SATECH-00-11

Evaluation of Situation Awareness in Flight Operations Employing Synthetic Vision Systems

Mica R. Endsley, Ph.D.
June 2000

Prepared for:

Dr. Lance Prinzel
NASA Langley Research Center
Hampton, VA 23681-2199
Under P.O. L-11403

SA technologies
4731 East Forest Peak
Marietta, GA 30066

1.0 Critical SA issues for Synthetic Vision Systems

Synthetic Vision Systems (SVS) are built on database derived information that is used to aid the pilot in visualizing the aircraft situation relative to information outside the cockpit (Koczo, Klein, Both, & Lamb, 1998; Regal & Whittington, 1994). It may incorporate terrain, obstacles, cultural features, weather and/or traffic information. As an additional and intuitive source of information, the SVS concept may aid pilot situation awareness (SA) in many ways. Due to the limitations of display technology, however, it may also lead to certain SA difficulties, particularly if it is used in place of out-of-the-window viewing under no or low visibility conditions.

While it can be easily argued that the information provided by the SVS is better than the very limited information available today under such conditions, the desire to increase aircraft throughput under these conditions by use of the SVS (creating significant efficiency gains), demands that any potential SA problems be detected in the evaluation process and corrected for prior to its implementation in flight operations

In the aviation domain, maintaining a high level of situation awareness is one of the most critical and challenging features of the flight crew's job. Situation awareness can be thought of as an internalized mental model of the current state of the flight environment. This integrated picture forms the central organizing feature from which all decision making and action takes place. A vast portion of the flight crew's job is involved in developing SA and keeping it up to date in a rapidly changing environment

Situation awareness is defined as "The perception of the elements in the environment within a volume of space and time, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1988a). Various types of situation awareness are needed within the aviation context, including geographical SA, spatial/temporal SA, systems SA, environmental SA and tactical SA (Endsley, 1999). These are described in Table 1. As shown in this table, there is a lot of information that pilots need to keep up with, integrate together to form a coherent picture of what that information means within the context of their goals (comprehension) and project into the future in order to make good decisions.

A number of situation awareness factors may be impacted by SVS (either positively or negatively). These will each be discussed separately. Subsequently, methods for measuring the impact of SVS on pilot SA in simulation and aircraft flight studies will be presented, in order that any possible positive or negative impacts of SVS can be determined prior to implementation.

Table 1 Elements of SA in Aviation (Endsley, 1996)	
Geographical SA Location of own aircraft, other aircraft, terrain features, airports, cities, waypoints and navigation fixes; position relative to designated features; runway & taxiway assignments; path to desired locations; climb/descent points.	**Spatial/Temporal SA** Attitude, altitude, heading, velocity, vertical velocity, G's, flight path; deviation from flight plan and clearances; aircraft capabilities; projected flight path; projected landing time.
System SA System status, functioning and settings; settings of radio, altimeter and transponder equipment; ATC communications present; deviations from correct settings; flight modes and automation entries and settings; impact of malfunctions/system degrades and settings on system performance and flight safety; fuel; time and distance available on fuel.	**Environmental SA** Weather formations (area and altitudes affected and movement; temperature, icing, ceilings, clouds, fog, sun, visibility, turbulence, winds, microbursts; IFR vs VFR conditions; areas and altitudes to avoid; flight safety; projected weather conditions.

1.1 Controlled Flight Into Terrain (CFIT)

Controlled Flight Into Terrain (CFIT) represents one of the leading causes cited for aviation accidents and fatalities each year. In the commercial sector, an analysis of worldwide accident data from 1988 to 1993 revealed that CFIT was blamed in 36.8% of all accidents and 53.6% of all fatalities (Graeber, 1996). Data indicate that enroute CFIT incidents can frequently be attributed to a lateral path error (Corwin, 1995). Many civilian CFIT incidents occur during the descent or approach phase. An analysis of 40 CFIT accidents and incidents involving commercial jets in the 1986 to 1990 period shows that the vast majority of flight crews were on the proper heading for the approach, however, they descended too soon or too steeply towards the terminal area (Graeber, 1996). Both flight management system mode errors and non-precision approaches have been cited as significant factors in these incidents (Corwin, 1995).

CFIT is one possible outcome of a loss of situation awareness. No-one has suggested that pilots intentionally hit the ground, but rather that for various reasons, their perception of the situation is flawed, leading to the accidental outcome. The main limitations humans have in achieving SA is a limited amount of attention and limited working memory for processing and integrating the cues perceived (Endsley, 1995c). These limitations are why many CFIT accidents occur under conditions which are very demanding of attention and in which the pilot's attention is drawn towards other aspects of the problem (e.g. locating some feature or dealing with an equipment malfunction). To get around these problems, experienced pilots can draw on pattern matching skills which allow them to more quickly process information, match recognized patterns to stored memory representations of similar situations and correct actions. None-the-less, even highly experienced pilots can fall prey to lapses in attention and memory, particularly under stressful conditions. Where direct outside visual cues are impoverished (night & IMC), the attention demands are even greater than normal, as it is more demanding to process instrument based displays, and the external cues available for detecting errors in that process are not available.

CFIT can arise due to a failure of SA in a number of the categories in Table 1. One failure may occur in terms of geographical SA. In this regard pilots need to know where they are, not just in an absolute sense, but in relation to other relevant things (airports, runways, geographical features, etc...) For this type of SA, the challenge is to match self-referenced information (route knowledge) with world referenced information (such as map or survey knowledge). This process can be very demanding of limited working memory and is prone to error if pilots mismatch the cues in the environment to the map to form an erroneous picture of where they are or where other objects are. This creates a lateral path error which has been implicated in many incidents of CFIT (Corwin, 1995; Endsley, 1995b). Another error occurs in which the pilot is aware of being off-path geographically, yet is not aware of exactly what the aircraft's relationship to surrounding terrain is, such as happened in the crash of the B-757 in Cali, Columbia (Endsley & Strauch, 1997). Displays such as SVS that help to align these two sources of information (self-referenced route knowledge and world referenced knowledge) and *directly depict* the spatial relationship of the pilot's aircraft to reference information should be helpful in preventing this class of CFIT incidents.

CFIT may more frequently arise due to failures in the spatial/temporal area, however. Particularly in instances where the aircraft crashes in flat terrain or water, it is more likely that the SA failure occurred as a result of a misperception of the aircraft's trajectory in the vertical plane. Accidents that occur in the descent/landing phase are also more likely to fall into this category. Documented problems in this area include improperly setting the glide-slope (frequently co-occurring with systems SA errors), misperceiving the relationship between the aircraft' vertical trajectory and the slope of the terrain (particularly with gently sloping terrain), and misperceptions of aircraft altitude (frequently co-occurring with errors in setting altimeters) (Flight Safety Foundation, 1996). In addition, CFIT incidents that are the effect of spatial disorientation affects fall into this category. Displays which directly present a vertical picture of aircraft trajectory relative to the ground are most appropriate for this class of CFIT problems. While a 3-D display such as that incorporated into SVS may be more helpful than a God's-eye display, it remains to be seen whether it will provide sufficient information for the detection of vertical navigation problems.

In summary, the following SA factors relevant to CFIT may be affected by SVS. Pilot awareness of these factors should be examined in testing of the SVS concept.

- Awareness of adherence to horizontal flight path
- Awareness of adherence to vertical flight path (& glide slope)
- Awareness of terrain and obstacles on flight path (current and alternates)
- Clearance of terrain and obstacles on flight path
- Projected touch down point on runway

- Projected stopping point on runway (based on relevant aircraft parameters)

1.2 Aircraft Flight Vector

A central aspect of a pilot's job is maintaining basic flight control (Endsley, Farley, Jones, Midkiff, & Hansman, 1998a). While standard instruments will likely be a part of any cockpit, SVS has the potential to augment pilot awareness of flight control information, or possibly surplant these traditional displays in providing this information. Even if traditional speed, altitude and attitude displays remain in the cockpit, the visually compelling nature of the SVS concept may mean that pilots' will increasingly attend to these displays over other information available in the cockpit. Due to problems with accurate depth perception in 3-D graphic displays, pilots' perceptions of speed, altitude, heading and location may not be accurate with these displays (Endsley, 1989; Wickens, Liang, Prevett, & Olmos, 1996).

Many of the SVS concepts include superimposed symbology for basic flight control information (e.g. speed, altitude, heading). While the format of these displays may be very similar to traditional displays, the fact that they are superimposed over what may be fairly busy colored displays may affect their readability under some conditions. More importantly, the compelling nature of the SVS 3-D graphic displays may lead pilots to be influenced more by their perceptions of this graphic information, than by the digitally displayed information in the superimposed symbology.

On the other hand, the design of the SVS may allow for the direct presentation of certain flight control information that is not currently well presented. For instance, project path as affected by winds, winds at different altitudes, or acceleration/deceleration vectors may be directly depicted on the displays,

In summary, the following SA factors relevant to flight control may be affected by SVS. Pilot awareness of these factors should be examined in testing of the SVS concept.

