Aviation Human Factors Division
Institute of Aviation

University of Illinois
at Urbana-Champaign
1 Airport Road
Savoy, Illinois 61874

Compensation for Display Enlargement in Flight Control and Surveillance

Emily K. Muthard and
Christopher D. Wickens

Technical Report
AHFD-04-3/NASA-04-1

March 2004

Prepared for

NASA Ames Research Center
Moffett Field, CA

Contract NASA NAG 2-1535

Compensation for Display Enlargement in Flight Control and Surveillance

EMILY K. MUTHARD and CHRISTOPHER D. WICKENS

ABSTRACT

Two experiments were conducted to assess the impact of display size on flight control, airspace surveillance, and goal-directed target search. In Experiment 1, 16 pilots completed a basic flight control task under single and dual axis control. Pilots exhibited less path error and greater stick activity with a large-scale display, suggesting that larger depictions of error lead to greater urgency in correcting deviations. Experiment 2 scaled-up findings from Experiment 1 for flight control, hazard surveillance, and target search tasks. Results from Experiment 1 were replicated for the flight control task. Size, however, did not affect surveillance or search because pilots were adaptive in altering scanning patterns in response to increases in display size. Practical implications of these results are that displays should not be minified to fit within cramped cockpits, as such changes will hinder flight control.

With the influx of aviation technological advances and limited cockpit "real estate," designers have been faced with the dilemma of reducing the sizes of displays in order to preserve their presence in the cockpit. There are also competing pressures for full-scale panoramic displays that synthetically present the world as it might be seen out the cockpit window (e.g., Prinzel, Comstock, Glaab, Kramer, and Arthur, 2004; Schnell, Kwon, Merchant, and Etherington, in press). Though these trends are leading designers to vary display sizes, it remains unclear how such changes in size may impact pilot performance in carrying out an array of task responsibilities. Three common tasks that the pilot must complete when flying with an integrated hazard display include (1) deviation judgment and resulting flight control (e.g., tracking); (2) goal-directed search for target hazards, such as looking in the airspace for a specific aircraft identified by air traffic control; and (3) the surveillance of the airspace for traffic aircraft that may pose conflicts or weather systems that may impact flight safety.

Flight Control and Tracking

A major goal of many aviation displays is to support pilot's estimation of deviations between the state of his or her own aircraft ("ownship"), and some target value or position. When such a deviation is to be corrected by a compensatory control action, the pilot is engaged in tracking behavior, although such estimation may also be required in the absence of control. The deviations and movement of the aircraft is, of course in "world units" (WU) such as meters, nautical miles, or true heading angles. However, the expression of these deviations is in display units (DU), such as pixels, centimeters or degrees of visual angle. Our current investigation asks what happens when the display changes the size of display units with which it represents world units, while the ratio of control displacement to world units remains constant. That is, for example, a given deflection of the control yoke of an aircraft produces the same aircraft movement in world units; expressed in such terms as a meters/sec deviation from a flight path.

Figure 1 shows four different ways in which this ratio of display units to world units might be reduced. In Figure 1a, the physical dimensions of a display are made smaller. In Figure 1b, the field of view of a given display is made larger (or the map scale is made smaller). In Figure 1c, the display is viewed from a greater distance. Finally, in Figure 1d, two axes of a volume of space are viewed on a 3D display, and that axis of world space that is parallel to the line of sight of the observer to the display is compressed, relative to the two axes that are parallel to the display surface (only the vertical axis is shown in the figure).

Our research contrasts three possible hypotheses regarding how the pilot tracks dynamic systems with different ratios of DU/WU. In the **ecological hypothesis,** the pilot is able to make differences in display scale totally transparent, tracking the aircraft as it exists in the world (the ecology). Hence tracking performance, as reflected for example by RMS error and by control activity, will be totally unaffected by scale differences. While the ecological hypothesis would seemingly define an optimal pilot, data from other domains suggest that operator performance may be influenced ("contaminated") by surface features of a display. In particular, evidence from research on 3D displays suggests that people underestimate differences on a compressed axis, an effect that is expressed in the tendency to mentally rotate perceived vectors more parallel with the display viewing surface than they are (McGreevy and Ellis, 1986; Perrone, and Wenderoth, 1993; Wickens, 2002). This effect has been attributed as the cause of increased tracking error

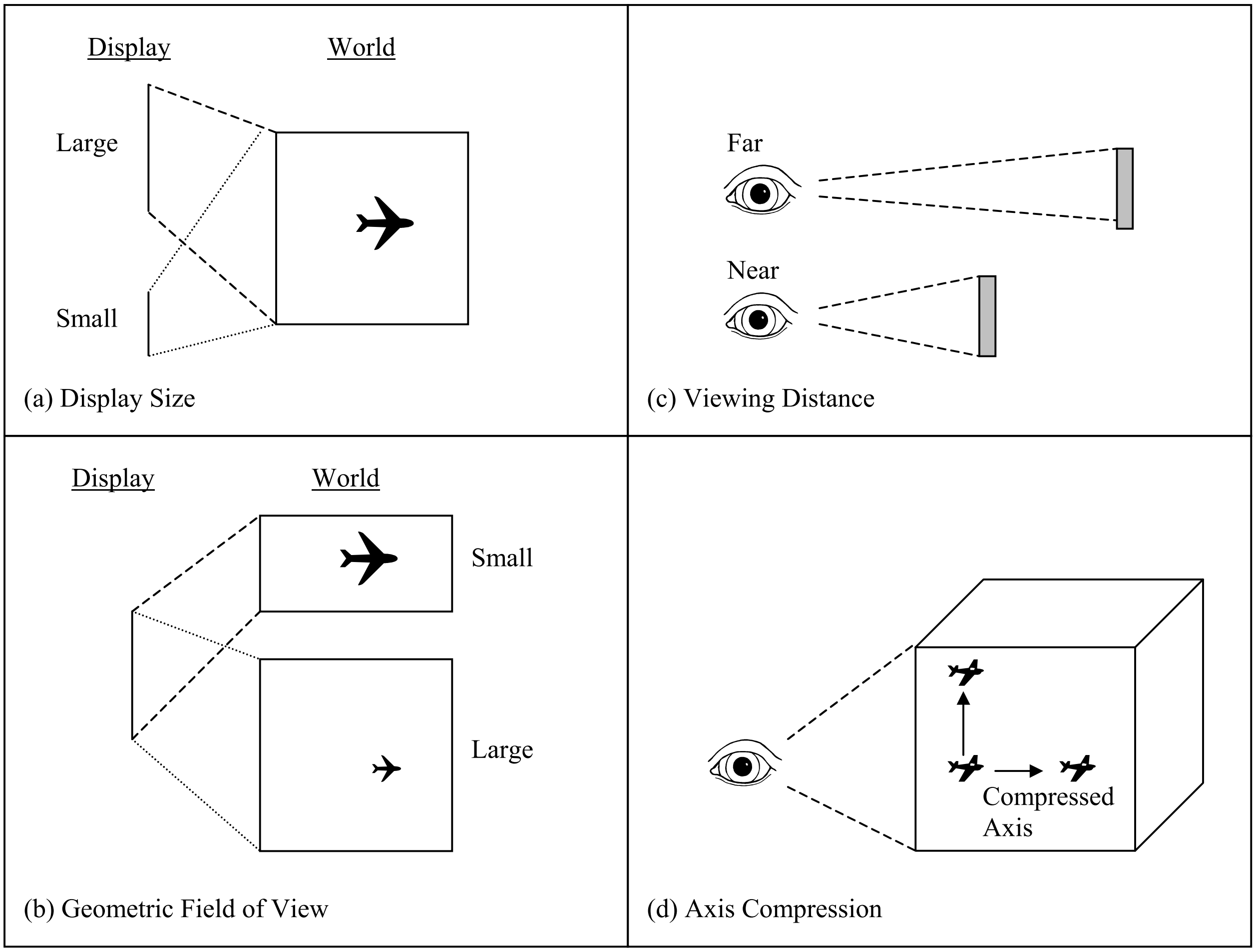

Figure 1. Differences in display size due to (a) physical size, (b) geometric field of view, (c) viewing distance, and (d) the compression of an axis in three-dimensional displays. In (d) the compressed axis is represented by fewer display units and a smaller visual angle than the uncompressed (here the vertical) axis.

along a compressed axis within a 3D display (Haskell and Wickens, 1993; Wickens, Liang, Prevett, and Olmos, 1996).

