DOT/FAA/AM-05/4

Office of Aerospace Medicine
Washington, DC 20591

Complexity and Automation Displays of Air Traffic Control: Literature Review and Analysis

Jing Xing
Carol A. Manning
Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

April 2005

Final Report

U.S. Department
of Transportation

**Federal Aviation
Administration**

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-05/4		

4. Title and Subtitle	5. Report Date
Complexity and Automation Displays of Air Traffic Control: Literature Review and Analysis	April 2005
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Xing J, Manning CA	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125	11. Contract or Grant No.
12. Sponsoring Agency name and Address	13. Type of Report and Period Covered
Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

Work was accomplished under approved task BHRR522

16. Abstract

This report reviewed a number of measures of complexity associated with visual displays and analyzed the potential to apply these methods to assess the complexity of air traffic control (ATC) displays. Through the literature review, we identified three basic complexity factors: numeric size, variety, and rules. Essentially, all the complexity measures could be described by these factors. Through the analysis of available complexity measures, we showed that neither information complexity that focused on the system nor cognitive complexity that aimed at observers could provide a complete description for ATC application. The great variety in complexity measures reflected the fact that the contribution of each of the three factors to overall complexity depended on how information is processed by users. We generalized that complexity is the integration of the observer with the three basic factors. Therefore, to develop objective complexity measures for ATC displays, the methods presented in this report need to be integrated with the ATC display specifications.

17. Key Words	18. Distribution Statement
Information Complexity, Cognitive Complexity, Visual Displays, Evaluation, Interface Design	Document is available to the public through the Defense Technical Information Center, Ft. Belvior, VA 22060; and the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	24	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENTS

The authors thank Drs. Lawrence Bailey, Scott Shappell, and David Schroeder for their valuable comments on the manuscript. We also thank the editor, Mr. Mike Wayda, and other FAA internal reviewers whose effort greatly improved the quality of this paper.

Complexity and Automation Displays of Air Traffic Control: Literature Review and Analysis

INTRODUCTION

Traditionally, air traffic controllers use a radar screen and flight progress strips as separate representations of aircraft. The radar screen shows the spatial position, altitude, and progress of aircraft, while the strips contain discrete information about the origin, destination, route, aircraft type, and requested altitude of aircraft. In the course of their work, air traffic controllers cognitively integrate these two representations and then make decisions accordingly (Moertl, Canning, Gronlund, Dougherty, & Johansson, 2002). Occasionally such tasks have the potential to create an overload condition as the complexity of air traffic increases. To help controllers manage the increasing volume of air traffic, many automation tools have been provided, such as the User Request Evaluation Tool, Center-Tracon Automation System.

Air traffic control (ATC) is a dynamic environment where controllers constantly receive a large volume of information from multiple sources to monitor the changes in the environment, make decisions, and perform effective actions in a timely manner. While ATC automation tools are designed with the objectives of increasing capacity and reducing workload, controllers need to combine information from automation displays with information from the radar screen to plan their activities. Those activities must be synchronized with rapid information evolution. With automation tools, new tasks of interface management and consultation are added to traditional control tasks. Moreover, the use of new tools requires that controllers integrate the interaction demands of the new system into the management of their cognitive resources (Bressolle, Benhacene, Boudes, & Parise, 2000). Not surprising then, the introduction of new systems can introduce additional complexity to ATC task management. What's more, if information provided by the tools overwhelms controllers' cognitive capacities, critical information could be either missed or misinterpreted and put performance at risk.

The importance of understanding the complexity of ATC tasks has been widely acknowledged. While many studies have been conducted to assess the complexity of air traffic control (Mogford, Guttman, Morrow, & Kopardekar, 1995; Guttmann 1995; Laudeman, Shelden, Branstrom, & Brasil, 1998), little effort has been devoted to assessing the complexity of ATC automation displays. Given the fact that many new automation tools are being developed and are projected to be fielded over the next several years, it is necessary to develop methods to assess the complexity of the tools. In this report, we will review the studies on complexity and analyze their application to ATC displays. The ultimate objective of the report is to identify methods from the literature that are applicable to ATC displays. To accomplish this, we organized the report into two main sections: first, we will review the literature about complexity measures and analyze the potential to apply these methods to assess the complexity of ATC displays; second, we will discuss several issues in the evaluation of ATC tools.

DEFINITIONS AND MEASURES OF COMPLEXITY

In this section we will review some definitions of complexity and methods for measuring it. Note however, that the review is not exhaustive. Rather we intend to review only the approaches that are generically relevant to the concept of complexity and visual displays. One exception is air traffic complexity. We will introduce air traffic complexity because it is relevant to air traffic control and has been studied with respect to controller workload. This section is organized into four parts: We first discuss some concepts and definitions of complexity to provide a basic understanding about what complexity is. We then introduce the two major threads of the issue: information complexity and cognitive complexity followed by a presentation of complexity measures related to visual displays. Finally, we will summarize the definitions and measures.