- Velocity
- Altitude
- Attitude (pitch & roll)
- Wind effects
- Heading
- Projected path
- Acceleration/Deceleration
- Location

1.3 Other Traffic

Currently the pilots' perception of the location and projected flight path of other traffic comes from direct visual detection (in VMC conditions), air traffic control reports and party-line information through radio transmissions, or more recently through TCAS displays. While radio information and TCAS displays can both provide pilots with the information needed to form a visual picture of the traffic situation, these methods both have their shortcomings. Monitoring radio traffic and developing a mental picture of this information is very mentally demanding. The rise of data link may also create a situation where much of this information would no longer be available in the future (Midkiff & Hansman, 1992).

While TCAS has provided pilots with far more information on air traffic, this information will be far more useful if it is integrated with SVS displays. Not only would this allow for better SA of critical information without the extra demands of searching for and monitoring separate displays (conflicting displays can take extra time to process and can induce more errors), but it also should significantly reduce the likelihood that pilots' might gain a false sense of security from viewing an SVS display that shows no other aircraft present. The tendency to drop visual scans of information under task load or stress might exacerbate this possibility.

The way in which SVS concepts integrate and portray other traffic through TCAS, datalink or other technologies, will be very important in its ultimate affect on SA. The following SA factors relevant to air traffic may be affected by SVS and should be examined in testing of the SVS concept.

- Location of other aircraft (range & bearing)
- Rate of closure
- Relative flight paths
- Projected minimum separation

1.4 Use in Flight Operations

Piloting is not just a perceptual task, but also a cognitive one undertaken in a very dynamic environment. Pilots need to actively evaluate and react to changes within the air traffic control system, often very rapidly and within dynamic contexts. An example of this challenge is the need to rapidly change approaches and runways in response to a new ATC clearance. Tasks such as this have been found to be very demanding, particularly when coupled with advanced flight management systems operating at higher levels of automation. The ability of the SVS system to aid the pilot in making strategic and tactical assessments such as the following will be critical.

- Evaluation of changes in runway and approach
- Evaluation of new ATC vector/clearance
- Evaluation of aircraft spacing
- Evaluation of timing and fuel usage on path
- Awareness of poor weather conditions on route
- Ability to determine the veracity of displayed information

1.5 Other Operational Considerations

A number of other factors relevant to flight operations should be considered with regard to the SVS display. Each of these factors may act to reduce its utility or to provide degraded or erroneous SA to pilots to who are using the SVS system. Methods for dealing with these factors in SVS implementation need to be developed and these factors need to be incorporated into testing of the SVS system to detect any potential problems.

- Non-normal operations (e.g. SVS display or system outage)
- Operations in wind-shear/micro-burst and other weather hazards
- Operations in holding patterns
- Use with mixed fleets (e.g. aircraft who forget to turn on transponders)
- Presence of non-sensored vehicles (e.g. autos or catering trucks on runway)

2.0 Overview of Situation Awareness Measurement Techniques

High level performance measures (as collected during the limited conditions of simulation testing) may not be sufficiently granular or diagnostic of differences in SVS design concepts. Thus, while one design concept may be superior to another in providing the pilot with needed information in a format that is easier to assimilate with pilot needs, the benefits of this may go unnoticed during the limited conditions of simulation testing or due to extra effort on the part of pilots to compensate for a design concept's deficiencies. If situation awareness is measured directly, it should be possible to select concepts that promote SA, and thus increase the probability that pilots will make effective decisions and avoid poor ones. Advantages of particular design options can be ascertained and problems with situation awareness can be detected so that corrective changes made to improve the design.

To better represent the issues involved in selecting measures of SA, the process model in Figure 1 is presented. This model shows the stages involved in the perception-action sequence. While they are shown as separate stages for simplicity in narration, it should be noted that these stages may be very closely coupled. Moderating factors that may influence each stage are shown on the bottom. Across the top, classes of measures appropriate to each stage are shown. Some of these will provide indirect indices of SA and others will be more direct. Measures at each stage will be discussed, including advantages and disadvantages of each.

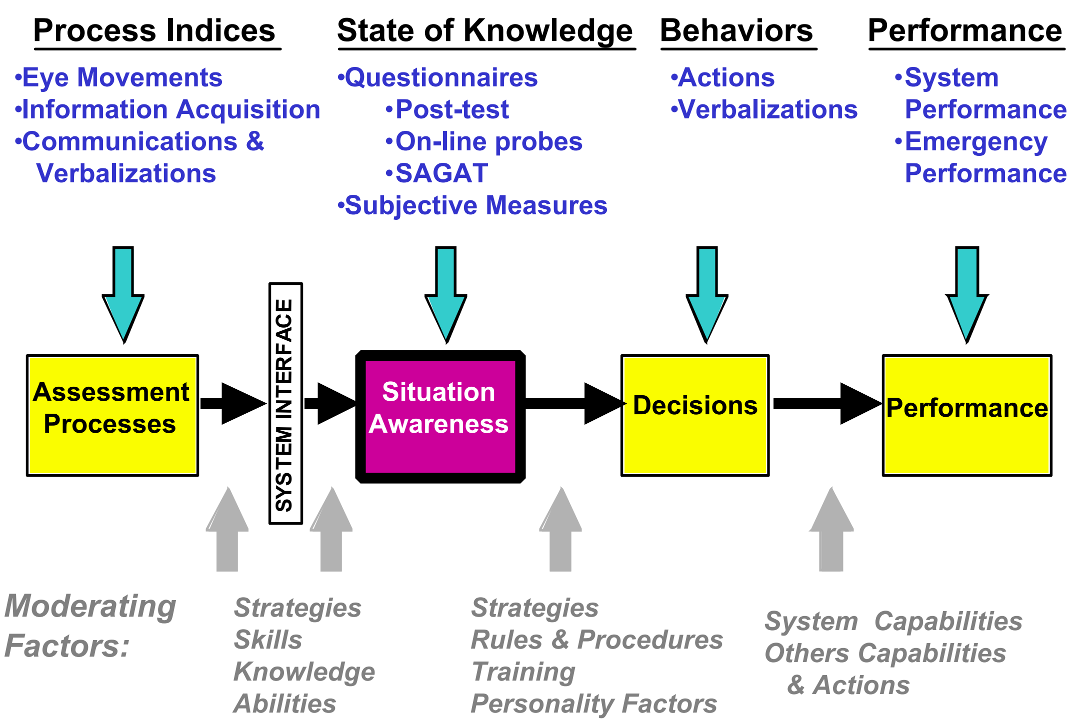

Figure 1. Measures of Situation Awareness (adapted from Endsley, 1996)

2.1 Process Indices

Many characteristics of individuals will influence the assessment processes used in acquiring information from the environment. There is evidence that some people are better at developing situation awareness from a given system design than others (Endsley & Bolstad, 1994). Differences in underlying abilities have been shown to contribute to this finding, including spatial, attention, memory, perceptual and cognitive abilities. Individuals will also form strategies, skills and knowledge with experience and training that will contribute to their selection of assessment processes and to the situation awareness they derive from those processes.

An examination of processes used for acquiring situation awareness may be useful in some contexts. It can provide information about how pilots allocate their attention in using a particular system design. This may indicate information about the relative priority of different types of information or the relative utility of information sources. It also can provide information about individual differences in these processes that may be useful for developing training strategies. In general, process measures may be useful for certain test objectives, but provide only an indirect indication of pilot situation awareness.

Several measures may be considered, many of which may be useful in conjunction with each other. Process tracing tools which have been applied to the study of decision making may be applicable to the study of SA processes. Eye-trackers and methods for assessing information acquisition (such as covering information so that overt actions are required for observation) may provide useful assessments of how attention is deployed (or not deployed) in the process of acquiring SA, typical scan patterns, and relations between elements. This information may provide useful insights into the process of acquiring SA or into the types of mental models directing this process. Studying the communications process between pilots may also provide useful information on the types of information that is lacking from displays, verbal techniques used for acquiring SA and differences in SA strategies between individuals.

Verbal protocols may provide some useful information on not only what is attended to, but also may provide a certain degree of insight into how that information is integrated and used in the process. Sullivan and Blackman (1991), for instance, used verbal protocols to investigate the relationship between working memory and long-term memory in maintaining SA. Significant difficulties in processing and using the data provided by verbal protocols must be dealt with by the experimenter, however, if this technique is to be used successfully.

Each of these techniques can be viewed as providing useful partial information on SA processes from which some inferences may be possible. Because verbal communications and verbal protocols take place in a very limited time frame, however, they cannot be regarded as complete representations of what pilots attend to or process. Eye-trackers and information acquisition methods are more likely to capture the SA acquisition process, but will not provide any information on how that information is used or combined to form higher level SA.

Sarter and Woods (1991) have proposed the use of a scenario manipulation method, wherein the simulation is altered by changing displayed information in some unpredicted way. While the artificial manipulation of parameters in a simulation may influence pilot SA too much to provide an accurate quantification of SA (as it will artificially affect their attention and situation awareness in the process), it may provide useful insights into the SA process. By systematically manipulating displayed or communicated information, useful patterns may emerge. Tenney, Adams, Pew, Huggins and Rogers (1992) discuss using this technique to lead subjects "down the garden path", thus investigating factors that may lead directly to misassessments of situations. Due to its intrusiveness and artificial impact on SA, scenario manipulation should not be used during simulation testing in which simultaneous assessments of situation awareness, workload or performance are to be made.