This interpretation suggests two possible sources of "contamination" to the ecological perception of display error. The **resolution** hypothesis posits that with smaller displays and smaller display units, the smallest distances or movements will simply be sub-threshold, not noticed, and hence, fail to be corrected at all. The **urgency** hypothesis, suggests that for all distances (and movements), not just those around threshold, smaller display units will suggest smaller errors, and hence, less urgency to correct those errors, just as it is found that within a display of fixed size, larger errors will be corrected with more urgent (greater velocity) control movement (McRuer and Jex, 1967; Onstott, 1976; Wickens, 1986). While the ecological hypothesis predicts no effect at all of display size, the extreme view of the urgency hypothesis predicts that perceived urgency is directly proportional to display error size; pilots will simply track display units and ignore world units, and hence tracking error measured in world units will be inversely proportional to display size. Because the resolution hypothesis only applies to the perception of small (near threshold) errors, it would not predict as great an error increase with decreasing size, as would the urgency hypothesis.

There is another discriminating feature between these hypotheses. If all deviations (in world units) are expressed as equally important, then neither the ecological hypothesis, nor the resolution hypothesis will predict tracking performance to be influenced by how scale differences are created. For the resolution hypothesis, this results because the creation of sub-threshold movements and distances will be the same, no matter which way these are achieved in the context of Figure 1. However an urgency hypothesis posits a cognitive influence which will be considerably more malleable to differences in the cause of display differences (e.g., differences between each of the panels in Figure 1). In a dual task (dual tracking) environment, it suggests that increased urgency or a tracking axis should be associated with the allocation of resources, relative to a less urgently perceived (ie., smaller display units) axis.

Empirical data on the effects of size on tracking behavior and performance is relatively sparse. Consistent with the urgency hypothesis, Abbott and Moen (1981) found that pilots maintained a more accurate spacing interval behind a lead aircraft when a cockpit display of traffic information was increased in size. In contrast, Alexander, Wickens, and Hardy (2003) found an influence of scale (as rendered by field of view in Figure 1b) on flight control performance that was opposite to the urgency hypothesis; larger scale produced greater error. Prinzel and colleagues (2004), however, reported no effect of display size. Finally, as noted above, Haskell and Wickens (1993) found in multi-axis flight simulations, that pilots showed greater tracking error in a compressed axis within a 3D display.

The objective of Experiment 1 was to try to examine the relative contribution of the mechanisms postulated in the 3 hypotheses (ecological, resolution, and urgency) to tracking performance, as display size is varied with the two manipulations shown in Figure 1a (display size) and 1d (3D compression), in both single task conditions, when resources were adequate, and in dual task conditions when they were scarce, and therefore might be allocated differentially between axes, to the extent that urgency differences are operating. Experiment 2 was designed to measure the impact of display size on flight control, while also examining how display size also affects hazard surveillance and search with an integrated hazard display.

EXPERIMENT 1

In Experiment 1, the effects of display size on tracking error and control activity were assessed under single and dual axis control in a first-order, compensatory tracking task. Display size was manipulated by varying axis size in the 2D display and through axis compression, inherent in the 3D display. The impact of display size was measured by examining tracking error and control activity within the framework of the three proposed hypotheses.

Under the assumptions of the ecological hypothesis, the operator would be expected to account for differential depictions of world units as display units, and thus we would expect to find no difference in tracking with variations in display size. To the extent that control performance is found to be inferior with the smaller display, however, the resolution or urgency hypothesis is validated. Recall that the resolution hypothesis postulates that some errors depicted in the small display are below the threshold for detection. Under the resolution hypothesis, we should also expect to find that performance differences associated with the reduction in display size with the 2D display are identical to those found with axis compression in the 3D display, to the extent that these display size manipulations are equal in magnitude. To the extent then that errors are somewhat larger and control activity is also less with display minification and that performance is affected equally by size in the 2D and 3D displays, we can attribute these differences in performance to reductions in depicted error size and thus validate the resolution hypothesis. If, however, the magnitude of the performance decrement is fully equal to the difference in display size, if the performance decrement is amplified under dual axis control, or if size manipulations differentially affect performance with the 2D relative to the 3D displays, then these differences would support the urgency hypothesis, which argues that smaller displays lead to perceptions of less error and thus less urgency to correct the deviation. Further support for the urgency hypothesis is provided to the extent that differences between axes, which can be attributed to the manner in which axis compression is rendered (2D vs. 3D), are amplified when resources are scarce (dual axis condition).

Methods

Participants. Participants were 16 aviation students from the University of Illinois Institute of Aviation, who ranged in age from 18 to 32 years (*M* = 19.9 years). Thirteen participants had their private piloting licenses, while the remaining three participants had student licenses. All participants were right-handed.

Displays. A 2D and 3D display were used (see Figures 2 and 3). The 2D display depicted a vertical and a horizontal axis in two separate panels that were separated by 17 degrees of visual angle (19 cm). The vertical axis was always represented in the left panel, while the horizontal axis was always shown in the right panel. The cross in the center of each panel represented the target location for each axis. The sizes of the scales were varied to be small (about 2 degrees of visual angle or 2 cm) or large (about 4 degrees of visual angle or 4 cm). A scale, presented in the lower left-hand corner of each panel, and tick marks, shown along the edges of each panel, were used to denote distance from the target point within each display panel. Distance on the screen was depicted as pixels, but representative of meters in the world (i.e., world units).

The 3D display depicted both the horizontal and depth axes in a single, integrated panel. Lines, which converged with increasing distance, were added to further convey the sense of display depth. As was the case in the 2D display, a cross was located in the center of the panel to represent the target location for the axes and a scale and tick marks were included to denote distance from the target location. Size was not explicitly manipulated, though the depth axis was the same size as the small axis in the 2D display (about 2 degrees of visual angle or 2 cm), due to the compression associated with representing depth in only two dimensions. The horizontal axis was the same size as the large display in the 2D display, as this axis was not compressed. Thus, the degree of axis compression in the 3D display was equal to the degree of size reduction in the 2D display.

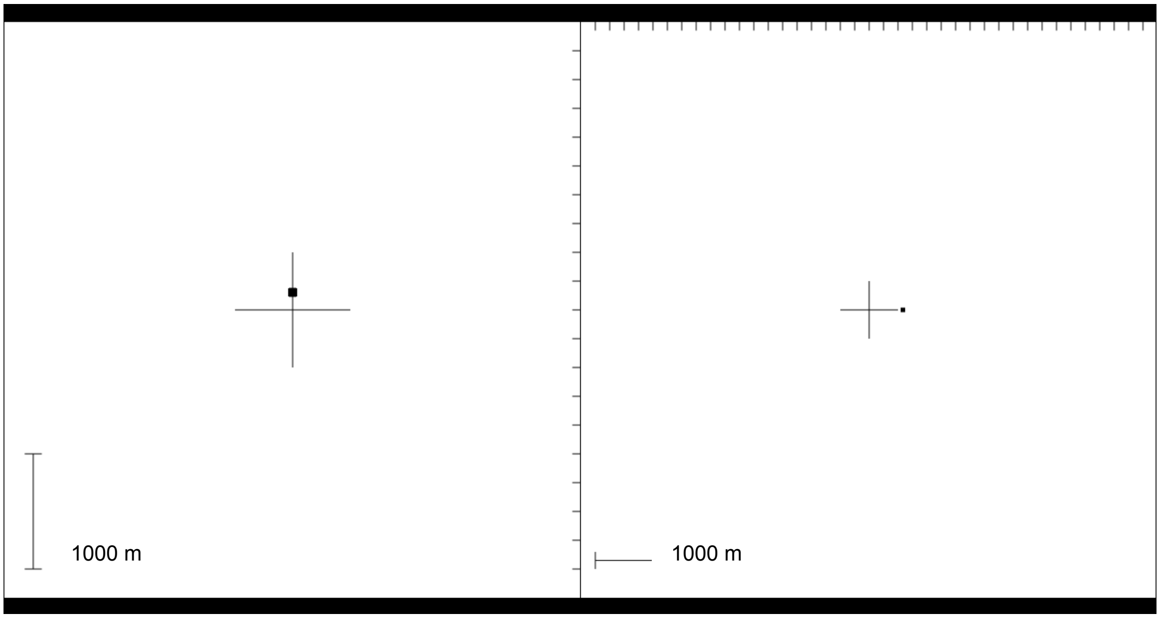

Figure 2. Two-dimensional display with large vertical axis in left panel and small horizontal axis in right panel.

