Situation awareness measurement: a review of applicability for C4i environments

Salmon, Paul M; Stanton, Neville A; Walker, Guy H; et.al.

Salmon, P. M., Stanton, N. A., Walker, G. H., & Green, D. (2006). Situation awareness measurement: a review of applicability for C4i environments. Applied Ergonomics, 37(2), 225–238.
https://doi.org/10.1016/j.apergo.2005.02.001
Document Type: Accepted Version

Link to Published Version: https://doi.org/10.1016/j.apergo.2005.02.001

Situation Awareness Measurement: A review of applicability for C4i environments

Paul Salmon*, Prof Neville Stanton, Dr Guy Walker & Dr Damian Green
Defence Technology Centre for Human Factors Integration (DTC-HFI)
Brunel University
BIT Lab, School of Engineering and Design, Uxbridge, Middlesex, UB8 3PH
Tel: 01985 265546
*Corresponding author, email: paul.salmon@brunel.ac.uk

Abstract

The construct of situation awareness (SA) has become a core theme within the human factors (HF) research community. Consequently, there have been numerous attempts to develop reliable and valid measures of SA. Despite this, it is apparent that there are a lack of techniques that have been developed specifically for the assessment of SA in C4i (command, control, communication, computers and intelligence) environments. During the design, development and evaluation of novel systems, technology and procedures, valid and reliable situation awareness measurement techniques are required for the assessment of individual and team SA. SA is assessed in order to determine the improvements (or in some cases decrements) resulting from proposed design and technological interventions. The following paper presents a review of existing situation awareness measurement techniques conducted in order to assess their suitability for use in the assessment of SA in C4i environments. Seventeen SA measures were evaluated against a set of HF methods criteria. It was concluded that current SA measurement techniques are inadequate for use in the assessment of SA in C4i environments, and a multiple-measure approach utilising different approaches is recommended.

Keywords: Situation Awareness, measurement, C4i systems

Introduction

C4i (command, control, communication, computers and intelligence) systems are comprised of both human and technological agents and are designed to gather information and facilitate the accurate communication of this information between multiple agents dispersed across multiple locations. White and Harris (1987) offer the following definition of C4i systems;

> "Command, control, communication, computers and intelligence is the management infrastructure for defence and war or any other large or complex, dynamic resource system […] it is intrinsically a diverse range of activities" (White and Harris 1987)

Examples of C4i systems range from simplistic command and control infrastructures such as those seen in the emergency services (police and fire brigade) to more sophisticated networks, such as military network enabled capability infrastructures. An example of a simple C4i system would be a gold (fixed command post), silver (mobile command units) and bronze (agents in the field) command hierarchy. According to Smith (2003), the goal of C4i systems is to provide more appropriate information, more quickly and in the correct format, to the relevant agents. Furthermore Smith (2003) argues that one of the most challenging aspects for C4i systems and concepts is determining whether and by how much they increase the effectiveness of the military force that utilises them. Of course, the design of any novel system requires rigorous scientific testing to establish the performance gains (or decrements) associated with its use. The design and development of a novel C4i system raises pertinent HF issues, including mental workload, error and situation awareness (SA), all of which should be assessed thoroughly throughout the design process. The level of SA that the C4i system provides to the agents involved in its use is a particularly crucial aspect, with potentially catastrophic consequences associated with poor or loss of SA. Within the military, the problem of 'blue-on-blue' or friendly fire is becoming more and more prominent, and so the level of awareness that agents possess during battlefield operations is placed under increasing scrutiny. During the first gulf war conflict, 35 (24%) of the 146 casualties suffered by US forces were caused by friendly fire incidents (Ripley 2003) and 9 (38%) of the 24 casualties suffered by British forces were attributed to friendly fire (Cooper 2003). British forces have also suffered tragic losses caused by friendly fire during the current Gulf conflict. Two British soldiers were killed when their Challenger II tank was mistakenly fired upon by another British tank during fighting outside of Basra, whilst two RAF pilots were killed when their GRU Tornado was erroneously shot down by a US patriot missile (BBC News, 2003). The problem of friendly fire and operator SA may be linked, in that a lack of awareness of both own and enemy forces may result in such mishaps, and so the accurate assessment of SA during the design of such systems is imperative.

A number of techniques designed to assess operator SA already exist and have been used extensively in the past, such as SAGAT (Endsley 1995b) and SART (Taylor 1990). However, a valid and reliable technique designed to measure operator SA in C4i environments is yet to emerge. This paper presents a review of existing SA measures that was conducted in order to determine their suitability for use in the assessment of SA in C4i systems.

Situation Awareness

The construct of SA has received considerable attention from the psychology and HF communities over the past twenty years. Although the original impetus for research into the construct came from within the military aviation domain, the construct has since developed into a critical research theme in almost any domain that involves humans performing tasks in complex, dynamic systems. SA research is widespread and ongoing in a variety of domains, including military operations (Endsley et al 2000, Matthews et al 2000), aviation (Kaber et al 2002, Keller et al 2004), air traffic control (ATC) (Hauss & Eyferth 2003, Endsley and Smolensky 1998), automotive (Zheng, McConkie and Tai 2004), and C4i environments (Walker et al 2004).

Despite numerous attempts, a universally accepted definition and model of SA is yet to emerge. According to Stanton, Chambers & Piggott (2001), three definitions and their associated theoretical perspectives dominate. These are the three-level model (Endsley, 1995a), the perceptual cycle model (Smith & Hancock 1995) and the activity theory model (Bedny & Meister 1999). The three-level model (Endsley 1995a) depicts SA as a product comprised of three hierarchical levels. The perceptual-cycle model (Smith & Hancock 1995) describes the cyclical process of achieving and updating SA, suggesting that SA resides through the interaction of the person with the world (Smith & Hancock 1995). The activity theory of SA (Bedny & Meister 1999) describes SA through an activity theory perspective, defining SA as an individual's conscious dynamic reflection on the situation. The main point of contention between theoretical perspectives lies in whether SA refers to the *processes* employed in achieving and maintaining it or the end *product* of SA, derived as a result of these processes. The three-level model proposed by Endsley (1995a) describes SA as a *product* comprised of three hierarchical levels, separate from those *processes* (labelled situation assessment) used to achieve it. Alternatively, the perceptual cycle model proposed by Smith & Hancock (1995) purports that SA resides through the interaction of the person with the world (Smith & Hancock 1995) and describes SA both in terms of the cognitive *processes* used to engineer it and also the continuously updating *product* of SA. The process versus product debate has a huge impact upon the measurement of SA in terms of what it is that should be measured when considering SA. Process oriented models would argue that the processes used to achieve and maintain SA should be measured, whilst the product models would suggest that the level or amount of SA achieved should be measured.

The three-level model of SA proposed by Endsley (1995a) is by far the most commonly cited and used model of SA, and for the purposes of this methods review is adequate for describing the construct, since the majority of SA measurement techniques are based upon the model. Endsley (1995a) defines SA as,

> "The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley 1995a)

The three-level model describes SA through an information processing approach, and proposes that operator SA is an internally held product comprising three hierarchical levels that is separate to those processes (termed situation assessment) used to achieve it. According to Endsley (1995a), operator SA consists of level 1 SA (the perception of the elements in the environment, level 2 SA (the comprehension of their meaning) and level 3 SA (the projection of their future status). The three-level model of SA is presented in figure 1. The model depicts SA as an essential component of human decision-making activity. The achievement and maintenance of SA is influenced by both individual and task factors, such as experience, training, workload and also interface design. The model is intuitive and neat, and the three hierarchical levels are particularly useful for the measurement of the construct. As a result of this, the majority of existing SA measurement techniques are based upon the three-level model.