2.2 Direct Measures of SA

Several measures have been developed for assessing situation awareness directly. These can be grouped into subjective techniques and questionnaires.

2.2.1 Subjective Techniques.

Subjective estimation of SA may be made by individual pilots or by experienced observers. The subjective assessment of SA is very attractive in that it is fairly inexpensive and easy to administer. In addition to allowing evaluation of design concepts in simulation studies, subjective techniques can be easily applied in less controlled real-world settings. Certain limitations are present, however, which limit the interpretation of subjective evaluations of SA.

Self-ratings of SA usually involve a subjective estimation of how much SA a particular pilot feels he or she has when using a given system design. Self-ratings of SA may not necessarily provide an accurate quantification of SA, however, as pilots may not know about their own inaccuracies or what information they are unaware of. They have a limited basis for making such judgments. In addition, subjective self-ratings may be highly influenced by self assessments of performance, and thus become biased by issues that are beyond the SA construct. These self-ratings may be useful, however, as they can be considered to provide an assessment of pilots' degree of confidence in their SA.

Subjective estimation has been used in several studies to measure the SA provided by system designs during simulation testing. An early study at McDonnell Douglas (AMRAAM, 1982) employed subjective estimation of SA by pilots and observers of a simulation of a new weapon system. Arback, Schwartz and Kuperman (1987) examined the use of a six item scale to evaluate a panoramic cockpit display concept. Using a five-point subjective SA rating scale, Kuchar and Hansman (1993) found that pilots subjectively rated a smoothed-contour display of terrain as providing better SA than a spot-elevated display, supporting performance differences between the displays.

One of the best known subjective scales is the Situational Awareness Rating Technique (SART) developed by Taylor (1990). SART has pilots rate a system design based on the amount of demand on attentional resources, supply of attentional resources and understanding of the situation provided. As such, it considers pilots' perceived workload (supply and demand on attentional resources) in addition to their perceived understanding of the situation.

While SART has been shown to be correlated with performance measures (Selcon & Taylor, 1990), it is unclear whether this is due to the workload or the understanding components. Selcon, Taylor and Koritsas (1991), for instance, showed SART to be sensitive to changes in task demands, correlating with the NASA-TLX measure of workload. Crabtree, Marcelo, McCoy and Vidulich (1993) examined the sensitivity of SART to various display manipulations. They found SART to be sensitive to most of the manipulations, particularly the attentional demand subscale.

As another approach to developing a standardized subjective measure of SA, Vidulich and Hughes (1991) used a modified version of the Subjective Workload Dominance (SWORD) technique to obtain subjective evaluations of the SA provided by displays. SA-SWORD has subjects provide a comparative preference for displays on a nine-point scale, based on their beliefs about the amount of SA provided by each. They found the technique discriminated between two display formats and had inter-rater reliability.

SA may also be assessed by subjective ratings by outside observers. As an advantage, trained observers may have more information than the subject about what is really happening in a given simulation, thus their knowledge of reality may be more complete. As a shortcoming, observers will have only limited knowledge about what the pilot's concept of the situation is, however. Pilot actions and verbalizations may provide useful diagnostic information on explicit SA problems (misperceptions or lack of knowledge) and provide an indication that certain information is known, supporting observer judgments. Actions and verbalizations cannot be taken to provide a complete representation of a pilot's SA, however. They may know many things they do not mention or make an immediate response to as they are performing other tasks, for instance. Observer ratings therefore provide only a partial indicant of pilot SA. Efforts to elicit more information (by asking questions or providing artificial tasks) may augment natural verbalizations, however, this may alter the subject's distribution of attention, thus altering SA.

Waag and Houck (1994) developed the Situational Awareness Rating Scales (SARS), to obtain subjective evaluations of the SA of operational aircrew. SARS involves the subjective rating of aircrew SA on a six-point scale on each of 31 items which fall into eight general categories: general traits, tactical game plan, system operation, communication, information interpretation, tactical employment - beyond visual range, tactical employment - visual, and tactical employment - general. These items were derived from an evaluation of air combat operations to include factors that are observable in day-to-day military aircraft flight operations. In evaluating SARS, Waag and Houck found a high level of correlation between self evaluations, peer evaluations and supervisor evaluations of F-15 aircrew using SARS. They also found a significant relationship between SARS scores and several measures of individual experience level. Bell and Waag (1995) found that the SARS ratings obtained from a pilot's squadrons correlated moderately (R^2 = .314) with the SARS ratings of experts viewing simulations involving these pilots. This indicates that while the squadron's ratings may have been at least partially influenced by personal knowledge of the individual, they may have also been at least partially measuring something additional.

Interpretation and use of a measure like SARS for system evaluation is problematic, however. SARS includes assessment of many factors relevant to air combat performance, some of which relate to a subjective impression of a person's personality traits and basic abilities (e.g. discipline, reasoning ability), decision making abilities and flight skills (e.g plan formation, system proficiency, targeting decisions) and performance (e.g. maintain track of bogeys/friendlies, defensive reaction), as well as their knowledge of the state of the environment (e.g. interpreting vertical situation display, integrating overall information). A composite SA score is developed from ratings on each item. As such, SARS combines assessments on many dimensions *besides situation awareness* that are theoretically relevant to performance in this setting, and can be seen to include measures at many levels of the process model in Figure 1. Interpretation of its results in terms of SA per se, is therefore difficult. As scales are closely tied to a particular aircraft type and mission they probably are not applicable in other domains, such as commercial aircraft operations, nor is it clear whether the measure could be used to evaluate systems as opposed to individuals. It does, however, provide an example of the use of a subjective measure in an operational environment.

2.2.2 Post-Test Questionnaires.

Questionnaires allow for detailed information about subject perceptions to be collected which can then be evaluated against reality, thus providing an objective assessment of SA on a detailed level. This type of assessment provides a direct measure of SA and does not require subjects or observers to make judgments about situational knowledge on the basis of incomplete information, as subjective assessments do. This type of information can be gathered in one of three ways: post-test, during simulations or during interruptions in the simulation.

A detailed questionnaire can be administered after the completion of each simulated trial, allowing ample time for subjects to respond to a lengthy and detailed list of questions. Memories of dynamic situation awareness will be less reliable with time, however. People have been shown to over-rationalize and over-generalize about past mental events (Nisbett & Wilson, 1977). Early misperceptions may be quickly forgotten as the situation unfolds over time. Post-test questionnaires will reliably capture SA only at the very end of a trial, therefore. Kibbe (1988) used this technique to evaluate SA as affected by automation of a threat recognition task. She found a retrospective recall measure to be insensitive to the automation and problematic.

2.2.3 Real-Time Probes

One way of overcoming this deficiency is to ask pilots about their SA while they are carrying out their simulated tasks. While it would be very easy for pilots to examine the displays for information which is probed, it might be possible to measure reaction time as an index of SA. Durso,et. al. (1998) recently investigated the use of this technique in an air traffic control task. They found that controller reaction time to probes about the current status of events in the simulation (Level 1 SA) were correlated with a subject matter expert's subjective ratings of controller performance (R^2 = .53). Probes about the future (Level 3 SA) were correlated (R^2 = .12) at a much lower level with a measure of how many actions the controllers still had left to complete at the end of the simulation. In general from this study it is difficult to tell whether this approach measures SA or whether it provides an index of workload, as a secondary task technique. There is also a concern that it may alter SA and be intrusive in ongoing task performance if subjects need to add the task of answering questions on top of their normal duties. More recent studies have not found these two issues to be a problem (Endsley & Jones, 1999; Endsley, Sollenberger, Nakata, & Stein, 2000).

2.2.4 Situation Awareness Global Assessment Technique

To overcome the limitations of reporting on SA after the fact, a frequently used a technique involves freezing the simulation at randomly selected times, at which point the system displays are blanked and the simulation is suspended while study participants quickly answer questions about their current perceptions of the situation. Pilot perceptions are then compared to the real situation based on simulation computer data bases to provide an objective measure of SA. The collection of SA data in this manner provides an objective, unbiased assessment of SA that overcomes the problems incurred when collecting data after the fact, yet minimizes biasing of pilot SA due to secondary task loading or artificially cueing their attention. The primary disadvantage of this technique involves the temporary halt in the simulation.

Several studies have used this technique to collect measures of SA on select parameters. Marshak, Kuperman, Ramsey, & Wilson (1987) administered queries on target location, altitude and status in evaluating various map display formats. Fracker (1989; 1990) used queries to measure subject knowledge of target identification and location in several studies. Mogford & Tansley (1991) used queries regarding aircraft location and status in a study of air traffic controllers. One potential shortcoming of obtaining an indication of SA by using probes on a few predefined elements is that this may have an effect on pilots' attention during testing. If they know they will be queried on certain factors, they may pay more attention to these factors than they normally would (possibly at the expense of other important factors).