General definitions of complexity

Although the term "complexity" has proven to be difficult to define, many attempts exist in the literature. The difficulty exists because complexity depends on which aspect you are concerned with. Moreover, complexity only makes sense when considered relative to a given observer (Edmonds, 1999). With this in mind, the objective of this report is to evaluate the complexity of ATC displays composed mainly of graphical symbols and text. It is with these displays that air traffic controllers acquire information to help them make predictions about future situations and identify actions that should be taken. With regard to complexity, however, we are not simply concerned with the complexity of the interface itself. Rather, we are interested in the complexity that the interface imposes on controllers. Thus, the complexity of an ATC display makes sense only when it is specified relative to controllers.

In perhaps a very straightforward way, complexity has been associated with concepts such as numeric size of basic elements, variety, and internal structure. However, while to some extent a larger numeric size corresponds to a higher degree of complexity, size, nevertheless, is a weak definition of complexity. Edmonds (1999) pointed out that, by using size for complexity, the parts of the system are neither inter-related nor interconnected. One example demonstrating that size cannot quantify complexity would be counting peas in a basket. While it takes more time to count peas in a full basket than a half basket, the complexity of the task remains the same. That is, the task of "counting peas" is the same in both situations.

Variety has also been used to describe complexity. In fact, the concept of variety or disorder has been widely used in various applications as the measure of complexity. Yet variety alone, like numeric size, is not sufficient to describe complexity. Several studies in different areas have made the same comment that complexity lies somewhere between order and disorder (Drozdz, Kwapien, Speth, & Wojcik, 2002). One example would be Grassberger's study of image complexity (Grassberger, 1991). Figure 1 shows the three images Grassberger used. The disorder or variation increases from the left to the right. However, human eyes perceive the image in the middle as the most complex. The reason is that humans interpret the image on the right as representing a situation with no rules.

Indeed, the structural rules of a system seem to contribute to its complexity. That is, individual parts of a system are held together through rules of internal structure. Rules determine the interconnections between parts of an object. According to the Random House dictionary, something that is complex is defined as being "composed of interconnected parts." So images like the one on the right in Figure 1, although graphically complex, are not perceived as such by humans because there appear to be no structural rules. In contrast, a chess pattern may be viewed as quite complex because of many rules embedded in it.

Edmonds (1999) analyzed various concepts that are generally assumed to be associated with complexity. He proposed a more sophisticated definition of complexity. Specifically, he defined complexity as "That property of a language expression which makes it difficult to formulate its overall behavior, even when given almost complete information about its atomic components and their inter-relations." This is a very general definition that can have different interpretations in different contexts. Here "language" is meant in a general sense while "atomic components" refer to irreducible signs in a chosen language of representation. This definition relates the difficulty in formalization of the whole to that of its fundamental parts. For air traffic control, this definition suggests that complexity reflects the difficulty to formulate an accurate representation of the situation, given many sources of information about aircraft, sectors, and flight rules.

Ultimately, the concept of complexity is multi–dimensional and cannot be sufficiently described with a single measure. Such conclusions are not unique as Burleson and Caplan (2002) defined complexity as the "diversity of forms, to emergence of coherent patterns out of randomness and also to some ability of frequent switching among such patterns." Likewise, Drozdz et al. (2002) viewed complexity as a trinity of coherence, chaos, and the transition between them. In this definition, coherence constitutes the essence as it makes patterns and structures; chaos is needed in a system as it allows switching one pattern of activity to another; the gap allows the structures to be identifiable. All three are needed in parallel to describe complexity.

In a sense then, this trinity corresponds to the three factors of complexity we reviewed above: coherence corresponds to the numeric size of basic elements, chaos corresponds to variety, and gap corresponds to structural rules. As we introduce additional complexity definitions in the following sections, it will become apparent that nearly all the definitions are concerned with some or all of the three factors.

Information complexity in information theories

Definitions of information complexity

Complexity has been extensively studied within the field of information theory, where the term "information complexity (IC)" is frequently used to describe complexity from the perspective of a system. There have been many attempts to quantify IC theoretically. Below we list some widely used complexity measures. These measures do not necessarily exclude each other. Instead, they emphasize different aspects of complexity and are somewhat complementary.

Kolmogorov complexity. According to information theories, the most straightforward definition of complexity is the minimum description size. Hence Kolmogorov complexity is defined as the minimum possible length of a description in some language (Casti, 1979). For instance, if a description can be greatly compressed without loss of meaning, then it is considered simpler than one that cannot. By this definition, highly ordered expressions appear as simple and random while maintaining maximal complexity. For example, the numeric string (1 1 1 1 1) is less complex than the string (1 5 3 2 4) because the former can be easily compressed into a description "five ones." Unfortunately, this definition corresponds to the difficulty of compressing a representation with little direct connection to the practical aspects of a functioning organism. Indeed, it is only concerned with the numeric size factor of complexity.

Topological complexity. Crutchfield and Young (1989) extended the concept of Kolmogorov complexity by defining complexity as the minimal size of a model representation of a system that can statistically reproduce the observed data within a specified tolerance. Consider, for example, two air traffic cases. In the first case, ten aircraft are flying on two fixed routes that have one intersection. In the second case, ten aircraft are flying off the routes, which can create many potential conflicts. A controller can build a model of the first case that has two flows of aircraft and one crossing point, while a model of the second case has to be composed of many flows and crossings. Thus the topological complexity of Case A is less than Case B. This definition takes into account both the minimal size and the fixed hierarchy or structural rules of a system. One shortcoming of the definition is that it does not provide a unique measure of complexity for a system because there is not necessarily a "minimal" model for it (Pressing, 1999). That is, users may construct different models of the same system. In addition, neither this nor the definition above is sufficient to describe complexity because they only emphasize the storage resource that it takes to solve a class of problems. In reality, the resource is not always a sensible measure of complexity (Holm, 1993).

