Spatial Awareness: Comparing Judgment-based and Subjective Measures

Matthew L. Bolton, Member, IEEE, and Ellen J. Bass, Senior Member, IEEE

***Abstract*— Spatial awareness is important in domains where safety hinges on human operators keeping track of the relative locations of objects in their environments. While a variety of subjective and judgment-based measures have been used to evaluate spatial awareness, none have probed all three of its levels: 1) identification of environmental objects, 2) their current locations relative to the operator, and 3) their relative positions over time. This work compares new judgment based measures of spatial awareness that probe all three levels of spatial awareness to conventional subjective measures. In the evaluation of 14 configurations of Synthetic Vision Systems head down displays (7 terrain textures and 2 Fields of View (FOVs)), 18 pilots made 4 types of judgments (relative angle, distance, height, and abeam time) regarding the location of terrain points displayed in 112 5-second, non-interactive simulations. They also provided subjective awareness and SA-SWORD measures. ANOVA analyses revealed that comparable results were found between display configurations that produced the minimum error in judgments and those that received the highest subjective ratings. However, none of the subjective measures were correlated with judgment error. Thus, given that the judgment based measures were explicitly designed to measure all three levels of spatial awareness, the subjective measures may not be measuring spatial awareness.**

I. Introduction

SPATIAL awareness (SpA) is an aspect of situation awareness (SA) [1] that encompasses the extent to which a person notices objects in the environment (Level 1), his understanding of where these objects are (Level 2), and his understanding of where they will be in the future (Level 3) [2]. SpA is important in domains such as air traffic control and aviation where safety hinges on operators keeping track of the relative location of objects in the environment. Within these domains, engineers need measures to appropriately evaluate SpA in the systems they intended to support.

A. Subjective Measures

Subjective measures have frequently been used to evaluate awareness, workload, clutter, and other cognitive metrics. The use of subjective measures typically involves exposing participants to a particular design under varying operational conditions and having them rate how they felt (in the case of awareness, how aware they thought they were) using Likert scales. A common subjective awareness measurement technique involves having participants rate the awareness provided by displays on either 7 or 100 point scales [3].

One domain where subjective measures have frequently been used to evaluate awareness, both SpA and general SA, is Synthetic Vision Systems (SVS). Controlled Flight Into Terrain (CFIT), where a fully functional aircraft is inadvertently flown into the ground, water, or other terrain obstacle, has been the cause of more than 25% of all fatal accidents in worldwide commercial aviation since 1987 [7]. Such accidents are often characterized by a loss of SA in low level flight and low visibility conditions [8]. SVS are technologies that address this problem by using onboard terrain databases and Global Positioning System (GPS) data to create a synthetic, clear-day view of the world in front of ownship.

Likert scale subjective measures have been used to evaluate SpA in SVS in [10]. Situation Awareness – Subjective WORkload Dominance (SA-SWORD) is another type of subjective measure that has been used to evaluate awareness in SVS [11]. SA-SWORD, which was adapted to measure awareness from the Subjective Workload Dominance technique, allows participants to make pair-wise comparisons between displays on a seventeen point scale concerning the relative amount of SA provided by each [4]. For *n* displays, SA-SWORD requires that $\binom{n}{2}$ comparisons be made. Values from each are then used to calculate scores for all of the displays [5].

Despite their wide spread use, subjective-based measures do not always match performance. While Likert scale measurement techniques have been correlated with performance measures [12], they have also been shown to be correlated with operator confidence [13]. Thus, these metrics may be measuring confidence rather than awareness.

Additionally, studies utilizing SA-SWORD have found strong correlations between pilot display preference and SA-SWORD scores [14]. It is possible that preference influences SA-SWORD rankings or vice versa.

B. Judgment-based Measures

Judgment based metrics provide another approach to measuring SpA. Yeh [15] used ordinal distance judgments to assess spatial perception for stereoscopic and perspective displays. Several studies have utilized azimuth and elevation angle judgments of the relative position of two objects over

Manuscript received July 12, 2007. This work was supported in part by Synthetic Vision Systems - Rising Star Fellowship (M. Bolton), UVA-03-01, Sub-Award Number: 4817-UV

M. L. Bolton is a Ph.D. student in Systems Engineering at the University of Virginia, Charlottesville, VA 22904 USA (phone: 434-243-2280; fax: 434-982-5393; e-mail: mlb4b@virginia.edu).

E. J. Bass, is an Assistant Professor in the Department of Systems and Information Engineering at the University of Virginia, Charlottesville, VA 22904 USA (email: ejb4n@virginia.edu).

synthetic terrain [16][17]. Point position replication has also been used [18][19]. None of these have measured all three levels of SpA.