The Situation Awareness Global Assessment Technique (SAGAT), is a global tool developed to assess SA across all of its elements based on a comprehensive assessment of pilot SA requirements (Endsley, 1987; Endsley, 1988b; Endsley, 1990b) . As a global measure, SAGAT includes queries about all pilot SA requirements, including Level 1 (perception of data), Level 2 (comprehension of meaning) and Level 3 (projection of the near future) components. This includes a consideration of system functioning and status as well as relevant features of the external environment. The approach minimizes possible biasing of attention, as subjects cannot prepare for the queries in advance since they could be queried over almost every aspect of the situation to which they would normally attend.

SAGAT has been used to perform evaluations of avionics systems (Endsley, 1988b), display designs (Bolstad & Endsley, 1990; Endsley, 1989), and display hardware configurations (Endsley, 1989), thus supporting test and evaluation during design concept development across a variety of considerations. In addition, it has been useful in conducting research on factors related to SA, including an investigation of the relationship between SA and workload (Endsley, 1993) and an investigation of factors leading to individual differences in SA (Endsley & Bolstad, 1994).

The SAGAT technique has thus far been shown to have a high degree of validity for measuring SA. SAGAT has been shown to have predictive validity, with SAGAT scores indicative of pilot performance in a combat simulation (Endsley, 1990a). Content validity was also established, showing the queries used to be relevant to SA in a fighter aircraft domain (Endsley, 1990b). Empirical validity has been demonstrated through several studies which have shown that a temporary freeze in the simulation to collect SAGAT data did not impact performance and that such data could be collected for up to 5 or 6 minutes during a freeze without running into memory decay problems (Endsley, 1990b; Endsley, 1995a). A certain degree of measurement reliability has been demonstrated in a study that found high reliability of SAGAT scores for four individuals who participated in two sets of simulation trials (Endsley & Bolstad, 1994). This technique has been used extensively in military aviation, air traffic control and nuclear power simulations. More recently, a version of SAGAT has been created for commercial pilots, based on a detailed cognitive task analysis of SA requirements in commercial aviation (Endsley & Strater, 2000).

2.3 Behavior measures

Pilots might be expected to act in certain ways based on their situation awareness. Some information about SA may be determined, therefore, from examining behavior on specific subtasks that are of interest.

Several authors have sought behavioral indices that might indicate a subject's level of SA. For instance, Mosier and Chidester (1991) measured communication frequency in evaluating crew SA in a commercial cockpit. They found that high performing crews has fewer verbal communications than poorer performing crews. Rogers (1994) examined answers to on-line probes from the dispatcher and first officer. He found this measure to be sensitive to design issues surrounding implementation of an automated system. Other behavioral indices might include time to make a response (verbal or non-verbal), and correct or incorrect SA as identified from pilot verbalizations and appropriateness of a given behavior for a particular situation.

As mentioned previously, however, assessments of SA based on these types of behavioral measures need to be viewed with caution, as they assume what appropriate behavior will be, given SA or lack of it. These assumptions may not necessarily be warranted. For instance, a pilot may choose not to immediately verbalize or respond to a given event, thus confounding this type of measure.

2.4 Performance Measures

In general, performance measures provide the advantage of being objective and are usually non-intrusive. Computers for conducting simulation testing can be programmed to record specified performance data automatically, making the required data relatively easy to collect. Several limitations exist in using performance data to infer SA, however. Global measures of performance suffer from problems of diagnosticity and sensitivity. While global measures of performance are very important, as measures of SA, they are somewhat limited. As many moderating factors can influence the link between situation awareness and performance (such as decision making, tactics, strategies, prioritization of tasks) global performance measures will only provide an indirect indication of SA.

Specific task performance has been examined as an indicant of SA in the form of testable response measures. Hansman, et al. (1992) for example, used detection of clearance amendment errors as a measure of aircrew SA in evaluating the use of an automated datalink system for updating the onboard flight management control system. The measure was sensitive to differences between manual and automated programming modes, but was not sensitive to the use of readback. In evaluating the SA provided by a three-dimensional perspective display in the cockpit, Andre, Wickens, Moorman and Boschelli (1991) measured navigation performance and aircraft control. They found that aircraft control was better with the three-dimensional display, however, navigation performance was not sensitive to the display change as anticipated. This illustrates some of the difficulty in predicting just how subjects will allocate their attention in achieving SA and tradeoffs in prioritizing tasks to achieve ultimate performance.

The above studies measured the performance of embedded naturally occurring tasks in the domain of interest. Other researchers have imposed artificial tasks in assessing SA. These tasks are usually associated with specific SA components that are expected to improve with some display manipulation. For instance, Wells, Venturino, and Osgood (1988) measured accuracy of target replacement following presentation of information on helmet mounted displays with different fields of view as an indicant of subject spatial awareness. Zenyuh, Reising, Walchli, and Biers (1988) measured performance on a target search task as an indicant of SA in evaluating a stereoscopic display. These types of measures are appropriate for part-task evaluations of design tradeoffs involving specific system components, however, they do not necessarily provide an indication of SA tradeoffs when using the component in the context of the larger system.

In addition, it is quite easy for subjects to bias their attention to a single issue which is under evaluation in a particular study if they figure out the purpose of the study. For example, Busquets, et. al. (1994) measured the time for aircraft to respond to a runway incursion as a measure of pilot SA. This is a good example of a situation in which a single response outcome can be expected if the person has good SA. The disadvantage is that only one such incursion can occur for a given study participant, as issues of predictability and response priming can easily confound the results. In the same study, Busquets, et. al., also instituted a technique where they disrupted the scenario and restarted it with the aircraft in a new location. They measured the time required for the pilot to return to the proper flight path. This approach is similar to studies where pilots are placed at an unusual attitude and asked to recover. While this measure provides interesting insight into a pilot's ability to reorient in the face of such disruptions, it is not clear how this relates to SA under normal conditions. It is highly likely that under such test conditions, the participants may skew their attention towards the displays supporting these measurements.

In a different approach, Jones (1997) inserted errors into air traffic control scenarios. These errors involved a wrong aircraft type listed on the flight strip, a wrong destination listed and a readback error. Each of these errors led to the aircraft performing actions (cues) that should have indicated to the controller that an error was present, and each of which necessitated a corrective action on the controller's part. She measured how long after the cue it took for the controller to take a corrective action or make a statement indicating that the error had been noticed. This approach allowed her to look at the subtle effects of wrong mental models on SA. As a measurement approach, however, it is difficult to determine whether a lack of response by a participant indicates that they did not understand that an error was present (poor SA), or whether they noticed, but did not bother to do anything about the error, even though they should have.

While finite task measures may readily present themselves for evaluating certain kinds of systems, for others determining appropriate measures may be more difficult. An expert system, for instance, may influence many factors in a global, not readily predicted manner. The major limitation of this approach stems from the interactive nature of situation awareness sub-components. A new system to provide SA on one factor may simultaneously reduce SA on another, not measured, factor. For instance, Wickens (1995) found that a three-dimensional pathway display improved pilot performance for routine tasks (adherence to the flight path). He also found in subsequent studies that pilot performance was worse on other types of tasks with those displays, such as responding to an unanticipated event and needing to reroute to a new airport. He hypothesized that these two tasks required very different types of SA, one local awareness and the other more global awareness. From a measurement stand-point, this exemplifies the problems of relying only on performance measures. The results are likely to be limited by the scenario conditions in the test. The effects of certain types of SA problems may not be brought out by the test conditions.

In general, performance measures provide observable and therefore easily measured indications of SA. Situation awareness must always be inferred from them, however, and many other factors will influence the degree to which they provide a clear indication of SA. Overall, as improved SA in one area may easily result in decreased SA in others, relying exclusively on the measurement of performance on specific parameters can yield misleading results, and should be viewed within the context of other measures of SA.

2.5 Comparisons of Measures

While most studies have sought to examine the utility or validity of a single measure of SA, some recent studies have sought to directly compare some of these measures. Vidulich (2000) recently completed an extensive meta-analysis of SA measures. His study reviewed 65 published studies which involved SA measurements, many of which used more than one measure. Of these, 61 included measures of performance, 22 involved a direct probe or query technique, and 20 involved a subjective rating. He examined the studies to see which measures were sensitive to the display manipulation examined in the study at a .05 level of significance. Roughly 70% of the performance measures were sensitive by this criterion. While direct queries of SA were not very sensitive when only a narrow range of questions was included (30%), they were sensitive 80% of the time when a wide range of questions was included, such as with SAGAT. Subjective measures were also found to be sensitive, with SART showing a difference 92% of the time and other subjective measures 75% of the time. While this analysis is largely limited by the fact that is assumes that the measures should have found a difference in each of the studies, which in fact may not have been the case, it does provide an interesting comparison of the sensitivity of the various techniques for measuring SA. The study also does not address the issue of validity. That is, it is impossible to tell whether the differences found by the measures where due to SA or to some other factor from such a meta-analysis.

In another study, SART was directly compared to SAGAT to provide more insight on the comparability and validity of these techniques (Endsley, Selcon, Hardiman, & Croft, 1998b). In this study, an enhanced aircraft display that showed enemy aircraft launch success zone envelopes was employed in a study involving military fighter pilots. In certain trials both SAGAT and SART were measured concurrently. No correlation between the SAGAT scores (either individual query accuracy or combined accuracy) and the SART scale (or its sub-scales) was found. SART was highly correlated with a subjective measure of performance (Pearson correlation = .554) and a subjective measure of pilot confidence level (Pearson correlation = .678), however. This indicates that subjective and objective measures of SA may be picking up on very different things.