Mutual information. Complexity is indicated by levels of mutual information that measure the correlation between information at sites separated by time and space (Langton, 1991). This definition describes the computational power requirement. For example, if each controller only needs to handle aircraft within one's sector regardless of traffic in the next sector, the task would be less complex because traffic in the next sector is not relevant to his or her problem space.

Logical depth. Logical depth is defined as the computational cost (time and memory) taken to calculate the shortest program that can reproduce a given object (Bennett, 1990). By this definition, complexity is the difficulty of computation from a random starting point to the resulting state. This measure is aimed at the complexity of the process and not the results. That is, it is a combination of both storage and computational power. Thus, the definition is concerned with all three factors: numeric size, variety and structural rules. An increase in any of these three dimensions may result in greater difficulty of computation. In air traffic control, this measure would reflect how difficult it is for a controller to make projections of air traffic situations in his or her mind based on the current situation.

Kauffman's complexity. Kauffman (1993) defined complexity as the "number of conflicting constraints." The definition represents the difficulty of specifying a successful task within the constraints or "rules" imposed. For example, an airspace can be made less complex by removing air traffic constraints such as military zones, bad weather, etc. Note, however, that the definition is only concerned with the complexity factor of structural rules.

Hieratical complexity. This definition is also concerned with structural rules. A complex system is often constructed hierarchically. That is, it is composed of structures on several scales or levels. These may be scales of space or time, or levels within a domain-specific functional space. For example, an ATC display may be composed of several windows, consisting of different types of text and graphical regions, and each text region (such as a datablock) containing several types of information. With this in mind, Bates and Shepard (1993) assumed that a system is composed of elementary units with local structures and the interconnections between the local structures are governed by rules. They suggested that complexity is manifested as variability in the convergence and divergence of interconnections. Then the dimensionality of local structures, number of local structures, and the range of connections all contribute to the global complexity. Moreover, if local regions possess certain computational abilities, then multiple regions can interact to achieve greater complexity.

Methods of computing IC

Entropy as a measure of complexity

Within information theory, entropy, denoted as H, is a measure of the redundancy contained in sets of information in binary data strings (Scott, 1969). In a more general sense, H represents the number of independent dimensions that a person uses to describe something. (i.e., it describes the numeric size factor of complexity). Therefore, complexity is greater when a person views an object as having many aspects and must make fine distinctions among those aspects. H can be computed according to the following formula:

$$H = log_2 n - (1/n) (\Sigma_i n_i log_2 n_i)$$

where n is the total number of attributes and n_i is the number of attributes that appear in a particular combination of the descriptions of self aspects. To use this formula, one has to model the system with three parameters: the number of basic elements (attributes), the number of groups (classes), and the attributes of each class. In addition, when a system is partitioned into several subsystems or classes, the information shared among the subsystems also contributes to the system complexity. Cha, Chung, and Kwon (1993) developed an excess entropy metric to measure such shared information.

Psychophysicists have used H to assess cognitive complexity. For example, Linville used H to quantify the complexity of persons (Linville, 1985, 1987). Individuals with greater complexity used different words to describe themselves in their social roles while individuals with less complexity used the same words repetitiously to describe their social roles.

Complexity computed with Random Matrix Theory

An approach to complex systems is typically based on analyzing large multivariate ensembles of parameters. For this reason, one efficient way to quantify the variety associated with complexity is the use of matrices. Toward this end, Random Matrix Theory provides an appropriate reference for quantifying various characteristics of complexity. Drozdz et al. (2002) identified some principal variants within the matrix that are common and typical to natural complex dynamic systems. Among the matrix variants, the correlation and eigenvalue of a matrix are dominant components of complexity. These variants reflect the degree of agreement and the deviation of a system, which correspond to the variety and numeric size factors of complexity described earlier. In particular, deviation can be quantified in the term of reduced dimensionality, which can be computed as the eigenvalue of the matrix.

Cognitive complexity

Definitions of cognitive complexity

Another line of complexity studies involves cognitive complexity. While complexity studies generated by information theory focus on the complexity of a system itself, studies of cognitive complexity focus on observers: complexity from the perspective of the observer, i.e., the users. Since air traffic control involves cognitive tasks such as monitoring the situation, resolving conflicts, issuing instructions, etc., it is important to understand how cognitive complexity is measured to assess the complexity of ATC displays.

Cognition may best be thought of a construct system composed of constructs and elements (Kelly 1955). The constructs are transparent templates that a person uses to comprehend the world. In a sense then, humans create the templates and fit the perception of the world to them. Elements are more concrete and can be placed on construct dimensions. Presumably, elements that belong to the same construct are more closely related to each other than elements in different constructs. Like most dynamic systems, a person's construct system is dominated by two processes: integration of constructs within and between subsystems (i.e., numeric size) and differentiation (variety) among subsystems (Adams-Webber, 1996). Differentiation serves the specialization of subsystems, whereas integration serves the unity of each subsystem to keep the entire system as an operational whole. These two processes constitute the basis of cognitive complexity. It is obvious that a more differentiated set of constructs would constitute a more complex system. On the other hand, consistency or integration has to supplement differentiation in the definition of complexity. Without consistency, the measures of complexity become a simple assessment of the randomness of the system.