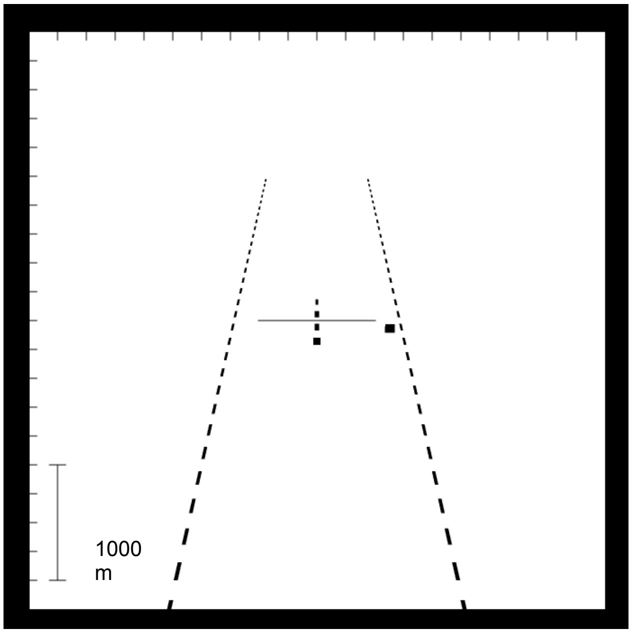

Figure 3. Three-dimensional display.

Task. Participants completed a first-order, single or dual-axis compensatory control task using the 2D or 3D display. Participants were asked to keep the icon, representing the pilot's ownship, within 500 meters of the target location at all times. The bandwidth of the disturbance was 0.5 Hz, and participants were told that the icon would move randomly in all possible directions. When using the 2D display, participants were asked to control either the vertical axis alone, the horizontal axis alone, or the two axes simultaneously. Control in the 2D display was accomplished through the use of two joysticks. The vertical axis, which was depicted in the left panel, was controlled with the left hand through the left joystick. The horizontal axis, conversely, was controlled with the right hand and the right joystick. When both axes were controlled simultaneously, both joysticks were used.

When using the 3D display, participants were asked to control the depth axis alone, the horizontal axis alone, or the two axes simultaneously. The right joystick, manipulated with the right hand, was used to control both axes. An integrated control format was chosen for the integrated 3D display and separate controls were chosen for the 2D display in order to maximize display-control compatibility within each display type (Fracker and Wickens, 1989). Participants completed a practice trial with both the 2D and 3D display and nine experimental trials, each of which lasted two minutes. The entire experimental session lasted thirty minutes.

Experimental Design. Display dimensionality, axis size, and task load were manipulated and counterbalanced in a within-subjects design. Dependent variables included root mean square (RMS) error and RMS control velocity, the latter measuring control effort. Axis direction, in the 2D display, was also manipulated as a within-subjects variable. In the 3D display, axis direction was not directly manipulated, but differed with axis size (i.e. the small axis was always the depth axis, while the large axis was always the horizontal axis). Each participant performed one trial within with each display condition.

Results

Figure 4 presents the tracking error data, for the 2D and 3D displays, as a function of display size (or compression, in the case of the 3D display), collapsed over single and dual axis conditions. Shown in the figure also are the predicted error differences under the pure ecological hypothesis (E), and the pure urgency hypothesis (U). Figure 5 presents the same data for control activity. On the one hand, the data clearly do not fully support the ecological hypothesis, as there was a significantly larger error for the smaller display size for both 2D and 3D displays ($F_{(1,\,15)} = 30.7$, $p < 0.001$; $F_{(1,\,15)} = 39.0$, $p < 0.001$, respectively). Size then did, to some extent "contaminate" the estimation of world error. On the other hand, this difference was relatively small (a ratio of small/large = 1.2), compared to the ratio of display sizes (2.0) reflected in the predicted error scores under the urgency hypothesis. The amount of display "contamination" was slight. These findings are also somewhat mirrored by control activity, as shown in Figure 5 ($F_{(1,\,15)} = 22.7$, $p < 0.001$ for 2D; $F_{(1,\,15)} = 67.8$, $p < 0.001$ for 3D; ratio large/small = 1.1 for 2D and 1.5 for 3D).

Further support for the expression of urgency in the data is provided by contrasting the size induced effects for 2D and 3D displays. The differences in these effects, supported by a significant dimensionality X size interaction ($F_{(1,\,15)} = 9.8$, $p = 0.007$) is in the direction such that a size reduction due to axis compression within the 3D display creates a greater increase in error, than does the reduction of the same magnitude within the 2D display. If the ecological hypothesis were exclusively operating size would exert no effect in either display. If the resolution hypothesis were in operation, then size should impose the same effect independent of the cause of the resolution reduction (display scale or axis compression). The data suggest that subjects' greater effort investment for the expanded axis was amplified when this expansion was accomplished by rotating the axis in a 3D display, than when it was accomplished by increasing the size of the display. This difference was also reflected in the significant interaction between

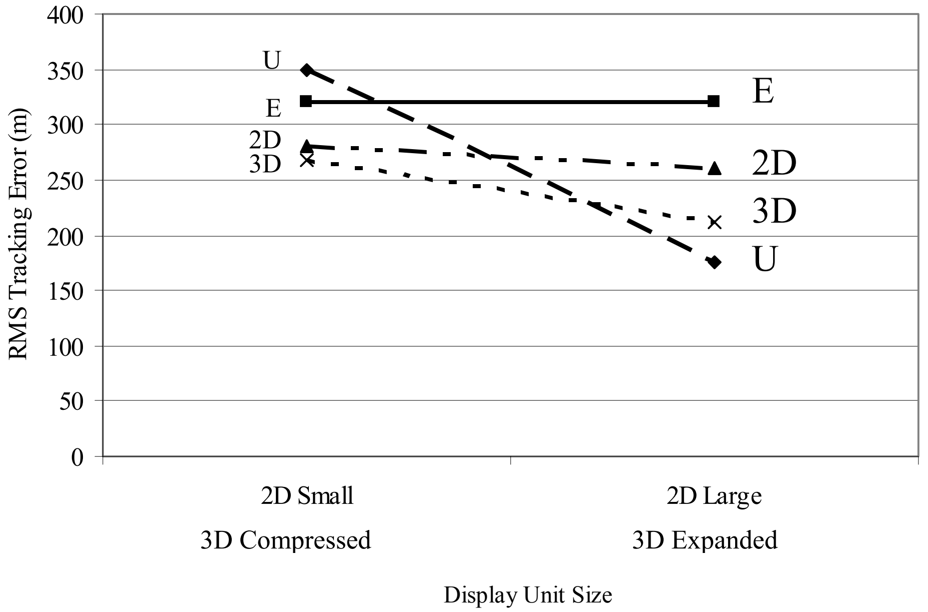

Figure 4. RMS tracking error as a function of axis compression and display dimension. The light solid and dashed lines are the error differences that would be predicted under the pure ecological (E) and urgency (U) hypotheses, respectively. Note that the difference in error under urgency is a 2:1 ratio, associated with the ratio of size differences employed.