In a more recent study, the use of real-time probes to measure controller SA was examined in comparison to the SAGAT technique (Endsley, et al., 2000). No evidence was found to support the concern of real-time probe intrusiveness in this study. The controllers subjectively rated its intrusiveness as low and reported that they often responded to such questions from other controllers. When the time to respond to each real-time probe was compared to mean accuracy on the SAGAT query for the same information in that trial, however, no relationship was found. A low but significant correlation was found to exist between reaction time to the real-time probes and reaction time to a simple workload measure that occurred within one minute of the probe, indicating a weak relationship with workload.

In a later test of this technique, Endsley and Jones (1999) found that the SART, SAGAT and real-time probe measures of SA all showed sensitivity to the differences between two scenarios in an air-sovereignty mission simulation. While the SART and real-time probe measures could only provide overall differences between the scenarios, the SAGAT measure provided diagnostic detail as to the types of SA differences that existed between the two scenarios. While the real-time probe measures did not show this level of sensitivity in the study, it is believed that this was probably due to the comparatively fewer number of queries of each particular type that were provided with the real-time probes. The fact that a weak, but significant correlation was found between the real-time probes and the SAGAT queries on the same content areas helps to support this contention. Both accuracy and response time associated with the real-time probes appeared to show value. No evidence was found to support the concern that the real-time probes might be measuring workload rather than SA. Subjectively, the probes were not noted by the participants to be intrusive or annoying.

2.6 Summary

Situation awareness is an important construct underlying successful performance in the complex and demanding flight environment. Evaluating the degree to which prospective designs for SVS actually provide benefits to situation awareness (and avoid pitfalls) is an important function of evaluation efforts. Assessment of situation awareness in a systematic fashion allows problems to be detected, potential solutions to be examined and insures that final designs provide pilots with the situation awareness needed to be successful in the performance of their many functions.

The evaluation of the Synthetic Vision System design concepts may be undertaken through two major methods: simulation and real-time operations. Different recommendations can be made for assessing situation awareness for SVS under these different conditions.

3.0 SVS Evaluation in Simulation Devices

Real-time operator-in-the-loop simulations of systems to be evaluated provide the most flexibility for examining a wide variety of human performance issues, including SA. Although certain aspects of operational reality may not be available in simulation devices, they allow for a wide variety issues to be carefully measured and very taxing and dangerous conditions to be examined which would not be possible in real-time operations (particularly where safety of flight is concerned). For this reason it is recommended that as much testing of situation awareness as possible be carried out in simulation devices.

When conducting evaluations of situation awareness in simulation devices, it is first recommended that available measures of performance be examined. Where there are direct unequivocal measures that indicate operator SA, these should be used. It may be difficult or impossible to find such measures, unfortunately, that are uniformly sensitive. They may also provide only part of the picture, additionally, indicating SA of some issues, but not others which may be equally important under certain circumstances.

For this reason, the use of SAGAT in simulation testing to provide detailed diagnostic information on operator SA is recommended as a needed addition to performance measures. SAGAT has been extensively validated in aircraft tasks and used successfully in a variety of other systems including air traffic control, power plant operations and driving tasks. It has provided sensitive measurement for a variety of display design, hardware, avionics, automation and operational concept questions. This direct measurement of SA allows for detailed diagnostic information indicating which aspects of SA are affected positively or negatively in a given system evaluation. While there is no known level of SA that can be considered to be a required minimum, such information can be compared to baseline system data to determine whether the system under evaluation is aiding or compromising SA.

3.1 Performance Based Testable Responses

Scenarios should be carefully constructed that will create a situation in which a discernable and observable set of actions is required of the pilot. In order to provide good data, a clear and unambiguous response must be available that every pilot would make if he/she has good SA (Pritchett & Hansman, 2000). In this case it needs to be clear that lack of action by the pilot indicates poor SA, and not just a choice to delay action or wait and see how the situation progresses. Care must be taken to insure that different actions than those expected are recorded in addition to the expected actions (Pritchett & Hansman, 2000). In addition it is important that pilot behavior with a variety of situations and test measures. While they may show improved SA in some situations, they may be negatively affected by the display in others. As a final issue, the scenarios need to be created so that pilots do not become sensitive to the problems that are imbedded in them. For example, if a runway incursion is experienced in each of several scenarios, pilots will become sensitized to expect such and it will skew the data, whereas these are usually very infrequent events. To guard against these shifts in expectancies is very difficult and will require that particular pilot subjects do not get repeated scenarios with the same types of problems. In all scenarios performance with alternate SVS display concepts should be compared to each other, as well as to a defined baseline condition (e.g. a B-757 without SVS). These tests should incorporate realistic flight conditions under which the SVS system will be used (e.g. VMC and IMC conditions, high and low traffic conditions, etc. as described in Section 5.0).

Particular testable responses for examining the SVS concept include:

(1) **Time to respond to traffic on the runway:**

These scenarios should include the following types of events:

- delays in aircraft moving off runway,
- crossing traffic,
- traffic with only part of the aircraft (e.g. a wing) intruding on runway,
- catering truck on the side of the runway creating an incursion

These measures would detect changes in SA of ground traffic on landing associated with the use of the SVS. These measures will be particularly important if the SVS system will be used to decrease runway spacing under poor visibility conditions.

(2) **Time to determine bad flight path:**

In these scenarios, at some point in the simulated flight ATC would provide an inappropriate clearance which should include:

- ATC clearance that vectors the aircraft into terrain or obstacles (horizontal or vertical),
- ATC clearance that vectors the aircraft into the path of other traffic

In other scenarios, a confederate co-pilot could be employed who would take actions that must be detected and corrected by the subject pilot. This would include:

- Entry of an incorrect clearance in error
- Determination of poor runway alignment/non-stable approach by the confederate co-pilot flying

These measures would detect changes in SA associated with knowledge of aircraft flight path in relation to desired objectives, traffic and terrain. It is expected that any improvements in SA of aircraft flight path associated with the SVS would be detected with these measures.

(3) **Time to reorient on proper flight path in response to SVS outage:**

These scenarios would incorporate a loss of the SVS system. The time required for the pilots to bring the aircraft to a simultaneously administered new ATC clearance would be measured as an indication of the degree of reliance on the SVS displays. A pilot who is too dependant on the SVS displays would have little internalized SA that would allow a quick recovery from such a problem. If the SVS enhances the pilot's mental picture, he/she should be able to rapidly switch to back up displays and continue on the flight unaffected. Problems in completing the flight to destination should also be recorded.

(4) **Time to determine/respond to traffic conflict:**

The following scenarios should be created that would institute various air traffic conflicts in the air:

- Traffic overtaking from rear
- Crossing traffic
- Loss of spacing on traffic in trail
- Traffic descending from above
- Traffic ascending from below

The affect of the SVS on enhancing SA of air traffic would be determined. While knowledge of air traffic is not a primary goal of the SVS concept, it is possible that (a) it may decrease SA of air traffic by shifting attention to other factors (such as terrain and flight path), (b) it may increase SA if air traffic information is successfully integrated with the terrain and flight path information. In particular the ability of pilots to discern which aircraft in a busy traffic situation will lead to conflicts would be assessed with these measures.

(5) **Flight path adherence:**

The ability of pilots to adhere to the desired flight path should be measured. (This can be done in separate scenarios or in conjunction with the scenarios testing other factors). In particular these scenarios should incorporate various levels of winds, both enroute and in final approach and landing scenarios. The ability of the pilots to detect and correct for winds to stay on the desired flight path will be an indicant of changes in SA of flight path due to SVS.

(6) **Recovery time from unusual attitude:**

In these scenarios, the pilot is presented with an unusual attitude (upon taking over control of the aircraft in mid-air). The time to detect and correct the aircraft attitude (straight and level) is measured. (Note: Pilots with experience in recovering from unusual attitude training, such as that provided in the military, should be used in this test, as well as pilots without this special training as SVS may affect these groups differently.)

(7) **Time to respond to new runway/approach:**

In these scenarios, ATC would provide a different runway and approach than that which had been briefed and prepared for. The time taken to enter and establish the aircraft on the new approach would be measured as an indicant of the SA related to new flight paths provided by the SVS display.

(8) **Knowledge of direction/location of alternate airport.**

This scenario would measure the time to determine and request vectors to an alternate airport in response to a simulated flight emergency. This measure would test global geographical SA as affected by the SVS concept.

(9) **Time to respond and compliance to TCAS alerts and GPWS alerts.**

These scenarios would incorporate conditions in which the GPWS or TCAS alerts are activated. As pilots will often seek confirming information in relation to such alerts prior to responding, it is possible that the SVS may hasten response time as confirmation may be quicker, or delay response time, particularly if the relevant factors (e.g. sink rate, or an obstacle not in the database) are not readily discernable from the display. This test should incorporate a wide range of conditions in different scenarios to adequately and realistically test pilot behavior in response to these alerts. As pilots may be able to detect the conditions leading to an alert prior to the alert going off, measurement will need to take this into account.

