

Using Videos Derived from Simulations to Support the Analysis of Spatial Awareness in Synthetic Vision Displays

Matthew L. Bolton, *Member, IEEE*, Ellen J. Bass, *Senior Member, IEEE*, James R. Comstock, Jr.

Abstract—The evaluation of human-centered systems can be performed using a variety of different methodologies. This paper describes a human-centered systems' evaluation methodology where participants watch 5-second non-interactive videos of a system in operation before supplying judgments and subjective measures based on the information conveyed in the videos. This methodology was used to evaluate the ability of different textures and fields of view to convey spatial awareness in synthetic vision systems (SVS) displays. It produced significant results for both judgment based and subjective measures. This method is compared to other methods commonly used to evaluate SVS displays based on cost, the amount of experimental time required, experimental flexibility, and the type of data provided.

I. INTRODUCTION

A human-centered system evaluation is concerned with three issues: compatibility (the ability of the system to present information and to expect control inputs within the limitations of human capabilities), understandability (the ability of the system to meaningfully communicate information), and effectiveness (the ability of the system to improve performance or accomplish a previously unrealized goal) [1]. Different categories of evaluation can be used to evaluate these criteria: paper evaluation, part-task simulator evaluation, full-scope simulator evaluation, and in-use evaluation (Fig. 1) [1].

Each of these evaluation techniques is used to address different goals of the systems evaluation process. Paper evaluations involve showing participants mockups and prototypes of the system being evaluated (on paper or electronically) for the purpose of identifying compatibility issues. Part-task simulations coarsely approximate the system being evaluated for the purpose of assessing specific user understandability issues. Full-scope simulations strive to simulate the actual system as accurately as possible for the purpose of evaluating the effectiveness of the system to accomplish its design goals. Thus they are primarily

concerned with determining system effectiveness. In-use evaluation involves using the system in its actual operating conditions and is also primarily concerned with effectiveness.

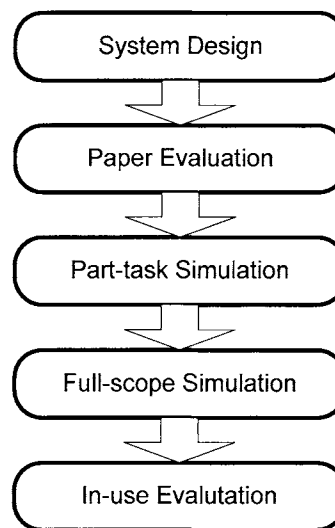


Fig. 1. Multiple methods of system evaluation, adapted from [1].

There are cost and schedule tradeoffs between each of these evaluation categories. A paper prototype or mockup will require fewer resources to develop than will a part-task simulator. Likewise, a part-task simulator that only simulates a portion of the system will require fewer resources to develop than a full-scope simulator that simulates the entire system. Finally, in-use evaluations require a fully working system and are thus more expensive than the other evaluation techniques, especially when the costs associated with redesigns are considered. Compatibility issues can be investigated using less costly paper evaluation methods. Understandability issues can be addressed using more expensive part-task simulations. Finally, expensive full-scope simulations and in-use evaluations are used to evaluate system effectiveness.

There are also tradeoffs in the flexibility of the experiments that can be supported by the different evaluation techniques. Simulation based evaluation techniques support more versatile experimental conditions than in-use ones. In addition, it may not be feasible or cost efficient to run an actual system at extreme operating conditions. Additionally, if the system being evaluated operates in a hazardous environment, then certain experimental conditions could result in injury or death. Simulation based evaluations allow extreme circumstances to be tested without the risk of such adverse consequences.

Manuscript received July 1, 2006. 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 Systems Engineering at the University Of Virginia, Charlottesville, VA 22904 USA (email: ejb4n@virginia.edu).

J. R. Comstock, Jr., is a Research Technologist in the Crew Systems and Aviation Operations Branch at the NASA Langley Research Center, Hampton, VA 23681 USA (e-mail: J.R.Comstock@larc.nasa.gov).