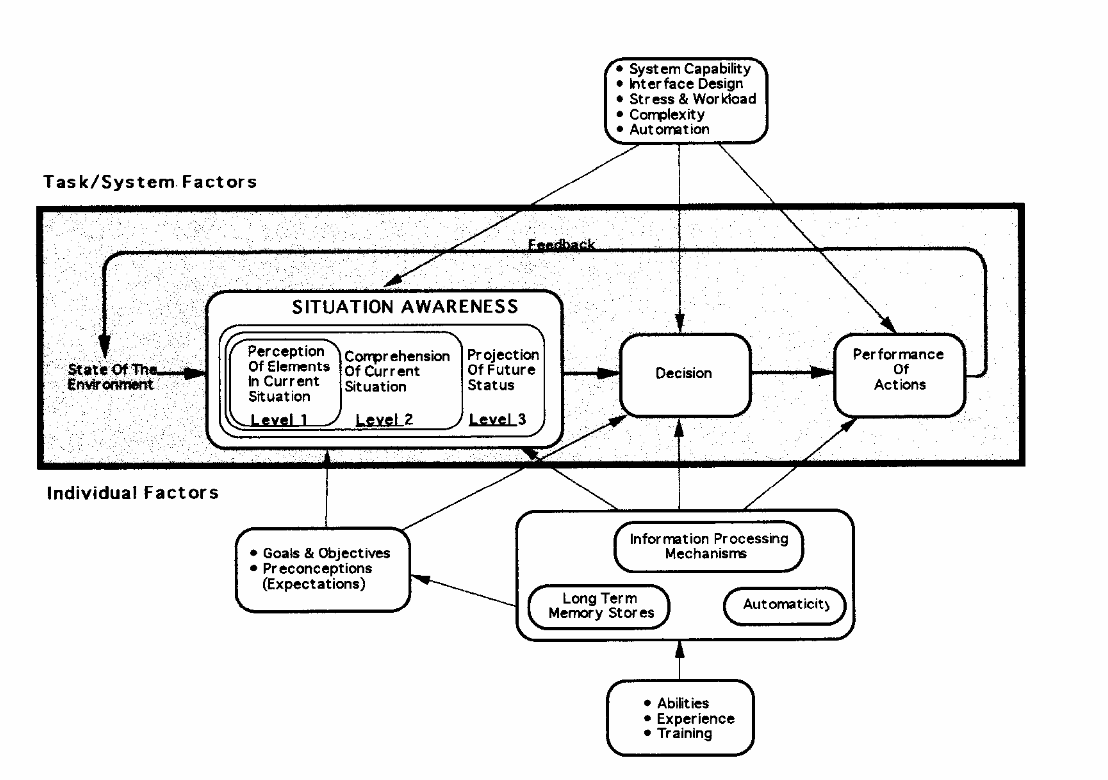

Figure 1. The three-level model of SA (Source: Endsley 1995a)

Team or shared SA

A more recent theme to emerge within the construct of SA is the concept of team or shared SA. The increased use of teams in complex environments has shifted the focus from individual operator SA onto the shared SA of teams of operators. Salas (2004) defines a team as consisting of two or more people, dealing with multiple information resources, who work to accomplish some shared goal. Cooke (2004) suggests that teams are required to detect and interpret cues, remember, reason, plan, solve problems, acquire knowledge and make decisions as an integrated and co-ordinated unit. The concept of team SA first began to emerge within the aviation community and has since received considerable attention from the military domain. Team or shared SA reflects the co-ordinated awareness that the team possesses as a whole unit. However, as Salas et al (1995) point out, there appears to be a lot more to the concept than merely combining individual team member SA. There have been a number of attempts to define team SA, although a universally accepted definition is yet to appear. Salas et al (1995) describe team SA as,

> "At least in part the shared understanding of a situation among team members at one point in time" (Salas et al 1995)

Endsley (1995a) refers to team SA as the degree to which every team member possesses the SA required for his or her responsibilities. According to Klein (2000) shared situation awareness refers to the degree to which the team members have the same interpretation of ongoing events. In their description of SA as distributed cognition, Artman & Garbis (1998) go further to suggest that the SA of a team is distributed not only throughout the agents comprising the team, but also in the artefacts that they use in order to accomplish their goals. According to Artman & Garbis (1998), existing models of SA are inadequate when considering team-based tasks. They also argue that in domains such as military command, teamwork is essential for success and thus a non-individual approach to the assessment of SA is necessary. Artman and Garbis (1998) define team SA as,

> “The active construction of a model of a situation partly shared and partly distributed between two or more agents, from which one can anticipate important future states in the near future” (Artman & Garbis 1998)

The concept of team SA is not yet clearly defined nor understood. Further, evidence of the accurate measurement of team SA in the literature is sparse. It is apparent that the majority of existing SA measurement approaches focus upon the assessment of individual operator SA. In a review of team SA literature, Salas et al (1995) discovered that of existing measurement approaches, most were designed to measure individual team member SA. Whilst the ability to assess individual team member SA is an important provision when considering team SA in C4i environments, the underlying aim has to be the measurement of overall or shared team SA during task performance. However, this has proved more elusive than the measurement of individual SA and an approach designed to assess team SA is yet to emerge. The individualistic approach to the assessment of SA used in the past may no longer suffice, and it is new team-based SA measures that require further investigation.

Measuring SA

The provision of valid and reliable methods for assessing SA is essential during system design and evaluation. One of the primary goals of any system design effort is to enhance the SA of the personnel involved in the operation of the system. In order to ensure that SA is improved and not degraded by the design effort, valid and reliable techniques that can accurately measure operator and team SA are required throughout the design process. The measurement of individual operator SA has been successfully undertaken across a wide variety of domains since the dawn of the construct.

Techniques such as SAGAT (Endsley 1995b) and SART (Taylor 1990) have been applied in a number of areas, including military aviation (Endsley 1995b), air traffic control (Endsley and Kiris 1995), military operations (Matthews et al 2000), driving (Walker, Stanton and Young 2004) and the process industry (Hogg et al 1995).

As the construct of SA has become more and more eminent, attempts have been made to develop more sophisticated SA measurement approaches. In a review of SA measurement techniques, Endsley (1995b) describes a number of different approaches, including physiological measurement techniques, performance measures (external task measures and imbedded task measures), subjective rating techniques (self and observer rating), questionnaires (post-trial and on-line) and the freeze technique (SAGAT).

Reliability and Validity

A recurrent issue when discussing the selection and application of human factors techniques is the problem surrounding the validation of the techniques. When using HF techniques during the design lifecycle, one has to ensure that the technique in question actually works. According to Stanton & Young (1999, 2003), despite the increased number of HF techniques, there is little evidence that the methods actually work, and the way in which to ensure that they work is to assess their reliability and validity. When measuring SA, reliability refers to the degree to which the measure will generate the same data when measuring SA under the same conditions over and over again. Validity refers to the extent to which the measure is actually measuring SA, and not some other psychological process or product. Endsley (1995b) reports that when considering SA measurement techniques, it is necessary to establish that the technique:

1) Measures SA and does not measure other processes or factors.

2) Possesses the required level of sensitivity i.e. the technique can accurately detect changes in SA caused by novel technologies and programmes.

3) Does not alter SA during the measurement procedure.