C. Objectives

Quantitative measures of SpA are necessary to inform the development of decision support tools in many domains. This paper compares subjective awareness measures to new judgment-based measures that probe all three levels of SpA: identifying an object in the environment probe Level 1 SpA; judgments of the relative position of the object probe Level 2 SpA; and a judgment of the time it would take to reach that object probe level 3 SpA. Both directional and absolute error terms for each judgment assess SpA accuracy at each level. Since these measures are theoretically motivated by the SpA literature, their comparison to subjective measures could reveal what aspects of SpA the subjective measures assess.

This comparison is conducted within the SVS domain because of its convention for utilizing subjective measures. The new SpA measures were evaluated with respect to a point indicated on the terrain (terrain point) of an SVS display. Identifying the terrain point probed Level 1 SpA; judgments of the relative angle, distance, and height of the terrain point probed Level 2 SpA (providing a three dimensional perspective of the pilot's perceived terrain location); and an abeam time judgment (the time it would take the pilot to fly abeam to the terrain point) probed level 3 SpA.

Two aspects of SVS that can affect SpA are terrain texture (the imagery drawn on the synthetic terrain) and FOV (the angular boundaries of the volume of space represented in the display). As part of a larger study to determine what textures and FOVs best facilitated SpA for SVS (see [20]) and identify SpA biases in SVS (see [21]) using the new SpA measures, this work compared the results obtained by the new SpA measures with terrain awareness and SA-SWORD subjective measures. This was done by comparing which textures and FOVs were recommended by the measure, and looking for correlations between measures.

II. Methods

A. Participants

Eighteen general aviation pilots were recruited to participate in the study. All participants had less than 400 hours of flight experience (median = 140, range = [65, 300]). They were not familiar with SVS displays.

B. Apparatus

Desktop computers running custom software [22] displayed simulation trials and collected participant judgments. SVS displays employed the symbology depicted in Fig. 1. In each trial, the location of the terrain point was indicated using a yellow inverted cone (d = 500 ft, h = 500 ft) which was rendered as part of the SVS environment. The tip of the cone intersected the terrain at the terrain point. All simulations depicted SVS displays in flight at 127 knots and

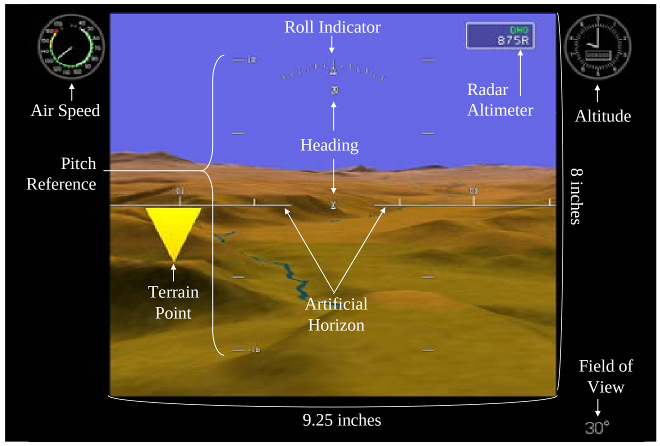

Fig. 1. The SVS display and symbology used in the experiment.

were displayed as 5 second, 836 pixel × 728 pixel, 30 frames per second, Windows Media Video (WMV) files.

C. Independent Variables

There were five within subject independent variables. These included texture, FOV, and three scenario geometry variables: the relative (azimuth) angle, distance, and height of the terrain point to ownship.

Seven textures (Fig. 2) were used in the experiment. There were three base textures: Fishnet (F), where a grid of 500ft by 500ft squares are drawn over solidly colored terrain; Elevation (E), where bands of color representing regularly spaced intervals of terrain elevation are depicted on the terrain; and Photo (P), where satellite photos of the actual terrain superimposed on the synthetic terrain. All combinations of these textures were also used: EF, PF, PE, and PEF. The three base textures were chosen because each had been used in SVS [10], and each facilitated different depth cues that persisted under combination [20].

Two FOVs (30° and 60°) were used in this experiment. These were selected because they were shown to be preferred by pilots in [10], [11], and [23].

In order to mediate the effect of known spatial biases ([2] and [21]), the location of the terrain point varied based on its relative position at the end of a simulation by changing the scenario geometry variables (relative angle, distance, and height). Each variable had two levels (TABLE I).

