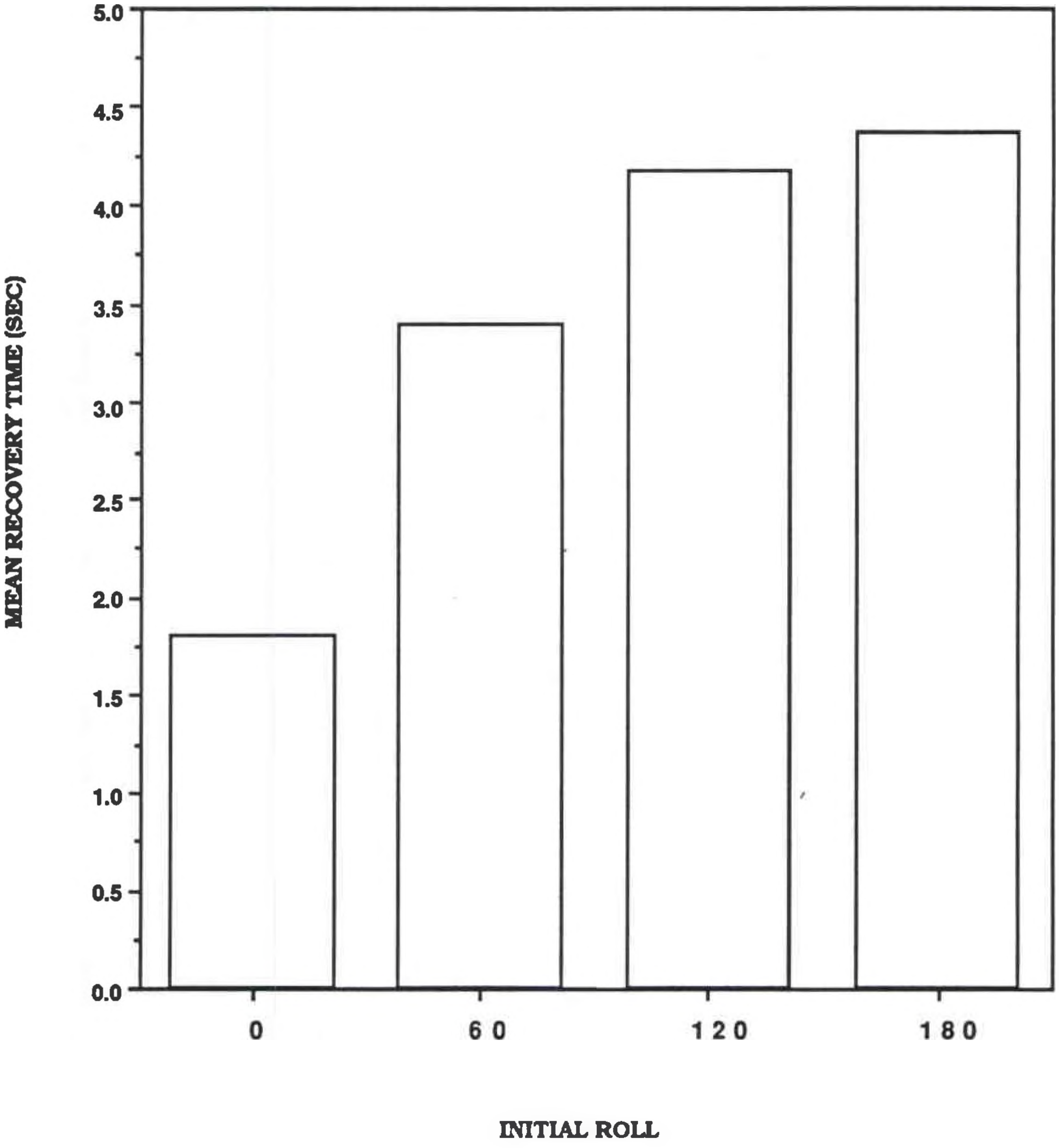


Figure 15a. Mean Recovery Time as a Function of Initial Roll During the Experimental Session.

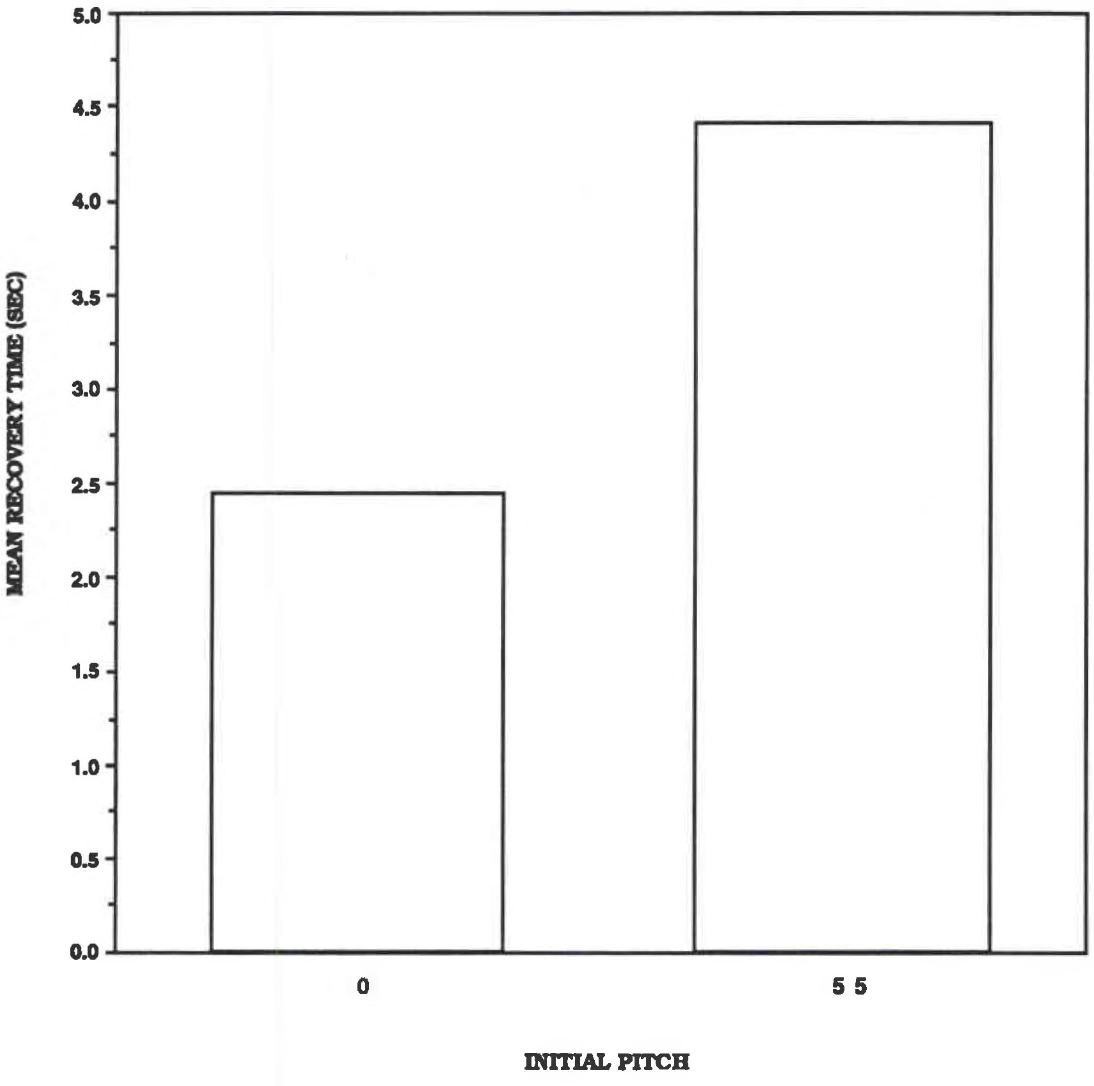


Figure 15b. Mean Recovery Time as a Function of Initial Pitch During the Experimental Session.

and initial pitch (F (3, 81) = 26.82, p = .0001). The interaction accounted for only 4.67% of the model variance (Figure 16). An analysis of simple effects found a significant effect of roll at the zero initial pitch level (F (3, 119) = 118.89, p = .0001). It was indicated that the zero degree initial roll level resulted in the quickest recovery time (mean = 0.14 sec). The 60 degree initial roll level (mean = 2.51 sec) resulted in a slower recovery time than the Zero degree level but a faster recovery time than either the 120 (mean = 3.53 sec) or 180 (mean = 3.62 sec) degree level. There was no difference between the 120 and 180 degree initial roll levels. At the 55 degree initial pitch level (F (3, 116) = 7.71, p = 0.0001) the zero roll level again resulted in the fastest recovery time (mean = 3.46 sec). The 120 (mean = 4.81 sec) and 180 (mean = 5.10 sec) degree levels were no different from one another but were both slower than the zero degree level. The 60 degree level (mean = 4.28 sec) recovery time was no different than the zero, 120, or 180 degree levels.

Initial input accuracy in both roll and pitch was recorded by determining whether or not the subject's first recovery input was in the direction of the closest horizon for the appropriate dimension. These data were converted to and analyzed as correct response percentages. An ANOVA determined that display format failed to reach statistical significance at the .05 level (F (2, 27) = 2.85, p = 0.075). Table 1 lists the mean accuracy percentages for the roll and pitch dimensions. An ANOVA indicated that display format did not result in a significant difference for accuracy in the pitch dimension. See Appendix A for the ANOVA summary tables.

Training Rerun Session

In all, a total of 15 subjects failed to reach the performance criterion within the 120 trials allotted per disturbance amplitude level. Of those subjects, 14 ran originally under the HUD display format. The remaining subject ran under the ADI display

