

Krebs, W.K. and Sinai, M.J. (2002). Psychophysical assessments of image-sensor fused imagery, *Human Factors*, 44, 257-271.

Psychophysical assessments of image-sensor fused imagery

William K. Krebs¹, and Michael J. Sinai

Naval Postgraduate School, Monterey, CA

ABSTRACT

The goal of this study was to determine the perceptual advantages of multi-band sensor-fused (achromatic and chromatic) imagery over conventional single-band nighttime (image-intensified and infrared) imagery for a wide range of visual tasks, including detection, orientation, and scene recognition. One hundred and fifty-one active-duty military observers' reaction time and accuracy scores were recorded during a visual search task. Data indicate that sensor fusion did not improve performance relative to that obtained with single-band imagery on a target detection task, but did facilitate object recognition, judgments of spatial orientation, and scene recognition. Observers' recognition and orientation judgments were improved by the emergent information within the image-fused imagery, i.e., combining dominant information from two or more sensors into a single displayed image. Actual or potential applications of this research include the deployment of image-sensor fused systems for automobile, aviation, and maritime displays to increase operators' visual processing during low-light-level conditions.

Keywords: Image-sensor fusion, fused imagery, color fusion, displays

¹To whom all correspondence should be sent: current address is Federal Aviation Administration, 800 Independence Avenue, S.W., AAR-100 (Room 907), Washington D.C. 20591. william.krebs@faa.gov

INTRODUCTION

Sensor fusion combines images from multiple sensors into a single display, with the aim of enhancing operators' target detection and situational awareness in high-workload environments. Military and civilian applications include enhancing commercial airline pilots' ability to land during low visibility (Nordwall, 1997); increasing observers' ability to detect targets in all weather and terrain environments using multispectral sensors (McDaniel, Scribner, Krebs, Warren, Ockman, and McCarley, 1998); enhancing helicopter pilots' night tactical terrain flight performance (Ryan and Tinkler, 1995); and improving airborne detection of vehicles (Bergman, 1996). The results of human performance tests of sensor-fused imagery, however, have been equivocal. While some studies have found performance improved with fused imagery (e.g., McCarley and Krebs, 2000; Toet, Ijspeert, Waxman, and Aguilar, 1997; Waxman, Gove, Seibert, Fay, Carrick, Racamoto, Savoye, Burke, Reich, McGonagle, and Craig, 1996; Essock, Sinai, McCarley, Krebs, and DeFord, 1999), others have not (Steele and Perconti, 1997; Krebs, Scribner, Miller, Ogawa, and Schuler, 1998). Although the discrepancies between the studies may be attributed to methodological inconsistencies, the potential benefit of image-sensor fusion is not overwhelmingly apparent.

Currently, two main varieties of night-vision imaging sensors are in widespread use: image-intensifiers (\hat{r}^2), such as the military's Night Vision Goggle (NVG), which amplify available light and near infrared in a nighttime scene; and long-wave infrared (IR) sensors, also in common military use, which convert invisible thermal energy into a visual display. Objects viewed by \hat{r}^2 and IR sensors will generally have the same spatial characteristics, but will appear to have dramatically different contrast levels (Krebs et al.,

1998). Consequently, each may offer certain advantages and disadvantages that can be exacerbated or minimized according to environmental conditions. For example, the resolution of the first-generation infrared sensor—the most common infrared sensor used in the military—is generally poorer than that of $\overset{2}{f}$ sensors, with the result that background details of a visual scene are generally more visible in $\overset{2}{f}$ than in thermal images (Steele and Perconti, 1997). However, the thermal contrast between heat-emitting objects and their cooler surroundings is typically much greater than the luminance contrast, allowing such objects to be seen more clearly in a thermal image than a visible image (O’Kane, Crenshaw, D’Agostino, and Tomkinson, 1992). Likewise, atmospheric conditions can affect these two sensors differently. For example, clouds that obscure moonlight and starlight will decrease the strength of the signal reaching an $\overset{2}{f}$ sensor and, in turn, reduce contrast within the $\overset{2}{f}$ image, but will leave the thermal contrast between objects unaffected. Conversely, changes in ambient temperature may alter the distribution of thermal contrast across a scene, while producing no concurrent change in illumination. Thus, the quality and information content of the imagery produced by IR and $\overset{2}{f}$ sensors are constrained in materially different ways.

By combining the output of two or more sensors within a composite image, sensor fusion offers a potential method of overcoming the limitations inherent to single-band imagery (e.g., Toet and Walraven, 1996; Therrien, Scrofani and Krebs, 1997; Scribner, Warren, Schuler, Satyshur, and Kruer, 1998; Waxman, Gove, Fay, Racamato, Carrick, Seibert, and Savoye, 1997; Aguilar, Fay, Ross, Waxman, Ireland, and Racamato, 1998; Das and Krebs, 2000). Such processing could enhance the quality of electronically-sensed images in at least two ways. First, fusion could simply combine information

conveyed by multiple input sources, allowing users to view multiple “images” within a single display and, perhaps, obviating the need to alternate one’s gaze between displays either electronically or via eye movements. More intriguingly, fusion could augment the information conveyed by single sensors with emergent information not available in any of the input images singly, but derived from the contrast between input images. Such contrast might provide information with which to embellish the spatial content of a fused scene (Therrien, Scrofani and Krebs, 1997; Peli, Peli, Ellis, and Stahl, 1999); or it might provide the basis for chromatic rendering of fused images, much as differences in the spectral sensitivity curves of the retinal receptors allow for biological color vision (Bowmaker and Dartnall, 1980; Baylor, D.A., Nunn, B.J., and Schnapf, 1987; Schnapf, Kraft, and Baylor, 1987). The most common types of single-band imagery typically fused are from $\bar{\lambda}$ sensors and long-wave IR sensors; however, mid-wave IR sensors and short-wave IR sensors have also been used (e.g., Toet, Ijspeert, Waxman, and Aguilar, 1997; Krebs, McCarley, Kozek, Miller, Sinai, and Werblin, 1999).

Unfortunately, the results of experiments assessing the value of sensor fusion to human perceptual performance have been equivocal. Some of the apparent inconsistencies in these results are likely due to differences in the environmental conditions under which input stimuli were collected and to the use of different algorithms for creating fused imagery (McCarley and Krebs, 2000). The discrepancy between the results of these studies also stems from the wide range of experimental tasks used to assess human performance with fused and non-fused imagery. While many studies have employed target detection tasks, the testing procedures have varied dramatically: in some tests, observers had to detect targets in small, briefly flashed patches of the original

scenes flashed (Essock, Sinai, McCarley, Krebs, and DeFord, 1999); other tests required observers to change the contrast of a small square embedded in the scene (Ahumada and Krebs, 2000; Waxman et al., 1996); and still others used video clips of the processed imagery (Steele and Perconti, 1997; Krebs, Scribner, Miller, Ogawa, and Schuler, 1998). It is not surprising that these very different methodologies have produced different results.

Past research, then, has been inconclusive in determining under what conditions and for what perceptual tasks sensor fusion facilitates human performance. What is clear, though, is that image-fusion does not always improve performance relative to that obtained with single-band imagery. We ascertain that improvements in performance may depend, in part, on the particular perceptual task to be performed. This task-dependent improvements may be related to the fact that performance with the single-band imagery itself is task-dependent. Steele and Perconti (1997), for example, found that performance for a target detection/recognition task was faster and more accurate with IR imagery than with i^2 imagery, but, conversely, that performance on a horizon-perception task was better with i^2 imagery than with IR imagery. Thus, due to the inherent differences in the information they convey, the quality of various forms of single-band imagery appears to be task-dependent, such that some tasks are performed better with one type of imagery and other tasks with another.