The validation of HF techniques such as SA measurement techniques, whilst inherently necessary, is often neglected. This is for a number of reasons, mainly the high cost and resources invested when conducting validation studies. Stanton & Young (1999) also point out that researchers tend to stick with methods that they know and trust (often methods that they developed themselves), and so validation is assumed, rather than tested. From a previous HF methods review (Salmon et al 2004) it was discovered that the majority of HF techniques available in the open literature are developed and then subjected to an initial validation study, and then discarded, never to be used again. Those techniques that are successful enough to be used elsewhere are often the techniques that are extensively validated. Typically, a handful of techniques within an application area emerge as the most commonly used and extensively validated. For example, SHERPA is by far the most commonly used human error identification (HEI) technique, and has a number of promising validation studies associated with it (Whalley & Kirwan 1989, Kirwan 1992, Baber & Stanton 1996, Stanton & Baber 2001, Stanton & Stevenage 2000). In the measurement of mental workload, the NASA-TLX (Hart & Staveland 1988) is the most commonly used and also the most widely validated of the various techniques available. The measurement of SA is no different, with the situation awareness global assessment technique (SAGAT) (Endsley 1995b) being by far the most commonly used approach, and also the technique with the most associated validation evidence. This highlights a potential problem: if new SA measurement techniques do not quickly catch on, there may not be any attempts to validate it. As a result, advances in the measurement of SA may become stilted, as practitioners will tend to use the familiar methods (Stanton & Young 1999).

SA measurement in C4i systems

The measurement of SA in C4i environments poses a great challenge to the human factors community. As described previously, the environment is typically complex, dynamic and information rich. Due to the collaborative and dispersed nature of C4i environments, an assessment of both individual and team (or shared SA) is required in order to provide accurate measures of SA. As a result, any method that is used to measure operator and team SA in C4i environments should possess three distinct capabilities. Firstly, the technique should be capable of measuring SA simultaneously at different geographical locations. In order to gain a true measure of team of shared SA, each agent involved should be simultaneously assessed for their SA. However, due to the dispersed nature of C4i environments, agents are typically remote from one another. Therefore, SA should also be assessed at different geographical locations. Therefore, any technique used to assess SA in C4i environments should be capable of simultaneous administration at different locations. For example, the level of SA at different command locations (command centre, mobile units and foot units) may need to be assessed to ensure that the team involved has an adequate level of shared SA and that task relevant information is communicated efficiently. This would require a concurrent assessment of SA at the command centre, the mobile units, and also commanders in the field. Secondly, the technique should be capable of measuring both individual and team or shared SA. Individual team members may possess individual goals, mental models and SA, whilst simultaneously pursuing team goals, and maintaining a level of team or shared SA. The SA technique used should be able to cater for the two different 'types' of SA, or at least offer separate approaches for each type of SA. Thirdly, the technique should be capable of measuring SA in real-time. Typically, simulations of scenarios are used in order to assess SA. However, due to the dynamic, collaborative and dispersed nature of C4i scenarios, it appears that this may not be possible, and exercises conducted 'in-the-field' may be used. As a result, simulations of task scenarios and querying SA during task 'freezes' may not be appropriate.

SA Methods Review

The aim of the SA methods review was to develop an understanding of existing SA measurement techniques and also to determine whether any of the existing approaches could potentially be used in the assessment of SA in C4i environments. An initial literature review was conducted in order to create an exhaustive database of existing SA measurement techniques. The literature review was based upon a survey of standard ergonomics textbooks, relevant scientific journals and existing HF method reviews. The literature survey identified over thirty existing SA measurement techniques. A screening process was then employed in order to select the most appropriate techniques for further analysis. The screening process was based upon technique availability, make-up and applicability to C4i, and was designed to quickly select or reject techniques from the initial database. As a result of the screening process, seventeen SA measurement techniques were selected for further analysis (see table 1). Each technique was then evaluated against a set of HF methods criteria adapted from Stanton et al (2004) designed to determine its suitability for use in the assessment of SA in C4i systems (e.g. domain of application, training and application times, tools needed, reliability and validity, advantages and disadvantages etc). A number of the criteria used were also descriptive, allowing the output to act as a user manual for each technique (e.g. background and applications, procedure and advice, flowchart etc). The following categories of SA measurement technique were assessed.

1) SA requirements analysis
2) Freeze probe techniques
3) Real-time probe techniques
4) Self-rating techniques
5) Observer-rating techniques
6) Performance measures
7) Process Indices (Eye Tracker)

A brief description of each of the different categories and the methods reviewed within each category is presented below.

SA requirements analysis

An SA requirements analysis represents the first step in any assessment of SA and is conducted in order to determine what actually comprises operator SA in the task or environment under analysis. Endsley (1993) describes a generic procedure for conducting an SA requirements analysis that involves the use of unstructured interviews with SME's (subject matter experts), goal-directed task analysis and questionnaires in order to determine the relevant SA requirements. The output of an

SA requirements analysis is then used during the development of the SA assessment technique used, in order to determine which elements comprise operator SA and thus, what should be assessed.

Freeze probe techniques

Freeze probe techniques involve the administration of SA related queries on-line during 'freezes' in a simulation of the task under analysis. Typically, a task is randomly frozen and a set of SA queries regarding the current situation are administered. The participant is required to answer each query based upon his knowledge of the situation at the point of the freeze. During these 'freezes' all operator displays and windows are typically blanked. A computer is used to select and administer the queries and also to record the responses. The primary advantage associated with the use of freeze probe techniques is their direct nature. However, freeze probe techniques are criticised for their intrusion upon primary task performance, and also can only be applied where there is a simulation of the task under analysis.

The situation awareness global assessment technique (SAGAT) is the most popular freeze probe technique that was developed to assess pilot SA across the three levels of SA proposed in the three-level model (Endsley 1995a). SAGAT comprises a set of queries of designed to assess participant SA, including level 1 SA (perception of the elements), level 2 SA (comprehension of their meaning) and level 3 SA (projection of future status). Although developed specifically for use in the military aviation domain, a number of different versions of SAGAT exist, including a specific air-to-air tactical aircraft version (Endsley, 1990), an advanced bomber aircraft version (Endsley, 1989) and an Air traffic control version (Endlsey & Kiris 1995). SALSA (Hauss & Eyferth

2003) is another freeze probe technique that was developed specifically for use in air traffic control. The SALSA queries are based upon fifteen aspects of aircraft flight, such as flight level, ground speed, heading, vertical tendency, conflict and type of conflict. The situation awareness control room inventory (SACRI) is an adaptation of SAGAT (Endsley 1995b) and uses the freeze technique to administer control room based SA queries. SACRI was developed as the result of a study investigating the use of SAGAT in process control rooms (Hogg et al 1995).

Real-time probe techniques

An alternative approach to the use of highly intrusive freeze probe techniques is the use of real-time probe techniques. Real-time probe techniques involve the administration of SA related queries on-line (during task performance), but with no freeze of the task under analysis. Typically, subject matter experts (SME's) develop queries either prior to or during task performance and administer them, without a freeze, at the relevant points during task performance. Answer content and response time are taken as a measure of participant SA. Based upon a comparison of real-time probes, SAGAT and SART when used to measure operator SA in war and peace scenarios, Jones and Endsley (2000) report that when there is no simulation of the system under analysis and the task cannot be frozen, real-time probes may provide a viable option for measuring SA. It is argued that the main advantage associated with these 'real-time' probe techniques are reduced intrusiveness, due to the fact that no freeze in the task under analysis is required. However, real-time probe queries may also serve to direct participant attention to the required elements in the environment, resulting in biased data. The alleged reduction in intrusion is also questionable, as receiving queries and responding to them still imposes a degree of intrusion upon the primary task.

The situation present assessment method (SPAM) (Durso et al 1998) is a real-time probe technique developed for use in the assessment of air traffic controllers SA. The technique involves the use of on-line real time probes to evaluate operator SA. The analyst probes the operator for SA using task related SA queries based on pertinent information in the environment (e.g. which of the two aircraft A or B, has the highest altitude?) via telephone. The query response time (for those responses that are correct) is taken as an indicator of the operators SA. Additionally, the time taken to answer the telephone acts as an indicator of workload. SASHA (Jeannot, Kelly & Thompson 2003) is a methodology developed by Eurocontrol™ for the assessment of air traffic controllers SA in automated systems. The methodology consists of two techniques, SASHA_L (real-time probe technique) and SASHA_Q (post-trial questionnaire). SASHA_L is based upon the SPAM technique (Durso et al 1998), and involves probing the participant on-line using real-time SA related queries. The response content and response time is recorded. Once the trial is completed, the participant completes the SASHA_Q questionnaire, which consists of ten questions designed to assess participant SA.