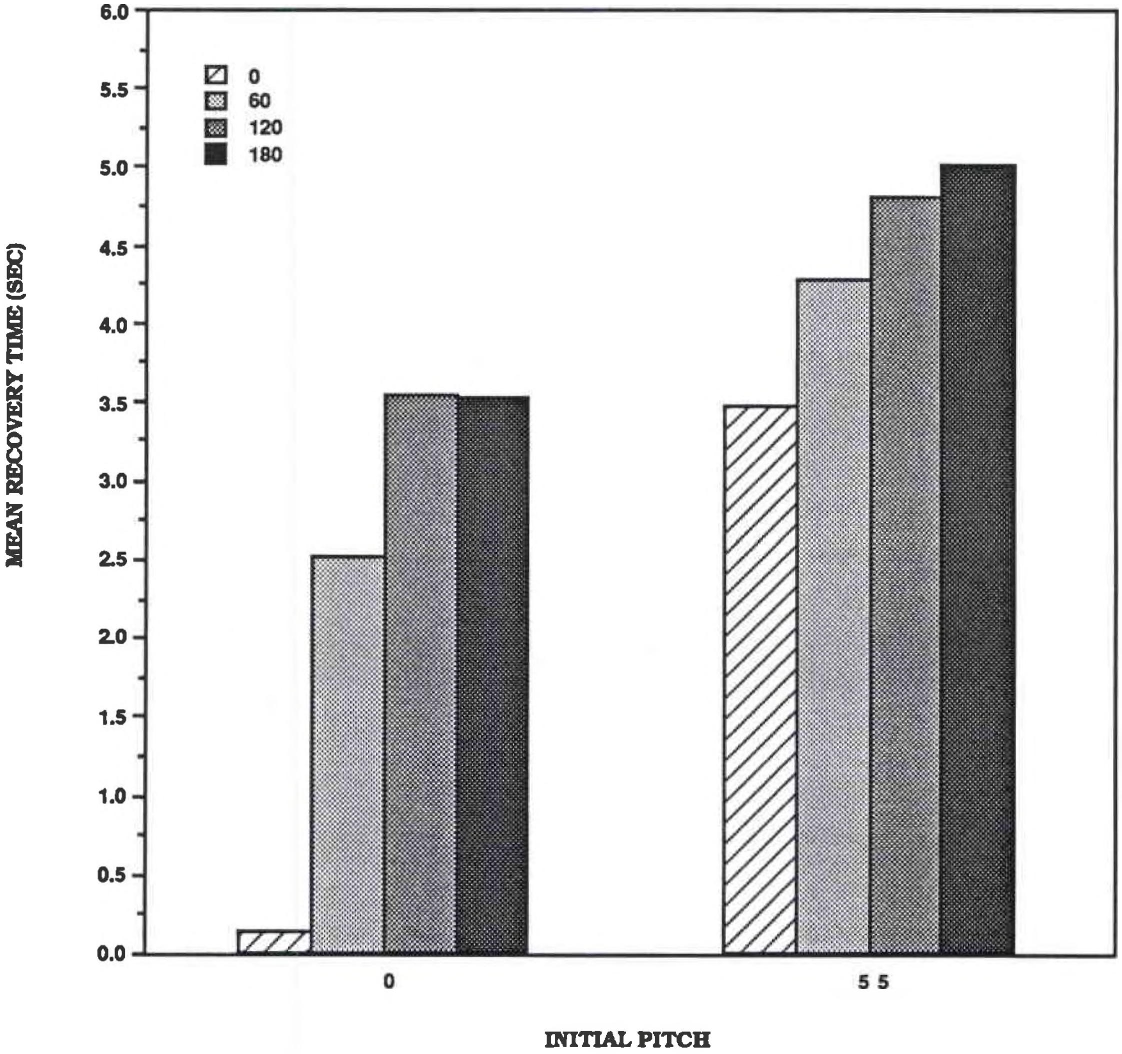


Figure 16. Mean Recovery Time as a Function of Initial Roll and Initial Pitch During the Experimental Session.

Table 1.

Percent Correct for Initial Recovery Input.

	G-HUD	ADI	HUD
ROLL	81.30%	70.93%	75.19%
PITCH	70.74%	60.37%	70.00%
AVG.	76.02%	65.65%	72.60%

format. In response to this disproportionate failure rate, nine available HUD failure subjects were returned to the lab in order to perform a Training Rerun Session. During the Training Rerun Session, subjects performed the original Training Session procedure under either the G-HUD or ADI display format. Four of the nine subjects were assigned to participate in the Training Rerun Session under the G-HUD and the remaining five used the ADI. The assignment of subjects to either the G-HUD rerun group or the ADI rerun group was alternated. The purpose of the Training Rerun Session was to determine if the HUD failure subjects were able to meet the performance criterion via one of the remaining display formats. All of the Training Rerun Session subjects failed to reach the performance criterion of the original Training Session under the HUD format before they were rerun under either the ADI or the G-HUD. The training rerun data were compared to the data from similar variable conditions of the original Training Session. This information is presented, when appropriate, within the following results.

Two dependent measures were of interest for the HUD rerun subjects. Included were: number of trials to reach criterion and MCH ratings.

Trials to reach criterion. Three ANOVA procedures were performed on the trials to reach performance criterion data in order to test the effects of display format and the disturbance amplitude levels for the HUD rerun subjects. The first ANOVA treated a comparison between the HUD (data collected during the original Training Session) and G-HUD (data collected during the Training Rerun Session) display formats. The analysis was performed as a completely within-subjects factorial. A significant main effect for display format was found (F (1, 3) = 223.91, p = 0.0006). The display format effect, shown in Figure 17, accounted for 25.87% of the variance. The results indicated that the subjects, when using the G-HUD (mean = 16.17 trials),

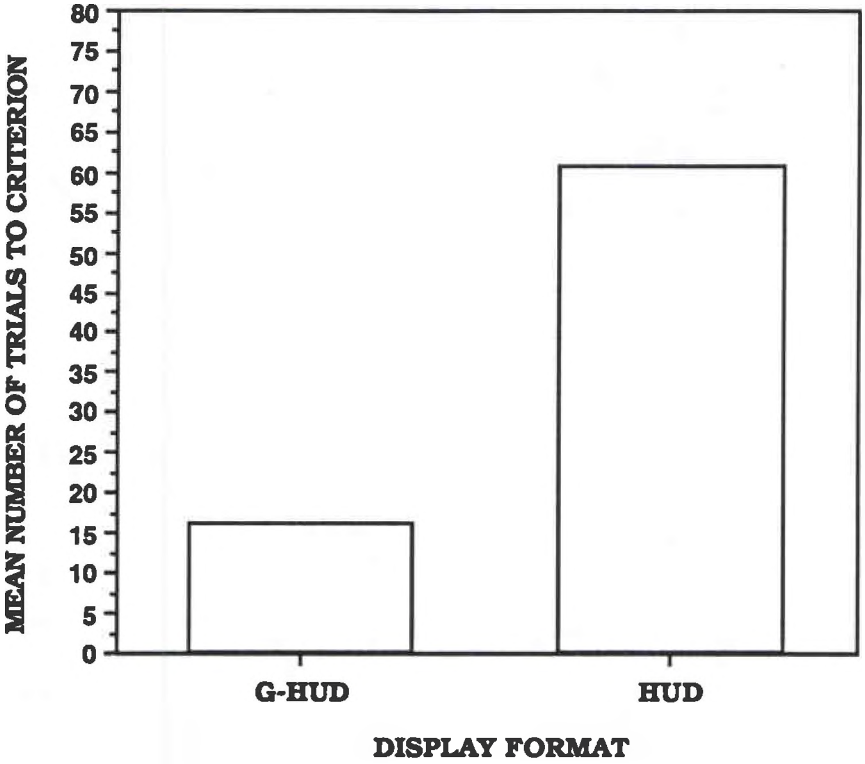


Figure 17. Mean Number of Trials to Reach Criterion as a Function of Display Format During the Training Rerun Session.

required fewer trials to reach the performance criterion than when using the HUD format (mean = 60.5 trials). In the original Training Session, G-HUD subjects met the performance criterion in about the same number of trials as required for the rerun subjects to meet criterion (mean = 17.4 trials). Original Training Session HUD subjects--those who met the performance criterion--required an average of 36.8 trials to meet criterion. A main effect of disturbance amplitude was also found (*F* (2, 6) = 22.56, *p* = 0.0016). A Tukey HSD comparison found significant differences between the disturbance amplitude levels such that the high disturbance amplitude (mean = 73.38 trials) required significantly more trials to reach criterion than either the medium amplitude (mean = 29.87 trials) or the low amplitude (mean = 11.75 trials). The medium and low amplitudes did not differ significantly from one another. This main effect accounted for 36.00% of the variance. Table 2 presents the original Training Session means so that they may be compared to the Training Rerun Session means. The analysis also found a significant interaction between display format and disturbance amplitude (*F* (2, 6) = 6.03, *p* = 0.0366). The interaction (Figure 18) accounted for 18.10% of the variance. An analysis of simple effects revealed that the only difference between display format was located at the high disturbance amplitude level (*F* (1, 6) = 39.85, *p* = 0.0007). Within the high disturbance amplitude, G-HUD subjects (mean = 26.75 trials) took significantly fewer trials to reach performance criterion than the HUD format (mean = 120 trials).

The second ANOVA treated a comparison between the HUD (data collected during the original Training Session) and ADI (data collected during the Training Rerun Session) display Formats. The analysis was performed as a completely within-subjects factorial. A significant main effect for display format was found (*F* (1, 4) = 46.43, *p* =

Table 2.

Comparison of Training Session and Training Rerun Session Mean Number of Trials to Reach Criterion.

	SUBJECTS WHO COMPLETED THE ORIGINAL TRAINING SESSION				SUBJECTS WHO RETURNED FOR THE TRAINING RERUN SESSION			
		Amplitude				Amplitude		
DISPLAY FORMAT		Low	Med	High		Low	Med	High
G-HUD	n=10	12.10	14.40	25.70	n=4	10.00	11.75	26.75
HUD	n=10	11.90	24.70	73.80	n=4	13.50	48.00	120.0
Mean	n=20	11.83	21.83	48.27	n=4	11.75	29.87	73.38
ADI	n=10	11.50	26.40	45.30	n=5	10.00	20.40	32.40
HUD	n=10	11.90	24.70	73.80	n=5	14.20	59.20	120.0
Mean	n=20	11.83	21.83	48.27	n=5	12.10	39.80	76.20
ADI	n=10	11.50	26.40	45.30	n=5	10.00	20.40	32.40
G-HUD	n=10	12.10	14.40	25.70	n=4	10.00	11.75	26.75
Mean	n=20	11.80	20.40	35.50	n=9	10.00	16.56	29.80

Note: The mean number of trials to reach criterion for the HUD format which appears on the Training Rerun section of Table 2 represents those subjects who failed to reach criterion performance during the original Training Session. These subjects were returned to the laboratory in order to participate in the Training Rerun Session under either the G-HUD or HUD format.