Bieri (1955) developed the first index of cognitive complexity. This index was aimed at measuring the numeric size factor of complexity. Two measures were used: number of constructs and matches between the constructs. Matches indicate that seemingly different constructs do not constitute different dimensions in cognition. The index increases with the number of constructs and decreases with the number of matches. Bieri et al. further pointed out that the relationship between construct dimensions could be described with Eucilidian geometry (Bieri, Atkins, Briar, Leoman, Miller, & Tripodi, 1966).

In a similar fashion, Crokett (1965) used the concept of "level of hierarchic integration of constructs" to define the complexity of a construct system. With this definition, cognitive complexity is associated with increasing differentiation (containing a greater number of constructs), articulation (consisting of more refined and abstract elements), and hierarchic integration (organized and interconnected). Notably, this definition includes all three basic components of complexity described earlier: numeric size, variety, and rules.

Methods of measuring cognitive complexity

Kelly's Repertory Grid technique

A popular method to reveal constructs and elements is Kelly's Repertory Grid method (Kelly, 1955). The method can be performed in several steps. First, subjects make a list of elements pertinent to the topic of the interview, then they determine the distance between the elements by comparing which pair of the elements is closer than other pairs. The constructs pertinent to the interview topic are thus elicited. The data collected with these two steps are then mapped to a matrix from which Bieri's index of complexity can be derived. A key issue in applying the data to Bieri's index is to determine the independent constructs. A number of numeric computational methods, such as principal components analysis and factor-analysis, can be used to reveal the independence of the elicited constructs (Bezzi, 1999; Woehr, Miller, & Lane, 1998). Moreover, principal components analysis can elucidate the degrees of both differentiation and integration among the elements. Recently, Morçöl (2002) applied this method to measure the creativity of persons in several social groups. The results indicated a high correlation between one's creativity and the computed value of cognitive complexity.

Sketch maps

Cognitive maps are mental models of the relative locations and attributes of phenomena in a spatial environment. Downs and Stea (1973; Downs, 1976) defined cognitive mapping as "a process of a series psychophysical transformations by which an individual acquires, codes, stores, recalls and decodes information about the relative locations and attributes of phenomena…" Cognitive maps are also made up of memories of objects and kinesthetic, visual, and auditory cues. The information stored in a cognitive map is especially interesting since it may correspond to the constructs in a cognitive system. For instance, Kuipers (1983) suggested that a cognitive map consists of five different types of information, each with its own representation: topological, metric, route description, fixed features, and sensory images.

One common method to reveal mental models is to have subjects sketch maps to represent their understanding of the objects. For example, Lynch (1960) used this method to measure subjects' representation of their local cities and found that sketch maps were more accurate when used for topographical rather then metric analysis. While sketching maps is easy to conduct, one challenge is analyzing the results. Billinghurst and Weghorst (1995) recommended three ways to score sketch maps: map goodness (accuracy), object class number, and the relative position ratio. They found that the three measures significantly correlate to subjects' sense of the virtual world. In addition, the results also indicated that sketching maps is more useful for relatively dense worlds than for sparse worlds. Overall, sketch maps reveal spatial relationships better than abstract, conceptual components of mental models.

Cognitive task analysis

Cognitive task analysis (CTA) refers to a set of methods for gaining access to cognition, mental events, and knowledge structures. The aim of CTA is to investigate the cognitive aspects of task performance and the knowledge needed for situation awareness, decision-making, planning, etc. This approach has been widely used in human-computer interface design (Jonassen, Tessmer, & Hannum, 1999). The CTA method typically includes three steps: knowledge elicitation, analysis, and knowledge representation. Knowledge elicitation is the process of extracting information through interviews and observations about cognitive events, structures, or models. Analysis is the process of structuring data—abstracting information, developing explanations, and extracting meaning. Knowledge representation is the process of displaying data and depicting relationships. Typically, the output of CTA is an ordered list of tasks with supplementary information about the cognitive requirements of the task structures.

One popular CTA method is GOMS: Goal, Operator, Methods and Selection (John, 1995; Card, Moran, & Newell, 1983). The method seeks to analyze and model the knowledge and skills a user must develop to perform tasks on a device or system (i.e., describes knowledge of procedures that users perform in a hierarchical arrangement). The result is a description of the Goals, Operators, Methods and Selection rules for any task. The tasks are broken down into a meaningful series of goals and sub-goals until one ends up with primitive psychomotor or mental acts. If there is more than one operation or method available to accomplish a goal, the GOMS model includes selection rules to choose the appropriate method depending on the context. Since this method aims at capturing knowledge representation that people have to complete a task, it has been proven to be very useful in identifying training needs and information requirements (Jonasson et al., 1999).