Self-rating techniques

Self-rating techniques are used to gain a subjective assessment of participant SA. Typically administered post-trial, self-rating techniques involve participants providing a subjective rating of their perceived SA via an SA related rating scale. The primary advantages of self-rating techniques are their ease of application (easy, quick and low cost) and their non-intrusive nature (since they are administered post-trial). However, subjective self-rating techniques are heavily criticised for a plethora of reasons, including the various problems associated with the collection of SA data post-trial (correlation of SA with performance, poor recall etc) and also issues regarding their sensitivity.

The situation awareness rating technique (SART) (Taylor 1990) is a subjective rating technique that was originally developed for the assessment of pilot SA. SART uses the following ten dimensions to measure operator SA: Familiarity of the situation, focussing of attention, information quantity, information quality, instability of the situation, concentration of attention, complexity of the situation, variability of the situation, arousal, and spare mental capacity. SART is administered post-trial and involves the participant rating each dimension on a seven point rating scale (1 = Low, 7 = High) in order to gain a subjective measure of SA. The ten SART dimensions can also be condensed into the quicker 3 dimensional (3-D) SART, which involves participants rating attentional demand, attentional supply and understanding. The situation awareness rating scales technique (SARS) (Waag & Houck 1994) is a subjective rating technique that was developed for the military aviation domain. When using the SARS technique, participants subjectively rate their performance on a six-point rating scale (from acceptable to outstanding) for 31 facets of fighter pilot SA. The SARS SA categories and associated behaviours were developed from interviews with experienced F-15 pilots (Waag & Houck 1994). The 31 SARS behaviours are divided into 8 categories representing phases of mission performance. The eight categories are: general traits (e.g. decisiveness, spatial ability), tactical game plan (e.g. developing and executing plan), communication (e.g. quality), information interpretation (e.g. threat prioritisation), tactical employment beyond visual range (e.g. targeting decisions), tactical employment visual (e.g. threat evaluation) and tactical employment general (e.g. lookout, defensive reaction). According to Waag & Houck (1994) the 31 SARS behaviours represent those that are crucial to mission success. The Crew awareness rating scale (CARS) (McGuiness & Foy 2000) technique has been used to assess command and control commanders SA and workload (McGuinness & Ebbage 2000). The CARS technique comprises two separate sets of questions based upon the three level model of SA (Endsley 1995a). The content subscale consists of three statements designed to elicit ratings based upon ease of identification, understanding and projection of task SA elements (i.e. levels 1, 2 and 3 SA). The fourth statement is designed to assess how well the participant identifies relevant task related goals in the situation. The workload subscale also consists of four statements, which are designed to assess how difficult, in terms of mental effort, it is for the participant in question to identify, understand, project the future states of the SA related elements in the situation. CARS is administered post-trial and involves participants rating each category on a scale of 1 (ideal) to 4 (worst) (McGuinness & Ebbage 2000). The mission awareness rating scale (MARS) (Matthews & Beal 2002) technique is a development of the CARS technique (McGuiness & Foy 2000) designed specifically for use in the assessment of SA in military exercises . The MARS technique was developed for use in 'real world' field settings, rather than in simulations of military exercises (Matthews & Beal 2002). The technique is normally administered post-trial, upon completion of the task or mission under analysis. The quantitative analysis of situational awareness (QUASA) technique (McGuinness 2004) combines participant self-ratings with on-line probes in order to assess actual and perceived SA in military command and control environments. Participants are probed for their SA during task performance and then simultaneously asked to rate their confidence in their answer to the probe in question.

QUASA uses true or false probes and a confidence ratings scale (Very low – Very high) in order to assess actual and perceived SA. The Cranfield situation awareness scale (C-SAS) (Dennehy 1997) is a subjective rating scale that is used to assess student pilot SA during flight training exercises. C-SAS is administered either during or post-trial and involves participants rating five SA related components on an appropriate rating scale. Each rating scale score is then summed in order to determine an overall SA score.

Observer rating techniques

Observer rating techniques are most commonly used to assess SA during tasks performed 'in-the-field'. Observer rating techniques typically involve a subject matter expert (SME) observing participants performing the task under analysis and then providing an assessment or rating of each participants SA. The SA ratings are based upon observable SA related behaviour exhibited by the participants during task performance. The main advantages associated with the use of observer rating scales to measure SA are their non-intrusive nature and their ability to be applied 'in-the-field'. However, the extent to which observers can accurately rate participant SA is questionable, and also multiple SME's are required. The situation awareness behavioural rating scale (SABARS) is an observer rating technique that has been used to assess infantry personnel situation awareness in field training exercises (Matthews, et al 2000, Matthews & Beal 2002). The technique involves domain experts observing participants during task performance and rating them on 28 observable SA related behaviours. A five point rating scale (1=Very poor, 5 =Very good) and an additional 'not applicable' category are used. The 28 behaviour rating items are designed specifically to assess platoon leader SA (Matthews et al 2000).

Performance measures

Using performance measures to assess SA involve measuring relevant aspects of participant performance during the task under analysis. Depending upon the task, certain aspects of performance are recorded in order to determine an indirect measure of SA. For example, in a military infantry exercise, performance measures may be 'kills', 'hits' or mission success or failure. In an assessment of driver SA, Gugerty (1997) measured hazard detection, blocking car detection, and crash avoidance during a simulated driving task. Whilst performance measures are simple to obtain and are non-intrusive as they are generated through the natural flow of the task, they are beset by a number of problems concerning the relationship between SA and performance. For example, an expert participant may be able to achieve acceptable performance even when his SA is inadequate. Similarly, a novice participant may possess superior levels of SA but still achieve inferior performance, due to other factors such as inexperience.

Process indices

Process indices can also be used to measure SA. Process indices involve recording the processes that the participants use in order to develop SA during the task under analysis. One example of using process indices to assess SA is the measurement of participant eye movements during task performance (Smolensky 1993). An eye-tracking device can be used to measure participant fixations during the task under analysis, which can then be used to determine how the participant's attention was allocated during the task under analysis. There are a number of disadvantages associated with the use of an eye-tracker device, including their in-direct nature (how do we no the participant perceived what they looked at?), an inability to be used outside of laboratory settings, the temperamental nature of the equipment, and also the problem of the ‘look but do not see’ phenomenon, whereby participants may fixate upon an environmental element but do not actually perceive it. Concurrent verbal protocol analysis (VPA) involves creating a written transcript of operator behaviour as they perform the task under analysis. The transcript is based upon the operator ‘thinking aloud’ as he conducts the task under analysis. VPA is used as a means of gaining an insight into the cognitive aspects of complex behaviours and is often used to indicate operator SA during task performance.

A summary of the methods review is presented in table 1. For the full methods review, readers should contact the authors of this paper.