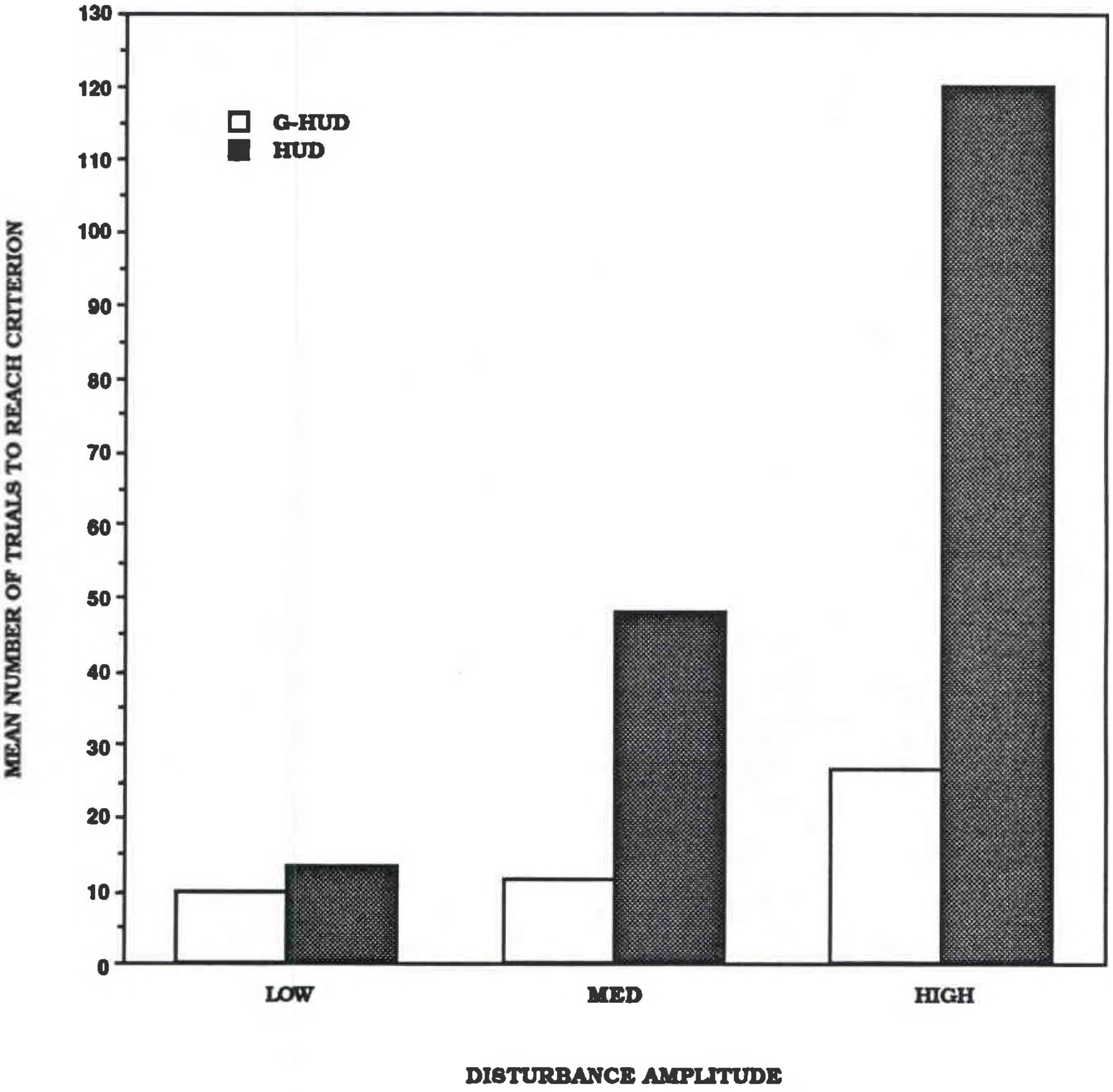


Figure 18. Mean Number of Trials to Reach Criterion as a Function of Display Format During the Training Rerun Session.

0.0024). The display format effect, shown in Figure 19, accounted for 23.00% of the variance. The results indicated that using the ADI (mean = 20.93 trials) required significantly fewer trials to reach performance criterion than when using the HUD format (mean = 64.47 trials). As with the G-HUD, ADI subjects in the original Training Session met the performance criterion in about the same number of trials as the Training Rerun subjects (mean = 27.73 trials). Again, original Training Session HUD subjects who met the performance criterion were able to do so, on average, in 36.8 trials. A main effect of disturbance amplitude was also found (F (2, 8) = 18.48, p = 0.0010). A Tukey HSD comparison found significant differences between the disturbance amplitude levels such that the high disturbance amplitude (mean = 76.20 trials) required significantly more trials to reach criterion than either the medium amplitude (mean = 39.80 trials) or the low amplitude (mean = 12.10 trials). The medium and low amplitudes did not differ significantly from one another. This main effect accounted for 33.44% of the variance. Table 2 presents the original Training Session means and the Training Rerun Session means so that they may be compared. Also found was a significant interaction between display format and disturbance amplitude (F (2, 8) = 7.03, p = 0.0173). The interaction (Figure 20) accounted for 14.20% of the variance. An analysis of simple effects revealed that the only difference between display format was located at the high disturbance amplitude level (F (1, 8) = 15.99, p = 0.0040). Within the high disturbance amplitude, ADI subjects (mean = 32.40 trials) took significantly fewer trials to reach performance criterion than HUD format subjects (mean = 120 trials).

The third ANOVA treated a comparison between the G-HUD and ADI display Formats (data for both formats were collected during the Training Rerun Session). The analysis was performed as a mixed factorial with display format manipulated

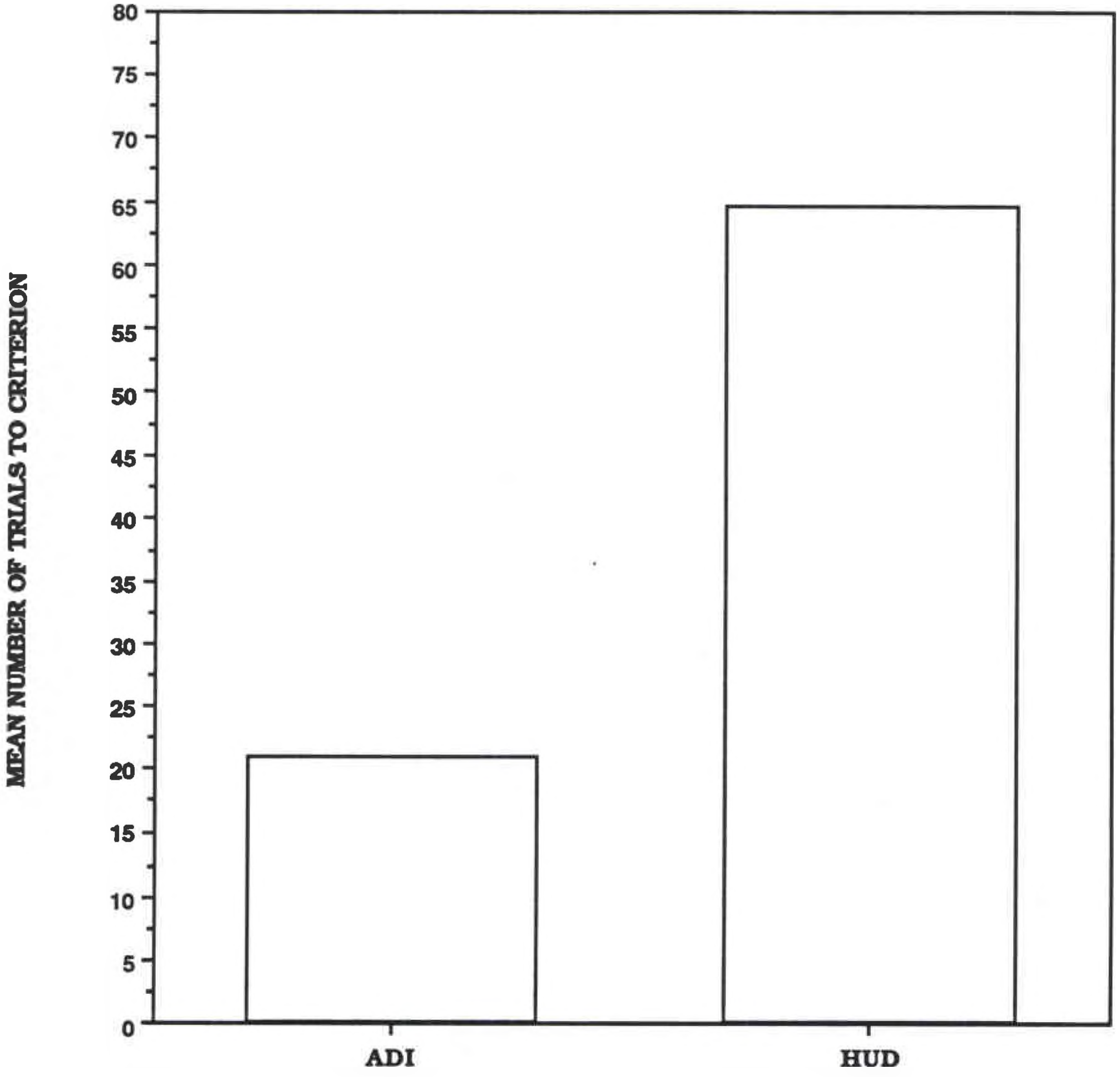


Figure 19. Mean Number of Trials to Reach Criterion as a Function of Display Format During the Training Rerun Session.

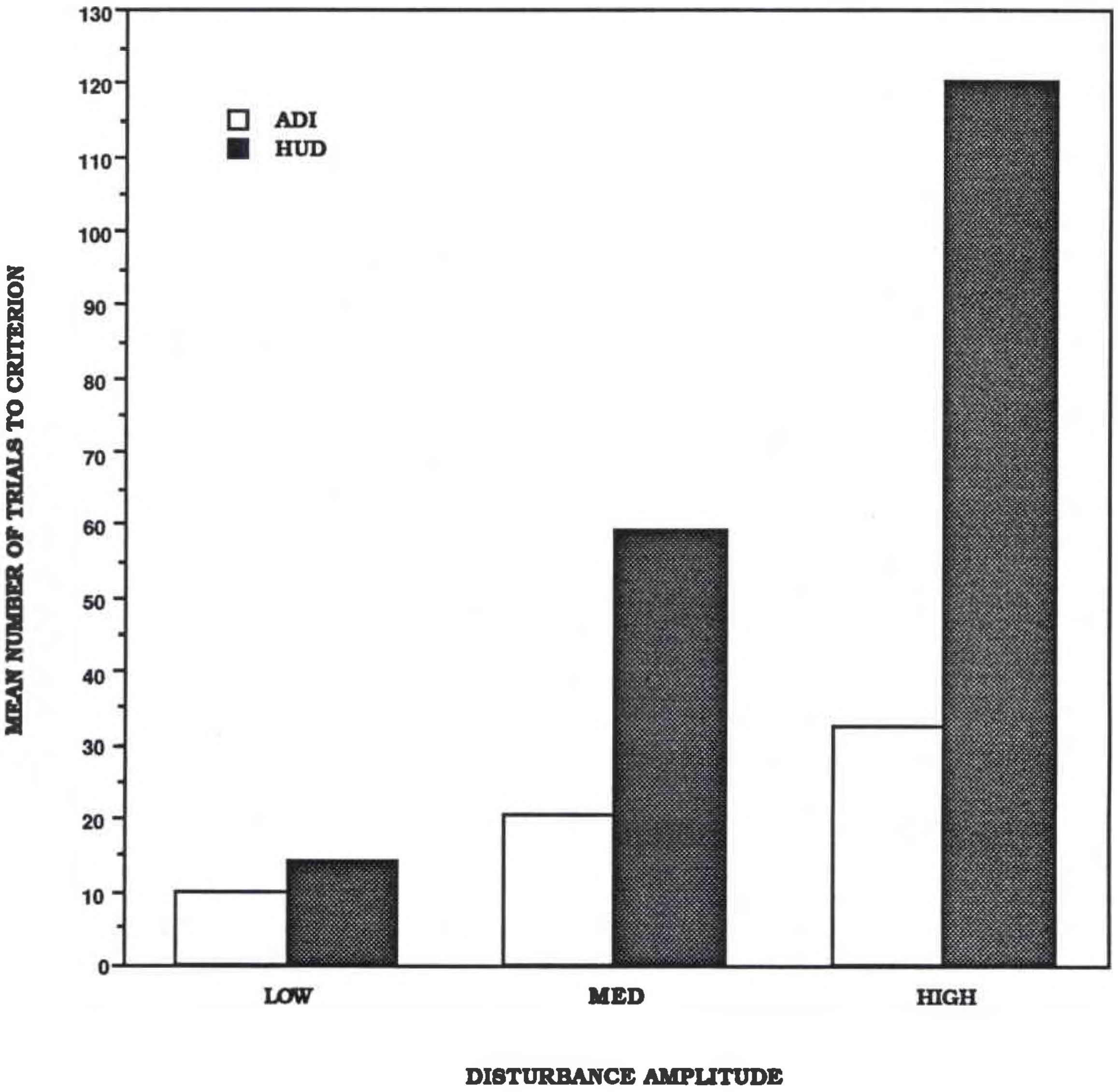


Figure 20. Mean Number of Trials to Reach Criterion as a Function of Display Format During the Trial Rerun Session.

between-subjects and disturbance amplitude manipulated within-subjects. Four subjects were rerun under the G-HUD display format and five subjects were rerun using the ADI. No statistically significant effects were indicated for display format or disturbance amplitude. These findings are consistent with those of the original Training Session. Table 2 presents the original Training Session means so that they may be compared to the means of the Training Rerun Session.