The goals of the present study were to determine, first, whether sensor fusion is, in general, beneficial across a range of tasks, and second, whether the benefits of sensor fusion vary substantially with the nature of the psychophysical task performed. We addressed these issues by comparing human performance across three very different

perceptual tasks using the same forms of single-band and sensor-fused imagery. Experiment 1 assessed performance on an object detection task. Experiment 2 employed a spatial-orientation task in which observers had to discern whether the image presented was upright or inverted. Finally, Experiment 3 required observers to view a briefly presented scene and then decide whether a second scene was the same or different, disregarding the image format type. Thus, the format type could change between the first and second scene, but the objective was to determine whether the scene itself had changed.

We chose these three tasks for two reasons. First, we wanted tasks that have, in the past, resulted in performance that was different for the two types of single-band imagery. For example, Steele and Perconti (1997) found that the IR imagery was superior to \hat{I} imagery for a target-detection task, but that \hat{I} imagery was superior to IR imagery for a type of spatial-orientation task. It was predicted that the results for the single-band imagery in Experiments 1 and 3 would follow this trend, and we hypothesized that the sensor-fused imagery would result in performance at least as good as the better of the two single-band sensors. This hypothesis would test whether scene information is maintained in the fused imagery (i.e., scene detail is not degraded through the fusion process), or whether fused imagery can enhance scene information, which would result in improved performance.

The second rationale for choosing these various tasks is they are likely to demand different types of scene information. This may be important because sensor fusion may improve imaging of some aspects of a scene more than others. Sensor fusion may be more useful, for example, for revealing prominent objects within a scene than for

enhancing a scene's background, or vice versa. If sensor fusion enhances image quality in such specific ways, this may be reflected in specific types of improvements in performance. Thus, we chose a wide variety of tasks that require observers to attend to different aspects or components of the scene in order to complete the task successfully (e.g., detecting an object as opposed to recognizing a scene). In this way, we could test whether sensor fusion is more or less beneficial for specific tasks that require specific types of scene information.

General Methods

Participants. A total of 151 active-duty military personnel participated in three experiments. All were naïve as to the purpose of the experiment, and all participants reported normal or corrected-to-normal visual acuity and normal color vision was tested with pseudo-isochromatic plates. All subjects signed an informed consent and were briefed on the ethical conduct for subject participation specified in the Protection of Human Subjects (Department of the Navy, 1984).

Apparatus. Stimuli were displayed by a VisionWorks computer graphics system (Vision Research Graphics, Inc., Durham, New Hampshire; Swift, Panish and Hippensteele, 1997) on a Nanao Flexscan F2.21 monitor. The monitor had a resolution of 800 x 600 pixels, a frame rate of 98.9 Hz, and a maximum luminance of 100 cd/m² with luminance linearized by means of a look-up table. Observers viewed the screen from 1.5 meters.

Stimuli. Stimuli were images collected at Fort Ambrose Powell Hill (commonly referred to as Fort A.P. Hill), Virginia, using an uncooled Lockheed Martin Infrared Systems LTC-500 long-wave infrared camera (327 x 245 pixels, spectral sensitivity range from 7.5 to 14 micrometers with a horizontal 33.7⁰ and vertical 26.5⁰ field of view) and a

Lockheed Martin Fairchild Fusion Low-Level-Light television i^2 camera (768 x 492 pixels, spectral sensitivity range from 0.61 – 0.95 micrometers with a peak sensitivity of 173 mA/W at 740 to 760 microns that was fiber-optically coupled to a Sony 768 by 492 pixel Charge Coupled Device (CCD) with a 40^0 horizontal and 30^0 vertical field-of-view). During post-processing, images were first spatially registered by means of an affine transformation in which the images were multiplied by a matrix that included elements for rotation, translation, and magnification (Scribner, Schuler, Warren, Satyshur, and Kruer, 1998). The images consisted of nighttime scenes from around the installation, including wooded areas, fields, roads, and buildings. Fusion of paired visible and thermal images was performed through the Principal Components fusion algorithm developed by Scribner, Warren, Schuler, Satyshur, and Kruer, (1998). The algorithm is described briefly here and explained in detail elsewhere (Scribner, Warren, Schuler, Satyshur, and Kruer, 1998). The composite stimuli approach is to assign each pixel a color vector defined by the detected power in the registered two-band imagery. A scatter plot of the image ensemble of colors frequently reveals pronounced anti-correlation between short and long wavelengths, consistent with Kirchoff's law (reflective objects that appear bright in the short-wave infrared typically have low emissivity and appear dark in the long-wave infrared). The algorithm maps the pair of grayscale visible and a thermal image onto a two-dimensional color space with the principal component corresponds to the major axis—luminance channel—and the orthogonal axis corresponding to the minor axis—color channel. The pixel values from the thermal component image are assigned to the red phosphor of a fused image, and the visible component image is assigned to the green and blue phosphors of a fused image. The

result is a composite image wherein colors range from saturated red (created by illumination of only the red phosphor) through gray to saturated cyan (created by illumination of only the green and blue phosphors). Orthogonally, colors range from dim to bright: pixels that are bright only in the thermal input appear red in the fused image; pixels that are bright only in the visible input appear cyan; and pixels whose values are approximately the same in both input images appear achromatic. Pixels whose mean thermal/visible input values are low appear dim, while those whose mean thermal/visible input values are large appear bright. That is, the brightness of a fused pixel is determined by the weighted mean value of corresponding visible and thermal pixels, and the hue is determined by the difference in value between corresponding visible and thermal pixels.

A total of six image formats were tested: single-band IR and i^2 formats, two chromatic fused formats, and two achromatic fused formats. One color-fused version of each scene was derived using IR imagery of white-hot polarity (*wh*), and one was derived using IR imagery of black-hot polarity (*bh*). Grayscale versions of these fused images (*whg* and *bhg*, respectively) were spatially identical to their chromatic counterparts, but were rendered achromatically. Single-band IR images were of white-hot polarity. A total of eighty scenes were used. Figure 1 shows examples of the imagery used. Here, an i^2 image is shown at the top, an IR image in the middle, and the achromatic white-hot image created by fusion of the top and middle images at the bottom. This scene contains two vehicles and a person among the trees. All images had dimensions of 625 x 400 pixels. The surrounding screen was maintained at 50 cd/m² throughout the experiment.

Insert Figure 1 about here

EXPERIMENT 1: TARGET DETECTION

Target detection is one of the most common tasks used by researchers to investigate the potential benefits of sensor fusion. Past research using target detection, however, has frequently found that image-sensor fused imagery does not significantly improve performance over single-band imagery (Sampson, Krebs, Scribner, and Essock, 1996; Steele and Perconti, 1997; Krebs et al., 1998). Exceptions to this trend have come from several reports: Waxman et al. (1996) found that color fusion could increase the visibility of a small contrast-modulated patch within a larger image; Toet, Ijspeert, Waxman, and Aguilar (1997) found that color-fused imagery improved subjects' ability to locate people within a scene; and Essock et al. (1999) found that fusion reliably improved observers' detection of designated target objects (e.g. 'people' and 'buildings') within small (1.4°), briefly flashed patches of a scene. The goal of Experiment 1 was to determine whether these apparent inconsistencies in past results might have been due, in part, to the different target items used by various researchers. The results of McCarley and Krebs (2000) indicated that, while fusion could substantially aid target detection, the utility of fusion varied with the environmental luminance of scenes depicted in the input imagery. It is possible that the benefits of fusion for target detection also vary with target characteristics (e.g., emissivity). To test this possibility, Experiment 1 measured target-detection performance for two classes of target item in various formats of single-band

and sensor-fused imagery. It was hypothesized that observers' target detection would be significantly improved by chromatic contrast produced by color fusion.