This paper describes an evaluation technique developed for the purpose of assessing the ability of Synthetic Vision Systems (SVS) displays to convey spatial awareness using a series of non-interactive, video-based simulations. This paper first introduces SVS and describes the procedures that have been used to evaluate them. It then provides the motivation for the development of the new evaluation technique and discusses its usefulness as indicated by the results of a human subjects experiment. It then puts this technique in context of the other techniques used to evaluate SVS, explaining how it can be used as an important tool in SVS development.

II. SYNTHETIC VISION SYSTEMS

SVS are cockpit display technologies currently being developed by NASA and industry to prevent incidents of Controlled Flight Into Terrain (CFIT), a condition where a normally functioning aircraft is inadvertently flown into the ground or other terrain feature [2]. SVS combats CFIT by using GPS data and onboard terrain databases to create a synthetic, clear-day, perspective view of the world surrounding the aircraft regardless of the visibility conditions.

Evaluation techniques from all four categories (paper evaluation, part-task simulation evaluation, full-scope simulation evaluation, and in-use evaluation) have been used to assess SVS displays. Paper evaluations have often been used to evaluate new ideas before testing them with other evaluation methods. Such analyses usually involve showing pilots working display prototypes on simulated cockpit displays and asking them to comment on their usefulness.

Part-task simulations have also been used. Schnell and Lemos conducted several experiments in which participants viewed either still shots or video of actual terrain and were asked to match them with SVS displays [3]. Scores were assigned based on the number of correct identifications. These procedures were used to evaluate different terrain resolutions, shadings, and texturing schemes. Experiments that utilized static images of actual terrain were conducted on desktop computers and did not produce any significant results. Experiments that used videos of actual terrain were conducted in flight simulators in which the videos were displayed in an out the window view and the SVS displays being matched were displayed on cockpit panels. These studies did find significant main effects for texture, shading, and terrain resolution.

Full-scope simulations have been used extensively [4] [5] [6]. In SVS full-scope simulations, pilots fly approach and departure paths in a flight simulator around terrain challenged airports. The dependent variables collected in these experiments typically include cross track error (the mean squared error of an aircraft from the optimally defined flight path) as well as a variety of subjective measures based on techniques including Situation Awareness Rating Technique (SART) [7], Situation Awareness – Subjective Workload Dominance (SA-SWORD) [8], and other Likert scale based

questionnaires. Pilots typically fill out questionnaires between experimental trials or after all trials have been completed. Because simulation experiments are conducted in a laboratory, more experimental possibilities are available than for experiments conducted in flight. For example, because simulations can be paused at controlled times, researchers are free to employ measurement techniques such as the Situation Awareness Global Assessment Technique (SAGAT) [9], a situation awareness measure that requires pausing simulations so that pilots can answer a battery of questions. Researchers are also afforded the ability to simulate conditions that may be too dangerous to attempt in actual aircraft. For example, Arthur, Prinzel, Kramer, Parrish, and Bailey used a full cockpit simulation to determine if pilots could detect a potential future CFIT incident while flying an approach [4].

SVS displays have also been assessed with in-use evaluations [5] [10] [11]. Such evaluations usually take the form of flight tests. As with full-scope simulations, flight tests typically measure cross track error while pilots fly approaches and departures using SVS displays in addition to having pilots fill out post-flight questionnaires to measure SART, SA-SWORD, and other subjective metrics. Flight tests allow data to be collected in an extremely operationally realistic environment, and are a necessary for a new technology to be introduced into a cockpit. However, they have limitations. Because tests are conducted in the natural environment, there is less experimental control than there would be in a laboratory environment [12]. Additionally, because flight tests are expensive and time consuming, most research programs can only afford a limited number of them [12]. Thus, researchers may not have the opportunity to test all the desired conditions [12]. Finally, because of safety concerns, researchers are prevented from having pilots fly scenarios that may put them into danger.

III. PROBLEM

The goal of this research was to evaluate the ability of seven textures and two fields of view (FOVs) to convey spatial awareness. Spatial awareness was defined as the extent to which a pilot noticed objects in the surrounding environment (Level 1), the pilot's understanding of where these objects were with respect to ownship (Level 2), and the pilot's understanding of where these objects would be relative to ownship in the future (Level 3) [13].