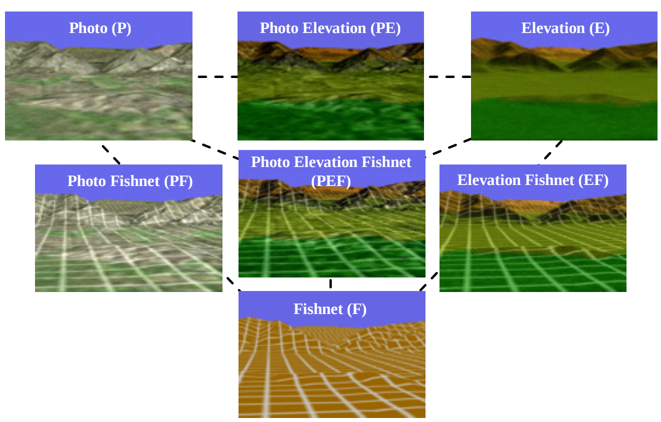

Fig. 2. Terrain textures evaluated in the experiment.

TABLE I
TERRAIN POINT RELATIVE POSITION LEVELS

Variable	Range	Distribution	Level
Relative Angle	[0°, 6.5°]	$N(\mu = 3.75, \sigma = 1.25)$	Small
	[8.5°, 15°]	$N(\mu = 11.25, \sigma = 1.25)$	Large
Relative Distance	[1 nmi,3.25 nmi]	$N(\mu = 2.25, \sigma = 0.417)$	Near
	[3.75 nmi,6 nmi]	$N(\mu = 4.75, \sigma = 0.417)$	Far
Relative Height	[-1000 ft,-100 ft]	$U(-1000, -100)$	Below
	[100 ft,1000 ft]	$U(100, 1000)$	Above

D. Dependent Measures

There were seventeen dependent measures in this study. Eleven are reported here (others are discussed in [20] and [21]). Eight of these were calculated from the four judgment values (relative angle (°), relative distance (nmi), relative height (ft), and abeam time (s)) from the three judgment tasks (TABLE II). Two dependent measures were associated with each of these judgment values: one for directional error and one for absolute error. Each directional error term represented both the direction and magnitude of the error in the judgment value. Absolute error terms represented the absolute value of the magnitude of the judgment error.

There were six subjective measure dependent variables in the study. Only Awareness and two SA-SWORD measures are discussed here. For each texture and FOV combination, Awareness ratings were collected on a one hundred point scale and were comparable to the SpA measure used in [10]. SA-SWORD scores were computed for each texture within a given FOV. Thus, there were two SA-SWORD dependent measures: SA-SWORD 30° and SA-SWORD 60°.

E. Procedure

Each experimental session lasted less than four hours. The participants completed consent forms and were briefed about the experiment. Each participant was randomly assigned to an experimental condition. For each trial, participants viewed five second simulations of an SVS heads down display in flight (Fig. 1). At the end of the five seconds, the display froze for one second and then the screen was cleared.

For each trial, participants made four judgments based on the relative position of the terrain point: relative angle, relative distance, relative height, and abeam time using the interface in Fig. 3. For the relative distance and angle judgments, participants placed a yellow X in the upper left section of the display corresponding to the lateral location of the terrain point relative to the aircraft. For the relative height judgment, the participant placed a yellow X on a vertical scale in the upper right of the display corresponding to the

TABLE II
DEPENDENT VARIABLE FORMULATIONS

Relative Measure	Actual Value	Judgment Value	Directional Error	Absolute Error
Angle	A_a	A_j	$A_e = \begin{cases} A_j - A_a & \text{if } A_a > 0 \\ -A_j + A_a & \text{otherwise} \end{cases}$	$\lvert A_e \rvert$
Distance	D_a	D_j	$D_e = D_j - D_a$	$\lvert D_e \rvert$
Height	H_a	H_j	$H_e = \begin{cases} H_j - H_a & \text{if } H_a > 0 \\ -H_j + H_a & \text{otherwise} \end{cases}$	$\lvert H_e \rvert$
Abeam Time	τ_a	τ_j	$\tau_e = \tau_j - \tau_a$	$\lvert \tau_e \rvert$