Method

Participants. A total of 84 active-duty military personnel volunteered in this experiment. There were 68 males and 16 females with a mean age of 25.57 and a standard deviation of 6.83. The 84 subjects comprised of 40 officers and 44 enlisted personnel from the United States Army (n=50), United States Navy (n=18), United States Marines (n=9), and foreign militaries (n=7). All observers had normal color vision, as tested with pseudo-isochromatic plates, and were naïve to the experimental hypothesis. All subjects signed an informed consent and were briefed on the ethical conduct for subject participation specified in the Protection of Human Subjects (Department of the Navy, 1984).

Procedure. The experimental task required observers to view a single image in each trial and to make a speeded response to indicate the presence or absence of a designated target therein. Observers were randomly assigned one of the six image formats (i^2 , *ir*, *wh*, *whg*, *bh*, *bhg*) for testing. Each then completed two blocks of thirty images each, one block with each of two designated target types (personnel and vehicle). The order in which blocks of different target types were presented was determined randomly for each subject, and subjects were informed of the target type before each block. Each block contained fifteen images with a target present and fifteen with the target absent. The first six trials of each block were considered practice, and data from these trials were discarded from analysis. The order in which images were presented within each block was randomly determined for each subject, with the constraint that the practice trials of

each block include three target-present and three target-absent images. No images were viewed more than once.

Subjects were instructed to press “1” on the experimental computer’s numeric keypad to indicate target presence and “2” to indicate target absence. They were instructed to make responses as quickly and accurately as possible. An alerting tone (100-millisecond) preceded presentation of each image, and the image remained visible until the subject responded. Reaction time (RT) and accuracy were recorded for each response. No feedback was given during the experiment. A two-second interval occurred between response for one trial and presentation of the subsequent stimulus. Between blocks, observers were allowed a brief break. Duration of the full experiment was approximately 15 minutes.

Results and Discussion

Two dependent variables were extracted from raw data for each combination of image format and target type, mean target-present RT and mean sensitivity (d'). In this and all experiments, only RTs for correct responses were considered. Additionally, because preliminary analysis indicated that effects of image format were not moderated by target presence/absence, and because target-absent RTs are known to be more susceptible to changes in response criteria than are target-present RTs (Chun & Wolfe, 1996), only target-present RTs were included in the analysis of the experiment reported here. Mean target-present RTs are displayed in Figure 2. Mean d' primes are displayed in Figure 3.

Insert Figure 2 about here

Insert Figure 3 about here

For statistical analysis, RTs were first inverted to reduce skew. Transformed RTs and sensitivity data were then submitted to separate 6 x 2 mixed ANOVAs, with image format as between-subjects factor and target type as within-subjects factor. Results of these analyses confirmed that performance as assessed by both RTs and d prime varied with type of target being sought, but was generally as good or better with single-band imagery as with sensor-fused multi-band images. Analysis of transformed RTs produced a reliable main effect of target type, $F(1, 78) = 101.71, p < .001$, indicating that responses were generally faster for personnel targets than for vehicles. Omnibus analysis failed to produce a reliable main effect of image format, $F(5, 78) = 1.216, p = .310$, but did reveal a reliable interaction of image format by target type, $F(5, 78) = 2.89, p = .019$. One-way ANOVAs with image format as a between-subjects factor were, therefore, conducted separately for personnel and for vehicle targets. This analysis of simple effects failed to reveal a reliable influence of image format on detection times for personnel targets, $F < 1$, but did indicate a reliable effect of image format for response times to vehicle targets, $F(5, 78) = 3.05, p = .014$. Post-hoc Tukey's HSD tests indicated that this effect was the result of reliable differences which obtained between RTs to IR imagery and RTs to chromatic black-hot, achromatic black-hot gray, and chromatic white-hot imagery, p 's <

.05. Detection times for targets rendered in any of these three sensor-fused formats, that is, were reliably slower than detection times for targets rendered in single-band IR imagery. Sensor fusion apparently degraded the single-band information that allowed rapid detection of targets in IR images.

Analysis of d' prime for targets embedded in imagery of various formats also produced no evidence for a benefit of sensor fusion. Omnibus ANOVA again revealed a reliable main effect of target type, $F(1, 78) = 68.32$, $p < .001$, indicating superior performance with personnel targets. Analysis also revealed a main effect of image format, $F(5, 78) = 2.4$, $p = .045$, indicating that image format did affect overall sensitivity. Given a reliable interaction, $F(5, 78) = 2.49$, $p = .038$, however, indicating that the effects of image format were modulated by target type, one-way ANOVAs with image format as between-subjects variable were again conducted separately for personnel and vehicle targets. This simple analysis again produced no reliable effect of image format on detection of personnel, $F(5, 78) = 1.83$, $p = .117$, but did reveal a reliable effect of format on sensitivity to vehicle targets, $F(5, 78) = 2.93$, $p = .018$, attributable here to the reliably lower sensitivity which obtained with sensor-fused achromatic white-hot imagery relative to that which obtained with IR and with sensor-fused chromatic white-hot imagery, Tukey's HSD, p 's $< .05$.

In summary, results suggest that neither of the sensor fusion methods employed here improved target-detection performance over conventional, single-band night-imaging sensors. When seeking personnel targets, observers were generally as fast and as sensitive when viewing single-band IR or single-band \hat{f} imagery as when viewing multi-band fused imagery. When seeking vehicle targets, observers were slower to detect

targets in fused images than in IR images, and were less sensitive to targets within achromatic white-hot fused images than to targets in IR images. Thus, while effects of sensor fusion were modulated by characteristics of the targets being sought, data indicate that fusion, at best, maintains the levels of performance attainable with either form of single-band imagery and, at worst, degrades performance relative to that obtainable with single-band imagery. This supports the finding of previous studies (Sampson, Krebs, Scribner, & Essock, 1996; Steele & Perconti, 1997; Krebs et al., 1998) that no reliable difference exists between performance with different sensor formats. However, it conflicts with the findings of a number of others, which revealed that sensor fusion, at least under some circumstances, can, in fact, improve visual performance (e.g., McCarley & Krebs, 2000; Toet et al., 1997). To determine the degree to which these apparent discrepancies might be attributable to differences in task demands between experiments, Experiments 2 and 3 examined performance on experimental tasks different from that of Experiment 1, but using the exact same imagery.

EXPERIMENT 2. SPATIAL ORIENTATION

The goal of this experiment was to evaluate performance on a task that would require a more global percept of the entire scene as opposed to the more local demands of the target detection task of Experiment 1. This task was designed to require the observer to attend to the global spatial relations within the entire scene, with the objective being to determine whether sensor-fusion in general and color-fusion in particular can help to improve performance over that obtained with single-sensor imagery. The task itself was to have the observers determine as quickly and accurately as possible whether a given image was upright or inverted.

Although color-fused images lack color constancy, the chromatic contrast produced by color fusion may aid observers in perceptually segmenting an image into coherent regions forming a recognizable scene (Nothdurft, 1993). Essock et al. (1999) had subjects respond to a briefly flashed, small circular patch of natural scenes cut from identical locations in spatially registered images from three different sensor types (infrared, image intensified, and fused-color). Subjects indicated whether or not the scene contained an object from a designated target category (e.g., building). Results showed that “color plays a predominate role in perceptual grouping and segmenting objects in a scene, and supports the suggestion that the addition of color in complex achromatic scenes would aid the perceptual organization required for visual search.” (Essock et al, 1999). Moreover, subjects’ performance with fused-color scenes was consistently equivalent to or better than performance with infrared and image intensified scenes, thus suggesting that color aids users’ ability to extract low-level information into an organized, coherent scene.