Memory-based metrics of cognitive complexity

Cognitive processes are associated with working memory (WM), also referred as to short-term memory. WM can be thought of as a container where a small number of concepts can be stored and associated to make inferences. While the capacity of WM has been a long debated issue, most recent studies have generally agreed that the capacity limit of WM is about four items on average (Cowan, 2001; Fisher, 1984). Broadbent (1975) also found that WM for understanding text is four concepts. With this capacity limit, if data are presented in such a way that too many concepts must be associated to make a correct decision or that the concepts are unfamiliar, the risk of error increases (Klemola, 2000a). Thus, the density of concept usage should be considered as a cognitive complexity metric. The principal challenge in using this measure is to determine which information is familiar or unfamiliar. If the object of comprehension is text, then the density of terms used to describe new information is a good indicator of comprehension error (Kintsch, 1998). Consider, for example, computer programming where identifiers represent concepts. If the program is unfamiliar to the programmer, then identifier density is a good predictor of error (Klemola, 2000a,b).

Halford, Wilson, and Phillips (1998) studied working memory limitations and proposed that they were best defined in term of the complexity of relations that can be processed in parallel. Consequently, they defined cognitive complexity as relational complexity, i.e., the number of interacting variables that must be presented in parallel to perform a process entailed in a task. Furthermore, Halford et al. argued that relational complexity reflects the cognitive resources required to perform a task. The greater the number of interacting variables that have to be processed in parallel, the higher both the cognitive demand and computational cost. Therefore, one way to measure complexity is to determine the level of relational complexity of cognitive tasks. For example, an equation $a = 3 * b$ is a binary relation, while a second equation $a/b = c/d$ is a quaternary relation, and thus, is more complex. Theoretically, any complex relation can be decomposed into low-ranked relations. Thus, complexity can be computed from the dimensions of low-ranked relations (Wilson & Halford, 1994; Humphreys, Bain, & Pike, 1989).

In recent years, neuroimaging techniques have been widely used to reveal brain activities related to ongoing cognitive processes while the human subject performs tasks. In this way, researchers have successfully identified several brain areas such as the prefrontal cortex that are involved in the execution of WM. There have been many attempts to determine task complexity features that trigger the executive functions of working memory. Christoff (1999) proposed that tasks that activate executive WM brain areas have the following features: 1) stimulus material needs to be analyzed along different dimensions and 2) multiple processing operations have to be carried out simultaneously during performance. Although those neuroimaging studies did not explore the issue of cognitive complexity explicitly, the results imply that the number of items to be maintained simultaneously, i.e., the number of connections between items, is an important metric for cognitive complexity. From the viewpoint of information processing, connections between components create dependencies that reduce the effectiveness of the system.

Methods of complexity measures related to displays

Complexity of human-computer-interface

A human-computer-interface (HCI) is a typical dialog system in which tasks are performed through interactions between the user and the system. The user must build up a mental representation of the system's structure and learn the appropriate "language" to evoke action sequences related to the task. Such a language includes the symbolic contexts about the system.

Automaton theories model a dynamic system as a deterministic finite automaton composed of system states and transitions between states, where state is defined as a possible status of the system, while transition is an action that moves the system from one state to another. For example, a computer window under the Microsoft system may have three states: open, closed, and minimized; a mouse click is a transition to move the window between states. The challenge here is to transform a complicated human-computer interface into the structure of an automaton. Many methods have been developed to perform the transformation automatically. One example is the automatic mental model evaluator developed by Rauterberg (1993), the detail of which is beyond the scope of this review since our concern is focused on how to measure the complexity of such a system. Described below are several complexity measures based on automaton models.

Structural complexity. In simplest terms, absolute structural complexity equals the number of states (Stevens, Myers, & Constantine, 1974). Relative structural complexity is the ratio of the number of transitions to the number of states, i.e., the number of transitions per state. For example, the computer window mentioned earlier has three states thus the absolute structural complexity is 3. On the other hand, such a window allows four transitions: open -> close, open -> minimize, minimize -> open, minimize -> close. Thus, the relative structural complexity is 4/3.

Cyclomatic complexity. McCabe (1976) defined cyclomatic complexity as the difference between the total number of transitions and the total number of states. By this definition, the complexity of the above example would be 4-3=1.

Structure density. Kornwachs (1987) proposed "structure density" as a measure of system complexity. This measure estimates the actual density of transitions compared with the maximal possible density. Let S be the number of all possible states of a system and T be the number of actual transitions. The maximal possible number of transitions is $S * (S-1)$. Then the structure density is defined as $T/(S*(S-1))$. By this definition, the structure density of the above example is 4/(3*(3-1))=0.66.

Rauterberg (1992) compared the above metrics by estimating the complexity of a database system. In the experiment, the user group was composed of beginners and experts. The users performed 12 database operation tasks. The users' behavior was then recorded in a "logfile" and converted to state / transition matrices. Those matrices were used to compute complexity values using the above four measures. Except for the structure density measure, the other three measures of complexity differentiated beginners and experts well. In particular, the value of cyclomatic complexity was independent of the task. Thus, it reflects the generic structure of the system. Although structure density did not reflect the difference between beginners and experts, it was highly correlated with the tasks. Thus, it is a good index for task complexity. Overall, the results showed that the four measures are of different value in measuring task and cognitive complexity, yet McCabe's cyclomatic complexity seems to be the best measure.