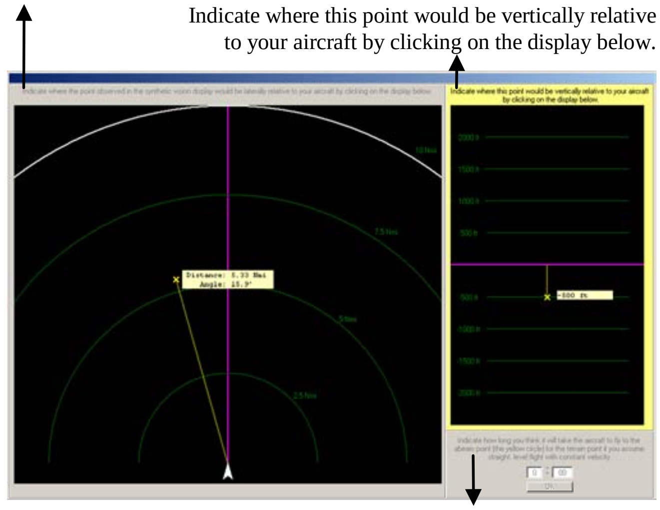

Fig. 3. The judgment collection interface.

relative height of the terrain point. For the abeam time judgment, the participant entered the time in minutes and seconds using the keyboard. To support this time judgment, a yellow dot on the relative distance and angle judgment collection interface indicated the location of the abeam point based on the relative distance and angle judgment. Participants were asked to perform these tasks as quickly and accurately as possible.

The data collection interface allowed judgments to be collected independently of FOVs used in the SVS displays. The height judgment was collected separately from the relative distance and angle judgments due to the difference in resolutions between the distance (nmi) and height judgments (ft) (see [22] for more information).

Each participant experienced 112 counterbalanced experimental trials (7 textures × 2 FOVs × 2 angles × 2 distances × 2 heights = 112) and 72 training trials. Participants saw all of the trials with one FOV before seeing any trials with the other. FOV presentation order was counterbalanced between participants. Base textures always appeared before their combinations to avoid confusion. Each participant saw two of the base textures, the combination of them, the third texture, and the rest of the combinations. Three texture orders were created so that no base texture was introduced in more than one ordered slot: {P, E, PE, F, PF, EF, PEF}, {E, F, EF, P, PE, PF, PEF}, and {F, P, PF, E, EF, PE, PEF}. Texture orders were counterbalanced between participants.

On completion of trials for each texture and FOV, Awareness ratings were collected (Fig. 4). After all of the trials for a FOV were completed, participants made SA-SWORD pair-wise comparisons between each texture with that FOV (Fig. 5).

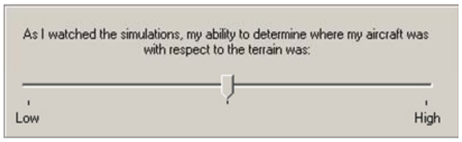

Fig. 4. The awareness and collection interface.

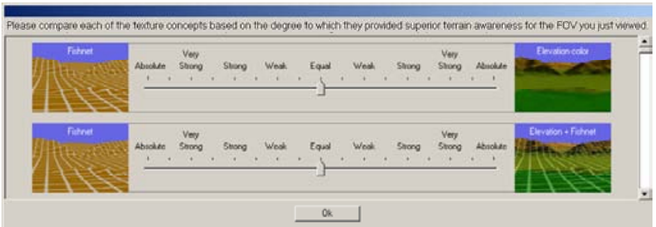

Fig. 5. The SA-SWORD collection interface.

F. Experimental Design and Data Analysis

The experiment employed a repeated measures design with 18 participants. Three participants were randomly assigned to each of the six combinations of the between subject variables (2 FOV orders × 3 texture orders = 6).

The Awareness and SA-SWORD dependent measures were subjected to post-processing before being evaluated in the subsequent analyses. Awareness measures were converted into z-scores based on the mean and standard deviation of a participant's ratings. SA-SWORD 30° and SA-SWORD 60° pair-wise comparisons were converted to numerical scores using the method described in [5].

The main and two-way interaction effects of the within and between subject factors on the dependent measures were assessed using a univariate repeated measures analyses of variance (ANOVA) [24]. A Mauchly's Test of Sphericity was performed in order to ensure that the assumptions for the repeated measure analysis were not violated. When sphericity was violated (p < 0.05) a Greenhouse-Geisser Epsilon correction factor was used [24]. Pearson's correlation coefficients were used to evaluate correlations between dependent measures.

III. Results

Results are presented using $\alpha = 0.05$ for significance.

A. Texture

1) Judgment-Based Measures. Texture was a significant main effect for $|A_e|$ and $|D_e|$ (TABLE III). A Bonferroni post hoc analysis revealed no significant differences between textures for $|A_e|$. However, a Tukey's HSD showed that participants committed significantly less distance error for the EF and PEF textures than they did for the F texture (Fig. 6a). Further, an examination of the error levels for which a texture × scenario geometry interaction was significant revealed that there were three textures, EF, PF and PEF, that were among the set of homogeneous subsets of textures (no significant differences between textures) that produce the minimum absolute error across error terms (see [20] for more detail).