Table 3. Summary of SA measurement techniques review

Method	Type of method	Domain	Team	SME's required	Training time	Application time	Tools needed	Validation studies	Advantages	Disadvantages
CARS McGuinness & Foy (2000)	Self rating technique	Military (infantry operations)	No	No	Low	Low	Pen and paper	Yes 2	1) Developed for use in infantry environments. 2) Less intrusive than on-line techniques. 3) Quick, easy to use requiring little training.	1) Construct validity questionable. 2) Limited evidence of use and validation. 3) Problems of gathering SA data post-trial e.g. correlation with performance, forgetting low SA periods.
MARS Matthews & Beal (2002)	Self rating technique	Military (infantry operations)	No	No	Low	Low	Pen and paper	Yes 2	1) Developed for use in infantry environments. 2) Less intrusive than on-line techniques. 3) Quick, easy to use requiring little training.	1) Construct validity questionable. 2) Limited evidence of use and validation. 3) Problems of gathering SA data post-trial e.g. correlation with performance, forgetting low SA periods.
SABARS Matthews & Beal (2002)	Observer rating	Military (infantry operations)	No	Yes	High	Med	Pen and paper	Yes 2	1) SABARS behaviours generated from infantry SA requirements exercise. 2) Non-intrusive.	1) Extent to which observers can accurately rate internal construct of SA is questionable. 2) The presence of observers may influence participant behaviour. 3) Access to SME's and field settings is required.
SACRI Hogg et al (1995)	Freeze on-line probe technique	Nuclear Power	No	No	Low	Med	Simulator Computer	Yes 1	1) Removes problems associated with collecting SA data post-trial. 2) Direct approach.	1) Requires expensive simulators. 2) Intrusive to primary task performance. 3) Cannot be applied 'in-the-field'.
SAGAT Endsley (1995b)	Freeze on-line probe technique	Aviation (military)	No	No	Low	Med	Simulator Computer	Yes 10+	1) Direct approach. 2) Subject to numerous validation studies. 3) Removes problems associated with collecting SA data post-trial	1) Requires expensive simulators. 2) Intrusive to primary task. Difficult to see how it would work in C4 environments. 3) Cannot be applied 'in-the-field' or in real-time.
SALSA Hauss & Eyferth (2003)	Freeze on-line probe technique	ATC	No	No	Low	Med	Simulator Computer	Yes 1	1) Removes problems associated with collecting SA data post-trial e.g. correlation with performance, forgetting etc.	1) Requires expensive simulators. 2) Intrusive to primary task performance. 3) Limited use and validation.
SASHA Jeannott, Kelly & Thompson (2003)	Real-time probe technique Post-trial quest	ATC	No	Yes	High	Med	Simulator PC Telephone Pen and paper	No	1) Offers two techniques for the assessment of SA. 2) Administering probes in real-time removes the need for task freezes, and allows the technique to be applied 'in the field'.	1) Probes may direct attention to required elements. 2) Generation of appropriate SA queries places great burden upon analyst/SME. 3) Limited evidence of use or validation studies.
SARS Waag & Houck (1994)	Self rating technique	Aviation (military)	No	No	Low	Low	Pen and paper	Yes 1	1) Quick and easy to use, requiring little training 2) Non-intrusive to primary task. 3) Low cost compared to other techniques.	1) Problems of gathering SA data post-trial e.g. correlation with performance, forgetting low SA. 2) Limited use and validation evidence.
SART Taylor (1990)	Self rating technique	Aviation (military)	No	No	Low	Low	Pen and paper	Yes 10+	1) Quick and easy to administer. Also low cost. 2) Generic – can be used in other domains. 3) Widely used in a number of domains.	1) Problems of gathering SA data post-trial e.g. correlation with performance, forgetting low SA periods. 2) Issues regarding sensitivity of the technique.

Method	Type of method	Domain	Team	SME's required	Training time	Application time	Tools needed	Validation studies	Advantages	Disadvantages
SA-SWORD Vidulich & Hughes (1991)	Self rating technique	Aviation	No	No	Low	Low	Pen and paper	Yes 2	1) Easy to learn and use. Also low cost. 2) Generic – can be used in other domains. 3) Useful when comparing two design concepts.	1) Post-trial administration – correlation with performance, forgetting etc. 2) Limited use and validation evidence. 3) Does not provide a measure of SA.
SPAM Durso et al (1998)	Real-time probe technique	ATC	No	Yes	High	Low	Simulator Computer Telephone	Yes 4	1) No freeze required. 2) Has shown promising results in validation studies. 3) Administering probes in real-time removes the need for task freezes, allowing the technique to be applied 'in the field'.	1) Low construct validity. 2) Limited use. 3) Attention may be directed to required SA elements.
SA requirements analysis Endsley (1993)	SA requirements analysis	Generic	No	Yes	Med	High	Video and audio recording equipment	No	1) The output specifies the elements that comprise operator SA in the scenario under analysis. 2) Output can be used to develop SA measure. 3) The procedure is generic and can be applied in any domain.	1) The procedure is a time consuming one, involving observation, interviews and task analysis. 2) Access to numerous SME's is required for a lengthy period of time. This may prove difficult to gain.
C-SAS Dennehy (1997)	Self rating technique Observer rating technique	Civil aviation	No	Yes	Low	Low	Pen and paper	No	1) Very quick and very easy to use, requiring very little training. 2) C-SAS scales are generic, and can be applied in any domain. 3) Can be used as a self-rating tool and an observer-rating tool.	1) Very unsophisticated measurement tool. 2) No validation evidence associated with the technique. 3) Problems of gathering SA data post-trial e.g. correlation with performance, forgetting low SA periods.
Performance Measures	Performance measure	Generic	No	No	Low	Low	Computer	No	1) Data collection is simplistic. 2) Provides an objective measure if SA. 3) Non-intrusive	1) May not reflect actual level of SA e.g. poor performance may still occur with accurate SA. 2) Indirect assessment of SA. 3) Suffers from diagnosticity and sensitivity problems.
Eye tracker	Process Indices	Generic	No	No	Med	High	Eye Tracking Device Relevant Software PC	No	1) Relatively unintrusive to primary task performance. 2) Can be used to determine which environmental elements are attended to. 3) Widely used.	1) Equipment is temperamental and difficult to operate, cannot be used 'in-the-field' and the data analysis procedure is very time consuming. 2) 'Look but do not see' phenomenon should be considered. 3) Offers only an indirect assessment of SA (Endsley et al 2000).
Verbal Protocol Analysis	Process Indices	Generic	No	No	Med	High	Audio recording equipment Observer Pro + PC	Yes	1) Verbalisations provide a genuine insight into cognitive processes. 2) VPA provides a rich data source (Walker In Press) 3) Simplistic procedure.	1) Data analysis procedure is extremely laborious and time consuming. 2) Prone to bias. 3) Verbal commentary can sometimes serve to change the nature of the task.
QUASA McGuinness (2004)	Probe/Self rating technique	Military	No	No	Low	Low	Pen and paper	Yes	1) Combines subjective ratings with SA probes. 2) Developed specifically for military command and control environments. 3) Provides an assessment of actual participant SA and also their perceived SA (confidence in their SA)	1) Intrusive to primary task performance. 2) Does not cater for teams. 3) Limited evidence of use and validation.

Summary

In summary, the results of the methods review demonstrate that (aside from the SA requirements analysis procedure which would be required prior to any form of SA analysis) in their current format existing SA measurement approaches are inadequate for the measurement of SA in C4i environments. There are two main reasons for this conclusion. Firstly, the SA measurement techniques reviewed all focused upon the assessment of individual SA. As outlined earlier in this report, this is problematic when considering the measurement of team or shared SA in C4i environments. The methods review highlighted a lack of specific team SA measurement approaches, and further investigation into the measurement of team or shared SA is required. Secondly, the review revealed that each of the different SA measurement approaches are beset by distinct flaws that could potentially hinder any SA data obtained. Freeze-probe techniques are intrusive and cannot be applied ‘in the field’ whilst real-time probe techniques are difficult to apply and are still intrusive to primary task performance. Self-rating techniques suffer from a host of problems associated with collecting SA data post-trial (correlation with performance, participant’s inability to rate low periods of SA etc) and the construct validity of observer rating techniques, measuring participant fixations and performance measures is questionable. A brief description of the conclusions for each category of SA measurement technique is presented below.