Four judgments measured spatial awareness with respect to a point on the synthetic terrain of an SVS display. Relative distance, angle, and height judgments evaluated how well a participant was able to assess the spatial location of the terrain point. A time to fly abeam judgment (how long it would take the airplane to fly to the point of closest approach for the terrain point) was used to assess a participant's understanding of the point's relative temporal location.

To prevent known spatial biases from affecting the results (see [13]), the experimenters wished to control the relative

position of the terrain point (scenario geometry) (Fig. 2). This was done by parameterizing the point's relative angle, distance, and height into two levels each. Angles could be large or small, distances could be near or far, and heights could be above or below the aircraft. Thus, in order to run a full factorial experiment where there were two distance levels, two angle levels, two height levels, two FOV levels, and seven texture levels, 112 ($2 \cdot 2 \cdot 2 \cdot 2 \cdot 7 = 112$) trials were required for each participant. Additionally, in order to familiarize participants with the experimental task, and to introduce each texture and FOV, there were 72 training trials. Thus there were a total of 184 trials for each participant.

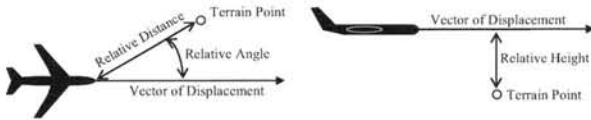


Fig. 2. Scenario geometry parameters

Eighteen participants were required in order to achieve the desired error in judgment means and to maintain balance between the experiment's between subject factors (see [14] and [15] for more details).

Thus, in order to best utilize the time used to run participants, the experiment needed to meet the following requirements: 1) Trials must be short so that participants could complete all trials in a single session; 2) The experimental apparatus must support rapid transition between trials; 3) Multiple participants must be run in parallel in order to reduce the amount of calendar time used to run the experiment.

Given these requirements and the nature of the data being collected, none of the SVS evaluation procedures that have been discussed were appropriate for this experiment. The paper evaluation methods and part-task simulations used by Schnell and Lemos would not facilitate the necessary data collection, and full-scope simulation tests and flight tests would not allow for rapid turnaround between trials or let multiple participants be run in parallel.

IV. APPROACH

In order to meet the experimental requirements, the experimenters developed a part-task simulation evaluation methodology. A part-task simulation evaluation seemed appropriate given that the researchers were only concerned with SVS displays' abilities to convey spatial awareness. Thus, the experiment was only concerned with evaluating understandability. The developed methodology had all of the following properties: 1) The experiment would be conducted on desktop computers which would allow multiple participants to be run in parallel; 2) Simulations of the SVS display would be stored as videos to avoid having to reconfigure and restart the SVS software while testing; 3) Videos would only run for five seconds in order to give participants enough time to identify the location of the terrain point, but not enough to use grid patterns in the textures to

measure the actual distance; 4) In order to facilitate rapid transitions between trials, custom software would display the simulation videos, collect user inputs, and transition between trials.

A. Participants

Eighteen general aviation pilots participated in the study. They had less than 400 hours of flight experience ($\mu = 157$, $\sigma = 75$). They were familiar with the out the window view of a cockpit, but not with SVS displays.

B. Apparatus

Experiments were run on desktop computers using software developed for this study [16]. These computers served to display each simulation and collect participant judgments. SVS displays used during simulation were 9.25 in. by 8 in. and employed the symbology depicted in Fig. 3. In simulations, 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 were displayed as 5 second, 836 pixel x 728 pixel, 30 frames per second, Windows Media Video (WMV) files. Each of the workstations ran the custom software which played the WMV files and collected participant responses.

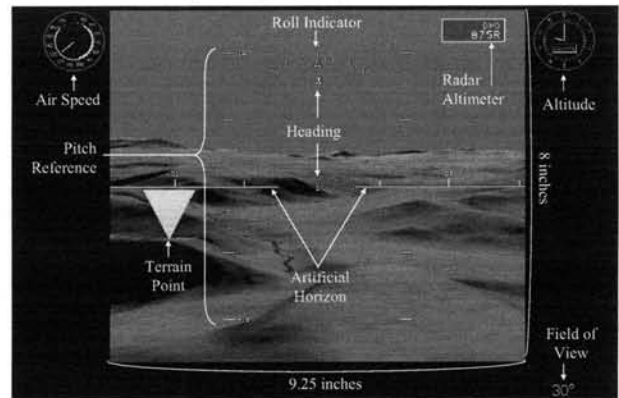


Fig. 3. A synthetic vision display used in the experiment (labels added).