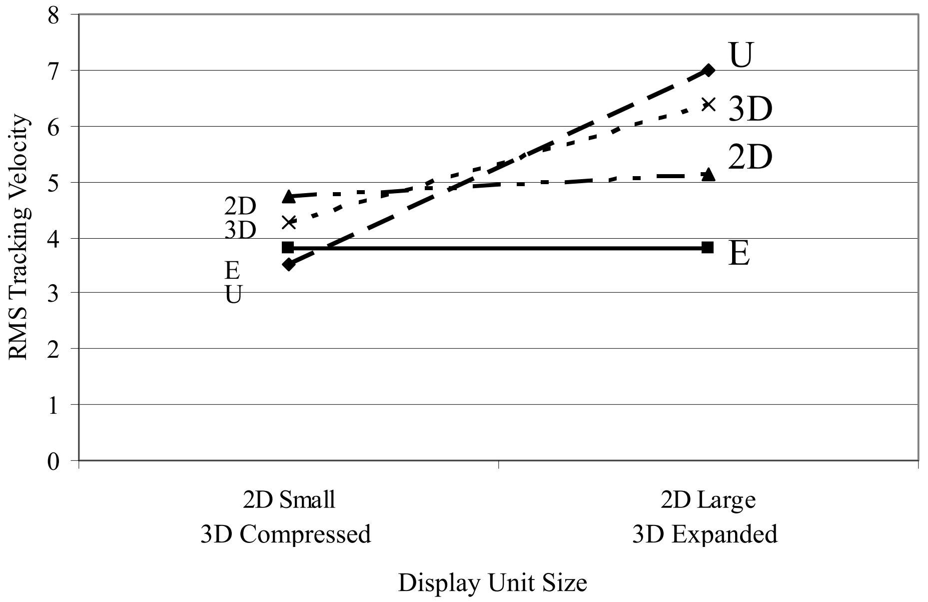

Figure 5. Control activity as a function of axis compression and display dimension. The light solid and dashed lines are the velocity differences that would be predicted under the pure ecological and urgency hypotheses, respectively. Note that the difference in velocity under urgency is a 2:1 ratio, associated with the ratio of size differences employed.

dimensionality and size in control activity ($F_{1, 15)} = 41.92$, $p < 0.001$), with a larger effect of size for the 3D than for the 2D display (Figure 5).

Finally, to the extent that the allocation of effort, associated with "urgency," is a resource limited process, then we might expect a greater effect of size in dual axis tracking when resources were at a premium, than in single axis tracking when they were less so. Surprisingly, this finding was not confirmed. Task load (single vs. dual axis control), while significant ($F_{(1, 15)} = 66.4$, $p < 0.001$), and while interacting significantly with display dimensionality ($F_{(1, 15)} = 80.7$, $p < 0.001$; greater cost in 2D than 3D), did not interact with size; either in the 2 way interaction, or in the 3 way size X dimensionality X load interaction ($p > 0.10$ for all analyses).

Discussion

The present experiment was designed to evaluate the ecological, resolution, and urgency hypotheses of the effects of size on simulated flight control. The results indicated that flight control became inferior with decreases in display size and with axis compression, thus eliminating the applicability of the pure ecological hypothesis, which posited that pilots would fully compensate for differences in the ratio of displayed units to world units which results from changes to display size. The results therefore provided support for the contributions of the resolution and urgency hypotheses. Evidence, however, that showed that control performance was affected by size differentially as a function of display dimensionality failed to corroborate the resolution hypothesis, thus suggesting that the manner in which size is manipulated has a direct influence on subjective urgency. Specifically, the evidence suggested that participants felt less urgency to correct deviations in small displays, particularly when minification was accomplished through axis compression in the 3D display. These data would then appear to validate the explanation offered by Wickens et al. (1996) as to why pilots showed larger error when tracking a compressed axis on a 3D aviation display in a more realistic simulation.

Though we found evidence in support of the urgency hypothesis, the reduction in error and increases in control activity with larger display units were small relative to the increases in display size, suggesting that effects of urgency were minor (though larger in the 3D than in the 2D display). Given that urgency is associated with effort and resource-driven processes, we would also expect that, if the size effect was in fact a result of urgency, this effect should be amplified when these resources are heavily tapped, as in the dual task condition. The lack of interaction of size with task load provides additional support to the finding that urgency plays only a moderate, but meaningful role in manual control.

Experiment 1 examined flight control in a low-fidelity simulation. Analyses revealed that perceived urgency in correcting path deviations serves as a driving factor for the performance decrement that results with display minification. Though flight control is the pilot's task of primary importance during flight, he or she is also responsible for monitoring and searching for hazards in the airspace. Experiment 2 was designed to examine how changes in display size affects flight control, along with surveillance, and search tasks while using an integrated hazard display in a low-fidelity flight simulation.

EXPERIMENT 2

Surveillance and Target Search

A noted benefit of an integrated hazard display, which overlays weather, air traffic, and terrain information in a single display, is that the display supports better awareness of the hazards in the environment and allows the pilot to more easily monitor the dynamic airspace for changes in the lateral and vertical behavior of all three classes of air hazards. This task may be directly mapped onto the paradigm examining "change blindness" (Carpenter, 2001; Rensink, 2002; Simons, 2000), which refers to the great difficulty people have in readily detecting changes that occur outside of the region of focused attention (Levin and Simons, 1997; Rensink, O'Regan, and Clark, 1997). Change blindness has been noted in such real-world, complex environments as vehicle driving (Pringle, Irwin, Kramer, and Atchley, 2001), battlefield monitoring (Wickens, Thomas, and Young, 2000), aviation automation monitoring (Mumaw, Sarter, and Wickens, 2001), and in monitoring changes to aircraft information on a traffic display (Muthard and Wickens, 2002). While integrated hazard displays assist pilots in simultaneously monitoring all air hazards for changes, they also provide pilots a platform for goal-directed search for a specific element or hazard in response to an information probe (e.g. an air traffic control request).

The tasks of surveillance and search can be affected by increases in display size (Figure 1a) through an enlargement of the area needing to be scanned as well as by increasing the physical effort required to access information with saccades from the area of focused attention to the data source, particularly when such displays are highly cluttered (Wickens, 1992). As research examining surveillance has noted, focused attention is necessary to detect environmental changes. To this end, enlarging the size of a display will also increase the display area that is located within the periphery, while decreasing the proportion of the display located in the fovea, and will subsequently require a greater number of saccades to the outer display regions in order to optimally survey the entire display. As display eccentricity has been shown to damage change detection performance (Pringle et al., 2001), display enlargements, which place information at greater eccentricities from the display center, can be predicted to hinder monitoring and surveillance performance further, via the mechanism of change blindness.

Goal-directed search may also be harmed by display enlargements, which is reflected in the information access cost model proposed by Wickens (1992). This model describes a function in which access costs increase nonlinearly with eye movements and head movements of increasing magnitude. Thus small displays, within which display elements are located closely to one another, will require only eye movements, for which the physical effort is low regardless of the size of the saccade (Martin-Emerson and Wickens, 1997). Search in large displays, that exceed twenty or more degrees of visual angle, may begin to require head movements (Bahill, Adler, and Stark, 1975), which become linearly more difficult as they increase in magnitude (Wickens, 1992). When the search process involves the comparison of more than one display element, and thus requires the use of divided attention, the negative influences of larger display size are further amplified (Kroft and Wickens, 2003; Yeh and Wickens, 2001), as access costs would be manifest in both the search and comparison phases. Such a comparison would also impose a cognitive cost beyond the physically effort described, as the larger separation of elements would hinder their mental integration (Wickens and Carswell, 1995).

The goal of Experiment 2 is to evaluate the feasibility of two contrasting hypotheses of the effects of display enlargements on surveillance and search. The review of the research supports the hypothesis of **effort conservation** which posits that the display periphery will suffer proportionally more with enlargements of the displays. Furthermore, the amount of "suffering", which can be defined as the ratio of change detection performance in the smallest display to that in the largest display, should be directly proportional to the ratio of display sizes (0.31:1 in the present experiment). This would be the case if scanning behavior, in visual angle units, was unaffected by display size. The opposing hypothesis of **strategic compensation** postulates, however, that participants recognize the need to adapt scanning patterns in response to display enlargements and thus expend additional effort to survey the larger display. This hypothesis would be supported to the extent that participants detect as many changes in the center and peripheral display regions across all three display sizes. Surprisingly, given the importance of this design issue, few studies have directly examined the effects of size on surveillance and search. Enoch (1959) examined search for a target in displays ranging from 3 to 51 degrees of visual angle and found that fixations were concentrated around the center of the display with maps larger than 9 degrees, though performance in the search task was not discussed. Kroft and Wickens (2003) found that search actually appeared to be inhibited on a smaller display, but attributed this inhibition to negative effects of clutter and small symbol size on the smaller display. Large displays (21 inch) where also found to improve performance in search of spreadsheets and tables (Simmons, 2001). In addition to evaluating these two hypotheses, this subsequent study was designed to replicate, in a more realistic presentation of a flight task, the findings from Experiment 1, which supported the hypothesis that larger displays help tracking performance by increasing subjective urgency.