3.2 SAGAT Queries

The number of scenarios that would need to be conducted to test all of the factors relevant to pilot SA by means of testable performance responses may be prohibitive, particularly given the need to minimize changing pilot expectancies through exposure to a high number of infrequent events in the scenarios. For this reason, among others, it is recommended that SAGAT be used to directly assess pilot SA with the SVS concept. In these trials, as many as 3 freezes in the simulation scenario would be inserted in any 30 minute period of simulated flight. During each freeze the following SAGAT queries will be provided via a laptop computer, for comparison to the actual state of each factor at the time of the freeze.

- What is the current heading of your aircraft?
- What is the current altitude of your aircraft?
- What is the indicated airspeed of your aircraft?
- What is the current attitude of your aircraft? (pitch, roll, AOA)
- What is the current climb/descent rate of your aircraft?
- What are the current winds?
- How does your current altitude compare to your planned altitude at this point?
- How does your current speed compare to your planned speed at this point?
- How does your current heading compare to your planned heading at this point?
- How does your current fuel compare to your planned fuel at this point?
- How does your current position compare to your planned position at this point?
- Are you in conformance with your current clearance?
- Is there any traffic on your current path?
    - Conflicting traffic is located at? (range and bearing)
    - Traffic is conflicting in what way?
    - How much time is available until you must maneuver?
- Is a change in path or altitude needed to avoid obstacles or terrain?
- Is a change in path or altitude needed to avoid restricted or special use airspace?
- Is there any hazardous weather along this phase of flight? (by type)
    - What impact will the hazardous weather have on your flight?
- Are you on the glide slope?
- Where on the runway do you think you will touch down?
- Where on the runway do you think you will stop the aircraft?
- What is your bearing and range to your next fix?
- What is your bearing and range to the destination airport?
- What is your current rate of acceleration or deceleration?
- What is your current rate of closure on the aircraft in front of you?

Format and presentation of the queries are shown in Endsley, Jones and Strater (2000). Phases of flight appropriate for each query are shown in Appendix A. The queries should be administered in a random order at each freeze point. Approximately 30-40 samples per query are needed across pilot participants per test condition. At the time of each freeze, corresponding data on the correct answer must be collected from the simulation computer and a subject matter expert in order to score the pilot participant's responses. Analysis of response accuracy across conditions (using non-parametric statistics) will provide a direct indication of the effect of the SVS display options on pilot situation awareness.

4.0 Evaluation of SVS in Real-time Operations

It will also be desirable to measure SA under the conditions of real-time operations to collect a more realistic picture of pilot performance. Under such conditions, the ability of the evaluator to control either the operational environment or the features of the test are far more limited. The ability to measure a cognitive construct such as SA is also far more limited. Any intrusions into the pilots' tasks must be done very carefully.

In real-time operations it is also recommended whatever performance measures available be collected and examined as indications of pilot SA. As it is difficult to thoroughly test the SA of pilots in this way in operational conditions, however, supplemental data collection will most likely be necessary. Subjective measures of SA will be the easiest to collect under these conditions, even though their interpretation must be taken with caution. Due to the limitations of such measures, an objective measure of SA is also desirable. The use of real-time probes may be explored for this purpose.

4.1 Subjective SA Measures

(1) SART

Pilots will provide ratings on each of the following scales for each concept tested (in Appendix B)

(a) Demand on Resources
- Instability of Situation
- Complexity of Situation
- Variability of Situation

(b) Supply of Resources
- Arousal
- Concentration of Attention
- Division of Attention
- Spare Mental Capacity

(c) Understanding of the situation
- Information Quantity
- Information Quality
- Familiarity with Situation

Ratings are provided on a bi-polar scale for each measure. Ratings can be filled out via paper and pencil or computer following each trial. The ratings are then combined into a single SART score by taking the subscales for each measure and combining them into an average score. The SART score is then calculated as the Understanding of the situation + the Supply of Resources – the Demand on Resources. SART scores can be analyzed via ANOVA to determine the significant differences between display conditions.

(2) SA SWORD

At the end of testing, each concept can be compared to the others in a pairwise fashion (in Appendix C). Each SA-SWORD form allows the degree of preference of the pilot participants to be indicated for each of the display concepts (based on the degree to which they feel it provides them with situation awareness).SA-SWORD scores are then calculated into a weighted preference for each display for each participant. These weighted preferences can be submitted to a non-parametric statistical technique for analysis across pilot participants.

(3) Confidence level in information

An important aspect of situation awareness the degree of confidence operators have in the information they receive. (The can affect how likely they are to act on a given piece of information as compared to the need to collect further confirming information, for example.) The SVS concept may affect pilot confidence level. For example, by virtue of having on-board pictorial representations of information as opposed to only map knowledge or verbally transmitted radio information, they may be more likely to accept more difficult clearances or reduce separation if they are more confident in their own awareness of the situation. Subjective confidence in one's knowledge of the situation has also been found to be highly correlated with subjective assessments of one's own SA (Endsley, et al., 1998b). Therefore, it would be useful for the SVS concept to obtain pilot participant ratings of confidence level associated with the following information.

- Confidence in knowledge of obstacle separation
- Confidence in knowledge of terrain separation
- Confidence in knowledge of traffic separation
- Confidence in awareness of aircraft state and flight vector
- Confidence in knowledge of adherence to clearances and assigned flight path
- Confidence in assessment of hazards
- Confidence in knowledge of relevant weather information

These ratings can be obtained on a bi-polar scale (as shown in Appendix D).

4.2 Real-time SAGAT Queries

As it would be difficult and potentially unsafe to "freeze" operations in real-time flight, it is recommended that real-time SAGAT queries be provided to obtain a direct indication of pilot SA under these circumstances. (Note: *While it might be possible to use a technique whereby the pilot participant is queried with his/her displays covered while a safety pilot takes over flight control, it is not known to what degree this would provide significantly different information than that obtained when using the technique in simulation conditions, and it would involve a greater degree of risk under these circumstances.)*

The real-time queries will correspond to the SAGAT queries listed in 3.2, however they will be asked *while* the pilot is flying the aircraft (without freezing or covering the displays). The queries should be administered one-at-a-time verbally to the pilot at randomly selected times during the trial. The pilot participants should be instructed to respond verbally as quickly and accurately as possible. Time to respond as well as accuracy will be measured. Audio recordings should be made to allow for later analysis of subject responses and response times.

This technique allows for only one query to be administered at a time. Queries can be provided as often as every 1-2 minutes on average (at randomly determined times), none-the-less, fewer queries can generally be administered using this technique as compared to the SAGAT technique (which allows for all the queries to be provided at each freeze point) over the same length scenario. Therefore it is likely that it will be more efficient to select a subset of queries (based on previous simulation studies) to be administered during real-time operations. Again 30-40 samples per query will be needed across pilot participants per test condition. At the time of each query, corresponding data on the correct answer must be collected from the test aircraft flight computer and an on-board subject matter expert in order to score the pilot participant's responses. Analysis of response accuracy across conditions (using non-parametric statistics) and response time (using ANOVA) will provide a direct indication of the effect of the SVS display options on pilot situation awareness under realistic flight conditions.

4.3 Performance Measures

A number of general performance factors should be measured under VMC and IMC (display only) conditions as indirect indications of pilot SA with and without the SVS concepts These factors should be recorded via the test aircraft's on-board computer (if possible) for later analysis via appropriate statistical techniques.

- Flight path adherence (RMS error associated with commanded flight path).
- Time to respond to a new runway/approach (based on new ATC clearance)
  - Number of actions required (to orient on new runway/approach)
- Knowledge of direction/location of alternate airport (in response to verbal query)
- Time to respond to SVS outage (to verbally provide current location relative to airport, altitude and airspeed, and to successfully complete approach without SVS)
- Flight parameters on landing
    - Location on center-line
    - Distance down runway
    - Flare height
    - Touchdown velocity
    - Touch down vertical velocity

4.4 Process indices

In addition to the above measures of performance, several aspects of the pilot's behavior may be affected by the presence of the SVS displays. These factors should be examined in order to determine any potential factors that might affect SA in the long-run.

- Time spent head-up vs head-down (as determinable via a camera or head-tracker).
- Time spent in examining each display (through eye tracker if available)
- FOV/Display range selected (over time)

4.5 Other measures

Other information that should be obtained from pilot participants following participation in the study might include:

(1) Subjective opinions on:
  - Utility of SVS as supplemental information
  - Utility of SVS to support operations in low visibility
  - Other information needed to support flight operations
  - Other information needed to rely on system in Cat III with reduced spacing
  - Implementation considerations

(2) Uses of system
  - E.g. reduction of spacing, flying around rather than over terrain

Real-time operational testing provides an opportunity to gather potentially useful information from pilots with regard to the SVS's use under operations conditions. This subjective input may be as useful in its development as other information obtained during the study.

5.0 General Scenario Considerations

The realisticness and complexity of the scenarios developed for the simulation testing and the real-time operational testing will be highly important in ensuring that the results of the studies provide useful and complete information about the potential advantages and problems associated with the use of the SVS display concepts. Some factors that need to be specifically considered and incorporated into scenario development include:

(1) Level of traffic- Scenarios should incorporate moderate to heavy traffic, particularly in and around airport areas.