Studies that have investigated whether sensor-fusion can improve the perception of spatial relations within a scene, however, have produced mixed results. Toet et al. (1997) devised a spatial orientation task in which observers attempted to locate the position of individuals present in a nighttime scene in relation to specific objects also within the scene (e.g., a fence). They compared accuracy with single-band imagery and with two types of fused imagery along with grayscale counterparts. They found that observers were significantly more accurate at identifying the location of individuals within a scene when viewing the scenes in an image-fused format, but did not find any significant differences between the color-fused and the grayscale-fused formats.

Alternatively, Steele and Perconti (1997) measured observer's ability to determine whether or not the horizon was level, and found no improvement with fused imagery. The goal of the current study was to further investigate the effects of image fusion on a spatial-orientation task, and to determine whether color provides more low-level information than single sensors do to perceptually organize the scene.

Method

Participants. A total of 48 active-duty military personnel volunteered in this experiment. There were 42 males and 6 females with a mean age of 27.68 and a standard deviation of 5.61. The 48 subjects comprised of 33 officers and 15 enlisted personnel from the United States Army (n=19), United States Navy (n=16), United States Marines (n=8), United States Air Force (n=1), and foreign militaries (n=4). All observers had normal color vision, as tested with pseudo-isochromatic plates, and were naïve to the experimental hypothesis. All subjects signed an informed consent and were briefed on the ethical conduct for subject participation specified in the Protection of Human Subjects (Department of the Navy, 1984).

Procedure. The experimental task required observers to view a single image in each trial and to make a speeded response to indicate whether the scene depicted was upright or was rotated 180 degrees from upright. Each trial began with an alerting tone followed immediately by presentation of the visual stimulus. Observers were instructed to press '1' on the numeric keypad to indicate an upright scene, '2' to indicate a rotated scene. Stimuli remained visible until response. Reaction time and accuracy were recorded for each response. No feedback was given. A two-second interval occurred between response for one trial and presentation of the subsequent stimulus. Each observer

performed ten practice trials and thirty experimental trials. The scenes were rotated on half of all trials. The order of stimuli was randomly determined.

Results and Discussion

For analysis, accuracy data were subjected to signal-detection analysis, with upright images arbitrarily defined as signal + noise stimuli and rotated images defined as noise stimuli. Mean RTs to upright and rotated stimuli were calculated separately. Mean RTs for orientation judgments are presented in Figure 4, mean d' primes in Figure 5.

 Insert Figure 4 about here

 Insert Figure 5 about here

For statistical analysis, sensitivity data were submitted to a one-way ANOVA with image format as a between-subjects factor, while inverse RT data were submitted to a mixed two-way ANOVA with image format as between-subjects factor and stimulus orientation (upright or rotated) as within-subjects factor. Analysis failed to reveal a reliable effect of image format on sensitivity for orientation judgments, $F(5, 42) = 1.94$, $p = .107$, but did indicate a reliable effect of stimulus orientation on inverse RTs, $F(1, 42) = 8.61$, $p = .005$, and more importantly, a reliable effect of image format on inverse RTs, $F(5, 42) = 3.29$, $p = .013$. Post-hoc analysis indicated that this reliable omnibus effect of format was the result of reliable differences between responses to IR stimuli and responses to bh, bhg, and whg stimuli, Tukey's HSD, all p 's < .05. More specifically,

RTs to IR stimuli were slower than those to stimuli in chromatic black-hot, achromatic black-hot, and achromatic white-hot stimuli.

EXPERIMENT 3. SCENE RECOGNITION

This experiment used a test of immediate scene recognition to assess the information shared by and unique to various renderings of a common scene. The psychophysical task asked observers to view a briefly presented image of a nighttime scene, rendered in one of the six image formats, and then to determine whether a subsequently presented test image depicted the same scene. The second image could be of the same format as the first or of a different format. Based on this methodology, performance should have been determined by the degree to which the sensor formats, in which the scenes to be compared are rendered, contain similar information. If two sensor formats tend to convey similar information, then observers should be able to determine easily whether images rendered in those formats depict the same scene. Conversely, if two sensor formats tend to convey information about different aspects of the depicted stimulus, then observers might be expected to perform poorly when asked to determine whether images rendered in those formats depict the same or different distal objects.

Under these assumptions, it was hypothesized that fused imagery, if it indeed conveys information derived from multiple single-band sources, would allow for easier matching against imagery of a different format than would single-band imagery. The usefulness of the emergent chromatic information provided by false color was investigated by comparing performance with the fused achromatic images with their false color counterparts.

Method

Participants. A total of 19 active-duty military personnel volunteered in this experiment. There were 19 males and 0 females with a mean age of 31.94 and a standard deviation of 3.89. The 19 subjects comprised of 18 officers and 1 enlisted personnel from the United States Army (n=2), United States Navy (n=10), United States Marines (n=3), and foreign militaries (n=4). All observers had normal color vision, as tested with pseudo-isochromatic plates, and were naïve to the experimental hypothesis. All subjects signed an informed consent and were briefed on the ethical conduct for subject participation specified in the Protection of Human Subjects (Department of the Navy, 1984).

Stimuli. Twenty scenes were each rendered in the six different formats described earlier. Post-stimulus pattern masks were checkerboard patterns comprised of 5' squares randomly assigned values from a look-up table comparable to that of the masked image. If the initial image was achromatic, then the squares in its checkerboard mask were randomly assigned values from the grayscale look-up table. If the initial image was chromatic, then the squares in its checkerboard mask were randomly assigned values from the look-up table corresponding to one of the chromatic images. This insured that the masking pattern contained similar colors to those in the images with which they were used.

Procedure. Subjects were instructed that they would view two images in each trial and that their task would be to judge whether or not the second image was of the same scene as the first, disregarding image format. Each trial began with a 250-ms presentation of a fixation cross. The first image was then presented at fixation for 50 ms, and was followed immediately by a 300- msec pattern mask. After this, the second image was presented and remained visible until the observer responded. Observers were instructed

to press '1' on the numeric keypad to indicate that the second image depicted the same scene as the first image, '2' to indicate that the second image depicted a different scene. Auditory feedback was given following incorrect responses.

After completing ten practice trials, each observer performed five blocks of 72 trials each. With six format types, 36 pair-wise permutations of image format were possible. Each of these pairings appeared twice within a block, once with the two scenes identical and once with them different. The first scene presented during each trial was chosen randomly from the full set of twenty possible scenes. The second scene, when different from the first, was randomly chosen from the remaining pool of 19 scenes. The order of trials within a block was randomly determined. The dependent variables were again reaction time and accuracy.

Results and Discussion

Mean RTs for same/different judgments are presented in Figure 6 and mean d primes in Figure 7. Values of d prime were calculated assuming an independent observation strategy of same/different performance. For statistical analysis, d primes and inverse RTs were submitted to separate omnibus 6 x 6 within-subjects ANOVAs, with format of the first image and format of the second image as factors.