Methods

Participants. Nineteen student pilots from the University of Illinois Institute of Aviation participated in the study. These pilots ranged in age from 19 to 23 years (*M* = 21 years) and all were male. The participants had an average of 226 flight hours of experience and six participants had their private license, while the remaining thirteen were instrument certified.

Display. Participants were presented with a 2D, top-down integrated hazard display that depicted a topographical map based on the National Oceanic and Atmospheric Administration's section aeronautical chart, which was overlain with traffic and weather information (see Figure 6). Ownship was depicted with an aircraft icon that was larger than the traffic aircraft icons and was always located in the center of the hazard display. Additionally, ownship always remained stationary within the display, while traffic, weather, and terrain moved relative to ownship. An attitude directional indicator, which depicted only pitch, was placed directly below ownship to assist in altitude control. Ownship aircraft heading could be deduced only from the directional heading (angle) of the icon. Three display sizes were presented. The small display measured 8.9 cm by 6.4 cm (10° by 7°). The medium and large displays measured 19.1 by 14.0 (20° by 15°) and 34.3 cm by 25.4 cm (36° by 27°), respectively. With changes in display size, the text and icon size of all information present on the display also changed proportionately.

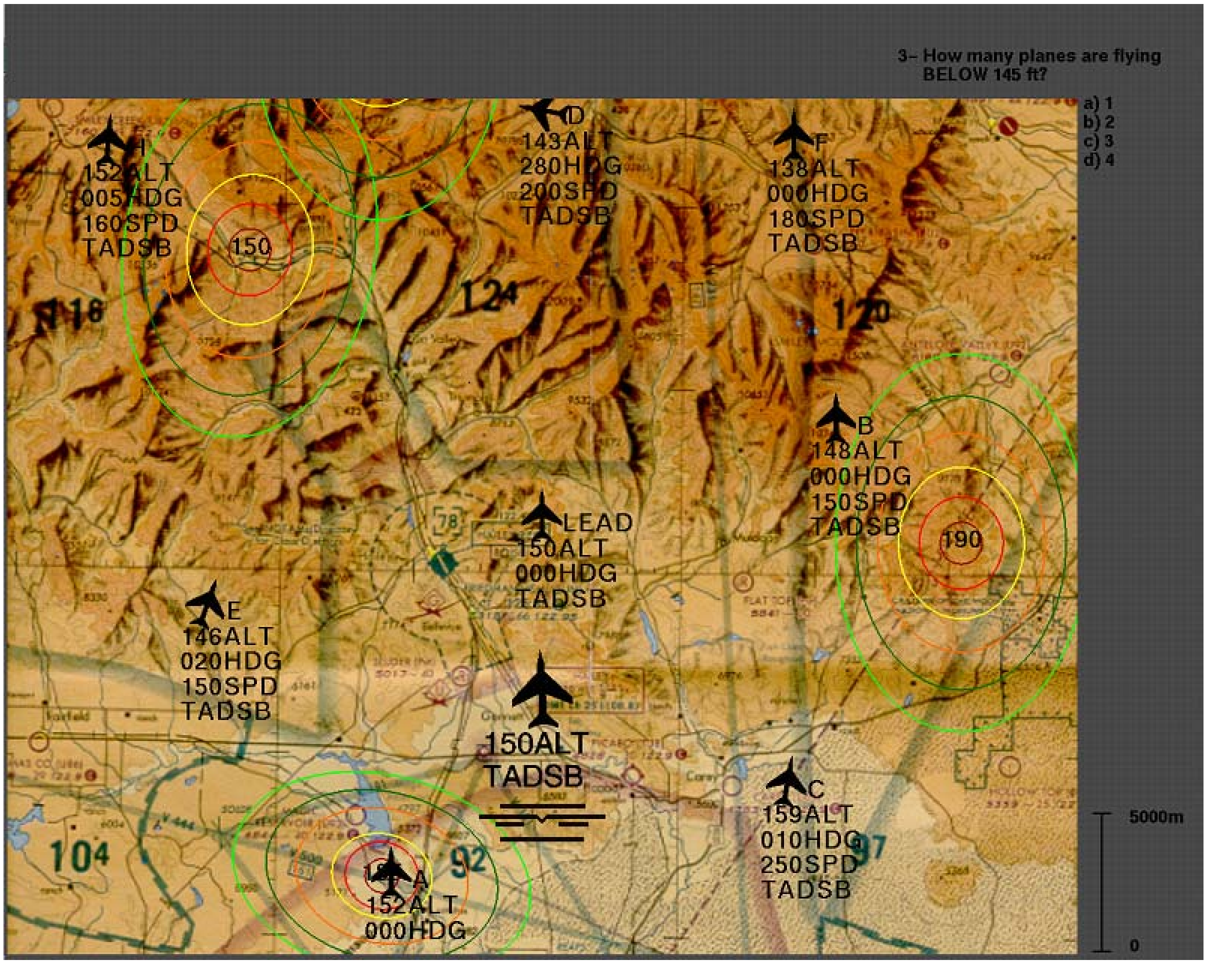

Figure 6. Integrated hazard display depicting terrain, weather, and air traffic databases. Ownship is the largest aircraft icon located near the center of the display. The aircraft immediately north of ownship is the lead aircraft. Questions were presented in the upper-right hand display corner.

Tasks. Participants were asked to complete three sets of tasks, the first of which centered on control of the pilot's ownship. Pilots were first asked to maintain a target altitude of 15,000 feet and the north-up heading of the aircraft through control inputs with a joystick. Altitude information was available directly through the datatag located below ownship and in terms of altitude change, via pitch, which could be deduced from the displacement of the attitude directional indicator (ADI). Heading information could only be determined from the orientation of the ownship icon. It is important to note that changes in display size varied the size of the text located in ownship's datatag as well as the size of the ADI and aircraft icon. Participants were also required to maintain a distance of 5,000 feet from a lead aircraft by increasing or decreasing ownship airspeed. The target distance of 5,000 feet was depicted in a scale that was located on the right-hand side of the display. The second class of tasks involved examining the airspace for changes in the altitude, heading, or airspeed of traffic aircraft and weather systems. Implemented changes occurred randomly every fifteen to seventy seconds, and seven changes occurred in each trial. Participants were instructed to press a key and verbally identify the hazard that changed and the nature of the change (e.g. "aircraft C changed heading").

In the guided search task, participants were asked to answer multiple choice questions about hazards in the airspace, generating goal-directed search. Questions appeared in the upper right hand corner of the display every 40 to 75 seconds. There were a total of eight questions in each trial. Some questions in each trial were designed to require distance or speed judgments (e.g. What is the distance between aircraft C and the center of the weather top at 15,000 ft?), while others only required the participant to focus on a single hazard (e.g. What is the heading of aircraft F?). Event timing was constrained so that a change was not presented while the pilot would be engaged in question answering. Participants completed one practice trial and six experimental trials. Each trial lasted six minutes, and the experimental session lasted for one hour.

Eye Movement Apparatus. Eye movement data were collected on twelve of the nineteen participants using a Model 501 Applied Science Laboratories integrated eye and head tracker. The eyetracker collected both eye and head position at 60 Hz.

Experimental Design. Performance was assessed as a function of display size in a repeated-measures design. Trials were grouped and counterbalanced by display size, and participants completed two trials with each display size.

Results

Distance Estimation. The effects of size on distance estimation were examined by measuring ownship heading, vertical, and airspeed tracking error, as these all reflect errors in estimating the distance between an a target state and the actual state, as shown in Figure 7. Analyses revealed that heading RMS error ($F_{(2, 36)} = 16.3$, $p < 0.001$), and altitude RMS error ($F_{(2, 36)} = 6.8$, $p = 0.003$) were significantly increased by decreasing display size. Thus, pilots had the greatest path deviation with small displays, relative to medium (heading: $t_{(18)} = 3.8$, $p = 0.001$; vertical: $t_{(18)} = 2.3$, $p = 0.03$). In light of the analyses in Experiment 1, it is important to note the nearly twofold increase in vertical tracking error, which directly mirrors the 50% reduction in the size of the ADI, which is used to estimate vertical error change. There was no benefit to lateral and vertical tracking, however, when displays were increased from medium to large ($p > 0.10$). Display size had no significant effect on the pilots' abilities to maintain the target distance from the lead aircraft by varying ownship speed ($p > 0.10$).