In a related study, Vikal (2000) analyzed complexity of autoflight systems. He used a term "apparent complexity" to refer to the complexity perceived by the operator of a system. Vikal developed a "hybrid automation representation" to model general autoflight systems with the elements of "mode" and "transition." The modes can be modeled using control block diagrams at various levels of loop closure. The transitions can be modeled with either transition diagrams or transition matrices. Based on the model, Vakil proposed three factors that affect the apparent complexity: the number of modes in the autoflight systems, the number of transitions among modes, and the nature of transitions among modes. To compute the factors one has to quantitatively specify the terms of "control," "transition," and "mode." Vikal conducted a survey of pilots to identify the autoflight mode transitions. The transitions were analyzed using MaCabe complexity to gain insight into the apparent complexity of the autoflight system from the perspective of pilots. Notably, mode transitions that had been identified by pilots as being complex were also found to have high McCabe complexity.

Image complexity

A digital image is numerically specified; thus, the information content can be easily computed using information theory. Many algorithms have been developed to compute image complexity. The standard Boltzmann-Gibbs entropy measure defines complexity with respect to a given size of a window of view. According to the definition, image complexity, measured as configurational entropy, is a function of the total number of distinguishable spatial arrangements within view windows of a given size. The statistical paradigms based on this measure have shown great success in quantifying image complexity. However, experiments have shown that information complexity computed in term of entropy does not correspond to perceived complexity. While entropy is a measure of image disorder and reflects the lack of spatial homogeneity, complexity is a combination of order and disorder. Indeed, Grassberger (1986, 1991) has shown that complexity is sometimes posited as a mid-point between order and disorder.

Similarly, Landsberg and Shiner (1998) proposed that image complexity could be expressed in terms of order/disorder. A simple form of complexity is expressed as:
$T = delta \times (1\text{-}delta),\ delta = S/Samx$

where T is denoted to complexity, S is Boltzmann configurational entropy and $Smax$ is the highest possible value of entropy at the given size of view window. Piasecki, Martin, and Plastino (2002) compared the measures of spatial inhomogeneity and the complexity index. The results showed that inhomogenity and complexity are correlated but vary differently with the size of the view window.

Pattern complexity

Unlike Boltzmann-based complexity, pattern complexity of an image is based on measures of visual features. Orland et al. developed an algorithm to measure pattern complexity (Orland, Weidemann, Larsen, & Radja, 1994). Pattern complexity includes measures of color, edges, fractal dimensions, deviation and entropy. While the measure is somewhat correlated to human judgment of image appearance, it is not a solid predictor of perceived complexity. Klinger and Salingaros (2000) proposed a pattern complexity index based on the following visual features: size, density, line curvature, color, symmetry, similarity of shapes, and correctness of form. In their algorithm, complexity is composed of two components: Harmony and Temperature. Harmony H measures the correlation of subunits via symmetries; Temperature T measures symbol variation. The temperature components for complex structures were: 1) intensity and size of details; 2) differentiation density; 3) line curvature; 4) color-intensity; and 5) color-contrast. Harmony is a similar five-part sum composed of the following symmetry values: 1) vertical and horizontal reflections; 2) translations and rotations; 3) shape-similarity; 4) form-connectedness; and 5) color-matching. Pattern complexity can then be computed as $C = T\,(H_{max} - H)$.

Patel and Holt (2000) tested Klinger and Salingaros' algorithm against human assessment of visual complexity on binary and natural images. They asked subjects to rate the complexity of images. The tested images were manipulated differently in size, grayscale, and format from the same original image. The results showed a high correlation ($r=0.899$, $p<0.01$) between human assessment and the complexity value calculated with Klinger and Salingaros's algorithm. Interestingly, the results indicated that perceived complexity is related to the image factors described above. For example, the complexity value of an image perceived by the observers increased with the size of the image. Moreover, the complexity of an image in JPEG format varied less with the image size than the same image in GIF format did. Therefore, comparisons of image complexity should be made only when images are equal in size and are treated in the same way.

Tulis' Display complexity

Perhaps the most useful tool to quantify the information and layout of screen elements is Tullis' metric of display complexity (Tullis, 1984, 1985, 1986). Tullis studied over a thousand computer-generated displays. He measured search time to locate items on the displays and collected subjective ratings of ease of use. The results revealed that four basic characteristics of display formats affect how well users can extract information from the displays:

1. **Overall density** — the number of characters displayed, expressed as a percentage of the total spaces available.
2. **Local density** — the number of other characters near each character.
3. **Grouping** — the number of groups and average group size, both describing the extent to which characters on the display form perceptual groups. The groups can be determined by considering the white space around them.
4. **Layout complexity** — the extent to which the arrangement of items on the display follows a predictable visual scheme, typically computed as the differences in view angles between the items.

Using these four display characteristics, Tullis was able to obtain correlation coefficients of .71 for predicting search time and .90 for predicting subjective ratings. The most important predictors for search time are two measures associated with the grouping of characters: the number of groups on a display and the average visual angle subtended by those groups. The shortest search times were associated with a range of about 19 to 40 groups, which corresponds to an average visual angle of about 4.9 to 2.4 degrees. Likewise, the most important predictors of subjective ratings were a measure of local density, which is essentially how "tightly packed" the display is, and a measure of layout complexity, which is essentially how well the items on the display are aligned with each other. Layout complexity can be computed from the number of distinct items (labels, data items, etc.) and item uncertainty (use of vertical/horizontal alignment).