(2) Presence of weather cells, visibility problems (e.g. low clouds, fog), and winds should be explicitly incorporated into the SVS test scenarios.

(3) ATC changes/ re-routes and last minute runway changes should be incorporated into at least some of the scenarios in order to represent the challenging real-world conditions under which the SVS will need to be used.

(4) Mixed equipage of other traffic and ground vehicles should be considered in these scenarios as it is unlikely that future operational environments will be homogenous in this respect.

(5) Various levels of workload (as affected by traffic, weather and ATC changes) can therefore be examined across different operational test scenarios.

By developing realistic and challenging testing conditions, the real advantages of the SVS concept over current day equipment will be more likely to be revealed. In addition, any potential problems associated with the system (either with regard to SA or performance) will be more likely to be brought to light, so that alternate design concepts can be developed for dealing with these issues.

Synthetic vision systems provide a significant potential for improving situation awareness in the cockpits of the future and helping to lower the aviation accident rates well below current levels. Meeting this goal, and that of improving traffic flow efficiency, will require that the SVS display design alternatives are carefully tested during its development. A graduated program of SA measurement is presented here for assisting in this evaluation at each stage of the design process. By carefully measuring the effect of the SVS on pilot SA, system development efforts can be guided in the most productive direction.

6.0 References

AMRAAM (1982). AMRAAM OUE Final Report. Unpublished Manuscript.
Andre, A. D., Wickens, C. D., Moorman, L., & Boschelli, M. M. (1991). Display formatting techniques for improving situation awareness in the aircraft cockpit. The international Journal of Aviation Psychology, 1, 205-218.
Arbak, C. J., Schwartz, N., & Kuperman, G. (1987). Evaluating the panoramic cockpit controls and displays system. In Proceedings of the Fourth International Symposium on Aviation Psychology (pp. 30-36). Columbus, OH: The Ohio State University.
Bell, H. H., & Waag, W. L. (1995). Using observer ratings to assess situational awareness in tactical air environments. In D. J. Garland & M. R. Endsley (Eds.), Experimental analysis and measurement of situation awareness (pp. 93-99). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.
Bolstad, C. A., & Endsley, M. R. (1990). Single versus dual scale range display investigation (NOR DOC 90-90). Hawthorne, CA: Northrop Corporation.
Busquets, A. M., Parrish, R. V., Williams, S. P., & Nold, D. E. (1994). Comparison of pilots' acceptance and spatial awareness when using EFIS vs. pictorial display formats for complex, curved landing approaches. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), Situational awareness in complex systems (pp. 139-167). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.
Corwin, W. H. (1995). Controlled flight into terrain avoidance: why the ground proximity warning system is too little, too late. In R. Fuller, N. Johnston, & N. McDonald (Eds.), Human Factors in Aviation Operations (pp. 155-160). Aldershot, UK: Avebury Aviation.
Crabtree, M. S., Marcelo, R. A. Q., McCoy, A. L., & Vidulich, M. A. (1993). An examination of a subjective sitautional awareness measure during training ona tactical operations simulator. In R. S. Jensen & D. Neumeister (Ed.), Proceedings of the Seventh International Symposium on Aviation Psychology (pp. 891-895). Columbus, OH: The Ohio State University.
Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1998). Situation awareness as a predictor of performance for en route air traffic controllers. Air Traffic Control Quarterly, 6(1), 1-20.
Endsley, M. R. (1987). SAGAT: A methodology for the measurement of situation awareness (NOR DOC 87-83). Hawthorne, CA: Northrop Corporation.
Endsley, M. R. (1988a). Design and evaluation for situation awareness enhancement. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 97-101). Santa Monica, CA: Human Factors Society.
Endsley, M. R. (1988b). Situation Awareness Global Assessment Technique (SAGAT). In Proceedings of the National Aerospace and Electronics Conference (NAECON) (pp. 789-795). New York: IEEE.
Endsley, M. R. (1989). Tactical simulation 3 test report: Addendum 1 situation awareness evaluations (81203033R). Hawthorne, CA: Northrop Corporation.
Endsley, M. R. (1990a). Predictive utility of an objective measure of situation awareness. In Proceedings of the Human Factors Society 34th Annual Meeting (pp. 41-45). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1990b). Situation awareness in dynamic human decision making: Theory and measurement. Unpublished doctoral dissertation, University of Southern California, Los Angeles, CA.

Endsley, M. R. (1993). Situation awareness and workload: Flip sides of the same coin. In R. S. Jensen & D. Neumeister (Ed.), Proceedings of the Seventh International Symposium on Aviation Psychology (pp. 906-911). Columbus, OH: Department of Aviation, The Ohio State University.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. Human Factors, 37(1), 65-84.

Endsley, M. R. (1995b). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston, & N. McDonald (Eds.), Human Factors in Aviation Operations (pp. 287-292). Aldershot, England: Avebury Aviation, Ashgate Publishing Ltd.

Endsley, M. R. (1995c). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32-64.

Endsley, M. R. (1996). Situation awareness in aircraft. In B. J. Hayward & A. R. Lowe (Eds.), Applied aviation psychology: Achievement, change and challenge (pp. 403-417). Aldershot, UK: Avebury Aviation.

Endsley, M. R. (1999). Situation awareness in aviation systems. In D. J. Garland, J. A. Wise, & V. D. Hopkin (Eds.), Human factors in aviation systems (pp. 257-276). Hillsdale, N.J.: Lawrence Erlbaum.

Endsley, M. R., & Bolstad, C. A. (1994). Individual differences in pilot situation awareness. International Journal of Aviation Psychology, 4(3), 241-264.

Endsley, M. R., Farley, T. C., Jones, W. M., Midkiff, A. H., & Hansman, R. J. (1998a). Situation awareness information requirements for commercial airline pilots (ICAT-98-1). Cambridge, MA: Massachusetts Institute of Technology International Center for Air Transportation.

Endsley, M. R., & Jones, D. G. (1999). Evaluation of a real-time probe measure of situation awareness (SATECH-99-12). Marietta, GA: SA Technologies.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998b). A comparative evaluation of SAGAT and SART for evaluations of situation awareness. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (pp. 82-86). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R., Sollenberger, R., Nakata, A., & Stein, E. (2000). Situation awareness in air traffic control: Enhanced displays for advanced operations (DOT/FAA/CT-TN00/01). Atlantic City, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Endsley, M. R., Jones, W. M. & Strater, L. (2000). Situation Awareness Global Assessment Technique User Guide for Commercial Aviation (SATECH-00-10). Marietta, GA: SA Technologies.

Endsley, M. R., & Strauch, B. (1997). Automation and situation awareness: The accident at Cali, Columbia. In R. S. Jensen & L. Rakovan (Ed.), Proceedings of the Ninth International Symposium on Aviation Psychology (pp. 877-881). Columbus, OH: Ohio State University.

Flight Safety Foundation (1996). World wide web page. In

Fracker, M. L. (1989). Attention allocation in situation awareness. In Proceedings of the Human Factors Society 33rd Annual Meeting (pp. 1396-1400). Santa Monica, CA: Human Factors Society.

Fracker, M. L. (1990). Attention gradients in situation awareness. In Situational Awareness in Aerospace Operations (AGARD-CP-478) (Conference Proceedings #478) (pp. 6/1-6/10). Neuilly Sur Seine, France: NATO - AGARD.

Graeber, R. C. (1996). Integrating human factors and safety into airplane design and operations. In B. J. Hayward & A. R. Lowe (Eds.), Applied aviation psychology: Achievement, change and challenge (pp. 27-38). Aldershot, UK: Avebury Aviation.

Hansman, R. J., Wanke, C., Kuchar, J., Mykityshyn, M., Hahn, E., & Midkiff, A. (1992, September). Hazard alerting and situational awareness in advanced air transport cockpits. Paper presented at the 18th ICAS Congress, Beijing, China.

Jones, D. G. (1997). Reducing situation awareness errors in air traffic control. In Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting (pp. 230-233). Santa Monica, CA: Human Factors and Ergonomics Society.

Kibbe, M. (1988). Information transfer from intelligent EW displays. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 107-110). Santa Monica, CA: Human Factors Society.

Koczo, S., Klein, J., Both, A., & Lamb, T. (1998). Preliminary system requirements for synthetic vision. Cedar Rapids, IA: Rockwell Collins.

Kuchar, J. K., & Hansman, R. J. (1993). An exploratory study of plan-view terrain displays for air carrier operations. International Journal of Aviation Psychology, 3(1), 39-54.

Marshak, W. P., Kuperman, G., Ramsey, E. G., & Wilson, D. (1987). Situational awareness in map displays. In Proceedings of the Human Factors Society 31st Annual Meeting (pp. 533-535). Santa Monica, CA: Human Factors Society.

Midkiff, A. H., & Hansman, R. J. (1992). Identification of important party-line information elements and implications for situational awareness in the datalink environment. In Proceedings of the SAE Aerotech Conference and Exposition . Anaheim, CA:

Mogford, R. H., & Tansley, B. W. (1991). The importance of the air traffic controller's mental model. Paper presented at the Human Factors Society of Canada Annual Meeting, Canada.