Guided Search. The effects of display size were assessed for performance in answering both focused attention and divided attention questions. While participants did answer focused attention questions more accurately ($F_{(1, 18)} = 183.9$, $p < 0.001$) and more quickly ($F_{(1, 18)} = 53.4$, $p < 0.001$) than divided attention questions, there was no significant difference in accuracy or response time ($p > 0.10$; $\phi = 0.13$, $\phi = 0.49$) as a function of display size, though power was low for the accuracy measure. Question type and display size did not interact for either performance measure ($p > 0.10$).

Recall that questions requiring estimates of speed and distance were also included as additional measures of the impact of display size. Analyses of these subsets of questions were conducted and showed that display size had no significant effect ($p > 0.10$) on question accuracy

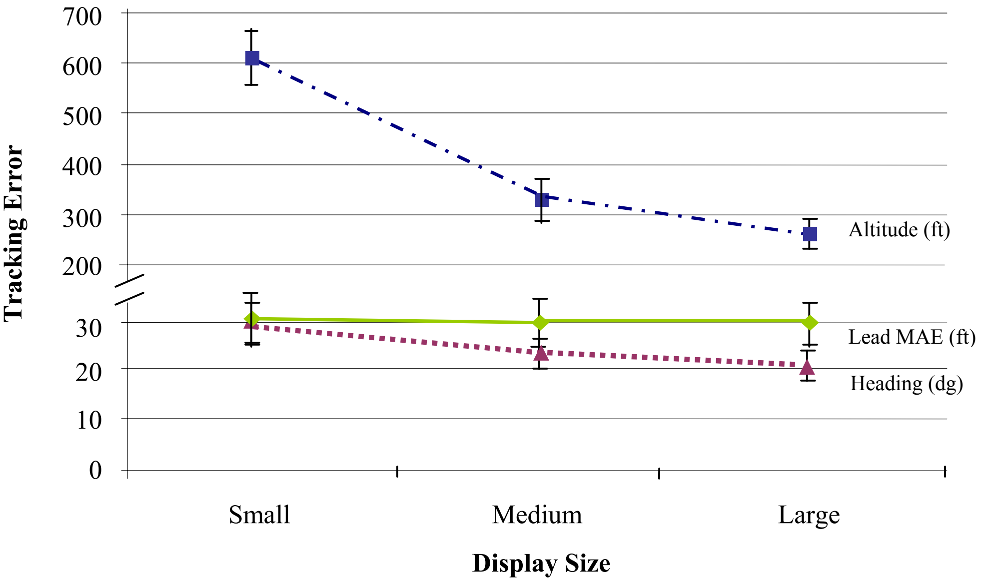

Figure 7. Altitude and heading RMS error and lead MAE. Note a significant decrease in altitude and heading error when display size is increased from small to medium.

or response time for either distance or speed questions, though again, power was low ($\phi < 0.25$ for all tests).

Surveillance. Surveillance was assessed through change detection performance measures. On average, participants detected 12.2% of changes at an average latency of 18.0 s. Analyses revealed that change detection accuracy and response time were unaffected by manipulations to display sizes ($p > 0.10$; $\phi = 0.48$, $\phi = 0.26$, respectively). The impact of change eccentricity on detection performance was also assessed using two different approaches. First, both accuracy and response time were evaluated as a function of eccentricity without accounting for the differences in display size. This approach yielded a significant negative correlation between eccentricity and detection accuracy ($r = -.49$, $p < .01$), suggesting that accuracy is significantly reduced with the increasing distance of the changing aircraft or weather system from ownship, independent of the relevance of the event to flight safety. Given that display size had no effect on surveillance performance, the analyses suggest that performance was degraded as changes occurred further from the center of the display, but display enlargements, which served to further increase the distance between the center of the display and the display perimeter, did not amplify this effect, suggesting that participants strategically compensated scanning patterns to account for the increased display area.

Eye Movement Data. Given that subjects maintain equivalent surveillance performance across increasing display size, we now ask how they compensate via changing eye movements; in particular, we ask if they maintain the same proportionality of scanning (surveillance) behavior across different regions of the display (ownship, midrange, and outer areas, as shown in Figure 8), independent of its size. Though head movement data was also collected, participants only rarely used head movements to access information, thus this data will not be discussed. Figure 9 presents the percentage dwell time (PDT) within each of the three concentric regions of the display, as a function of display size. The measure PDT is an effective measure of the

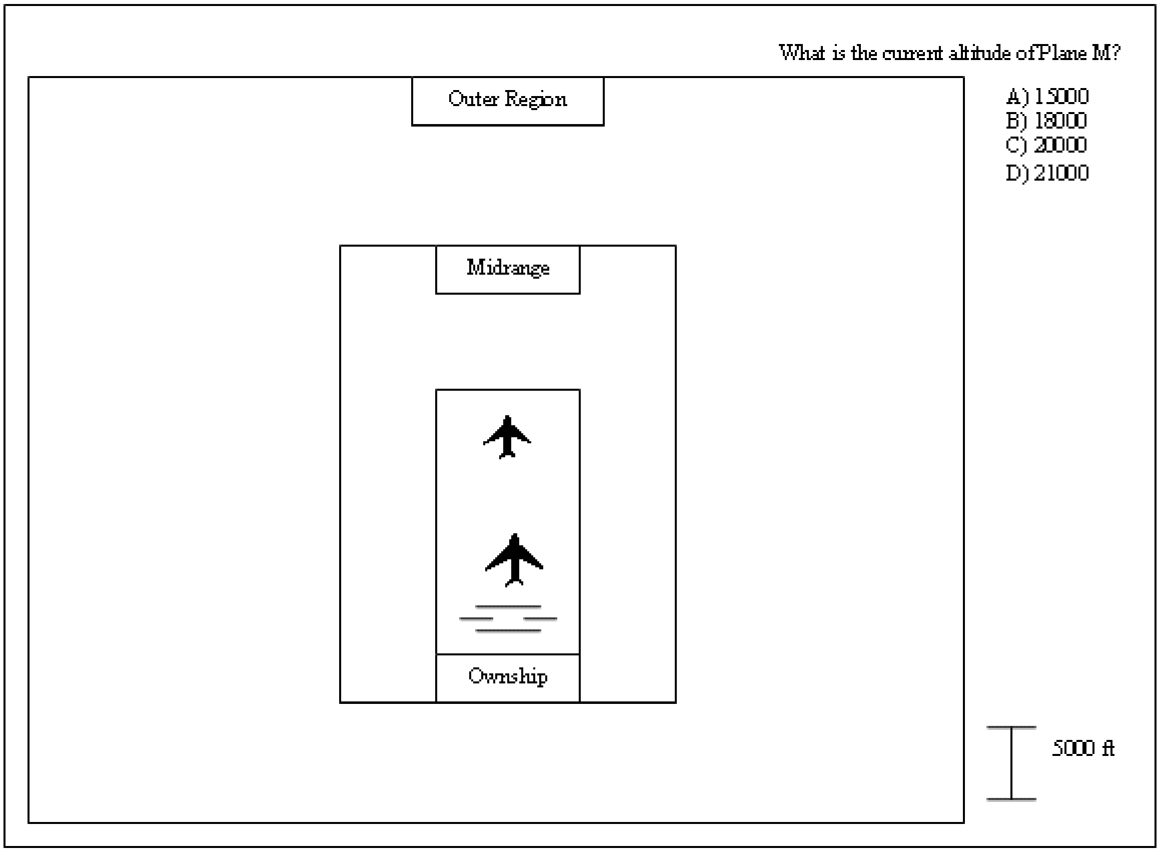

Figure 8. Ownship, midrange, and outer display regions that define the display areas of interest.

allocation of attention (Wickens, Goh, Helleberg, Horrey, and Talleur, 2003). The figure indeed indicates a significant shift in the allocation of attention from the small to the medium displays, with attention being drawn from the midrange area and allocated to the ownship and outer display regions ($F_{(4,\ 40)} = 4.12$, $p < 0.007$). There was a relatively constant proportionality of attention allocation across the three regions between the medium and large display, suggesting that scanning behavior strategically compensated for the increase in display size, by scanning across a greater area. The figure also suggests that more attention is located to the central ownship area than to the two peripheral areas ($F_{(2,\ 20)} = 56.13$, $p < 0.001$), reflecting the greater information content of that area, containing ownship information that was necessary for tracking in all three axes.