2) Subjective Measures. Texture was significant for Awareness (Table 3). While a Tukey's post hoc analysis indicated that there were no significant differences between the mean Awareness ratings for each texture (Fig. 6b), differences were found using a least significant difference post hoc analysis. This revealed that E, EF, PF, and PEF were amongst the four textures in the homogeneous subset of textures that received the highest scores (Fig. 6b).

Texture was significant for both SA-SWORD 30° and SA-SWORD 60°. Bonferroni post hoc analyses revealed that there were five textures that were in the homogeneous subsets of textures that received the highest scores for SA-SWORD 30°: P, EF, PF, PE, and PEF (Fig. 6c). For SA-SWORD 60°, there were four such textures: EF, PF, PE, and PEF (Fig. 6d).

B. FOV

1) Judgment-based Measures. FOV was significant for A_e and H_e (TABLE III). It was not significant for any of the absolute error terms. Participants overestimated angle judgments for both FOVs, but overestimated them more for the 30° FOV (M = 3.16, SD = 5.97) than the 60° FOV (M = 1.91, SD = 4.299). They underestimated relative height judgments for both FOVs, but overestimated them more for the 60° FOV (M = -108.4, SD = 322.2) than the 30° FOV (M = -54.8, SD = 324.7).

2) Subjective Measures. FOV was not significant for Awareness.

TABLE III
Repeated measures ANOVA results

Dependent Measure	Independent Variable: Texture	FOV	Texture × FOV
A_e	$F_{3.85,63.4} = 1.61^a$ $p = 0.19$	$F_{1,17} = 5.63$ $p = 0.03$*	$F_{2.40,40.76} = 2.68^b$ $p = 0.07$
D_e	$F_{6,102} = 1.73$ $p = 0.12$	$F_{1,17} = 3.30$ $p = 0.09$	$F_{3.22,54.75} = 0.48^c$ $p = 0.71$
H_e	$F_{2.82,63.40} = 1.60^d$ $p = 0.20$	$F_{1,17} = 5.87$ $p = 0.03$*	$F_{6,102} = 1.42$ $p = 0.21$
τ_e	$F_{2.60,63.4} = 1.50^e$ $p = 0.23$	$F_{1,17} = 1.16$ $p = 0.30$	$F_{2.33,39.57} = 0.69^f$ $p = 0.53$
$\|A_e\|$	$F_{3.74,65.87} = 3.21^g$ $p = 0.02$*	$F_{1,17} = 2.03$ $p = 0.17$	$F_{2.18,37.13} = 1.71^h$ $p = 0.19$
$\|D_e\|$	$F_{6,102} = 3.43$ $p < 0.01$*	$F_{1,17} = 0.00$ $p = 0.97$	$F_{6,102} = 1.73$ $p = 0.12$
$\|H_e\|$	$F_{6,102} = 1.39$ $p = 0.22$	$F_{1,17} = 0.03$ $p = 0.87$	$F_{3.72,63.16} = 1.55^i$ $p = 0.20$
$\|\tau_e\|$	$F_{2.15,65.87} = 2.95^j$ $p = 0.06$	$F_{1,17} = 0.86$ $p = 0.37$	$F_{1.96,33.30} = 0.99^k$ $p = 0.38$
Awareness	$F_{6,102} = 2.78$ $p = 0.02$*	$F_{1,17} = 0.03$ $p = 0.87$	$F_{6,102} = 0.63$ $p = 0.71$
SA-SWORD 30°	$F_{3.10,52.66} = 6.86^l$ $p < 0.01$*	NA	NA
SA-SWORD 60°	$F_{3.40,57.88} = 12.62^m$ $p < 0.01$*	NA	NA

[a] W = 0.05, $X^2(20)$ = 44.96, p < 0.01, ε = 0.64; [b] W = 0.02, $X^2(20)$ = 61, p < 0.01, ε = 0.40; [c] W = 0.06, $X^2(20)$ = 40.85, p < 0.01, ε = 0.54; [d] W = 0.02, $X^2(20)$ = 59.26, p < 0.01, ε = 0.47; [e] W = 0.04, $X^2(20)$ = 46.09, p < 0.01, ε = 0.43; [f] W = 0.02, $X^2(20)$ = 56.61, p < 0.01, ε = 0.39; [g] W = 0.10, $X^2(20)$ = 33.48, p = 0.03, ε = 0.62; [h] W = 0.02, $X^2(20)$ = 61.83, p < 0.01, ε = 0.36; [i] W = 0.1, $X^2(20)$ = 34.3, p = 0.03, ε = 0.62; [j] W = 0.01, $X^2(20)$ = 62.54, p < 0.01, ε = 0.36; [k] W = 0, $X^2(20)$ = 88.04, p < 0.01, ε = 0.33; [l] W = 0.04, $X^2(20)$ = 49.07, p < 0.01, ε = 0.52; [m] W = 0.04, $X^2(20)$ = 46.34, p < 0.01, ε = 0.57.