C. Independent Variables

There were five within subject variables: one for texture, one for field of view, and three for scenario geometry (relative distance, angle, and height of the terrain point) (Fig. 2). There were seven textures: three basic textures (fishnet, photo, and elevation) and four derivative textures (elevation fishnet, photo fishnet, photo elevation, and photo elevation fishnet) (Fig. 4). There were two FOVs: 30° and 60°. A justification for why these particular variable levels were chosen can be found in [14] and [15].

The location of the terrain point varied through changes to the scenario geometry variables. Each had two levels.

There were two between subject variables: FOV order and texture order. FOV order had two levels: 30° FOV first or 60° FOV first. Either a participant saw the 30° FOV trials first or the 60° FOV trials first.

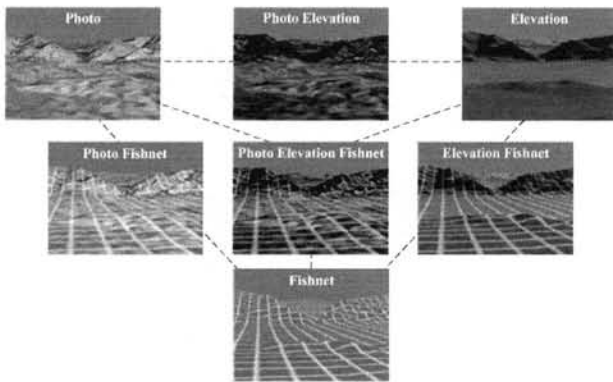


Fig. 4. The terrain textures evaluated in the experiment.

Textures always appeared before their derivatives. Each participant saw two 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 position.

D. Dependent Variables

There were eight dependent variables calculated from the four judgment values (relative angle ($^{\circ}$), relative distance (nmi), relative height (ft), and abeam time (s)). There were two dependent variables associated with each judgment: one for directional error and one for absolute error. Directional error was positive when a participant overestimated a judgment and negative when he underestimated it. Absolute error was calculated as the absolute value of the corresponding directional error.

Participants also provided subjective measures using Likert scales at various intervals in the experiment. There are four subjective measures that will be discussed in this paper: Demand, Awareness, Clutter, and SA-SWORD. The Demand measure asked participants to assess the demand placed on their attentional resources while watching the simulations. The Awareness measure asked participants to assess their ability to determine where their aircraft was with respect to the terrain while watching the simulations. The Clutter measure asked participants to assess the amount of clutter on the SVS displays. The Awareness measure was similar to the terrain awareness question used in [10] and [11]. The Demand and Awareness questions are directly comparable to the demand and awareness dimensions of a 3-D SART score.

SA-SWORD allows participants to make pair-wise comparisons between displays on a nine point scale about the relative amount of SA or spatial awareness provided by each [8]. For seven displays (corresponding to the seven textures),

SA-SWORD requires that $\binom{7}{2} = 21$ comparisons be made.

Values from each comparison are then used to calculate scores for each of the textures [8].

E. Experimental Design

The experiment employed a repeated measure design with eighteen participants. Three participants were randomly

assigned to each of the six combinations of the between subject variables.

All participants experienced 184 counterbalanced trials (112 experimental trials and 72 training trials). Trials were grouped together based on FOV and by texture within each FOV.

F. Procedure

Participants started by watching a presentation about the experiment. Each participant was then randomly assigned to a workstation and experimental condition. The software on each workstation administered the experiment. The beginning of a FOV block was introduced with eight training trials with feedback. All subsequent texture blocks were introduced with four training trials with feedback. After each trial participants provided judgments for the four spatial awareness measures (relative angle, relative distance, relative height, and abeam time) using the interface shown in Fig. 5.