Modified Cooper-Harper. Similar to the trials to reach criterion data, three ANOVA procedures were performed on the MCH data in order to test the effects of display format and the disturbance amplitude levels for the HUD rerun subjects. The first ANOVA tested a comparison between the HUD (these data were collected during the original Training Session) and G-HUD (theses data were collected during the Training Rerun Session) Formats. The analysis was performed as a completely within-subjects factorial. A significant main effect for display format was found (F (1, 3) = 66.67, p = 0.0038). This effect, shown in Figure 21, accounted for 38.35% of the variance. The results indicated that the G-HUD (mean = 2.42 MCH) was rated as resulting in less workload than the HUD format (means = 5.75). During the original Training Session, the mean MCH ratings for the G-HUD and the HUD were not significantly different from one another. For comparison, Subjects of the original Training Session rated the task under the G-HUD as resulting in a mean MCH rating of 4.27. The HUD format for subjects included in the original Training Session resulted in a mean MCH rating of 4.90. A main effect of disturbance amplitude was also found (F (2, 6) = 20.77, p = 0.0020). This main effect accounted for 34.85% of the variance. A Tukey HSD comparison found significant differences between the disturbance amplitudes such that the high disturbance amplitude (mean = 6.13 MCH) resulted in a significantly higher MCH rating than either the medium amplitude mean = 3.88 MCH) or the low amplitude (mean = 2.25 MCH). The medium and low amplitudes did not significantly differ from one another.

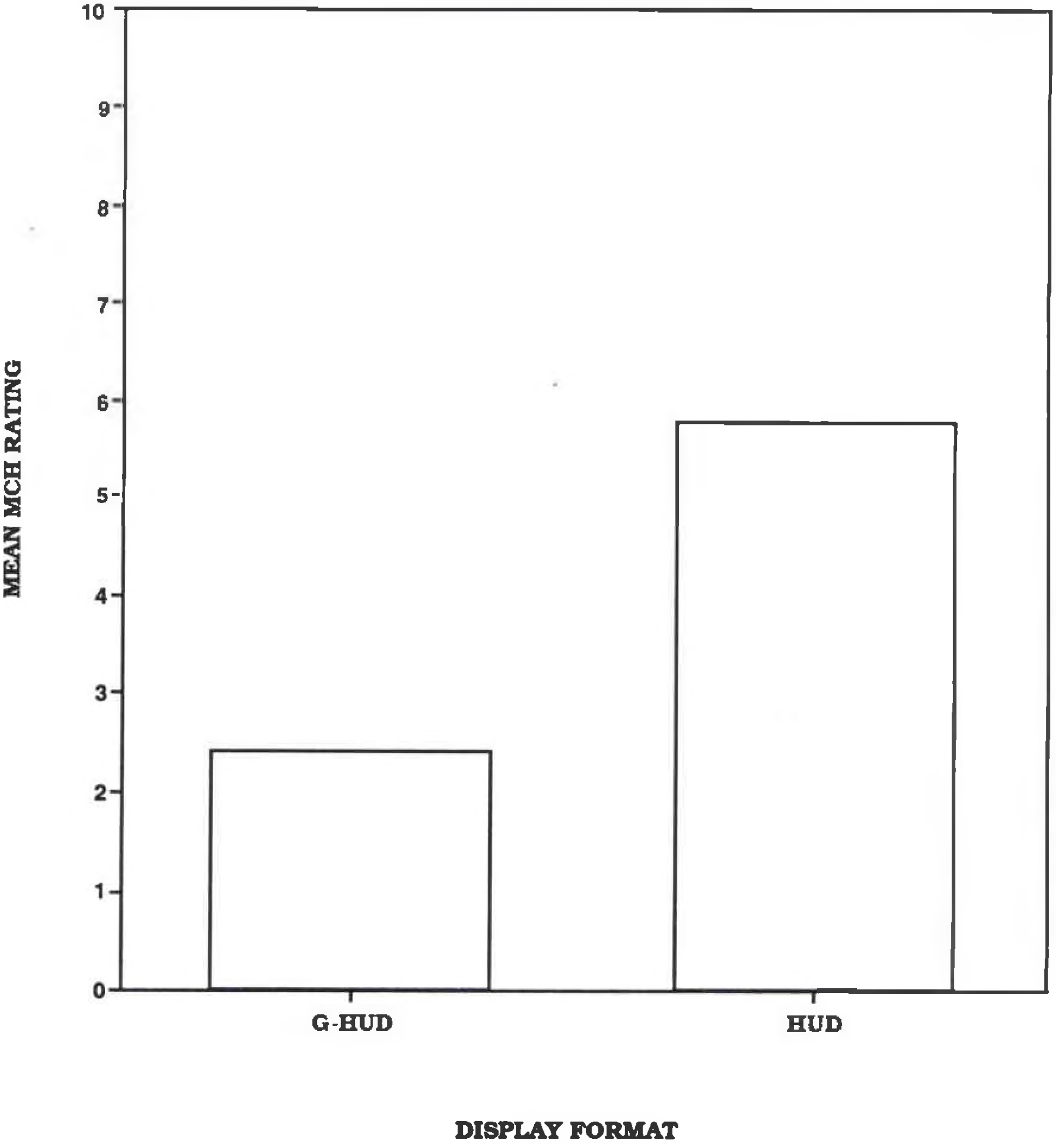


Figure 21. Mean Modified Cooper-Harper Rating as a Function of Display Format During the Training Rerun Session.

The analysis also indicated a significant interaction among display format and disturbance amplitude (F (2, 6) = 6.94, p = 0.0275). The interaction (Figure 22) accounted for 10.98% of the variance. An analysis of simple effects showed a significant difference between the display format manipulation at both the low and high disturbance amplitude levels. At the low amplitude the G-HUD (mean = 1.5 MCH) was assigned a significantly lower MCH rating than the HUD (mean = 3.0 MCH) format (F (1, 6) = 27.00, p = 0.0020). At the high amplitude the G-HUD (mean = 3.25 MCH) was also assigned a significantly lower MCH rating than the HUD (mean = 9.0 MCH) format (F (1, 6) = 83.53, p = .0001). The means suggest that the interaction stems from an effect size difference between the low and high disturbance amplitude levels. The interaction becomes disordinal when plotted with display format on the abscissa. At the G-HUD format, low disturbance amplitude resulted in a significantly lower MCH rating than at the high disturbance amplitude. Mean MCH rating for medium disturbance amplitude did not differ significantly from either the low or high disturbance level. At the HUD format, both the low and medium disturbance amplitude levels resulted in significantly lower mean MCH ratings than the high disturbance amplitude level. The low and medium level MCH ratings were not significantly different from one another. During the original Training Session, no significant interaction among display format and disturbance amplitude was indicated.

The second ANOVA compared the HUD (these data were collected during the original Training Session) and ADI (these data were collected during the Training Rerun Session) display Formats. The analysis was performed as a completely within-subjects factorial. A significant main effect for display format was found (F (1, 4) = 105, p = 0.0005). This effect, shown in Figure 23, accounted for 24.10% of the variance. The means indicated that the ADI (mean = 3.13 MCH) was assigned a significantly lower MCH rating than that of the HUD format (mean = 5.4 MCH). During

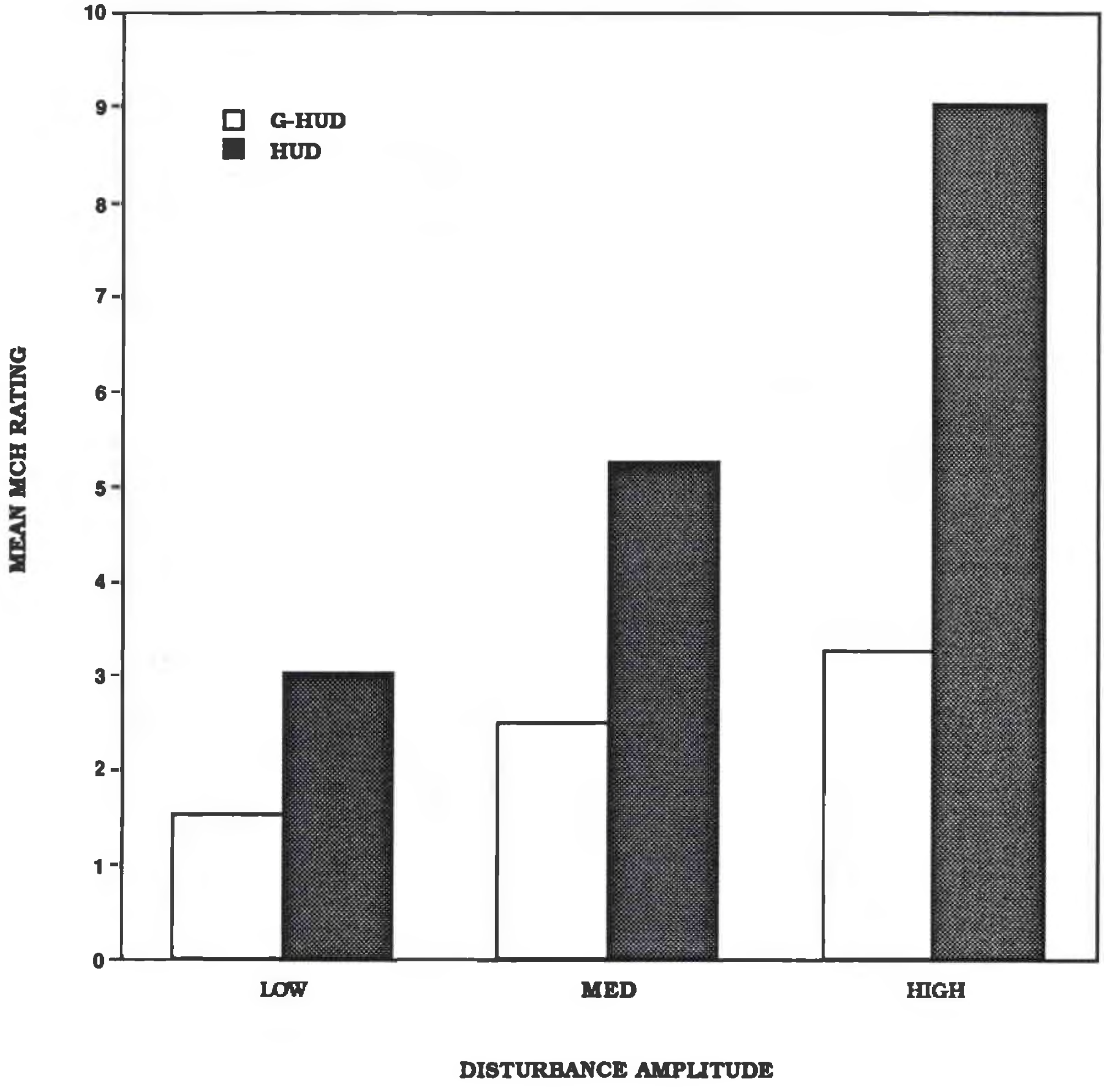


Figure 22. Mean MCH Rating as a Function of Disturbance Amplitude and Display Format During the Training Rerun Session.