Surprisingly, there was more attention allocated to the outer than to the middle region. However, this appears to be a partial artifact of the fact that the outer region simply contains much more space to monitor than does the middle region (see Figure 8). Indeed, when scanning behavior was normalized by a measure of "percent/cm^2" (i.e., percentage of time per area to be monitored), this measure declined monotonically and significantly from the inner to the middle to the outer ring ($F_{(2,\ 20)} = 273.95$, $p < 0.001$). Thus, while the outer region receives more total attention, the attention allocation is more sparsely distributed across space. This monotonic decline well accounts for the decrease in detection performance with greater eccentricity. PDT and change detection accuracy of events in the outer display region were also found to be highly

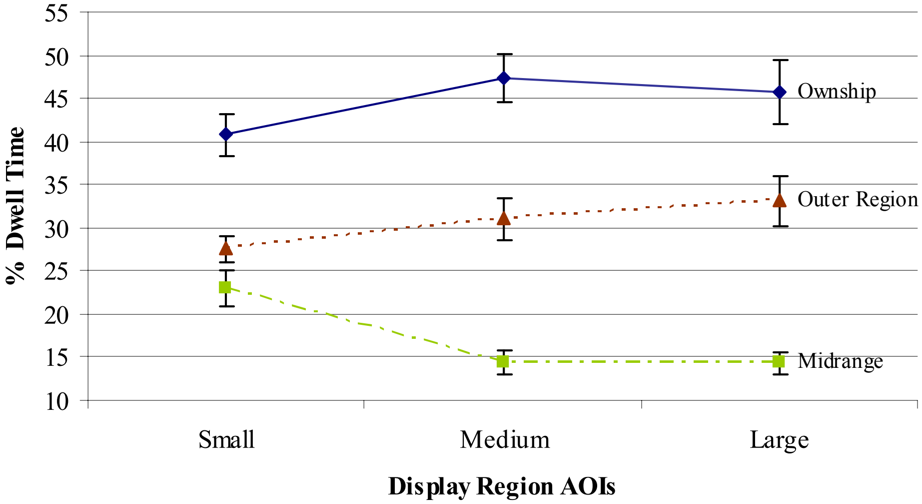

Figure 9. Constant proportionality of attention allocation across the display areas of interest as a function of display size.

correlated ($r = 0.57$, $p < 0.001$), thus lending support to the relationship between attention allocation and detection performance. The correlation between PDT and detection accuracy for the middle display region, however, was not significant ($p < 0.10$). The significant region X display size interaction in the data of Figure 9 appears to be particularly attributable to the relative increase in attention to the middle region in the smallest display, an issue we address below.

An increase in the PDT measure reflecting attention does not distinguish the extent to which this increase is the result of more visits (scans) to a particular region, or a longer dwell per visit, as the latter two combine multiplicatively to produce PDT. To establish which factors underlay the PDT differences, we also analyzed the mean dwell duration (MDD), which data are presented in Figure 10. The figure reveals a pattern of data that is somewhat similar to that in Figure 9, with a few noteworthy exceptions. First, the main effect of region ($F_{(2, 20)} = 132.80$, $p < 0.001$) appears to be totally due to longer dwells in the central ownship area, as opposed to differences between the middle and outer regions. Thus, the increase in attention to the outer region in Figure 9 was due to a greater number of visits, not to a greater dwell time per visit. Since longer scans require greater effort, and may involve head movements, this finding clearly indicates the role of some strategic effort compensation to deal with the larger display region. The fact that dwells were nearly four times as long in the central ownship display region reflects the much greater information content of that region, necessary to extract error data for the three axes of tracking (Bellenkes, Wickens, and Kramer, 1997).

Second, although there was a significant effect of display size on MDD, with longer dwells for the smaller display ($F_{(2, 20)} = 5.10$, $p = 0.02$), this effect was relatively small in its magnitude, compared to the display region effect, and appears to be primarily due to the specific increase in dwell duration on the middle region with the small display ($t_{(11)} = 4.99$, $p < 0.001$), which mirrors the interaction shown in the PDT data of Figure 9. Indeed it appears to be the case that a large amount of the increase in PDT to the middle region in the small display can be

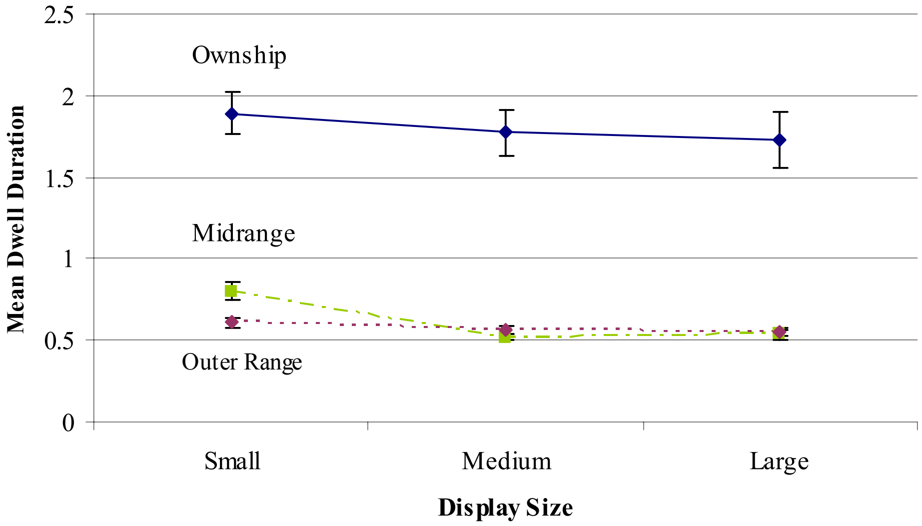

Figure 10. Significant interaction of display AOI and display size. Participants had significantly longer dwells on ownship than in other display regions. Additionally, longer dwells occurred with the smallest display size.

directly associated with the longer dwells there, rather than more visits. The specific interpretation of this effect will be discussed below.

Discussion

The results of Experiment 2 provide partial evidence for the effects of display size, and bear on the extent to which effort conservation versus strategic compensation took place as size was varied. Replicating the effects of size on tracking behavior observed in Experiment 1, we found that smaller size reduced tracking performance (increased error) in two out of the three tracking measures (Figure 7). It remains unclear the extent to which this was a consequence of resolution rather than urgency. Some evidence for the resolution hypothesis is provided by the fact that the increase in error was only seen in that tracking task (altitude) whose basis was assessing spatial differences of the pitch ladder of the ADI, and that the increase in error was only seen from the middle to the smallest display. Had urgency been the mediating factor, we would have anticipated error increases (with decreasing size) proportional to the size decrease across both the spatially defined headway tracking as well as the altitude tracking task, and across both display size intervals. In any case, we can postulate that participants did invest more effort when tracking altitude in the medium and large displays than in the smaller display, either because they better perceived smaller error deviations with the larger display (resolution) and therefore corrected them, or because they perceived all deviations as larger in world units (urgency), thereby improving their altitude tracking performance considerably.

Display size had almost no influence on search, on the spatial judgments that were derived from that search, nor on overall surveillance. Thus for those aspects of the task, it appears that pilots effectively compensated for the larger areas, by engaging in longer scans. They even increased PDT to the outer region with the larger displays (Figure 9). In essence, analogous to the "zoom lens" model of attention (Ericksen and St. James, 1986), as the display size increases, pilots broaden their attention spotlight to compensate for the wider area of monitoring responsibility. They do so apparently without performance cost, replicating effects from other visual search studies (e.g., Teichner and Morcharnuk, 1979). How this is accomplished is by maintaining roughly the same ratio of middle to outer display ring scans, a ratio that favors the middle region in terms of the effectiveness of the scan and accounts for the degraded detection with greater eccentricities. Interestingly, this wider scan pattern does not disrupt, and even slightly enhances performance of the tracking tasks, whose sources are in the display center.