Tullis' metric is very useful in the sense that it is sensitive to observable differences of a system and the relative values of the metric correspond to intuitive notions about the characteristics of a display system. However, there are several limitations to Tullis' model, as pointed out by Perlman (1987). First, Tullis used plain character displays with no quasi-graphic characters such as lines for drawing boxes. Second, Tullis' model does not make use of the information structure underlying a display. Third, the model was based on predictions about search time and subjective ratings of how easily information can be extracted. These two measures may not correspond to task performance.

In a separate study, Schwartz (1988) examined how well the display format effects described by Tullis (1984, 1985) could be generalized to other display situations. The results indicated that Tullis' metrics could not predict the situation where the tasks required the use of several pieces of information from predictable display locations. Thus, it is necessary for us to study Tullis' format dimensions more fully before using his equations to evaluate display designs for use outside the task situation in which the equations were developed.

Layout Appropriateness

Tullis' metrics are task independent. They are focused on the general appearance of an interface. Therefore, it is more useful for predicting user preference than user performance other than search time. In contrast, task-sensitive metrics are more useful in understanding what users do with an interface and how to make the interface more efficient. For instance, Sears (1994) proposed a measure, called Layout Appropriateness, to evaluate the efficiency of the organization of objects in an interface. This metric first computes the cost of a layout using the following formula:

Cost = sum (frequency of transition × cost of the transition)

A transition here is considered an action a user makes on a display such as moving the mouse or closing a window. The cost of that transition is measured as the distance that users must move a mouse and the size of the object they are selecting. However, if an interface is used only to display information, then the cost is better measured with eye fixation information. The frequency of each transition can be estimated through task analysis. Once the cost is computed, the next step is to identify an optimal layout. The optimal layout can be identified with any standard searching algorithm by searching for the minimal cost based on the current method of assigning costs. Given that, Layout Appropriateness (LA) is then specified as follows:

LA= 100 × (cost of the optimal layout / cost of the proposed layout).

Sears further validated the metric with experiments. He showed that the LA value highly correlated to task completion time and user preference ratings. He further suggested that combining both task-independent and task-sensitive metrics could be more powerful than using each set of metrics alone.

Air traffic complexity

Studies on air traffic complexity have focused on identifying factors that make an air traffic situation more complex and increase the workload. The studies of air traffic complexity provide us some useful methodologies on how to develop complexity measures with respect to controllers' workload. For instance, Mogford et al. (1995) presented a literature review of air traffic control complexity. He classified the methods of determining the complexity factors into two categories: 1) asking controllers to rate complexity factors in terms of how they made the traffic control tasks more or less difficult, and 2) having controllers make paired comparisons with respect to the complexity of different situations. From the data he formulated complexity factors with analytical techniques such as multidimensional scaling.

In a similar study, Laudeman et al. proposed a metric of dynamic density as the measure of air traffic complexity (Laudeman, Shelden, Branstrom, & Brasil, 1998). The factors identified by Laudeman et al. were grouped into three categories: density factors, transition factors, and conflict factors. The density factors captured local and overall numbers of aircraft; the transition factors represented changes in aircraft states; while the conflict factors reflected the complexity imposed by the presence of potential conflicts. Interestingly, these three categories correspond to three basic aspects of complexity described earlier: size, variety, and rules.

Yet another study explored how dynamic density factors influenced controller workload (Sirdhar, Seth, & Grabbe, 1998). Through regression analysis they determined the weight of each factor in its contribution to overall complexity. However, like the work of Mogford et al. and Laudeman et al., this effort did not take into account the intrinsic disorder of air traffic. Indeed, Delahaye and Puechmorel (2000) applied the Kolmogorov-entropy metric to measure the global disorder of aircraft systems. The results indicated that topographic entropy was an intrinsic measure of the complexity of the traffic geometry because traffic with crossing trajectories had higher entropy.

Summary of complexity definitions and measurement methods

We have briefly reviewed definitions and measures of complexity from several types of studies: general concepts, information complexity, cognitive complexity, and display complexity. While each of these is focused on different aspects of human or machine systems, there is tremendous overlap among these definitions. Essentially, each definition is either fully or partially concerned with three basic aspects of complexity: size, variety, and rules. These relationships can be better understood from Table 1, which lists the definitions with the source of the research and the factors contributing to complexity. With such an understanding, we can view complexity as a 3-dimensional entity comprised of numeric size, variety, and rules. The contribution of each dimension to the entity depends on how the observer processes information and which aspects the observer is concerned with. Recall that "complexity only makes sense when considered relative to a given observer" (Edmonds, 1999). This is the critical point in the development of a complexity measure for a given application. Nevertheless, the integration of the system and the observer is either obscure or missed in many complexity measures.

Table 2 summarizes complexity measurement methodologies with the research sources and parameters to be specified. Once again, we can see from Table 2 that all the methods are aimed at different forms of the same basic factors: numeric size, variety, and rules.