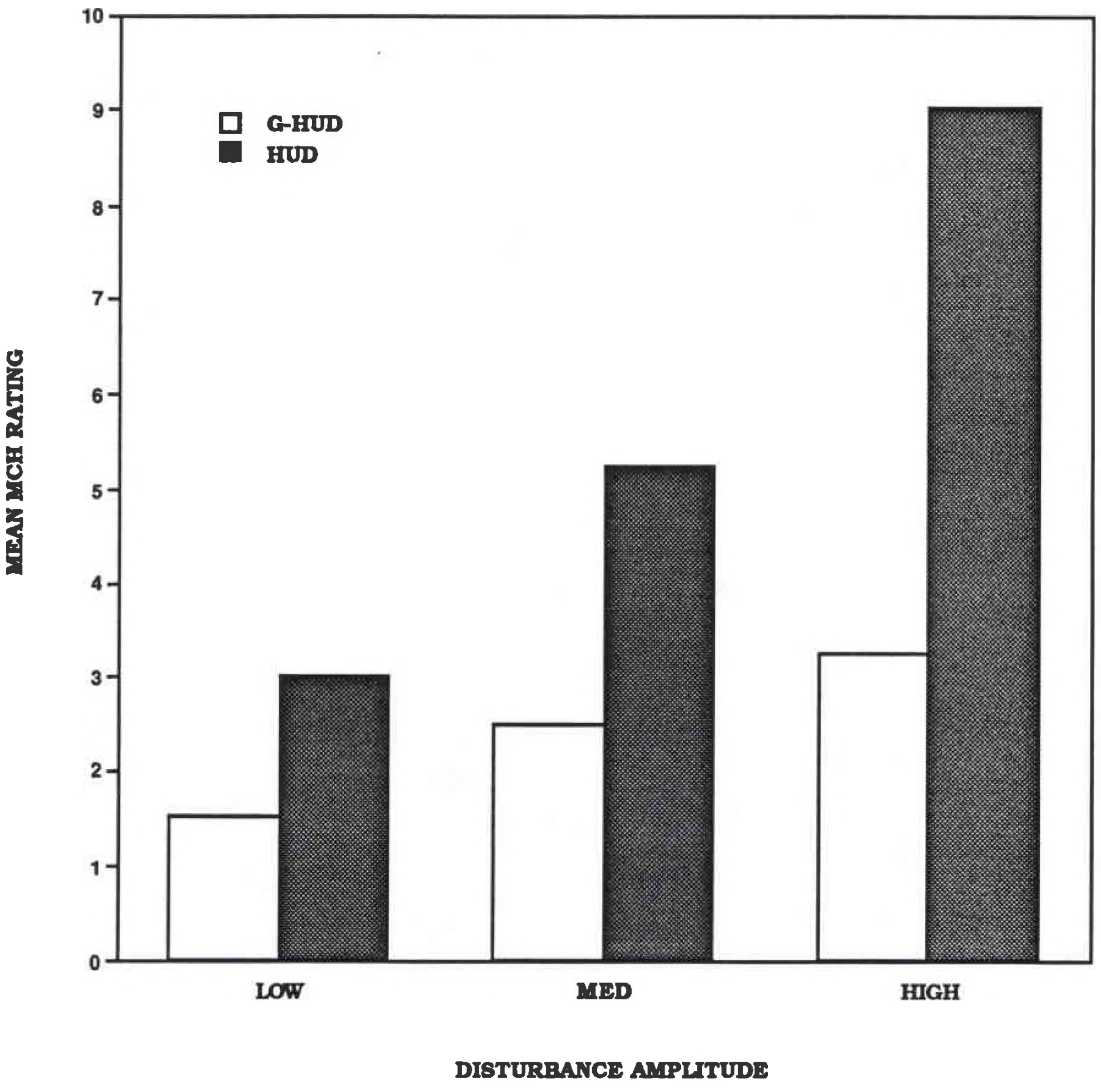


Figure 23. Mean Modified Cooper-Harper rating as a Function of Display Format During the Training Rerun Session.

the original Training Session, the mean MCH ratings for the ADI and the HUD were not significantly different from one another. For comparison, subjects of the original Training Session rated the task under the ADI as resulting in a mean MCH rating of 3.93. Again, the HUD format for subjects included in the original Training Session resulted in a mean MCH rating of 4.90. A main effect of disturbance amplitude was also found (F (2, 8) = 12.81, p = 0.0032). A Tukey HSD comparison found significant differences between the levels of disturbance amplitude such that the high disturbance amplitude (mean = 6.0 MCH) resulted in a significantly higher MCH rating than either the medium amplitude (mean = 3.8 MCH) or the low amplitude (mean = 3.0 MCH). The medium and low amplitudes did not significantly differ from one another. This main effect accounted for 7.17% of the variance. No interaction was indicated.

The third ANOVA tested a comparison between the G-HUD and ADI display Formats for differences existing in the MCH data (data for both formats were collected during the Training Rerun Session). The analysis was performed as a mixed factorial with display format manipulated between-subjects and disturbance amplitude manipulated within-subjects. Four subjects were rerun under the G-HUD display format and five subjects were rerun using the ADI. The only significant effect was that of disturbance amplitude (F (2,14) = 14.67, p = 0.0004). Once again, a Tukey HSD comparison found significant differences such that the high disturbance amplitude (mean = 3.67 MCH) resulted in a significantly higher MCH rating than either the medium amplitude (mean = 2.77 MCH) or the low amplitude (mean = 2.0 MCH). The medium and low amplitudes did not significantly differ from one another.

CHAPTER 4

DISCUSSION

The results of this experiment lend support to the hypothesis that an attitude display formed of the integration of ADI and HUD type symbology will demonstrate a performance benefit over a pure HUD format. That is, the integrated display would take on successful attitude conveyance characteristics of the ADI but would continue to be transparent in nature. Transparency is an important requirement for any display that is being considered for utilization in a head-up or helmet-mounted configuration. For the present study, the G-HUD represented an integration of ADI and HUD type symbology structures. The G-HUD was designed to be global in nature, to benefit from pictorial realism, to be a compressed format, and to be perceptually integral. The previous features are those thought to be present in the ADI. The G-HUD was also designed to be minimally cluttered and transparent with the potential for further de-cluttering and increased transparency. The merit of the G-HUD as an attitude indicator was evaluated in the present study but display transparency was not empirically tested.

The following discussion will first address the Training Session findings followed by a discussion of the Experimental Session findings. The discussion concludes with suggestions for further research.

Training Session.

During the Training Session, subjects performed an attitude maintenance task. The task consisted of maintaining a display representation as close to SL as possible while being buffeted by disturbance functions of varying severity. Subjects were required to reach a specific level of performance on three different disturbance function amplitudes (Figure 4) before they were eligible to participate in the Experimental Session of the study. Subjects moved through the disturbance levels in ascending order of amplitude and were required to meet a performance criterion before moving from one amplitude to the next and ultimately onto the Experimental Session.

The Training Session data suggest a clear learning difference between the evaluated displays. For the metrics collected, the G-HUD demonstrated better performance and quicker learning than the HUD format. G-HUD subjects met the performance requirement with fewer trials than with the HUD format. Overall maintenance task performance was better with the G-HUD than with the HUD format. Both metrics indicated a benefit for the G-HUD over the HUD but no difference was found between the ADI and the remaining formats. Statistically, subjects' Performance on the ADI was the same as performance on both the G-HUD and the HUD and Performance on the G-HUD was statistically better than that of the HUD. This seems to support the possibility that some unique or enhanced feature of the the G-HUD resulted in the significant performance benefit over the HUD format.

There are four main differences between the G-HUD and the ADI. Included is the difference between the OS, the contrast between the positive and negative pitch areas, the connected heading lines of the G-HUD, and the increased display compression of the G-HUD. Display compression refers to a relationship between a display and the real world which the display represents. Display compression will affect display rate-of-motion, resolution, and field-of-view. This relationship can be described as some ratio between the display and the real world. The HUD display compression ratio for the present study was 1:1. As the simulated aircraft nose traversed along the pitch dimension, the HUD display represented that movement across the same distance at the same velocity. The ADI is a compressed format in which the display/world ratio was greater than one. The ADI artificial horizon did not have to move as far as the natural horizon to represent the same deflection. This resulted in an apparent slowing of display translation when compared to the HUD motion. Display compression of the G-HUD was even greater than that of the ADI. It was most likely this increased display compression of the G-HUD that resulted in faster training and better performance than that of the HUD format. The compression of the G-HUD acted to slow the apparent display translation. The slower translation of the G-HUD enabled subjects to quickly interpret and correct for the disturbance function. The translation of the HUD may have been too fast for the subjects to keep up with. For the HUD, it took subjects more trials to learn how much force to use and how quickly to react to the disturbance function. The magnitude of ADI display compression was between that of the G-HUD and the HUD. This may explain why no differences were found between the ADI and the G-HUD or between the ADI and the HUD.

It is interesting to note that the lack of contrast between the positive and negative pitch areas of the G-HUD did not seem to result in an associated performance cost. The contrast between the positive and negative pitch areas of the ADI has traditionally been given much credit for the display's value as an attitude indicator (Burns and Lovering, 1988; Previc, 1989; Roscoe, Corl, and Jensen, 1981; Taylor, 1982, 1984, 1988). The results of the present study indicated that the contrast between the display hemispheres was not as critical for the integrity of pictorial realism as previously believed or that the contrast formed between the solid lines of positive pitch area and the broken lines of negative pitch area for the G-HUD was sufficiently realistic. That is, the G-HUD successfully--or at least to the same degree as the ADI--portrayed an intuitive upper sky area and an intuitive lower ground area. An alternative interpretation is that the G-HUD benefited from some other characteristic to the extent that it made up for the performance cost associated with the poor contrast of the G-HUD. Once again, this characteristic may be the increased display compression of the G-HUD.