Insert Figure 6 about here

Insert Figure 7 about here

Omnibus analysis of d prime scores revealed no significant main effect of the format of the first image presented in each trial, $F(5, 90) = 1.82$, $p = .116$, but did reveal a reliable effect of the second image's format, $F(5, 90) = 2.47$, $p = .038$, along with a reliable interaction of first and second images' formats, $F(25, 450) = 1.73$, $p = .016$. Given the 6 x 6 design of the current experiment, the interaction of the first image's format with the second image's format becomes somewhat difficult to interpret. One effect which likely contributed, however, was a reliable tendency toward better performance when both images within a trial were of the same format than when the two images were of different formats, mean d prime = 2.96 for two images of the same format, mean d prime = 2.63, $t(18) = 3.89$, $p = .001$. To allow closer examination of the reliable main effect of second image's format, post-hoc single-df tests were conducted to examine two comparisons of particular interest. First, to determine whether sensor fusion improved image quality generally, mean performance when the second image was of a fused format was compared to mean performance when the second image was of a single-band format. Second, to determine if chromatic information derived through sensor fusion improved was perceptually useful, performance when the second image was of a chromatic fused format was compared to performance when the second image was of an achromatic fused format. The first of these comparisons indicated a reliable effect; mean sensitivity when the second image was of a fused format was reliably higher than mean sensitivity when the second image was of either IR or f format, $F(1, 18) =$

4.67, $p = .044$. This improvement, moreover, was apparently engendered primarily by effects of fusion on spatial, not chromatic, image content; mean sensitivity when the second image was of a chromatic fused format did not differ reliably from mean sensitivity when the second image was of an achromatic fused format, $F = 1$, indicating that chromatic information derived through fusion did little to improve image quality as measured by the current task.

Omnibus analysis of inverse RTs failed to reveal either a main effect of the first image's format, $F(5, 90) = 1.64$, $p = .159$, or an interaction of the first image's format with that of the second image, $F(25, 450) = 1.24$, $p = .197$. Consistent with sensitivity data, however, analysis did produce a reliable main effect of second image's format, $F(5, 90) = 5.29$, $p < .001$. Post-hoc single-df tests conducted to explore this effect revealed a reliable difference between mean performance when the second image was of a single-band format and mean performance when the second image was of a fused format, $F(1, 18) = 18.014$, $p < .001$. Comparison of mean performance when the second image was of a chromatic fused format to mean performance when the second image was of an achromatic fused format, however, again revealed no reliable effect, $F < 1$, reaffirming the conclusion that chromatic information derived through fusion did little to improve image quality.

The experimental task required observers to view a briefly presented image and to then determine whether or not a second image, presented 300 msec later, depicted the same scene, regardless of the specific format type. Recognition was best when the first and second images were presented in the same format. Performance was best more generally, however, when the second image was displayed in a sensor-fused rather than a

single-band format. The format of the first image itself had little effect at all, suggesting that the primary benefits of sensor fusion in this task were in matching the content of the second image to a stored representation of the first, and not in processing the briefly viewed first image. These results suggest that fused images contained more perceptually accessible information than did single-band input images. Thus, fusion appeared to allow information from multiple single-band sensors to be effectively combined and, perhaps, improved. Notably, the benefits of fusion observed here appeared to result exclusively from changes to spatial image content, as chromatic rendering of fused images did little to improve performance relative to that obtained with achromatic rendering.

GENERAL DISCUSSION

The results in these three experiments show that the benefits of sensor fusion are, indeed, task-dependent. We found that sensor fusion does not improve performance over single-band imagery for target detection, but does lower RTs in a spatial-orientation task and improve accuracy in an unspeeded scene-recognition task. The significant improvements in performance observed with fused-imagery in these latter tasks indicate that perceptible information, indeed, can be effectively combined and perhaps even derived through fusion of single-band imagery.

As noted above, past studies of the perceptual utility of sensor-fused imagery have generally produced equivocal and even conflicting results, with some reports failing to reveal any benefits of sensor fusion (e.g., Sampson et al., 1996; Steele and Perconti, 1997; Krebs et al., 1998) and others suggesting that fused imagery can, in fact, benefit human performance (Essock, et al, 1999; McCarley & Krebs, 2000; Toet, et al, 1997; Waxman, et al, 1996). The current study points to one potential cause of these seeming

discrepancies. By testing multiple psychophysical tasks with the same type of imagery, and even the exact same scenes in many cases, the experiments described here reveal that the benefits of fusion can depend critically on the information to be extracted from an image—that is, on the perceptual task to be performed with an image. Here, fused imagery was of no benefit when the observers' task was to detect a designated target object within an image (Experiment 1), but aided performance when observers were asked to judge the orientation of a scene (Experiment 2) or to recognize a briefly viewed scene after a short delay (Experiment 3). Data thus indicate that fusion is probably not acceptable as a general-purpose method of improving degraded imagery, but will vary in its utility according to the demands of the task it is meant to support.

Unfortunately, it is still impossible to know a priori whether or not a given psychophysical task will benefit from fusion of single-band imagery. That is, there is no clearly defined set of tasks for which it can be assumed that image fusion will be helpful. For example, although fusion did not facilitate target detection in the current experiments, it has in other studies (Essock, et al, 1999; McCarley & Krebs, 2000; Toet, et al, 1997). It therefore cannot be true that fusion of single-band images is either invariably useless or invariably beneficial to target detection. Rather, the utility of sensor fusion to an observer performing a detection task will itself vary as a function of the environmental conditions under which input imagery is obtained, and of the algorithm by which fused imagery is produced. This point is illustrated well by the experiment described by McCarley and Krebs (2000). In that study, observers were asked to detect a pedestrian in a nighttime scene. Input imagery was degraded by three levels of glare (light from the headlights of an automobile), and composite imagery was produced through fusion of

two forms of input imagery through two different fusion algorithms. Performance was assessed with grayscale, and color imagery produced through both fusion algorithms, as well as with both formats of input imagery. The effects of both grayscale and color fusion varied strikingly as a function of fusion algorithm and with the level of glare characterizing the input imagery; while, under some conditions, fusion improved performance relative to either form of input imagery, under other conditions, it degraded performance relative to the better form of input imagery. Even color rendering, under some conditions, was detrimental to performance. In conjunction with those of the current study, such results demonstrate not only that sensor fusion provides a potentially valuable method for improving the quality of electronically-sensed images, but also that a sensor-fusion system should be carefully tailored to the circumstances under which it will be employed.

While sensor-fusion systems have until now been considered as general-purpose aids to nighttime visual performance, human performance data show that single-band systems perform as well as or better than multi-band or sensor-fused systems for a variety of tasks. The engineer's decision to implement a single- or dual-band system must therefore depend upon the operator's task. If the operator's task is to detect an object – searching for a luminance difference – the engineer should consider a single-band system, while recognition – discrimination between specific objects within a class of similar objects – the engineer should consider a dual-band system. If the dual-band system is selected, the engineer must consider what is the appropriate dual-band algorithm for the task? The engineer can select from an unlimited number of algorithms, however careful consideration should identify the advantages and disadvantages between

the different image processing techniques. Meaning, some algorithms enhance contrast (Therrien, Scrofani and Krebs, 1997) to aid human vision while a non-color constancy algorithm (Scribner, Warren, Schuler, Satyshur, and Kruer, 1998) is more suitable for missile threat detection.

Once the sensor configuration and algorithm have been selected for a particular task, the engineer must determine the appropriate method to display the information. For all systems, the method of presentation will be task dependent. The engineer should consider the single and dual-band imagery as an additional visual cue that may enhance operators' performance. For example, the United States Army's AH-64 Apache helicopter, the pilot can either select between infrared, image-intensified, or unaided eye to enhance nighttime visual performance. By allowing pilots to select between multiple sensors, pilots may be able to determine whether the object is a target or false alarm. As an alternative, displaying an image-fused scene to the AH-64 pilot may improve target detection sensitivity as well as increase safety and efficiency by eliminating the need to switch between two or more disparate scenes. Currently, the United States Army is considering replacing some warfighting platform single-band systems with dual-band systems.