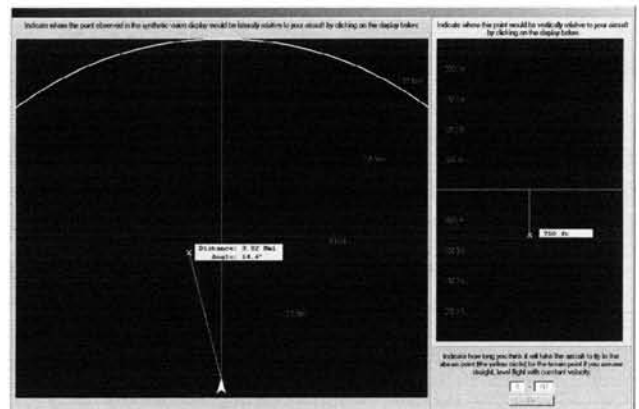


Fig. 5. The interface used to collect the spatial awareness judgments following each trial.

Demand, Awareness, and Clutter values were collected after each texture block. SA-SWORD pair-wise comparisons were collected after each FOV block.

G. Hypothesis

Because each of the three base textures (fishnet, elevation, and photo) convey different spatial information (see [14] and [15]), it was hypothesized that the highest level of spatial awareness would be achieved by combining all three texture types (the photo elevation fishnet texture).

V. RESULTS

The data collected in this study were analyzed using a repeated measures analysis of variance which looked for significant ($p \leq 0.05$) main and interaction effects and trends ($p \leq 0.10$) in each of the dependent variables.

There were significant effects and trends for a variety of main and interaction effects for the directional and absolute error terms (TABLE I). There were also significant effects and trends for the subjective measures: Texture, FOV * FOV order, and Texture Order * FOV Order were trends for awareness; Texture * FOV was a trend for Clutter; and Texture was significant for SA-SWORD scores collected

TABLE I
SIGNIFICANCE AND TRENDS FOR ERROR DEPENDENT VARIABLES

Independent Variable	Directional Error				Absolute Error			
	D	A	H	T	D	A	H	T
Texture					X	X		*
Angle			X					
Distance	X	X	X	X	*	X	X	X
Height	*	X		*	*	X		X
FOV	X	X	X					
Texture * Angle	X		X	X		X	X	X
Texture * Distance	X		X	X	X			X
Texture * Height			X	*	X		X	*
Texture * FOV		X						
Angle * Distance			X			X		
Angle * Height	X	*		X	X			
Angle * FOV	X	*	X	X			X	
Distance * Height					X	*		*
Distance * FOV			X		X		*	*
Height * FOV								
Texture Order								
FOV Order	X			X			X	
Texture * Texture Order								
Texture * FOV Order	X				X			
Angle * Texture Order								
Angle * FOV Order								
Distance * Texture Order		*			X			*
Distance * FOV Order					X			X
Height * Texture Order			*					
Height * FOV Order		X						
FOV * Texture Order								
FOV * FOV Order	X							
Texture Order * FOV Order		*	X		*		*	

D, A, H, & T stand for Distance, Angle, Height, and Time respectively.

X Indicates significance ($p \leq 0.05$).

* Indicates a trend ($p \leq 0.10$).

with both a 30° FOV and a 60° FOV. There were no significant effects or trends for Demand.

Additionally, an examination of the results using the appropriate post hoc analyses (least significant difference for variables with two levels, Tukey's for variables with more than two levels that did not violate sphericity, and Bonferroni for variables with more than two levels that did violate sphericity [17]) revealed that spatial awareness was best supported by two textures: elevation fishnet and photo elevation fishnet. This was indicated by the fact that, for each absolute error dependent variable for which texture or a texture * scenario geometry interaction was significant or a trend, elevation fishnet and photo elevation fishnet were the only textures that were among the group of textures that produced the least error for all of the different types of errors. This is compatible with the hypothesis of this research given that the elevation fishnet and photo elevation fishnet textures are combinations of base textures, and that photo elevation fishnet is the combination of all three base textures.

The results of the Awareness rating are also consistent with those found in the SVS literature. While a Tukey's post hoc analysis indicated that there were no significant differences between the mean awareness ratings for each texture, differences were found using least significant difference. Participants tended to give lower scores to the fishnet, photo, and photo elevation textures than for the elevation, elevation fishnet, photo fishnet, and photo elevation fishnet textures.