Freeze probe techniques

Freeze-probe techniques such as SAGAT (Endsley 1995a) are the most commonly used SA measurement techniques. There are two primary advantages associated with freeze probe approaches. Firstly, they offer a direct measurement of operator SA, which removes the various problems associated with collecting post-trial and subjective SA data (see self-rating techniques summary). Secondly, the SAGAT approach is the most widely used and validated of the SA measures available, and has consistently demonstrated reliability and validity in a number of domains. Along with the SART technique, SAGAT is the most widely validated of all SA techniques. According to Jones and Kaber (2004) numerous studies have been performed to assess the validity of the SAGAT and the evidence suggests that the method is a valid metric of SA. Endsley (2000) reports that the SAGAT technique has been shown to have a high degree of validity and reliability for measuring SA. According to Endsley (2000) a study found SAGAT to have high reliability (test-retest scores of .98, .99, .99 and .92) of mean scores for four fighter pilots participating in 2 sets of simulation trials. Collier and Folleso (1995) also reported good reliability for SAGAT when measuring nuclear power plant operator SA. Also, in a driving task study (Gugerty, 1997) reported good reliability for the percentage of cars recalled, recall error and composite recall error. Regarding validity, Endsley et al (2000) reported a good level of sensitivity for SAGAT, but not for real time probes (on-line queries with no freeze) and subjective SA measures. Endsley (1990) also reported that SAGAT showed a degree of predictive validity when measuring pilot SA, with SAGAT scores indicative of pilot performance in a combat simulation. The study found that pilots who were able to report on enemy aircraft via SAGAT were three times more likely to later kill that target in the simulation. However, whilst freeze probe techniques are the most popular of existing approaches, they are also seriously flawed when considering the measurement of SA in C4i environments. Firstly, the use of freeze-probe techniques ‘in-the-field’ is problematic. Freezing a ‘real’ scenario (with multiple information sources) and administering SA queries to multiple agents across multiple geographical locations appears to be almost impossible. This limitation alone poses serious questions regarding the use of a freeze probe technique in C4i environments. Secondly, the intrusion upon primary task performance caused by the task freezes is a major problem. If a novel way of using freeze-probe techniques in the field were developed, then the intrusion upon primary task performance would still presumably be high. There are alternative approaches that could be used to remove the various problems of using a SAGAT style approach 'in-the-field'. Instead of incorporating freezes, participants could be queried for their SA during low complexity portions of the task. Incorporating freezes whereby participants are queried for their SA into the natural flow of the task is also another possible approach. In considering the measurement of SA in infantry operations, Endsley et al (2000) report two alternatives designed to remove the problems associated with applying SAGAT 'in the field'. The 'St Peter Technique' involves querying participants who have been 'killed' during task performance, and the 'Angel of Death Technique' involves randomly selecting participants to be 'killed' and then immediately administering a series of SA queries. Both approaches, however, whilst allowing a freeze probe style approach to be applied 'in-the-field', are still problematic. The St Peter Technique may provide a bias measure of SA (Endsley et al 2000) as those participants who die during task performance may be those with poor SA, and so participants with higher levels of SA may not be subject to measurement. Furthermore, both approaches still carry a high level of intrusion to the task under analysis.

Therefore, the use of freeze-probe techniques to measure SA in C4 environments is questionable. It is apparent that a novel variation of the freeze-probe technique designed to cater for the dispersed, collaborative nature of C4i environments requires development. Incorporating freezes into ‘in-the-field’ exercises represents a major challenge, and is one that has not yet been met by the techniques available in the open literature.

Real-time probe techniques

Real-time probe techniques offer a way of circumventing the intrusion upon the task under analysis imposed by freeze-probe techniques. The main advantages associated with real-time probe techniques are the removal of the need for task freezes and also the ability to be applied ‘in-the-field’. However, the degree to which intrusion upon task performance is reduced is certainly questionable. Whilst no freeze is required, the SA queries are still administered during task performance, which still represents a level of intrusion upon the primary task. Furthermore, participant attention may be directed to the relevant SA information as a result of the query, which could bias the results obtained. Real-time probe techniques also suffer from a number of other major flaws. Due to the dynamic and unpredictable nature of C4i tasks, the SA queries would presumably be generated in real-time, and not prior to task performance. The generation of probes in real-time would potentially place a great burden upon SME’s used, and may prove too difficult for C4 environments. Also, when using a real-time probe approach in C4i environments, numerous SME’s would be required due to the amount of personnel involved. Furthermore, a measurement of team or shared SA would be difficult to obtain using such an approach.

Self-rating techniques

The use of self-rating techniques to measure SA in C4i environments is attractive for a number of reasons. Firstly, self-rating techniques are non-intrusive to task performance, as they are completed post-trial. Secondly, they are very quick and easy to use and require very little training. Thirdly, as a result of their simplistic nature, very little cost is incurred when using self-rating techniques. Fourthly, and perhaps most importantly, self-ratings of SA can be obtained from different team members (Endsley et al 2000) and so offer a potential avenue into the assessment of team SA. The majority of self-rating techniques are pen and paper tools, whereby participants rate their own SA upon completion of the task under analysis, and so there is no requirement for expensive simulators, SME's or a lengthy training process, all of which reduces the cost of the procedure considerably. The simplicity and low cost of self-rating techniques is reflected in their widespread use, with the SART technique (Taylor 1990) being especially popular. However, despite the encouraging advantages associated with the use of self-rating techniques, their use in measuring SA in C4i environments is questionable due to a number of distinct flaws. Firstly, whilst the majority of self-rating techniques are generic and can be applied in numerous environments, a specific team SA approach is yet to emerge. Consequently, a C4i specific self-rating technique would require development, incorporating the dispersed collaborative nature of C4i environments. Secondly, there are a host of problems associated with the collection of SA data post-trial that would appear to rule out the use of a self-rating tool (on its own at least) for assessing SA in C4i environments. SA ratings may be correlated with performance (Endsley 1995b) i.e. a participant who performs well in a trial automatically rates their SA as good. Also, participants are prone to 'forgetting' periods of the task when they possessed a poor level of SA, and more readily remember the periods when they possessed a superior level of SA. Endsley (1995b) reports that people are poor at reporting detailed information about past mental events and that post-trial questionnaires only capture participant SA at the end of the task under analysis. Thirdly, in various validation studies, the SAGAT (freeze probe) technique has proved to be superior in terms of reliability, validity and sensitivity when compared to the SART (self-rating) technique. Fourthly, as Endsley (1995b) points out, participant's ability to rate their own SA is questionable, as they may not be able to accurately rate their poor SA e.g. it is questionable how accurately they can rate poor SA as they may not realise that they have inadequate SA in the first place.

Observer-rating techniques

Observer-rating techniques are most commonly used when measuring SA 'in-the-field' due to their non-intrusive nature, and at first glance appear to be the most suited to measuring SA in C4i environments. However, upon further investigation, it is quickly apparent that observer-rating approaches are also beset by a number of crucial flaws that may restrict their usage. The primary disadvantage associated with observer-rating techniques concerns the construct validity of the measure. The extent to which observers can accurately rate the internal construct of SA remains a major doubt (Endsley 1995b). Whilst there are observable behaviours that may indicate certain things regarding participant SA, the actual internal level of SA cannot be accurately measured by observation alone. Observer-rating techniques may also be subject to bias, in that they may serve to alter participant behaviour. Knowing that they are being observed may change participant behaviour, in that they may strive to operate 'by-the-book' so to speak, and as a result the data obtained is subject to bias. Finally, observer-rating techniques require repeated access to multiple SME's over a long duration of time. This may prove problematic, if not impossible, especially when military personnel are required.

Performance measures

It is perhaps somewhat irrelevant to assess the potential use of performance measures in the measurement of SA as various performance measures are normally taken during task performance anyway. Performance measures are often very simplistic to take and are non-intrusive. The main problem associated with performance measures being used to measure SA is the assumption that efficient performance is achieved as a result of efficient SA and vice versa. It may be that efficient performance is achieved despite an inadequate level of SA, or that poor performance is achieved regardless of a high level of SA. The unstable nature of the relationship between task performance and SA serves to diminish the suitability of performance measures as indicators of participant SA. However, their use is not to be discounted as the procedure is often very simple, and the data may still have uses, particularly as a back-up SA measure to the other techniques employed.