In summary, careful and elaborate psychophysical testing must precede the deployment of any sensor-fusion system. Before an engineer integrates a single- or dual-band sensor into a system, the engineer must determine what is the operator's task. From this task description, the engineer can then determine whether a single- or dual-band system will be appropriate. Furthermore, sensor characteristics such as spectral sensitivity, field-of-view, field-of-regard, overlay of synthetic symbols on the scene, and

resolution must be optimized, as should be the selection of image fusion algorithm. If these variables are not considered, then operator performance may suffer.

ACKNOWLEDGEMENTS

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the United States Government.

This study was sponsored by the Office of Naval Research contracts #N0001497WR30091 and #N0001498WR30112.

REFERENCES

- Ahumada, A.J. & Krebs, W.K. (2000). Signal detection in fixed pattern chromatic noise. Investigative Ophthalmology and Visual Science, (SUPPL) 41, S3796.
- Aguilar, M., Fay, D. A., Ross, W. D., Waxman, A. M., Ireland, D. B., & Racamato, J. P. (1998). Real-time fusion of low-light CCD and uncooled IR imagery for color night vision. In J.G. Verly (Ed.), Proceedings of the SPIE – Enhanced and Synthetic Vision, (Vol. 3364, pp.124-135). Bellingham, WA: SPIE – The International Society for Optical Engineering.
- Baylor, D.A., Nunn, B.J., & Schnapf, J.L. (1987). Spectral sensitivity of cones of the monkey *Macaca fascicularis*. Journal of Physiology, 390, 145-160.
- Bergman, S.M. (1996). The utility of hyperspectral data to detect and discriminate actual and target decoy vehicles. Unpublished master's Thesis, Naval Postgraduate School, Monterey, CA.
- Bowmaker, J.K. & Dartnall, H.J.A. (1980). Visual pigments of rods and cones in a human retina. Journal of Physiology, 298, 501-511.
- Chun, M.M. & Wolfe, J.M. (1996). Just say no: how are visual searches terminated when there is no target present? Cognitive Psychology, 30(1), 39-78.
- Das, S. & Krebs, W.K. (2000). Sensor fusion of multi-spectral imagery. Institution of Electrical Engineers: Electronics Letters, 36, 1115-1116.
- Department of the Navy. (1984). Protection of human subjects. (SECNAV Instruction 3900.39B). Washington, D.C.: Chief of Naval Operations OP-098.

- Essock, E.A., Sinai, M.J., McCarley, J.S., Krebs, W.K., & DeFord, J. K. (1999). Perceptual ability with real-world nighttime scenes: image-intensified, infrared and fused-color imagery. Human Factors, *41*, 438-452.
- Krebs, W.K., Scribner, D.A., Miller, G.M., Ogawa, J.S., Schuler, J. (1998). Beyond third generation: a sensor fusion targeting FLIR pod for the F/A-18. In B.V. Dasarathy (Ed.), Proceedings of the SPIE - Sensor Fusion: Architectures, Algorithms, and Applications II, (Vol. 3376, pp.129-140). Bellingham, WA: SPIE – The International Society for Optical Engineering.
- Krebs, W.K., McCarley, J.S., Kozek, T., Miller, G.M., Sinai, M.J., & Werblin, F.S. (1999). An evaluation of a sensor fusion system to improve drivers' nighttime detection of road hazards. Proceedings of the 43rd Annual Meeting Human Factors and Ergonomics Society, *43*, 1333-1337.
- McDaniel, R., Scribner, D., Krebs, W., Warren, P., Ockman, N., McCarley, J. (1998). Image fusion for tactical applications. In B.F. Andresen & S.M. Strojnik (Eds.), Proceedings of the SPIE - Infrared Technology and Applications XXIV, (Vol. 3436, pp. 685-695). Bellingham, WA: SPIE – The International Society for Optical Engineering.
- McCarley, J.S., & Krebs, W.K. (2000). Visibility of road hazards in thermal, visible, and sensor-fused nighttime imagery. Applied Ergonomics, *31*, 523-530.
- Nordwall, B.D. (1997). UV sensor proposed as pilot landing aid. Aviation Week and Space Technology, *147*, 81.
- Nothdurft, H.C. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. Vision Research, *33*, 1937-1958.

- O’Kane, B.L., Crenshaw, M.D., D’Agostino, J.D., & Tomkinson, D. (1992). Human target detection using thermal systems. In J.S. Accetta & M.J. Cantella (Eds.), Proceedings of the SPIE - Aerospace/Defense Sensing, Simulation and Controls (Vol. 2075, pp. 75-88). Bellingham, WA: SPIE – The International Society for Optical Engineering.
- Peli, T., Peli, E., Ellis, K., Stahl, R. (1999). Multi-spectral image fusion for visual display. In B.V. Dasarathy (Ed.), Proceedings of the SPIE - Fusion: Architectures, Algorithms, and Applications III, 3376, 129-140. (Vol. 3719, pp. 359-368). Bellingham, WA: SPIE – The International Society for Optical Engineering.
- Ryan, D. & Tinkler, R. (1995). Night pilotage assessment of image fusion. In R.J. Lewandowski, W. Stephens, & L.A. Haworth (Eds.), Proceedings of the SPIE - International Symposium on Aerospace/Defense Sensing, Simulation and Controls (Vol. 2465, pp. 50-67). Bellingham, WA: SPIE – The International Society for Optical Engineering.
- Sampson, M. T., Krebs, W. K., Scribner, D. A. & Essock, E. A. (1996). Visual search in natural (visible, infrared, and fused visible and infrared) stimuli. Investigative Ophthalmology and Visual Science, (SUPPL), 37, S296.
- Schnapf, J.L., Kraft, T.W., & Baylor, D.A. (1987). Spectral sensitivity of human cone photoreceptors. Nature, 325, 439-441.
- Scribner, D.A., Schuler, J., Warren, P., Satyshur, M., Kruer, M.R. (1998). Infrared color vision: separating objects from backgrounds. In E.L. Dereniak & R.E. Sampson (Eds.), Proceedings of the SPIE - Sensor Fusion: Infrared Detectors and Focal Plane

Arrays V (Vol. 3379, pp. 2-13). Bellingham, WA: SPIE – The International Society for Optical Engineering.

Scribner, D.A., Warren, P., Schuler, J., Satyshur, M., Kruer, M.R. (1998). Infrared color vision: an approach to sensor fusion. Optics and Photonics News, 8, 27-32.

Steele, P.M. & Perconti, P. (1997). Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage. In W.R. Watkins & D. Clement (Eds.), Proceedings of the SPIE: Aerospace/Defense Sensing, Simulation and Controls (Vol. 3062, pp. 88-100). Bellingham, WA: SPIE – The International Society for Optical Engineering.

Swift, D.J., Panish, S. & Hippensteel, B. (1997). The use of VisionWorks in visual psychophysics research. Spatial Vision, 10, 471-477.

Therrien, C.W., Scrofani, J., & Krebs, W.K. (1997). An adaptive technique for the enhanced fusion of low-light visible with uncooled thermal infrared imagery. In R.M. Gray & B. Hunt (Eds.), Proceedings of the IEEE: International Conference on Imaging Processing, 405-408. Los Alamitos, CA: Institute of Electrical and Electronics Engineers.

Toet, A. & Walraven, J. (1996). New false colour mapping for image fusion. Optical Engineering, 35, 650-658.

Toet, A., Ijspeert, J. K., Waxman, A. M., & Aguilar, M. (1997). Fusion of visible and thermal imagery improves situational awareness. In J.G. Verly (Ed.), Proceedings of the SPIE – Enhanced and Synthetic Vision, (Vol. 3088, pp. 177-188). Bellingham, WA: SPIE – The International Society for Optical Engineering.