This is consistent with flight test data collected by Glaab and Hughes who found that participants tended to give photo, photo fishnet, elevation, and elevation fishnet textures higher terrain awareness ratings than the fishnet texture [10] (Glaab and Hughes did not test a photo elevation or photo elevation fishnet texture).

VI. DISCUSSION

The results of this experiment reveal several advantages of this evaluation procedure. Firstly, given the large number of significant effects (TABLE I), it is clear that the use of short, non-interactive video simulation trials can produce significant results. The validity of these results is helped by the fact that they support the hypothesis being evaluated in the experiment. Since there are other SVS display parameters that could potentially affect pilot spatial awareness (display size, other FOVs, terrain resolution, atmospheric perspective, etc.), this procedure could be used to evaluate them.

However, given that no previous SVS experiments have used these judgments, a validation procedure should be conducted in which spatial awareness judgments are collected as part of simulation or flight tests. If such tests produce comparable results to those found in this experiment, then experiments could safely test other display parameters using this new methodology.

The data also show that short non-interactive video simulations are capable of eliciting significant differences in subjective measures with results consistent with those found using flight tests (in-use evaluation). Thus, similar procedures may prove useful for quickly gathering subjective measure data for new display concepts.

However, the procedure did not result in significant effects or trends for the Demand subjective measure, and only a single interaction term indicated a trend for the Clutter measure. This is likely due to the simplicity of the task participants performed. Had participants actually been flying the aircraft, their attentional resources would have been in higher demand and they might have been able to assess differences in a texture's demand on attentional resources more acutely. They may also have found some textures or FOVs to be more cluttered than others.

The fact that there were significant main effects/trends for Awareness and SA-SWORD, and not for Demand and Clutter, illustrates a potential limitation of this procedure. The Awareness and SA-SWORD scores both attempted to measure spatial awareness, the same as the judgment values. Since Clutter and Demand were not directly related to the judgment task, it appears that subjective measures may only provide useful information when they are trying to measure values directly related to the judgment task.

This new procedure offers advantages in terms of experimental time. Each participant took approximately four hours to complete the entire experimental procedure. For 18 participants, this equated to 72 hours of experimental time. However, participants were run in parallel.

The use of workstations can also be an advantage. Since workstations are generally cheaper and more readily available than flight simulators and experiment ready aircraft, experiments using desktop simulations can be conducted in more locations and less expensively than simulator and flight tests.

While the procedure discussed in this paper used five-second videos, this time period should be investigated. Given that this particular experiment had not been conducted before, the use of five second videos was an informed guess (based on a consultation with multiple NASA researchers). Even though the experiment produced significant results, future work should go into investigating how long video should be shown to participants to best mimic pilot instrumentation utilization in cockpits.

Because this evaluation methodology is a part-task simulation, the type of data that can be collected using it is limited. There are several reasons for this. Because the experiment is conducted on computer workstations in a laboratory environment, there is a lack of realism. Significant results found under such conditions need to be validated in a more realistic environment before being implemented in the actual system. Similarly, the nature of the task is somewhat artificial given that pilots would not actually be making explicit spatial judgments while flying. Thus, if a display concept is found to have advantages over others using this procedure, it would need to be tested in other contexts to ensure that other important metrics were not being compromised. Third, the brevity of the scenarios limits the amount of data that can be collected. Thus, while this procedure is more efficient than flight and simulation tests in terms of experimental time, it may be less effective in terms of the range of data collected.

A researcher should select an experimental procedure based on the research goals and the hypotheses being tested. Additionally, before a new design can be introduced into its operating environment, it must be evaluated in a variety of capacities. In this context, the procedure discussed in this paper would not replace paper evolutions, other part-task simulations, full-scope simulations, or flight tests, but supplement them over the course of a human centered system evaluation of SVS.

Given its short and efficient nature, the fact that it does not depend on integration into a flight deck or full-scope simulation, and its capacity for producing significant results for both judgment values and subjective measures, the experimental procedure discussed in this paper would be appropriate as a cursory analysis of a new display concept's ability to convey spatial awareness. If a display proved to be beneficial in this context, it should then be evaluated in a full-scope simulation environment or flight test, thus ensuring that both understandability and effectiveness goals are met.