It is likely that the Effects of the Training Session were underestimated due to the experimental design. There was a disproportionate number of subjects who were not able to reach the Training Session performance criterion under the HUD format condition. The data from these subjects were not included in the Training Session analysis. That is, the Training Session analysis was based on data collected from the 10 HUD subjects who successfully met the performance criterion. 14 other HUD subjects failed to meet the performance criterion and their data were not included in the analysis. If the 14 unsuccessful subjects were included in the analysis, the HUD condition would have required considerably more trials to reach the performance criterion than are apparent in the present data. It is likely that the addition of the 14 unsuccessful subjects would have caused the HUD condition to demonstrate poorer mean performance than with only the successful subjects.

The findings of the present study are confounded by a subject selection bias. As mentioned above, the subjects who were replaced during the original Training Session were almost exclusively from the HUD format condition. In fact, 94% of the subjects that failed to reach criterion were running under the HUD format while the success rate for the G-HUD was 100% and 91% for the ADI. This suggests a systematic subject selection confound which was present throughout the experiment. The reason for the disproportionate failure rate is most likely due to the difficulty of the task under HUD format. In most simple terms, the features of the HUD acted to form a difficult task so that only those subjects who were most motivated and most proficient at the task were able to reach the performance criterion. The more easily utilized G-HUD and ADI formats enabled subjects of all levels of proficiency and motivation to meet the performance criterion. Therefore, the subjects representing the G-HUD and ADI displays formed a random distribution while only the most proficient and motivated subjects represented the HUD format. It should be noted that one ADI subject failed to reach the maintenance task performance criterion.

The Training Rerun Session was performed in order to determine if the subjects who failed to reach the performance criterion with the HUD format could reach criterion under either the G-HUD or the ADI. Nine of the 14 HUD failure subjects returned to the laboratory to run through the Training Session with one of the remaining display formats. Four of the subjects were rerun with the G-HUD and the remaining five subjects were rerun with the ADI. The results indicated that these subjects were, on average, able to reach criterion performance under both the G-HUD and the ADI. Although, one Rerun Session subject failed to reach criterion under the ADI format. It should be noted that these data were somewhat confounded due to ordering--all Rerun Session subjects experienced the HUD before experiencing either the G-HUD or the ADI. Despite the ordering confound, the Rerun Session results suggest that the reason for subject failure in the Training Session was that the task under the HUD format was more difficult than under either the G-HUD format or the ADI format. The influence of the display features was demonstrated in that the subjects were able to reach the criterion and, in addition, their mean performance under the G-HUD and the ADI was not much different than the performance of subjects under the same displays in the Training Session (Table, 2). Also, subjects failed almost exclusively under the HUD but easily met the performance criterion under the G-HUD and the ADI.

The findings within the Training Rerun Session parallel those of the Training Session proper. Subjects took fewer trials to meet the performance criterion with the G-HUD and the ADI than with the HUD. This fact suggests that the G-HUD and the ADI were easier to learn than the HUD format. It should be realized that this comparison is somewhat unfair given the ordering confound of the Training Rerun Session.

Although the objective findings for the Training Session data are clear, subjects did not make a subjective workload distinction between the display formats. According to the MCH results, subjects felt that the training task--overall--resulted in unacceptable mental workload and minor but annoying difficulty. See Appendix C. It is understandable that the average MCH rating was relatively high in that the task was designed to be challenging in order to assure that subjects did indeed learn their respective display symbologies and tasks. The within-subjects component of the design was sensitive to the three disturbance amplitude levels for both the subjective and objective data but subjects reported that the Training Session task was of similar workload regardless of display format.

The subjective workload ratings for the Rerun Session indicated that subjects found the training task to result in less workload via the G-HUD and the ADI than with the HUD format. Subjects reported that the task with the G-HUD was of an acceptable level of workload, easy, and desirable. For the ADI, subjects found workload to be acceptable, fair and mildly difficult. Training Rerun subjects reported that the HUD format resulted in an unacceptable workload level as well as very objectionable, but tolerable difficulty. See Appendix C. The subjective workload ratings during the Training Rerun Session were compared within-subjects.

The discontinuity between the subjective findings of the Training Session proper and the Training Rerun Session may be due to one or any combination of three inseparable influences. These influences include the different experimental designs, the quantity of practice that Rerun Session subjects experienced, and the subject selection bias. First, the within-subjects design of the Training Rerun session may have been more sensitive to subjective workload differences than the between-subjects design of the original Training Session due to context effects. Greenwald (1976), explains that performance may be affected by the context in which levels of a condition were experienced or compared in a within-subjects design. The Training Rerun subjects were able to rate the task under either the G-HUD or the ADI in the context of what they had already experienced with the HUD format. Secondly, Training Rerun Session subjects experienced more practice than their Training Session counterparts. It is possible that the Rerun subjects perceived less workload while performing under the second display format (either the G-HUD or the ADI) because they became more practiced and comfortable with the task. Thirdly, subject selection may have affected the subjective ratings in that only ratings from those subjects who met performance criterion during the original Training Session were included in the data. If the HUD subjects of the original Training Session were indeed more highly motivated or proficient than some of the subjects that represented the remaining displays, it would follow that the workload rating of the HUD condition was an underestimate of what is representative of a random sample.

Overall the Training Session of the experiment indicated that the G-HUD, formed of an integration of both ADI and HUD type symbology, was easier to learn than the HUD display format.

Experimental Session.

Only subjects who met the performance criterion of the Training Session returned to the laboratory to participate in the Experimental Session. The Experimental Session consisted of trials formed of both a maintenance task and an unusual attitude recovery task. The findings of the maintenance task will first be discussed followed by those of the unusual attitude recovery task.

The findings of the maintenance task were straightforward. Subjects were able to perform the maintenance task significantly better under the G-HUD and ADI symbology structures than under the HUD. This was true only on the pitch dimension during the maintenance task. The roll performance trend was in favor of the G-HUD and ADI but failed to reach statistical significance. Overall Experimental Session maintenance task performance paralleled that found in the Training Session. This was true despite the fact that all the subjects of the Experimental Session were trained to the same level of performance during the Training Session. This finding suggests that post-training performance for G-HUD and ADI subjects continued to improve while HUD subjects' performance reached a plateau at some lower level. This finding was consistent with the hypothesis that the G-HUD and ADI symbologies would demonstrate superior performance over that of the HUD format.

The Experimental Session recovery task method was essentially a replication of earlier work published by Kinsley, Warner, and Gleisner (1985) and Guttman (1986).

The findings of the present study did not completely replicate those of the studies on which it was based. Although the trends within the data were in the expected direction, no significant differences between display format were found for any of the recorded metrics except for the three-way interaction between initial roll, initial pitch, and display format for the decision time metric. The three-way interaction indicated that the G-HUD and the ADI resulted in significantly faster decision times than the HUD format at the zero degree initial roll level and the 55 degree initial pitch level (Figure, 14). Kinsley et al. (1985), found that decision time for a statically presented ADI was significantly faster than that for two different HUD formats. Kinsley et al. (1985) found no significant differences for input error between the displays for a static presentation. In a dynamic evaluation, Kinsley et al. (1985) found no main effect differences between displays for decision time but a display effect for recovery time was revealed. Once again, the ADI out-performed the HUD format. Guttman (1986), found no significant differences between display format for decision time but a display format main effect was found for recovery time. The analysis indicated that recovery time was significantly faster using the ADI compared to the HUD.

There are two possible explanations for the discrepancy between the findings of the past studies and the findings of the present study. The first explanation is that the differences found between the displays in the past studies could in actuality be evidence of a training effect. The present study utilized a more stringent training procedure then either Kinsley et al. (1985) or Guttman (1986). Kinsley et al. (1985) gave the static presentation subjects several trials in order to familiarize them with the experimental procedure. Practice trials were continued until subjects reached a performance criterion of 90 percent correct responses. For the dynamic recoveries, subjects flew the simulator until they reported being comfortable with the flight characteristics of the aircraft. Subjects then performed 36 practice recovery trials (12 trials of three displays) and were required to meet the 90 percent correct response criterion before performing the data collection trials. Guttman (1986) gave each aviator a preflight briefing which included an explanation of the flight controls to be used. Subjects were shown viewgraphs depicting each display format to be evaluated. Subjects were then given 18 practice (three of each display) recoveries before data were collected. For the present study, subjects were required to meet the Training Session performance criterion and were given 12 practice recoveries before the Experimental Session data were collected. It is possible that the extensive training of the present study negated an effect that may have appeared without it. More specifically, it is possible that the effects found in the past studies are actually evidence of a training effect. The case may have been that the subjects of the past studies were continuing to learn the various display symbology structures while the experimental data were being collected. Subjects in the present study learned their respective displays to such an extent that the maintenance task learning curves between the displays converged. This had to have occurred in order for the subjects to meet the Training Session performance criterion. The fact may be that subjects can be trained up to a level of performance where objective differences between the displays become transparent. If this is the case, differences between the displays would show up as time to train or performance differences during training. This may have been the case in the present study. Unfortunately, Experimental Session subjective workload ratings were not collected. This information may have indicated that, although recovery task performance did not differ, subjects were experiencing different levels of workload related to each display type.

The second explanation for the discrepancy between the findings of the past studies and the present study again revolves around the systematic subject selection confound that surfaced during the Training Session. Of the 15 subjects that were ineligible to participate in the Experimental Session, 14 were in the HUD group and one was in the ADI group. Out of all the subjects that attempted to reach the training performance criterion under the HUD format, only 42% performed well enough to participate in the Experimental Session. It is possible, if not likely, that the Training Session criterion set up a systematic selection of subjects who were more motivated and/or were more proficient at the task than those subjects who failed to meet the performance criterion. Only the more motivated and proficient HUD subjects were then advanced to the Experimental Session. As mentioned in the Training Session section, the G-HUD and ADI symbology structures enabled subjects of all proficiency and motivational levels to move onto the Experimental Session. Because the HUD group data are based on the most proficient and motivated subjects' performance, display effects were underestimated. Although, it should be kept in mind that the data trends were in the hypothesized direction.

The present study showed that the G-HUD was significantly easier to learn and demonstrated superior performance over that of the HUD format for an attitude maintenance task. From the present study it can be confidently concluded that the integrated symbology structure of the G-HUD did not produce a performance cost beyond that which is currently present in either the ADI or HUD format for the aid of unusual attitude recovery. Because of the subject selection in the present study, no statement can be made as to the performance benefit of the G-HUD in terms of aiding unusual attitude recovery. Further research is warranted to address the experimental design issues that have surfaced in the present study as well as to continue development of the integrated symbology concept into a viable flight instrument.

Further Research.

Follow-on studies should be designed so that they work toward eliminating the systematic subject selection confound that has surfaced in the present study. Following are two possibilities for accomplishing this: First, a Training Session similar to the present study should be utilized but subjects should not be deleted from one display condition without replacing an equal number of low scoring subjects with "high proficiency" subjects for the other display conditions. Also, the training performance criterion could be derived by pilot testing subjects on only the HUD format. This would help determine the criterion so that all subjects are able to reach the required level of performance. The findings of the present study indicate that this strategy is viable. It is unlikely that a criterion level derived from the HUD would be too difficult for the subjects of the remaining formats. The second method is to record some index of individual subject's motor ability. A tracking task could be performed to gain this data. Subjects could then be grouped and replaced according to skill level. Intersubject variability could be controlled further through an analysis of covariance. Also, if subjects failed to reach criterion performance, some insight into the cause could be gained. The motor skill index data could be utilized as an experimental co-variate to be used in subsequent analyses.