* p < 0.05.

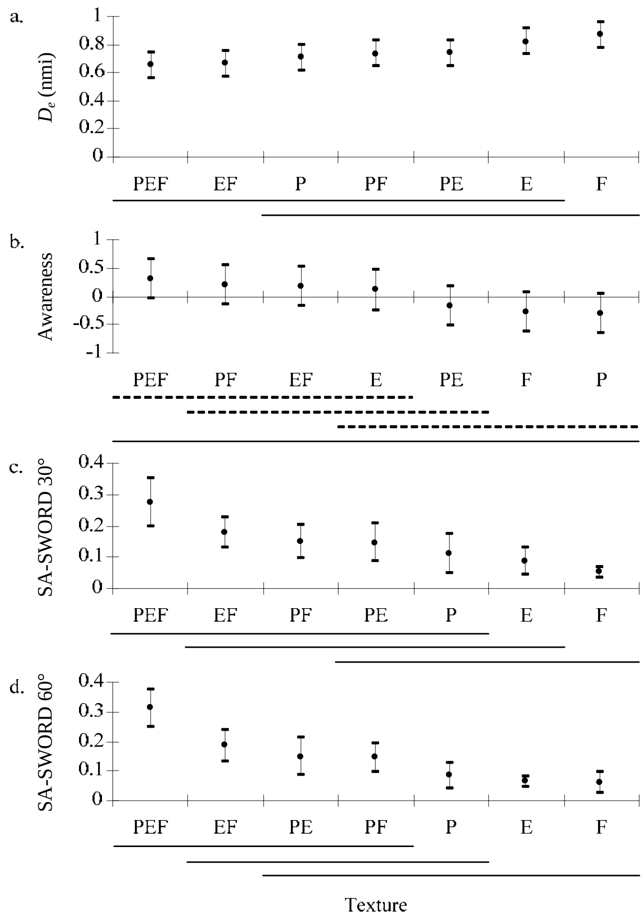

Fig. 6. Post hoc analyses results for the texture main effect. Filled circles indicate means. Bars around means in *a* and *b* represent Tukey's intervals as indicated by a Tukey's HSD with $\alpha = 0.05$. Bars around means in *c* and *d* represent 95% confidence intervals. Solid lines under textures indicate homogeneous subsets (no significant differences between textures) as indicated by a Tukey's HSD (*a* and *b*) or a Bonferroni post hoc (*c* and *d*). Dotted lines under the textures in *b* indicate homogeneous subsets as indicated by a least significant difference post hoc.

C. *Correlations Between Measures*

A two-tailed Pearson product moment correlation analysis was performed in order to determine if there were any significant correlations between the subjective measures and the judgment derived measures (a Bonferroni correction was used to account for multiple comparisons). This revealed that there were no correlations between Awareness and the judgment-based measures. Additionally, there were no significant correlations between SA-SWORD 30° and SA-SWORD 60° and any of the other measures.

IV. Discussion

This effort represents the first study comparing judgment-based measures that probed all three levels of SpA with subjective measures commonly used to evaluate display designs. However, in order to investigate a range of display options in a timely manner, experimental trials were short and non-interactive. In addition, the terrain point was indicated using an artificial object. As such, the use of comparable (judgment and subjective) measures in higher fidelity flight and simulation tests could help validated this experiment's findings. Despite these limitations, however, significant results were found.

The judgment-based measures introduced by this work have proven useful for evaluating SVS displays' ability to convey SpA. Because the judgments were grounded in the theoretical foundations of SpA, three textures (EF, PF, and PEF) were found to best support both level 2 SpA (the least $|A_e|$, $|D_e|$, and $|H_e|$) and level 3 SpA (the least $|\tau_e|$). Additionally, directional and absolute error terms provide information about biases in the perception of SpA information between FOVs (see [20] and [21] for a more in-depth interpretation of the results). However, while significant differences were found for subjective measures, the results provide little indication that they are actually measuring SpA.