In addition to compensating for a larger display, the scan data provided additional evidence for a form of strategic compensation, associated with the elevated values of PDT and MDD of the middle region on the smallest display. In interpreting this effect, it is noteworthy that, with the smallest display, pilots can fixate on the middle region and maintain the center of the inner region (the axis between ownship and the lead aircraft) within one to two degrees of visual angle, a distance that is clearly within the useful field of view, if not also within the foveal region. Thus, pilots might have chosen a strategy with this smallest display to fixate more often in the middle region, knowing that by doing so, they did not need to temporarily abandon the tracking tasks. However, such a strategy would be particularly harmful to that tracking task of highest bandwidth and greatest need for spatial resolution, the ADI-based altitude tracking task, as was shown to be the case in Figure 7.

It is important to note that the effects of display size in Experiment 2 were not linear across the three sizes. In almost all aspects of performance, the medium and large displays supported equivalent performance, whereas the small display suffered in supporting tracking, either because of the low resolution or because this display fostered scanning strategies that were sub-optimal for tracking. Finally it is noteworthy that very few head movements were made, even in the largest display, which spanned 36 degrees laterally. This suggests that estimates of the size of the eye field offered by Bahill and colleagues (1975) at 20 degrees may be slightly underestimated.

GENERAL DISCUSSION

Across two experiments we manipulated display size, examining its effects on three piloting tasks, as mediated by two processing mechanisms: strategic compensation and effort conservation. These mechanisms were manifest differently in the different tasks. For tracking, in both experiments, we found that larger displays yielded better tracking performance by encouraging the strategic compensation of investing more effort. In Experiment 1, we partially relate this effect to the greater urgency of larger depicted error, whereas in Experiment 2, it may be more related to the greater resolution with which smaller pitch errors are depicted on the larger displays, errors which are therefore left uncorrected on the smaller display.

For surveillance, the strategic compensation was accomplished through broader scanning that entirely offset any costs associated with the wider area to be monitored. Thus for surveillance, we can definitely reject an "effort conservation" view, which would have predicted relatively worse peripheral performance with the larger display. For tracking, in Experiment 1, effort conservation has a slightly different meaning, predicting the same control activity (control effort), and therefore tracking error independent of display size, as pilots track true error (ecological) and not displayed error. The fact that this equivalence was *not* observed, as was furthermore modulated by the form of display size changes in Experiment 1, indicates that pure effort conservation was not in effect, although the extent to which it was modulated by urgency versus resolution cannot be fully ascertained from the current data.

In conclusion, the present results suggest that display size can have important effects on performance of spatial tasks, above and beyond their effects on reading and discriminating fine detail, issues not directly examined here. As contradictory pressures for both miniaturization and for large screen displays are provided by designers, these issues must be better addressed by human factors practitioners.

ACKNOWLEDGMENTS

The authors wish to thank Sharon Yeakel for her programming assistance and Roger Marsh for assisting with the installation of the eyetracker and filtering the eye data.

REFERENCES

Abbott, T. S., & Moen, G. C. (1981). *Utilization of traffic situation display for self-spacing task* (Technical Paper 1885, Technical Report 81-B-4). Hampton, VA: NASA Langley Research Center, AVRADCOM Research and Technology Laboratories.

Alexander, A. L., Wickens, C. D., & Hardy, T. J. (2003). Examining the effects of guidance symbology, display size, and field of view on flight performance and situation awareness. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

Bahill, A. T., Alder, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology, 14,* 468-469.

Beck, D. M., Rees, G., Frith, C. D., & Lavie, N. (2001). Neural correlates of change detection and change blindness. *Nature Neuroscience, 4,* 645-650.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine, 68,* 569-579.

Carpenter, S. (2001). Sights unseen. *APA Monitor, 32,* 54-57.

Enoch, J. M. (1959). Effects of the size of a complex display upon visual search. *Journal of the Optical Society of America, 49,* 280-286.

Ericksen, B. A., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40,* 225-240.

Fracker, M. L., & Wickens, C. D. (1989). Resources, confusions, and compatibility in dual axis tracking: Displays, controls, and dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 80-96.

Haskell, I. D., & Wickens, C. D. (1993). Two-and-three-dimensional displays for aviation: A theoretical and empirical comparison. *International Journal of Aviation Psychology, 3,* 87-109.

Kahneman, D. (1973). *Attention and effort.* Prentice Hall: Englewood Cliffs, NJ.

Kahneman, D., Peavler, W. S., & Onuska, L. (1968). Effects of verbalization and incentive on the pupil response to mental activity. *Canadian Journal of Psychology, 22,* 186-196.

Kantowitz, B. H. (1987). Mental workload. In P. A. Hancock (Ed.), *Human Factors Psychology.* Amsterdam: North Holland.

Kroft, P., & Wickens, C. D. (2003). Displaying multi-domain graphical database information: An evaluation of scanning, clutter, display size, and user interactivity. *Information Design Journal, 11,* 44-52.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4,* 501-506.

Martin-Emerson, R., & Wickens, C. D. (1997). Superimposition, symbology, visual attention, and the head-up display. *Human Factors, 39,* 581-601.

McConkie, G. W., & Loschky, L. (2001). Attending to objects in a complex display. *Advanced Displays and Interactive Displays Consortium, Proceedings of the ARL Federated Laboratory 4th Annual Symposium* (pp. 21-25).

McGreevy, M. W. & Ellis, S. R. (1986). The effect of perspective geometry on judged direction in spatial information instruments. *Human Factors, 28,* 439-456.

McRuer, D. T., & Jex, H. R. (1967). A review of quasi-linear pilot models. *IEEE Transactions on Human Factors in Electronics, 8,* 231.

Mumaw, R. J., Sarter, N. D., & Wickens, C. D. (2001). Analysis of pilots' monitoring and performance on an automated flight deck. *Proceedings of the 11th International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University.

Muthard, E. K., & Wickens, C. D. (2002). Change detection after preliminary flight decisions: Linking planning errors to biases in plan monitoring. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

Onstott, E. D. (1976). Task interference in multi-axis aircraft stabilization. *Proceedings of the Twelfth Annual Conference on Manual Control.* Champaign, IL: University of Illinois at Urbana-Champaign.

Perrone, J. A., & Wenderoth, P. (1993). Visual slant underestimation. In S. R. Ellis, M. Kaiser, & A. J. Grunwald (Eds.), *Pictorial Communication in Virtual and Real Environments.* London: Taylor & Francis.

Pringle, H. L., Irwin, D. E., Kramer, A. F., & Atchley, P. (2001). The role of attentional breadth in perceptual change detection. *Psychonomic Bulletin & Review, 8,* 89-95.

Prinzel, L. J., Comstock, R. C., Glaab, L. J., Kramer, L. J., & Arthur, J. J. (2004). The efficacy of head-down and head-up synthetic vision display concepts for retro- and forward-fit of commercial aircraft. *The International Journal of Aviation Psychology, 14,* 53-77.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53,* 245-277.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368-373.

Schnell, T., Kwon, Y., Merchant, S., & Etherington, T. (in press). Improved flight technical performance in flight decks equipped with synthetic vision information system displays. *International Journal of Aviation Psychology.*

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7,* 1-15.

Simmons, T. (2001). What is the optimum computer display size? *Ergonomics in Design, fall,* 19-25.

Teichner, W. H., & Mocharnuk, J. B. (1979). Visual search for complex targets. *Human Factors, 21,* 259-275.

Wickens, C. D. (1986). The effects of control dynamics on performance. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and performance* (vol. II, pp. 39-1-39-60). New York: Wiley.

Wickens, C. D. (1992). *Computational Models of Human Performance* (Technical Report No. ARL-92-4/NASA-A3I-92-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Wickens, C. D. (2002). *Spatial Awareness Biases* (Technical Report No. AHFD-02-6/NASA-02-4). Savoy, IL: University of Illinois, Aviation Human Factors Division.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37,* 473-494.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45,* 360-380.

Wickens, C. D., Liang, C. C., Prevett, T. T., & Olmos, O. (1996). Egocentric and exocentric displays for terminal area navigation. *International Journal of Aviation Psychology, 6,* 241-271.

Wickens, C. D., Thomas, L. C., & Young, R. (2000). Frames of reference for display of battlefield information: Judgment-display dependencies. *Human Factors, 42,* 660-675.

Yeh, M., & Wickens, C. D. (2001). Attentional filtering in the design of electronic map displays: A comparison of color coding, intensity coding, and decluttering techniques. *Human Factors, 43,* 543-562.