Mosier, K. L., & Chidester, T. R. (1991). Situation assessment and situation awareness in a team setting. In Y. Queinnec & F. Daniellou (Eds.), Designing for Everyone (pp. 798-800). London: Taylor and Francis.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84(3), 231-259.

Pritchett, A. R., & Hansman, R. J. (2000). Use of testable responses for performance-based measurement of situation awareness. In M. R. Endsley & D. J. Garland (Eds.), Situation awareness analysis and measurement. Mahwah, NJ: Lawrence Erlbaum.

Regal, D. M., & Whittington, D. H. (1994). Synthetic vision display evaluation studies (NASA-CR-194963). Hampton, VA: NASA Langely.

Rogers, W. (1994). The effect of an automated fault management decision aid on pilot situation awareness. Paper presented at the First Automation and Human Performance Conference, Washington, DC.
Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. The International Journal of Aviation Psychology, 1(1), 45-57.
Selcon, S. J., & Taylor, R. M. (1990). Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design. In Situational Awareness in Aerospace Operations (AGARD-CP-478) (pp. 5/1 -5/8). Neuilly Sur Seine, France: NATO - AGARD.
Selcon, S. J., Taylor, R. M., & Koritsas, E. (1991). Workload or situational awareness?: TLX vs SART for aerospace systems design evaluation. In Proceedings of the Human Factors Society 35th Annual Meeting (pp. 62-66). Santa Monica, CA: Human Factors Society.
Sullivan, C., & Blackman, H. S. (1991). Insights into pilot situation awareness using verbal protocol analysis. In Proceedings of the Human Factors Society 35th Annual Meeting (pp. 57-61). Santa Monica, CA: Human Factors Society.
Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational Awareness in Aerospace Operations (AGARD-CP-478) (pp. 3/1 - 3/17). Neuilly Sur Seine, France: NATO - AGARD.
Tenney, Y. T., Adams, M. J., Pew, R. W., Huggins, A. W. F., & Rogers, W. H. (1992). A principled approach to the measurement of situation awareness in commercial aviation (NASA Contractor Report 4451). Langely, VA: NASA Langely Research Center.
Vidulich, M. (2000). Testing the sensitivity of situation awareness metrics in interface evaluations. In M. R. Endsley & D. J. Garland (Eds.), Situation Awareness Analysis and Measurement. Mahwah, NJ: Lawrence Erlbaum.
Vidulich, M. A., & Hughes, E. R. (1991). Testing a subjective metric of situation awareness. In Proceedings of the Human Factors Society 35th Annual Meeting (pp. 1307-1311). Santa Monica, CA: Human Factors Society.
Waag, W. L., & Houck, M. R. (1994). Tools for assessing situational awareness in an operational fighter environment. Aviation, Space and Environmental Medicine, 65(5, supl), A13-A19.
Wells, M. J., Venturino, M., & Osgood, R. K. (1988). Using target replacement performance to measure spatial awareness. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 1429-1433). Santa Monica, CA: Human Factors Society.
Wickens, C. D. (1995). The tradeoff of design for routine and unexpected performance: Implications of situation awareness. In D. J. Garland & M. R. Endsley (Eds.), Experimental analysis and measurement of situation awareness (pp. 57-64). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.
Wickens, C. D., Liang, C. C., Prevett, T., & Olmos, O. (1996). Electronic maps for terminal area navigation: Effects of frame of reference and dimensionality. The International Journal of Aviation Psychology, 6(3), 241-271.
Zenyuh, J. P., Reising, J. M., Walchli, S., & Biers, D. (1988). A comparison of stereographic 3-D display versus a 2-D display using an advanced air-to-air format.

In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 53-57). Santa Monica, CA: Human Factors Society.

Appendix A: SAGAT Queries for Each Phase of Flight

	Phase of Flight	Preflight/ start-up	taxi out	take off	climb out	cruise	descent	approach	landing	taxi in	shut down
1	Heading			■	■	■	■	■	■		
2	Altitude			■	■	■	■	■	■		
3	Airspeed			■	■	■	■	■	■		
4	Climb/descent			■	■	■	■	■	■		
5	Attitude			■	■	■	■	■	■		
6	Winds		■	■	■	■	■	■	■		
7	Fuel state		■	■	■	■	■	■	■	■	
8	Planned altitude				■	■	■	■			
9	Planned speed			■	■	■	■	■	■		
10	Planned heading			■	■	■	■	■	■		
11	Planned position			■	■	■	■	■	■		
12	Clearance conformance		■	■	■	■	■	■	■	■	
13	Conflicting traffic		■	■	■	■	■	■	■	■	
14	Traffic location		■	■	■	■	■	■	■	■	
15	Traffic problem		■	■	■	■	■	■	■	■	
16	Obstacles/terrain			■	■	■	■	■			
17	SUA			■	■	■	■	■			
18	Weather			■	■	■	■	■	■		
19	Weather impact			■	■	■	■	■	■		
20	Glide slope							■	■		
21	Runway touchdown								■		
22	Runway stop								■		
23	Bearing/Range to next fix				■	■	■				
24	Bearing/Range to airport						■	■	■		
25	Acceleratin/deceleration			■	■	■	■	■	■		
26	Closure rate			■	■	■	■	■	■		

Appendix B: SART

Instability of Situation

How changeable is the situation? Is the situation highly unstable and likely to change suddenly (high), or is it very stable and straight forward (low)?

Low |———————————————| **High**

Complexity of Situation

How complicated is the situation? Is it complex with many inter-related components (high) or is it simple and straightforward (low)?

Low |———————————————| **High**

Variability of Situation

How many variables are changing in the situation? Are there are large number of factors varying (high) or are there very few variables changing (low)?

Low |———————————————| **High**

Arousal

How aroused are you in the situation? Are you alert and ready for activity (high) or do you have a low degree of alertness (low)?

Low |———————————————| **High**

Concentration of Attention

How much are you concentrating on the situation? Are you bringing all your thoughts to bear (high) or is your attention elsewhere (low)?

Low |———————————————| **High**

Division of Attention

How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (high) or focussed on only one (low)?

Low **High**

Spare Mental Capacity

How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (high) or nothing to spare at all (low)?

Low **High**

Information Quantity

How much information have you gained about the situation? Have you received and understood a great deal of knowledge (high) or very little (low)?

Low **High**

Information Quality

How good is the information you have gained about the situation? Is the knowledge communicated very useful (high) or is it a new situation (low)?

Low **High**

Familiarity with Situation

How familiar are you with the situation? Do you have a great deal of relevant experience (high) or is it a new situation (low)?

Low **High**

Appendix C: SA-SWORD

Please compare each of the four display concepts you experienced flying with based on the degree to which it provided superior situation awarenes. in comparison to the other display concepts for each phase of flight. Make a check mark in the column corresponding to your preference for one display for providing superior situation awareness as compared to another in each row.

CLIMB OUT	Absolute		Very Strong		Strong		Weak		Equal		Weak		Strong		Very Strong		Absolute	
SVS Display A																		SVS Display B
SVS Display A																		SVS Display C
SVS Display A																		SVS Display D
SVS Display B																		SVS Display C
SVS Display B																		SVS Display D
SVS Display C																		SVS Display D

CRUISE	Absolute		Very Strong		Strong		Weak		Equal		Weak		Strong		Very Strong		Absolute	
SVS Display A																		SVS Display B
SVS Display A																		SVS Display C
SVS Display A																		SVS Display D
SVS Display B																		SVS Display C
SVS Display B																		SVS Display D
SVS Display C																		SVS Display D

APPROACH	Absolute		Very Strong		Strong		Weak		Equal		Weak		Strong		Very Strong		Absolute	
SVS Display A																		SVS Display B
SVS Display A																		SVS Display C
SVS Display A																		SVS Display D
SVS Display B																		SVS Display C
SVS Display B																		SVS Display D
SVS Display C																		SVS Display D

DESCENT	Absolute		Very Strong		Strong		Weak		Equal		Weak		Strong		Very Strong		Absolute	
SVS Display A																		SVS Display B
SVS Display A																		SVS Display C
SVS Display A																		SVS Display D
SVS Display B																		SVS Display C
SVS Display B																		SVS Display D
SVS Display C																		SVS Display D

LANDING	Absolute		Very Strong		Strong		Weak		Equal		Weak		Strong		Very Strong		Absolute	
SVS Display A																		SVS Display B
SVS Display A																		SVS Display C
SVS Display A																		SVS Display D
SVS Display B																		SVS Display C
SVS Display B																		SVS Display D
SVS Display C																		SVS Display D

Appendix D: Confidence Rating Form

Indicate how confident you are in your knowledge of your **separation from obstacles**

Low **High**

Indicate how confident you are in your knowledge of your **separation from terrain**

Low **High**

Indicate how confident you are in your knowledge of your **separation from air traffic**

Low **High**

Indicate how confident you are in your knowledge of your aircraft's **flight parameters and flight vector**

Low **High**

Indicate how confident you are in your knowledge of your **adherence to clearances and assigned flight path**

Low **High**

Indicate how confident you are in your **assessment of flight hazards**

Low **High**

Indicate how confident you are in your **knowledge of relevant weather information**

Low High