Waxman, A.M., Gove, A.N., Seibert, M.C., Fay, D.A., Carrick, J.E., Racamato, J.P., Savoye, E.D., Burke, B.E., Reich, R.K., McGonagle, W.H. & Craig, D.M. (1996). Progress on color night vision: Visible/IR fusion, perception and search, and low-light CCD imaging. In J.G. Verly (Ed.), Proceedings of the SPIE – Enhanced and Synthetic Vision, (Vol. 2736, pp.96-107). Bellingham, WA: SPIE – The International Society for Optical Engineering.

Waxman, A.M., Gove, A.N., Fay, D.A., Racamato, J.P., Carrick, J.E., Seibert, M.C., & Savoye, E.D. (1997). Color night vision: Opponent processing in the fusion of visible and IR imagery. Neural Networks, 10, 1-6.

Figure Legends

Figure 1. Sample stimulus of the same scene shown in LL (top), IR (middle), and achromatic fused (bottom) formats.

Figure 2. Observers' reaction times were generally faster for personnel targets than for vehicles, however image format did not affect performance nor was there a reliable interaction of image format by target type. Error bars represent +/- 1 standard error of the mean. In this and all proceeding graphs, format types are identified as follows: BH is color fused imagery with the IR input being black hot; BHG is the same imagery represented in grayscale; WH is color fused with IR input being white hot; WHG is the grayscale version; IR is infrared; and LL is image-intensified low light.

Figure 3. Mean sensitivity (d') values for the detection task. Observers were significantly better at detecting personnel targets than vehicle targets. Image format was significant with single-band IR or single-band i^2 imagery producing better detection sensitivity than multi-band fused imagery.

Figure 4. Mean reaction times for the spatial-orientation task. Observers' response times showed a significant effect for image format. Observers were slower to IR stimuli than those to stimuli in chromatic black-hot, achromatic black-hot, and achromatic white-hot stimuli.

Figure 5. Sensitivity (d') values for the for the spatial-orientation task. D' prime was calculated by defining the upright images as the signal + noise distribution and the rotated images as the noise distribution. The signal-detection analysis failed to show a significant image format difference on observers' orientation judgments.

Figure 6. Mean RTs for same/different judgments in the scene-recognition task. Subjects' task was to view two sequentially presented images then determine whether or not the second image was of the same scene as the first, disregarding image format. The image format main effect was non-significant, but there was a reliable main effect for the second image format.

Figure 7. Mean d' primes for same/different judgments in the scene-recognition task. Consistent with the reaction time results, observers' were more sensitive when the second image was of a fused format than when the second image was of either IR or I^2 ; however, increased sensitivity to fused formats was due to spatial information rather than chromatic information.