The present study has demonstrated the potential value of a G-HUD type integration as an aircraft attitude display. A follow-on research efforts should concentrate on refining the the G-HUD format for the flight environment. The study should include the evaluation of expanded G-HUD formats. In other words, the display size in terms of visual angle should be held constant and G-HUD displays of various fields-of-view could be tested. As the display is expanded the operator is presented with a partial view or a reduced display field-of-view. This in affect will act to decrease display clutter and compression proportionally. The evaluation should determine the point--percent of display field-of-view--at which a performance decrement arises. G-HUD transparency should also be empirically evaluated along with HUD type symbology.

In the interest of future attitude indicator research, it is the author's suggestion that the symbology structure of the ADI be studied at the elemental level. It is important that we systematically investigate the ADI structure at the feature level. This research should be accomplished similar to the method utilized by Taylor (1988) for HUD symbology variations. It is important that we determine the critical ADI display features that affect attitude awareness and attitude information conveyance. By learning the eccentricities of the displays we can better develop the optimal integration of display features and thereby develop an attitude indicator that is highly useful and affords natural and intuitive attitude cues. The optimized display will deliver maximum performance benefit while perceptual cost is minimized.

REFERENCES

Burns, R.K., and Lovering, P.B. (1988). F-15E EADI evaluation: a comparison of three formats. ASD-TR-88-5030 Final Report ASD AFSC Wright-Patterson AFB, OH.

Greenwald, A.G. (1976). Within-subjects designs: to use or not to use. Psychological Bulletin. 83(2), 314-320.

Guttman, J. (1986). Evaluation of the F/A-18 head-up-display for recovery from unusual attitudes. NADC-86157-60 Final Report, Naval Air Development Center. Warminster, PA.

Hoffman, J.E. (1980). Interaction between global and local levels of a form. Journal of Experimental Psychology: Human perception and Performance. 6(2), 222-234.

Johnson, S.L., and Roscoe, S.N. (1972). What moves, the airplane or the world? Human Factors. 14 (2), 107-129.

Kinsley, S. A., Warner, N. W., and Gleisner, T. P. (1985). A comparison of two pitch ladder formats and an ADI ball for recovery from unusual attitudes. Aircraft and Crew Systems Technology Directorate, Naval Air Development Center, Warminster, PA. (Report NADC-86012-60).

Navon, D. (1977). Forest before the trees: the precedence of global features in visual perception. Cognitive Psychology. 9, 353-383.

Osgood, R.K., and Venturino, M. (1990). Information representations for aircraft attitude displays. Proceedings of the Human Factors Society 34th Annual Meeting. pp 1542-1546 Human Factors Society, Santa Monica, Ca.

Previc, F. H. (1989). Towards a physiological based HUD symbology (Tech Report USAFSAM-TR-88-25). Brooks Air Force Base, TX: Human Systems Division.

Reising, J., Zenyuh, J., and Barthelemy (1988). Head up display symbology for unusual attitude recovery. Proceedings of the IEEE 1988 National Conference. pp 926-930 IEEE, New York, NY.

Roscoe, S.N., Corl, L., and Jensen, R.S. (1981). Flight display dynamics revisited. Human Factors. 23 (3), 341-353.

Roscoe, S.N. (1968). Airborne displays for flight and navigation. Human Factors,10. 21-332.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18. 643-662.

Summers, P.I. (1985). Multi-function display vs. the dedicated ADI. In: Aircraft Attitude Awareness Workshop Proceedings. Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio.

Taylor, R.M. (1988). Aircraft attitude awareness from visual displays. Displays. April. 65-75.

Taylor, R. M., (1984). Some effects of display format variables on the perception of aircraft spatial orientation. In AGARD-CP-371 NATO AGARD, France.

Taylor, R.M. (1982). Human factors aspects of aircraft head-up display symbology: the presentation of attitude information. In: Electronic Displays 1982. Session 3. Display Applications. Conference Proceedings. Network Exhibitions Ltd.Buckingham. ISBN: 0 904999 99 8: pp. 41-61.

Treisman, A.M., and Gelade, G. (1980). A feature-integration theory of attention. Cognitive Psychology, 12. 97-136.

Tukey, J.W. (1977). Exploratory data analysis. MA:Addison-Wesley.

Wickens, C.D. (1986). The object display: principles and a review of experimental findings. MDA903-83-K-0255 Interim Report, Army Research Institute. Alexandria, VA.

Wierwille, W.W., and Casali, J.G., (1983). A validated rating scale for global mental workload measurement applications. Proceedings of the 1983 Human Factors Society Annual Meeting. pp 129-133 Human Factors Society, Santa Monica, Ca.

Zenyuh, J.P., Reising, J.M., and McClain, J.E. (1987). Advanced head-up display (HUD) symbology: aiding unusual attitude recovery. Proceedings of the Human Factors Society 31st Annual Meeting. pp 1067-1071 Human Factors Society, Santa Monica, Ca.

APPENDICES

A. ANOVA Summary Tables

TRAINING SESSION

ANOVA Summary Table for Trials to Reach Criterion

Source	df	SS	MS	*F*	*p* =
Display	2	5653.42	2826.71	4.58	0.0193
Subject(Display)	27	16650.53	616.69		
Amplitude	2	21261.09	10630.54	32.82	0.0001
Amp*Sub(Disp)	54	17490.47	323.90		
Disp*Amp	4	6891.78	1722.94	5.32	0.0001
Amp*Sub(Disp)	54	17490.47	323.90		

TRAINING SESSION

ANOVA Summary Table for Roll RMS Error

Source	df	SS	MS	*F*	*p* =
Display	2	108.89	54.44	3.94	0.0314
Subject(Display)	27	372.79	13.80		
Amplitude	2	208.72	104.36	27.23	0.0001
Amp*Sub(Disp)	54	206.98	3.83		
Disp*Amp	4	2.75	0.69	0.18	0.9482
Amp*Sub(Disp)	54	206.98	3.83		

ANOVA Summary Table for Pitch RMS Error

Source	df	SS	MS	*F*	*p* =
Display	2	78.59	39.29	3.70	0.0380
Subject(Display)	27	286.82	10.62		
Amplitude	2	350.82	175.41	86.86	0.0001
Amp*Sub(Disp)	54	109.05	2.02		
Disp*Amp	4	13.37	3.34	1.66	0.1738
Amp*Sub(Disp)	54	109.05	2.02		

TRAINING SESSION

ANOVA Summary Table MCH Rating

Source	df	SS	MS	F	p=
Display	2	14.47	7.23	1.06	0.3608
Subject(Display)	27	184.43	6.83		
Amplitude	2	91.47	45.73	24.53	0.0001
Amp*Sub(Disp)	54	206.98	3.83		
Disp*Amp	4	3.87	0.97	0.52	0.7224
Amp*Sub(Disp)	54	100.67	1.86		

EXPERIMENTAL SESSION

ANOVA Summary Table for Roll RMS Error

Source	df	SS	MS	*F*	*p* =
Display	2	17.89	8.94	1.30	0.2882
Subject(Display)	27	185.30	6.86		

ANOVA Summary Table for Pitch RMS Error

Source	df	SS	MS	*F*	*p* =
Display	2	116.93	58.46	5.15	0.0128
Subject(Display)	27	306.61	11.36		

EXPERIMENTAL SESSION

ANOVA Summary Table for Decision Time

Source	df	SS	MS	*F*	*p*=
Display	2	0.26	0.13	0.27	0.7650
Subject(Display)	27	12.94	0.48		
Roll	3	9.37	3.12	76.26	0.0001
Roll*Sub(Disp)	81	3.32	0.04		
Pitch	1	4.78	4.78	141.74	0.0001
Pitch*Sub(Disp)	27	0.91	0.03		
Roll*Pitch	3	6.68	2.23	111.58	0.0001
Rol*Pit*Sub(Disp)	81	1.61	0.02		
Roll*Display	6	0.32	0.05	1.28	0.2742
Roll*Sub(Disp)	81	3.32	0.04		
Pitch*Display	2	0.004	0.002	0.05	0.9476
Pitch*Sub(Disp)	27	0.91	0.03		
Roll*Pitch*Disp	6	0.53	0.09	4.39	0.0007
Rol*Pit*Sub(Disp)	81	1.62	0.02		

EXPERIMENTAL SESSION

ANOVA Summary Table for Recovery Time

Source	df	SS	MS	F	p=
Display	2	21.14	10.57	2.15	0.1364
Subject(Display)	27	12.94	0.48		
Roll	3	244.08	81.36	100.20	0.0001
Roll*Sub(Disp)	81	65.77	0.81		
Pitch	1	230.68	230.68	170.31	0.0001
Pitch*Sub(Disp)	27	0.91	0.03		
Roll*Pitch	3	38.78	12.83	26.82	0.0001
Rol*Pit*Sub(Disp)	81	38.74	0.48		
Roll*Display	6	4.24	0.71	0.87	0.5204
Roll*Sub(Disp)	81	3.32	0.04		
Pitch*Display	2	7.56	3.78	2.79	0.0791
Pitch*Sub(Disp)	27	0.91	0.03		
Roll*Pitch*Disp	6	4.22	0.70	1.47	0.1992
Rol*Pit*Sub(Disp)	81	38.74	0.48		

EXPERIMENTAL SESSION

ANOVA Summary Table for Roll Accuracy

Source	df	SS	MS	*F*	*p* =
Display	2	0.05	0.03	2.85	0.0755
Subject(Display)	27	0.26	0.01		

ANOVA Summary Table for Pitch Accuracy

Source	df	SS	MS	*F*	*p* =
Display	2	0.07	0.03	2.28	0.1200
Subject(Display)	27	0.40	0.01		

B Experiment Instructions

1.

EXPERIMENT INSTRUCTIONS

We are currently conducting a study to evaluate various aircraft attitude display formats. Aircraft attitude displays represent the relationship between the aircraft being flown and the world over which it is flying. The attitude display has been referred to as an artificial horizon which is, in fact, what it is. An attitude display enables the pilot to stay oriented to his surroundings without having to see the actual horizon (night-flight or flight in clouds). An attitude display represents the aircraft's nose position above or below the horizon (pitch), right/left or inverted wing-bank angle (roll), and change in direction (heading). The present study is concerned with evaluating which of three displays is most able to represent aircraft attitude information.

You will be performing a number of tasks with only one of the three display formats involved in the evaluation. Please take a few minutes to turn to page four and familiarize yourself with the display you will utilize throughout the remainder of the experiment.

Your participation in the study will be divided into three distinct sessions of which the first two will be run today (DAY 1) and the third on your return visit (DAY 2).

2.

(DAY 1)

The first session consists of a "free-flight" period intended simply to familiarize you with the interaction between the control stick and the movement of the display. Session 1 will last for about five minutes. Session 2 is comprised of an attitude maintenance task intended to train you to accurately read and react to the attitude display. Basically, your task is to use the control stick to manipulate the display so that it represents as close to straight and level flight (zero pitch and zero roll) as possible while we disturb your aircraft's attitude by simulating turbulence. In other words, an updraft tends to push your nose up so you must counter the gust by pushing forward on the stick in order to keep your nose down toward the horizon.