ACKNOWLEDGEMENTS

This work was supported in part by the NIA and NASA.

The authors would like to thank those making suggestions for this work including: Trey Arthur, Daniel Burdette, Sheila Conway, Stephanie Guerlain, Regina Johns, Lynda Kramer, Kara Latorella, Neil O'Connor, Hayes Press, Valerie Plyler, Lawrence Prinzel, Paul Schutte, Jason Sweeters, Mohammad Takallu, and Steve Williams. The authors would like to thank Mohammad Takallu, Brent Weathered, and the three anonymous SMC reviewers for their feedback on earlier versions of this paper.

REFERENCES

- [1] W. B. Rouse, *Design for Success: A Human-Centered Approach to Designing Successful Products and Systems*, New York: John Wiley and Sons, 1991.
- [2] R. V. Parish, D. G. Baize, and M. S. Lewis, "Synthetic Vision," in *The Avionics Handbook*, C. R. Spitzer, Ed. Boca Raton : CRC Press, 2001, pp. 16-1-16-8.
- [3] T. Schnell and K. Lemos, "Terrain sampling density and texture requirements for synthetic vision systems," Rockwell Collins Final Report, Cedar Rapids, IA, 2002.
- [4] J. J. Arthur, L. J. Prinzel 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)," Tech. Rep. NASA/TP-2004-213008, Hampton, VA, 2004.
- [5] J. J. Arthur, L. J. Prinzel III, L. J. Kramer, R. E. Bailey, and R. V. Parrish, "CFIT prevention using synthetic vision," in *2003 Proc. SPIE, Enhanced and Synthetic Vision Conf.*, pp. 146-157.
- [6] J. R. Comstock, 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. 11th International Symposium on Aviation Psychology*. Dayton, OH, 2002.
- [7] R. M. Taylor, "Situation Awareness Rating Technique (SART): The development of a tool for aircrew systems design," *Situational Awareness in Aerospace Operations*, Neuilly-sur-Seine, France, 1990, pp. 3/1-3/17.
- [8] M. A. Vidulich, "Testing a subjective metric of situation awareness," in *Proc. Human Factors and Ergonomics Society 35th Annual Meeting*. San Francisco, CA, 1991.
- [9] M. R. Endsley "Situation awareness global assessment technique (SAGAT)," in *Proc. IEEE 1988 NAECON*, pp. 789-795.
- [10] L. J. Glaab, and M. F. Hughes, "Terrain portrayal for head-down displays flight test," in *Proc. 22nd Digital Avionics Systems Conference*. Indianapolis, IN, 2003.
- [11] R. E. Bailey, R. V. Parrish, J. J. Arthur, and R. M. Norman, "Flight test evaluation of tactical synthetic vision display concepts in a terrain-challenged operating environment," in *Proc. 2002 SPIE Enhanced and Synthetic Vision*, Bellingham, WA, pp. 178-189.
- [12] M. A. Bonner and G. F. Wilson, "Heart rate measures of flight test and evaluation," *The International Journal of Aviation Psychology*, 12(1), 63-77.
- [13] C. D. Wickens, "Spatial awareness biases," Tech. Rep. ARL-02-6/NASA-02-4, Savoy, IL, 2002.
- [14] M. L. Bolton, E. J. Bass, and J. R. Comstock "Using relative position and temporal judgments to assess the effects of texture and field of view on spatial awareness in synthetic vision systems displays," *The 50th Annual Meeting of the Human Factors and Ergonomics Society*, San Francisco, CA, October 16-20, 2006.
- [15] M. L. Bolton, "An investigation of the effects of texture and field of view on spatial awareness in synthetic vision head-down displays," M.S. thesis, Dept. Systems and Information Eng., Univ. of VA, Charlottesville, VA, 2006.
- [16] 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 *2006 IEEE Systems and Information Engineering Design Symposium*, Charlottesville, VA, pp. 58-63.
- [17] J. P. Stevens, *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2002.