A. *Awareness*

A less robust post-hoc analysis (least significant difference) showed that three textures in the set that produced the highest Awareness ratings (EF, PF, and PEF) were among the textures that produced the least $|A_e|$, $|D_e|$, and $|\tau_e|$ [20]. Given that the E, EF, and PF textures were among the set of textures that produced the highest Awareness ratings, these results are consistent with results obtained by Glaab and Hughes [10] who found that participants tended to give PF, E, and EF textures higher terrain awareness ratings than the F texture. However, Awareness was not correlated with any of the judgment-based measures, suggesting that judgment accuracy was not what they were using to assess their awareness.

The inability of Awareness to produce significant results for the ANOVA or the corresponding post hoc analyses may be due to the nature of the task. Had participants actually been flying the aircraft, their attentional resources would have been in higher demand and they might have been able to better assess texture's ability to convey SpA. Further, the displays used in this experiment utilized a reduced set of instrumentation than is often employed in SVS displays.

This illustrates a limitation of Awareness. While the judgment-based measures were able to produce significant results and convey significant differences with the given procedure, Awareness was not. Thus, since similar subjective measures have produced significant results for flight and simulation tests, Awareness may only be applicable in more realistic flight and simulation tests.

B. *SA-SWORD*

The ANOVA results for SA-SWORD did produce significant results, indicating that participants thought some textures enhanced SpA more than others. The textures that produced the highest SA-SWORD scores were somewhat consistent with the textures shown to produce the least amount of judgment error. Both EF and PEF were among the group of statistically similar textures that produced the highest scores for both SA-SWORD 30° and SA-SWORD 60° as well as being in the group of statistically similar textures that produced the least judgment error in both the main and interaction effects [20]. However, there were no correlations between SA-SWORD scores and other dependent measures. Thus, it is not clear if SA-SWORD is measuring a component of SpA.

C. Conclusions

It is important to find theoretically-based performance measures of awareness so that awareness can be comprehensively evaluated. This work introduced judgment-based measures of SpA designed to probe all three levels of SpA knowledge and compared them with subjective measures commonly used to evaluate SVS. While similarities were found between the results of the subjective and judgment-based measures, the lack of correlation between subjective scores and judgment error makes it difficult to determine what dimensions of SpA the subjective measures are assessing. Since subjective measures are being used to evaluate safety critical systems such as SVS, it is important that designers understand their value when interpreting data.

While this research focused on SpA with SVS displays, there are other potential applications. Subjective measures have been used to evaluate awareness for air combat systems [26], tunnel in the sky displays [9], human-robot interaction [27], and many other areas. Thus, given that the subjective awareness measures were unable to show correlations with judgments related to SpA knowledge, awareness measures in other domains may exhibit similar behavior.

Acknowledgement

The authors would like to thank: R. Comstock, T. Arthur, D. Burdette, S. Conway, S. Guerlain, R. Johns, L. Kramer, K. Latorella, N. O'Connor, H. Press, V. Plyler, L. Prinzel, P. Schutte, J. Sweeters, M. Takallu, S. Williams, and the anonymous reviewers.

References

[1] C. D. Wickens, "Situation Awareness and Workload in Aviation," *Current Directions in Psychological Science,* vol. 11, pp. 128-133, 2002.

[2] C. D. Wickens, "Spatial awareness biases," NASA Ames Research Center, Moffett Field, CA, Tech. Rep. AHFD-02-6/NASA-02-4, 2002.

[3] R. M. Taylor, "Situational awareness rating technique (SART): The development of a tool for aircrew systems design," in *AGARD, Situational Awareness in Aerospace Operations 17,* 1990, pp. 3/1-3/17.

[4] M. A. Vidulich and E. R. Hughes, "Testing a subjective metric of situation awareness," in *Proceedings of the Human Factors Society 35th Annual Meeting,* Sep 2-6, 1991, pp. 1307-1311.

[5] M. A. Vidulich, G. F. Ward and J. Schueren, "Using the subjective workload dominance (SWORD) technique for projective workload assessment," *Human Factors,* vol. 33, pp. 677-691, 1991.

[6] R. E. Bailey, L. J. Kramer and L. J. Prinzel III, "Crew and display concepts evaluation for synthetic/enhanced vision systems," *Proceedings of SPIE*. Vol. SPIE-6226, pp. 154-171, 2006.

[7] Boeing Commercial Airplanes, "Statistical summary of commercial jet airplane accidents: World wide operations 1959 – 2005," 2006. Available: http://www.boeing.com/news/techissues/pdf/statsum.pdf

[8] Flight Safety Foundation, "Killers in aviation: FSF task force presents facts about approach-and-landing and controlled-flight-into-terrain accidents," *Flight Safety Digest,* vol. 17, 1999.