Roll RMS and pitch RMS are two measures associated with the maintenance task. RMS (root mean squared error) computes your average distance (in degrees) from straight and level. Your task for session 2 is to obtain a 10 trial average score of 14 degrees or less in roll and 13.5 degrees or less in pitch. You will be performing this task under different turbulence conditions (low, moderate, and high respectively). You must reach criterion performance before you can move onto the next turbulence condition and ultimately DAY 2.

Between each turbulence condition the experimenter will ask you to answer a questionnaire in order to determine how much work you felt it took to perform the task.

We realize that this is a lot of information to digest but the procedure is easier than it sounds. Feel free to ask questions at any time during the experiment. Just relax, try to do your best and have fun! Think of this as a video game we pay you to play.

The instructions for the questionnaire will be given to you between session 1 and 2 and the instructions for DAY 2 (session 3) will be given to you just prior to your participation.

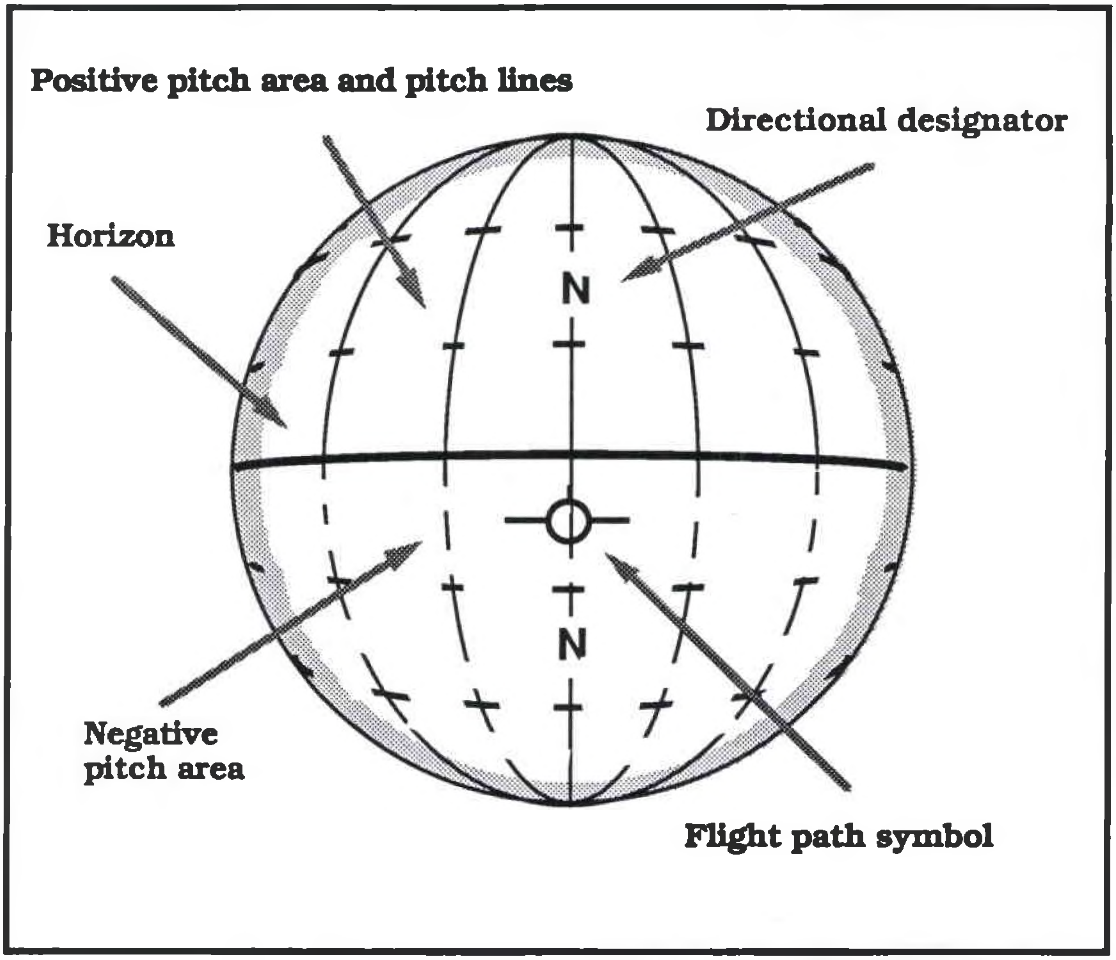
Session 1 procedure:

1. To begin, pull trigger on the control stick.
2. Study the display dynamics (up, down, right, left, and inverted).
3. Experimenter will end session.

Session 2 procedure:

1. To begin a trial, pull the trigger on the control stick at the status panel.
2. Try to maintain a straight and level indication by countering the turbulence with stick inputs. Each trial will last about 17 seconds.
3. Check performance at the status panel. You must perform 10 trials before the RMS average will appear. You are trying to get an average roll RMS of 14 or less degrees and an average pitch RMS of 13.5 degrees or less.
4. Perform subjective workload rating after reaching RMS criterion after each turbulence condition.
5. Repeat.

4.


Positive pitch area and pitch lines
Directional designator
Horizon
N
N
Negative pitch area
Flight path symbol

(DAY 2)

Today's session will utilize the same display as DAY 1 (see page four). Each trial you will have today will be divided into two phases. For the first phase you will perform an attitude maintenance task and during the second phase, an attitude recovery task. The attitude maintenance phase is performed just as it was on DAY 1. Your task is to maintain a straight and level display representation by cancelling the effects of simulated turbulence. After the attitude maintenance phase the monitor will blank and display a "WAIT" or "ZERO FORCE STICK" message. The message indicates whether or not you need to lighten your grip on the stick before the computer will begin phase 2. Please DO NOT take your hand off the stick. The computer is just attempting to determine that you are not making a purposeful input before the beginning of the trial. When the display comes back on the monitor it will represent a STATIC attitude. It is your task to decide what, if anything, should be done to make the display represent a straight and level attitude. Then, using the stick, move the display as quickly as possible (via the shortest route) so that it represents straight and level. For this phase of the trial, straight and level is within +/- five degrees in both roll and pitch. You should try to get as close to zero degrees as possible. After you been "recovered" for three seconds the trial will end.

After each trial, you will be given performance feedback for both phases of the trial. For the attitude maintenance phase you will be given roll and pitch RMS and the running 10 trial average of each. For the recovery phase you will be given your reaction time (time from the display presentation to your first stick input), recovery time (time from the display presentation until you first entered the recovery envelope), and whether or not your initial input was in the shortest direction (correct or incorrect for both roll and pitch). Please try to be as fast and as accurate as you can be. DAY 2 data collection will last for about one hour. You may rest between trials. Have fun!

Procedure:

1. Press trigger on control stick to begin phase 1 of the trial.
2. Perform attitude maintenance task.
3. Press trigger to begin phase 2 of the trial.
4. Zero input on the force stick and wait for display presentation.
5. Perform recovery to straight and level as quickly and accurately as possible. Straight and level is area within +/- five degrees in both roll and pitch.
6. Examine status panel for feedback information.
7. Repeat.

Please feel free to ask questions at any time.

C MCH Scale Instructions

MODIFIED COOPER-HARPER RATING SCALE

INSTRUCTIONS:

Overview

After each of the following turbulence conditions, you will be asked to give a rating on a Modified Cooper-Harper Scale for workload. This rating scale is shown on the sample which I have given to you on page six of this instruction packet. Before you run any trials, we will review:

1. The definition of the terms used in the scale,
2. The steps you should follow in making your rating on the scale, and
3. How you should think of the ratings.

If you have any questions as you review these points please ask the experimenter.

Important Definitions:

To understand and use the Modified Cooper-Harper Scale properly, it is important that you understand the terms used on the scale and how they apply in the context of this experiment.

First, "instructed task" is the attitude maintenance task you have been assigned to perform in this experiment. It includes manipulating the display within specified levels of accuracy and performing all duties that are requested of you during the time interval designated by the experimenter.

Second, the "operator" in this situation is YOU. Because the scale can be used in different situations, the person performing the ratings is called the operator. You will be operating the system and then using the rating scale to quantify your experience.

Third, the "system" is the complete group of equipment you will be using in performing the instructed task. Together you and the system make up the "operator/system". (For the present experiment, the system is composed of the computer monitor, the display, and force stick control.)

Fourth, "errors" include any of the following: mistakes, incorrect actions or responses, blunders, and incompletions. In other words, errors are any appreciable deviation from desired "operator/system" performance.

Finally, "mental workload" in the integrated mental effort required to perform the instructed task. It includes such factors as level of attention, depth of thinking, and level of concentration required by the instructed task.

Rating Scale Steps:

On the Modified Cooper-Harper Scale you will notice that there is a series of decisions which follow a predetermined logical sequence. This logical sequence is designed to help you make more consistent and accurate ratings. Thus, you should follow the logic sequence on the the scale for each of your ratings in the experiment.

9.

The steps which you will follow in using the rating scale logic are as follows:

1. First you will decide if the instructed task can be accomplished most of the time; if not, then your rating is 10 on the rating scale.
2. Second, you will decide if adequate performance is attainable. Adequate performance means that the errors are small and inconsequential in performing the instructed task. If they are not, then there are major deficiencies in the system and you should proceed to the right. By reading the descriptions associated with the numbers 7, 8, and 9, you should be able to select the one that best describes the situation you have experienced. You would then circle the most appropriate number.
3. If adequate performance is attainable your next decision is whether or not your mental workload for the instructed task is acceptable. If it is not acceptable, you should select a rating of 4, 5, or 6. One of these three ratings should describe the situation you have experienced, and you would circle the most appropriate number.
4. If mental workload is acceptable, you should then move to one of the top three descriptions on the scale. You would read and carefully select the rating 1, 2, or 3 based on the corresponding description that best describes the situation you have experienced. You would circle the most appropriate number.

Remember you are to circle only one number, and the number should be arrived at by following the logic of the scale. You should always begin at the LOWER LEFT and follow the logic path until you have decided on a rating. In particular, do not skip any steps in the logic. Otherwise, your rating may not be valid and reliable.

How You Should Think of the Rating:

Before you begin making ratings there are several points that need to be emphasized. First, be sure to try to perform the instructed task as instructed and make all your evaluations within the context of the instructed task. Try to reach adequate performance as specified for your task.

Second, the rating scale is not a test of your personal skill. On all of your ratings, you will be evaluating the system for a general user population, not yourself. You may assume you are a member of that population. You should make the assumption that problems you encounter are not problems you created. They are problems created by the system and the instructed task. In other words, do not blame yourself if the system is deficient, blame the system.

Third, try to avoid the problem of nit picking an especially good system, and of saying that a system which is difficult to use is not difficult to use at all. These problems can result in similar ratings for systems with quite different characteristics. Also, try not to overreact to small changes in the system. This can result in ratings which are extremely different when the system themselves are quite similar. Thus, to avoid any problems, just always try to "tell it like it is" in making your ratings.

If you have any questions, please ask the experimenter at this time.

11.

MODIFIED COPPER-HARPER SCALE