Process Indices (Eye tracker)

The most commonly used process index is the measurement of participant eye fixation using an eye-tracking device. Eye tracker data can be used to assess which situational elements the participant(s) fixated upon during task performance, and has been extensively used in SA assessment exercises. However, the use of an eye-tracking device ‘in-the-field’ is not possible, and so it is not recommended in this case. Furthermore, typical eye-tracking devices are temperamental in their operation, and the data analysis procedure is a lengthy one, requiring great patience on behalf of the analyst. Another problem associated with the use of eye-tracking devices surrounds the ‘look-but-failed-to-see’ phenomenon (Brown 2001). Whilst the eye-tracker data can point to which elements in the environment the participant fixated upon, there is no assurance that the element in question was accurately perceived.

Conclusions

Existing SA measurement techniques are inadequate for use in the assessment of SA in C4i systems. Whilst each class of technique (freeze-probe, real-time probe, self-rating etc) possess distinct flaws (as described above) which would hinder the data collected, the techniques also fail to meet the requirements specified earlier in the paper, namely that any technique used to assess SA in C4i environments should be able to assess SA across multiple locations at the same time, assess both individual and team SA for the same task and also assess SA in real-time. The methods review also produced a number of more general conclusions regarding the measurement of SA. Firstly, the SAGAT approach (Endsley 1995a) is the most commonly applied approach when assessing SA. Secondly, validation of the techniques remains a problem. Aside from SAGAT and SART, there is limited validation evidence associated with the existing SA measurement techniques.

Recommendations

The measurement of SA in C4i environments poses a considerable but exciting challenge to the HF community. The concept of team or shared SA requires much further investigation in itself, which in turn requires the provision of reliable and valid measurement procedures. It is apparent that, in their current format, existing SA measurement approaches are inadequate for this purpose, and a novel approach is required. As highlighted previously, the main issues surrounding the measurement of SA in C4i environments are the need to assess both individual and team SA in real-time, and simultaneously at different locations. From the categories of measurement technique available in the literature, not one can boast an ability to achieve this without incurring serious flaws that may hinder the data collected. There are two solutions to this problem. The first solution would be to develop a novel approach to the assessment of SA that could satisfy these requirements. This is a daunting prospect, and one that requires a great deal of further investigation. The second solution would be to combine the most successful SA measurement techniques in order to form a battery or 'toolkit' of SA measures. Consequently, it is proposed that a multiple-measure or toolkit approach of measurement techniques may be the most appropriate way to measure SA in C4i environments. This recommendation is made on the basis of two key factors. Firstly, the lack of a single technique that can cope with both individual and team SA across multiple geographical locations in real-time ensures that multiple approaches must be utilised. Secondly, a multiple measure approach ensures that SA data can be effectively crosschecked between measures in order to ensure reliability and accuracy. The concept of using a battery of HF methods to achieve more efficient performance is not a new one. For example, in conclusion to a review of thirty-eight existing human reliability analysis (HRA) and human error identification (HEI) techniques (Kirwan 1998a), Kirwan (1998b) suggested that as none of the techniques available satisfied all of the fourteen criteria against which they were evaluated, a framework or toolkit approach using a mixture of independent HRA/HEI tools may be the most suitable approach to error analysis. It is also common to use a battery of methods (e.g. physiological measures, primary and secondary task performance measures and subjective measures) for the assessment of operator workload. A multiple measure approach has no doubt been used previously to measure SA, and it is not offered as a novel procedure, rather it is offered as a solution to the considerable challenge faced when measuring SA in C4i environments. The make-up of such an approach is unclear, and considerable investigation is required in order to determine the logistics of such an approach. The first stage required in this process is to conduct an SA requirements analysis for the C4i environment in question, in order to determine those elements that comprise both individual and team SA. Next, the various components of the multiple measure approach require development. Such an approach may utilise a number of different SA measurement approaches, in order to cater for both individual and team SA. One such battery or toolkit of SA measurement techniques may include performance measures, a freeze probe technique (adapted for C4i environments), a post trial subjective rating technique and an observer rating technique.

Acknowledgement

The work reported in this paper was carried out under the UK Defence Technology Centre (DTC) for Human Factors Integration (HFI), funded by the UK Ministry of Defence and the Defence Science and Technology Laboratory (DSTL).

Bibliography

Artman, H., & Garbis, C., 1998. Situation awareness as distributed Cognition. Proceedings of ECCE' 98, Limerick.

Baber, C., and Stanton. N.A., 1996. Human error identification techniques applied to public technology: predictions compared with observed use, *Applied Ergonomics,* 27 pp. 119-131.

BBC News (2003). Inquiry into friendly fire deaths. www.bbc.co.uk, Wednesday, 2nd April, 2003.

Bedny, G., Meister, D., 1999. Theory of activity and situation awareness, *International Journal of Cognitive Ergonomics*, 3, 1, pp. 63-72.

Brown, I.D., 2001. A Review of the 'Looked-But-Failed-To-See' Accident Causation Factor, Department of Environment, Transport, and the Regions Conference on Driver behaviour, University of Manchester.

Collier, S.G., Folleso, K., 1995. SACRI: A measure of situation awareness for nuclear power control rooms. In D. J. Garland and M. R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness*, pp. 115-122, Daytona Beach, FL: Embry-Riddle University Press.

Cooke, N.J., 2004. Measuring Team Knowledge. In N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick. (Eds.), *Handbook of Human Factors methods*. Boca Raton, USA, CRC Press.

Cooper, P., 2003. Coalition deaths fewer than in 1991: 'We became stronger while Saddam became weaker', *www.cnn.com*, June 25th, 2003.

Dennehy, K., 1997. Cranfield – Situation Awareness Scale, User Manual. Applied Psychology unit, College of Aeronautics, Cranfield University, COA report No. 9702, Bedford, January.

Durso, F.T., Hackworth, C.A., Truitt, T., Crutchfield, J., Manning, C.A., 1998. *Situation awareness as a predictor of performance in en route air traffic controllers*. Air Traffic Quarterly, 6, pp. 1-20.

Endsley, M.R., 1989. Final report: Situation awareness in an advanced strategic mission (NOR DOC 89-32). Hawthorne, CA: Northrop Corporation.

Endsley, M.R., 1990. 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., 1993. A survey of situation awareness requirements in Air-to-Air Combat fighters. *The International Journal of Aviation Psychology*, Vol 3, pp. 157-168.

Endsley, M.R., 1995a. Towards a theory of Situation Awareness in Dynamic Systems, *Human Factors*, Vol. 37, pp. 32-64.

Endsley, M.R., 1995b. Measurement of Situation Awareness in Dynamic Systems, *Human Factors*, Vol. 37, pp. 65-84.

Endsley, M.R., 2000. Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley and D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates.

Endsley, M.R. & Kiris, E.O., 1995. Situation awareness global assessment technique (SAGAT) TRACON air traffic control version user guide. Lubbock TX: Texas Tech University.

Endsley, M.R., Smolensky, M., 1998. Situation Awareness in Air Traffic Control: The Picture. In M. Smolensky and E. Stein (Eds.) *Human Factors in Air Traffic Control*, pp. 115-154, New York: Academic Press

Endsley, M.R., Sollenberger, R., & Stein, E., 2000. Situation awareness: A comparison of measures. *In Proceedings of the Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium*. Savannah, GA: SA Technologies, Inc.

Endsley, M.R., Holder, C.D., Leibricht, B.C., Garland, D.C., Wampler, R.L. & Matthews, M.D., 2000. Modelling and measuring situation awareness in the infantry operational environment. (1753). Alexandria, VA: Army Research Institute.