Figure 1



Figure 2

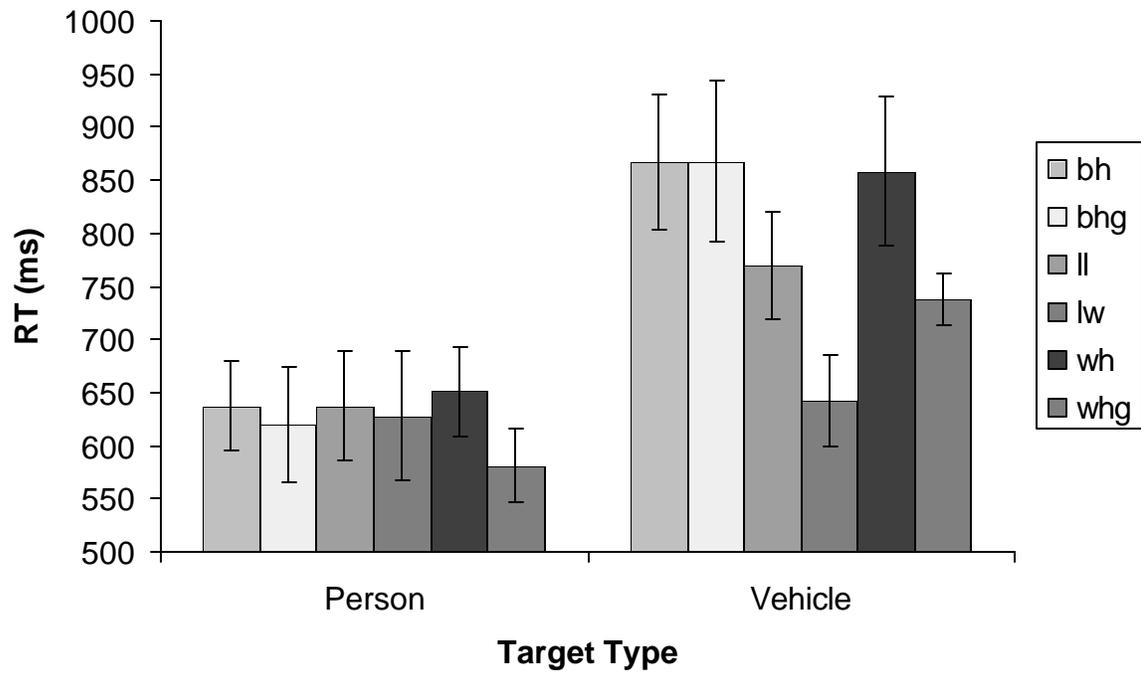


Figure 3

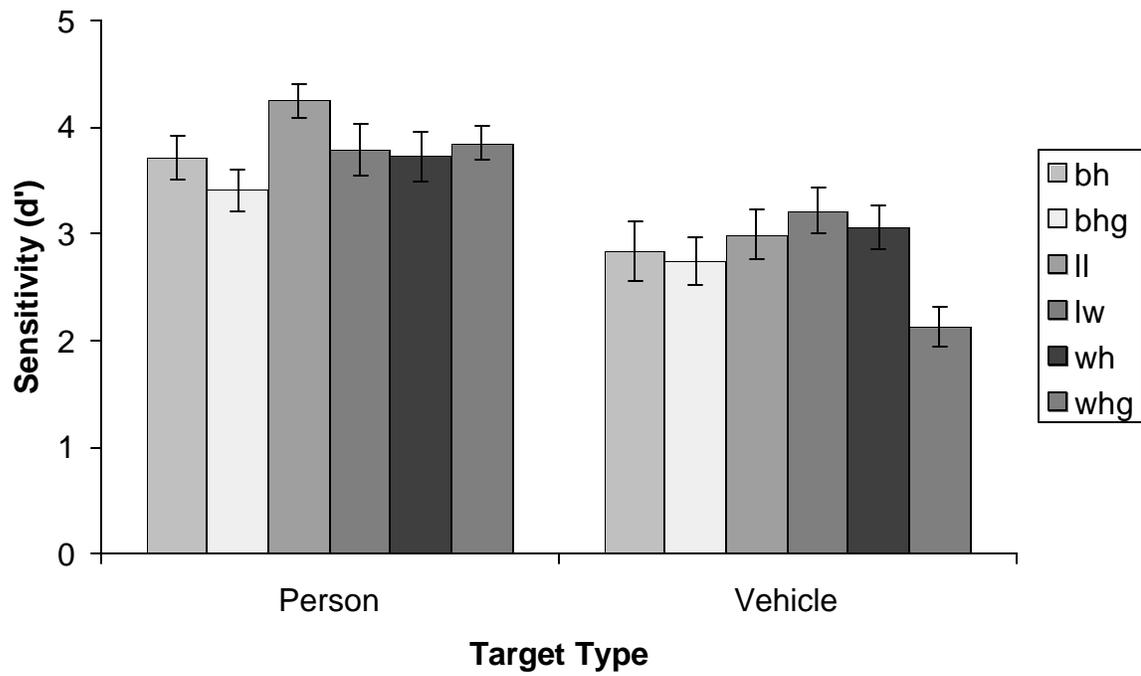


Figure 4

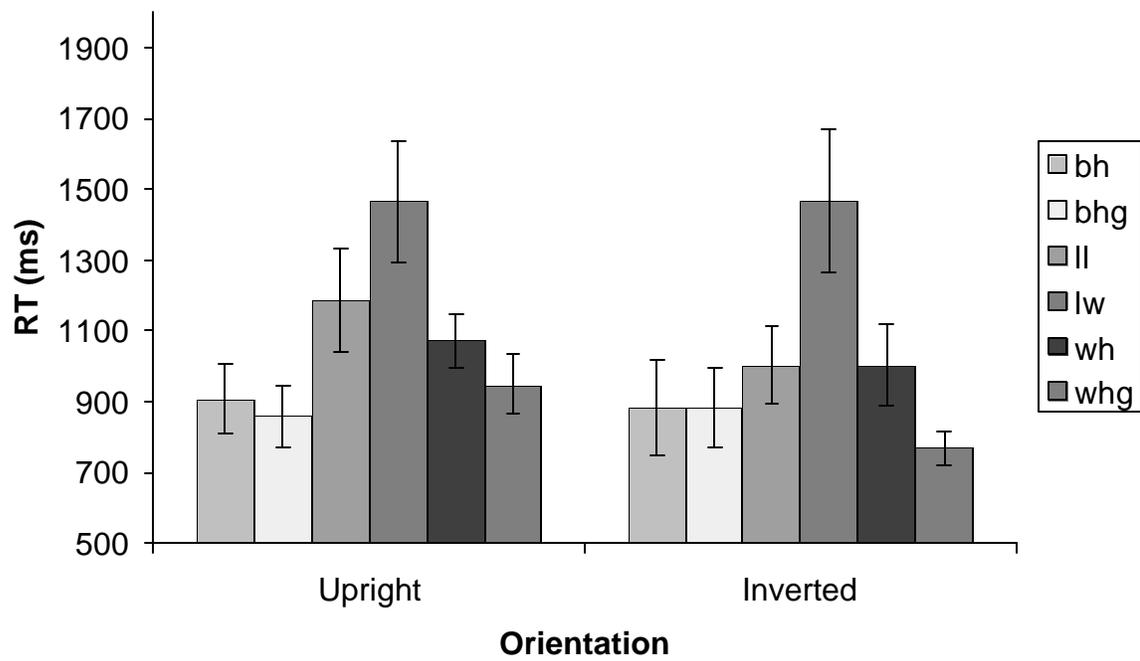


Figure 5

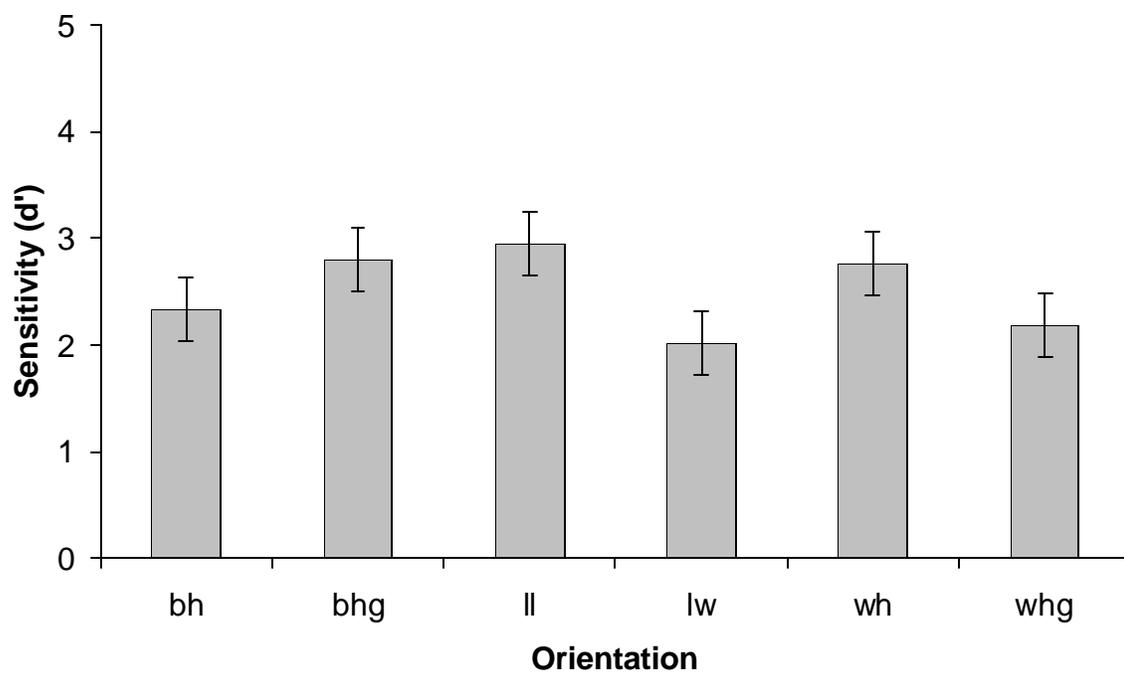


Figure 6

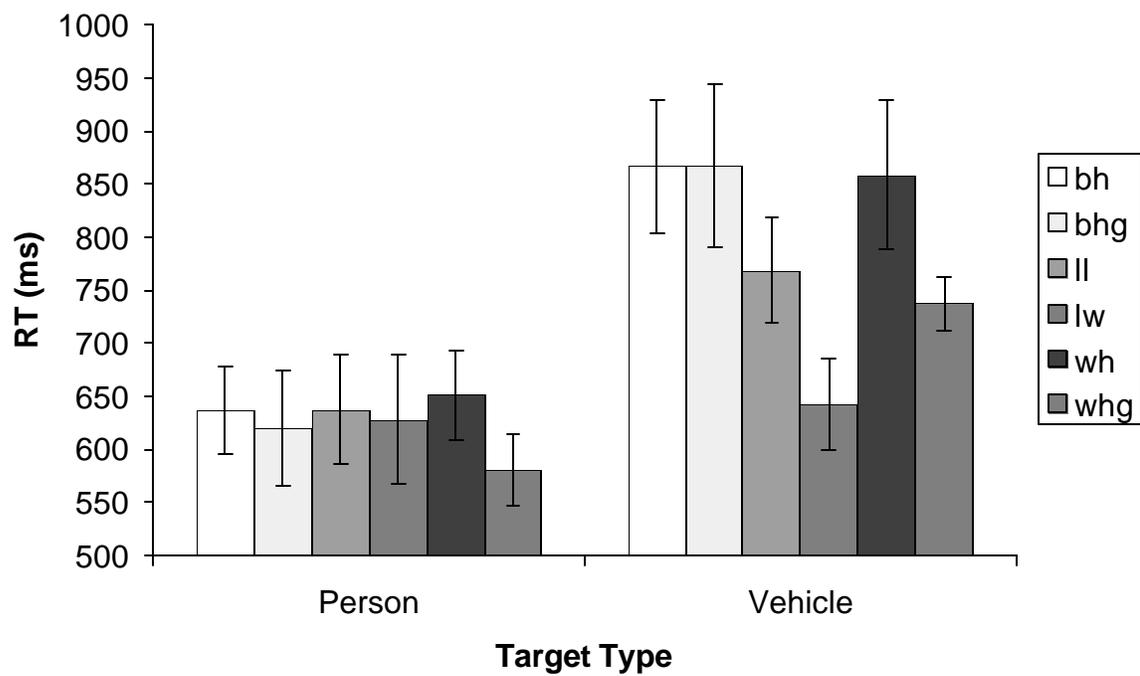


Figure 7