[9] M. A. Takallu, D. T. Wong, A. P. Bartolone, M. F. Hughes and L. J. Glaab, "Interaction between various terrain portrayals and guidance/tunnel symbology concepts for general aviation synthetic vision displays during a low en-route scenario," in *Proc. of the 23rd Digital Avionics Systems Conference*, 24-28 Oct., 2004, pp. 4-3.

[10] L. J. Glaab and M. F. Hughes, "Terrain portrayal for head-down displays flight test," in *Proc. of the 22nd Digital Avionics Systems Conference,* 12-16 Oct., 2003.

[11] J. J. Arthur, Prinzel, L. J., III, L. J. Kramer, R. V. Parrish and R. E. Bailey, "*Flight simulator evaluation of synthetic vision display concepts to prevent controlled flight into terrain (CFIT),*" NASA, Hampton, VA, Tech. Rep. NASA/TP-2004-213008, 2004.

[12] S. J. Selcon and R. M. Taylor, "Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design," in *Proc. of AGARD, Situational Awareness in Aerospace Operations 8,* 1990, pp. 5/1-5/8.

[13] M. R. Endsley, S. J. Selcon, T. D. Hardiman and D. G. Croft, "A comparative analysis of SAGAT and SART for evaluations of situation awareness," in *Proc. of the Human Factors and Ergonomics Society 42nd Annual Meeting,* Chicago, IL, 5-9 Oct., 1998.

[14] M. R. Endsley, “Measurement of Situation Awareness in dynamic systems. *Human Factors,* vol. 37, pp. 65-84, 1995.

[15] Y. Yeh and L. D. Silverstein, "Spatial judgments with monoscopic and stereoscopic presentation of perspective displays," *Human Factors,* vol. 34, pp. 583-600, 1992.

[16] B. D. Adelstein and S. R. Ellis, "Rotation and direction judgment from visual images head-slaved in two and three degrees-of-freedom," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*. Vol. 30, no. 2, pp. 165-173., 2000.

[17] M. W. McGreevy and S. R. Ellis, "The effect of perspective geometry on judged direction in spatial information instruments," *Human Factors*, vol. 28, pp. 439-456, Aug. 1986.

[18] M. Wells, R. Osgood and M. Venturino, "Using target replacement performance to measure spatial awareness in a helmet-mounted simulator," in *Proc. of the Human Factors Society 32nd Annual Meeting,* Anaheim, CA, 24-28 Oct., 1988.

[19] M. Fracker, "Attention gradients in situation awareness," in *Situational Awareness in Aerospace Operations AGARD-CP-478,* 1990, pp. 6/1-6/10.

[20] M. L. Bolton and E.J. Bass, and J. R. Comstock, “Spatial awareness in Synthetic Vision Systems: using spatial and temporal judgments to evaluate texture and field of view,” *Human Factors*, to be published.

[21] M. L. Bolton and E.J. Bass, “Using relative position and temporal judgments to identify biases in spatial awareness for synthetic vision systems,” *International Journal of Aviation Psychology*, to be published.

[22] M. L. Bolton, E.J. Bass, and J. R. Comstock, “A toolset to support the development of spatial and temporal judgment experiments for synthetic vision systems,” in *Proc. of the 2006 IEEE Systems and Information Engineering Design Symposium*, Charlottesville, VA, April 28, 2006, pp. 58-63.

[23] J. R. Comstock Jr., L. J. Glaab, L. J. Prinzel and D. M. Elliott, "Can effective synthetic vision system displays be implemented on limited size display spaces?" in *Proc. of the 11th International Symposium on Aviation Psychology*, Columbus, OH, 5-8 Mar., 2001.

[24] N. Brace, R. Kemp and R. Snelgar, *SPSS for Psychologists : A Guide to Data Analysis using SPSS for Windows*, 2nd ed. Mahwah, NJ: L. Erlbaum Associates, 2003.

[25] J. Stevens, *Applied Multivariate Statistics for the Social Sciences*. Mahwah, N.J.: Lawrence Erlbaum Associates, 2002.

[26] M. R. Endsley, "Supporting situation awareness in aviation systems," in *Proc. of the IEEE International Conference on Systems, Man, and Cybernetics,* Orlando, FL,12-15 Oct., 1997.

[27] R. Parasuraman, S. Galster, C. Miller, "Human control of multiple robots in the RoboFlag simulation environment," in *Proc. of the IEEE International Conference on Systems, Man and Cybernetics*, 5-8 Oct., 2003, pp. 3232-3237.