Gugerty, L.J. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory, *Journal of Experimental Psychology: Applied*, 3, pp. 42-66.

Hart, S.G., & Staveland, L.E., 1988. Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam. The Netherlands. Elsevier.

Hauss, Y. & Eyferth, K., 2003. Securing future ATM-concepts' safety by measuring situation awareness in ATC. *Aerospace Science and Technology*. In press

Hogg, D.N., Folleso, K., Strand-Volden, F., & Torralba, B., 1995. Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics*, Vol 38 (11), pp 2394-2413.

Jeannott, E., Kelly, C., Thompson, D. 2003. The development of Situation Awareness measures in ATM systems. EATMP report. HRS/HSP-005-REP-01.

Jones, D.G., and Endsley, M.R., 2000. Can real-time probes provide a valid measure of situation awareness. *Proceedings of the Human Performance, Situation Awareness and Automation: User Centred Design for the New Millennium* Conference, October 2000.

Jones, D.G., and Kaber, D.B., 2004. In N. Stanton, Hedge, Hendrick, K. Brookhuis, E. Salas (Eds.) *Handbook of Human Factors and Ergonomics Methods.* Boca Raton, USA, CRC Press.

Kaber, D.B., Endsley, M.R., Wright, M.C. and Warren, H., 2002. The effects of levels of automation on performance, situation awareness, and workload in an advanced commercial aircraft flight simulation. Final Rep.: NASA Langley Research Center Grant #NAG-1-01002). Hampton, VA: NASA Langley Research Center.

Keller, J., Lebiere, C., Shay, R., and Latorella, K., 2004. Cockpit System Situational Awareness Modelling Tool. *In Proceedings of the fifth human performance, situation awareness and automation conference,* Daytona Beach, FL, March 22nd-25th 2004.

Kirwan, B., 1992. Human error identification in human reliability assessment. Part 2: detailed comparison of techniques, *Applied Ergonomics*, 23, pp. 371-381.

Kirwan, B., Ainsworth, L.K., 1992. A guide to Task Analysis, Taylor and Francis, London, UK.

Kirwan, B., 1998a. Human error identification techniques for risk assessment of high-risk systems – Part 1: Review and evaluation of techniques. *Applied Ergonomics,* 29, pp.157-177.

Kirwan, B., 1998b. Human error identification techniques for risk assessment of high-risk systems – Part 2: Towards a framework approach. *Applied Ergonomics,* 5, pp. 299 – 319

Klein, G., 2000. Cognitive Task Analysis of Teams. In J. M. Schraagen, S. F. Chipman, V. L. Shalin (Eds). *Cognitive Task Analysis,* pp. 417-431. Lawrence Erlbaum associates.

Matthews, M.D., Beal, S. A., 2002. Assessing Situation Awareness in Field Training Exercises. U.S. Army Research Institute for the Behavioural and Social Sciences. Research Report 1795.

Matthews, M.D., Pleban, R.J., Endsley, M.R., & Strater, L.D., 2000. Measures of Infantry Situation Awareness for a Virtual MOUT Environment. *Proceedings of the Human Performance, Situation Awareness and Automation: User Centred Design for the New Millennium Conference*, October 2000.

McGuinness, B., 2004. Quantitative Analysis of Situational Awareness (QUASA): Applying Signal Detection Theory to True/False Probes and Self-Ratings. *Command and Control Research and Technology Symposium: The Power of Information Age Concepts and Technologies*, San Diego, California, 15th – 17th June 2004.

McGuinness, B. & Ebbage, L., 2000. Assessing Human Factors in Command and Control: Workload and Situational Awareness Metrics. *Unpublished.*

McGuinness, B. & Foy, L., 2000. A subjective measure of SA: the Crew Awareness Rating Scale (CARS). *Presented at the Human Performance, Situational Awareness and Automation Conference*, Savannah, Georgia, 16-19 Oct 2000.

Ripley, T., 2003. Combatting friendly fire. www.*ft.com*, January 23rd, 2003.

Salas, E., 2004. Team Methods. In N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick. (Eds.), *Handbook of Human Factors methods*. Boca Raton, USA, CRC Press.

Salas, E., Prince, C., Baker, P.D., & Shrestha, L., 1995. Situation awareness in team performance, *Human Factors*, 37, 1, pp. 123-126.

Salmon, P.M., Stanton, N.A., Walker, G., & Green, D., 2004. Human Factors Design and Evaluation Methods Review. Defence Technology Centre for Human Factors Integration, Report No. HFIDTC/WP1.3.2/1.

Savoie, E.J., 1998. Tapping the power of teams. In R. S. Tindale, L. Heath, et al. (Eds.), *Theory and research on small groups. Social psychological applications to social issues, Vol. 4,* pp. 229-244. New York, NY: Plenum Press.

Smith, J. D., 2003. Assessment of C4i systems. *Defence Management Journal,* Issue 23, pp. 74-76.

Smith, K., & Hancock, P.A., 1995. Situation awareness is adaptive, externally directed consciousness, *Human Factors*, 37, 1, pp. 137-148.

Smolensky, M. W., 1993. Toward the physiological measurement of situation awareness: The case for eye movement measurements. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, Santa Monica, Human Factors and Ergonomics Society.

Stanton, N.A. and Baber, C., 2001. Error by design: methods for predicting device usability, *Design Studies, 23 (4)* pp. 363-384

Stanton, N.A., & Stevenage, S.V., 2000. Learning to predict human error: issues of acceptability, reliability and validity. In J. Annett & N. A. Stanton (Eds.) *Task Analysis*. Taylor and Francis, London.

Stanton, N.A., Young, M., 1999. What price ergonomics? *Nature* 399, pp. 197-198.

Stanton, N.A., Young, M., 2003. The application of ergonomics methods by novices. *Applied Ergonomics* 34, pp. 479-490.

Stanton, N.A., Chambers, P.R.G., Piggott, J., 2001. Situational Awareness and safety. *Safety Science*, 39, pp189-204.

Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., & Hendrick, H., 2004. *Handbook of Human Factors methods*. Boca Raton, USA, CRC Press.

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) pp3/1 –3/17, Neuilly Sur Seine, France: NATO-AGARD.

Vidulich, M.A., Hughes, E.R., 1991. Testing a subjective metric of situation awareness. *Proceedings of the Human Factors Society 35th Annual meeting*. Pg 1307 – 1311.

Waag, W.L., Houck, M. R., 1994. Tools for assessing situational awareness in an operational fighter environment. *Aviation, Space and Environmental Medicine*. 65(5) A13-A19.

Walker, G. H., 2004. Verbal Protocol Analysis. In N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick. (Eds.), *Handbook of Human Factors methods*. Boca Raton, USA, CRC Press.

Walker, G.H., Stanton, N.A., Young, M.S., 2004. Examining the Effects of Vehicle Characteristics on Driver Situational Awareness Using and Enhanced form of the SAGAT Method. *Unpublished*.

Walker, G.H., Stanton, N.A., Salmon, P., Green, D., 2004. Measuring and Predicting SA in C4I: Development and Testing of a Refined SA Measurement Technique, and a New Concept for SA Prediction. *In Proceedings of the fifth human performance, situation awareness and automation conference*, Daytona Beach, FL, March 22nd-25th 2004.

Whalley, S.P. and Kirwan, B., 1989. An evaluation of five human error identification techniques. *Paper presented at the 5th International Loss Prevention Symposium*, Oslo, June 1989.

White, I. and Harris, C. J., 1987. Advances in command, control and communications systems. London; Peregrinus.

Zheng, X.S., McConkie, G.W., Tai, Y., 2004. Dynamic Monitoring of Traffic Flow: The Driver's Situation Awareness. *In Proceedings of the fifth human performance, situation awareness and automation conference*, Daytona Beach, FL, March 22nd-25th 2004.