ANALYSIS OF THE METHODS WITH RESPECT TO THE ATC ENVIRONMENT

While each of the methods reviewed in this report is more or less related to complexity of visual displays, it seems that none of them can be directly applied to ATC displays and allow an evaluation with respect to ATC task performance. ATC displays have unique features that differentiate them from other applications. Listed below are some typical characteristics of ATC automation displays:

1) They contain mainly text and binary graphical patterns (symbol, charts, etc.), whereas spatially continuous digital images are very rare.
2) Text and graphical patterns are usually compressed. For example, a datablock contains many pieces of iconic information.
3) ATC displays are dynamic; the information is regularly updated with the evolution of the traffic situation.
4) Unlike most human-computer-interaction systems, ATC automation tools are presented as aids, not objects that controllers have to operate on. Controllers use the aids only when they are helpful – i.e., the benefit is greater than the cost. Controllers may choose to ignore the aids and still perform their tasks. Indeed, one of the issues about the new tools is whether the benefit is greater than the cost to controllers and whether controllers will use or ignore them.

Next we will analyze the feasibility of applying the reviewed methods of measuring complexity with ATC displays. The analysis is based on our understanding of cognitive information processing in ATC. Table 2 lists the methods and summarizes the results of the analysis.

Entropy

Entropy computes redundancy in a system. Information theories state that redundancy reduces information content. An ideal engineering system is presumably designed toward reducing redundancy. However, this concept does not apply to air traffic control. First, signal processing in the human brain requires some amount of redundancy; second, the whole air traffic control system is built on redundancy to minimize operational errors. In fact, there is a great deal of redundancy in the way that ATC workstations are set up. In addition, entropy computation is based on the probabilities of all inputs that might be encountered. Unfortunately, the ATC environment tends to be much more dynamic and fuzzy.

Random Matrix Theory

By mapping the elements of a display and their relationships to a matrix, we can use Random Matrix Theory to compute independent dimensions of the elements and quantify the interconnections. On the surface, this technique seems plausible with ATC displays. For instance, we can have subjects identify the elements on an ATC display and specify their relationships. Note that the method may only apply to displays with a limited number of elements because the number of the relationships to be specified in a matrix increases as the square of the number of elements. In reality, this may make the use of Random Matrix Theory difficult, if not impossible, to employ in the complex ATC environment.

Kelly's grid technique

Kelly's grid technique, in principle, is similar to the Random Matrix Theory method. Subjects specify the elements and compare the similarity of elements in order to determine the "distance" between them; then the "independent constructs" will be derived through techniques such as principal component analysis or multidimensional scaling. The constructs can be elicited based on the notion that the distance between elements associated with the same constructs is shorter than the distance between elements associated with different constructs. A modified version of this method is to have the subjects identify the elements and describe their features (Nielsen, 1996). The "distance" can be inferred from the feature description although the inference process could be difficult to do. By doing so, the subjects are not required to specify a large number of "distances." The disadvantage is that the result could be biased with the choice of the feature description. Regardless, this technique may have some promise with ATC.

Sketch map

While the sketch map is probably the easiest method to implement, the reliability of the results is questionable. Various experiments have demonstrated that controllers have a low success rate of recalling the details of air traffic situations. Gronlund et al. reported that controllers' memory for detailed flight data was poor (remembering only important items associated with the flight) even though they exercised many actions on the flights (Gronlund, Ohrt, Dougherty, Perry, & Manning, 1998). One possible modification of the method would be to have subjects sketch their mental maps of a display in an "online" manner in which controllers are free to watch the display while they sketch what is relevant to their ATC tasks. Nevertheless, this method has a number of shortcomings. For example, controllers express their "mental thoughts" differently even though they are presumed to perform the same task at a similar performance level. Moreover, Bressolle et al. (2000) reported that controllers adapt different strategies when using an automation tool. Thus, the sketched maps could vary dramatically from controller to controller. In addition, it is difficult to quantify sketch maps. As a result, this method may only be useful for some initial pilot studies, such as a study to obtain some clues about how controllers describe a display and what display features they find important.

Cognitive task analysis

While the methods of cognitive task analysis were not originally targeted to assess information complexity, they can be very powerful in the evaluation of the completeness and efficiency of a design. Among the methods, GOMS (Goal, Operator, Methods, and Selection rules) is the one most pertinent to design evaluation. The results of GOMS include a series of steps of actions that the users have to perform to complete the tasks and the selection rules associated with the actions. The actions can be defined at various levels of abstraction. For example, one can decompose large tasks into units in terms of time to complete, or one can analyze the tasks to the level of keystrokes or eye movements to complete the task. Once the tasks are decomposed into units, we can apply other complexity measures to the sets of units and selection rules. The disadvantage of the method is that the results rely on levels of user experience and subjective interpretation.

Working memory metrics

Klemola (2000a) used the number of "unfamiliar" words to assess the complexity of text. In text reading, "unfamiliar" words are those that can't be comprehended automatically and require additional information to be understood. Although such a description of complexity is not sufficient, it nevertheless gives a straightforward assessment of the "size" factor of complexity. It seems that this method (counting the number of unfamiliar items on a display as the index of complexity) can be easily applied to ATC displays. However, the term "unfamiliar item" is not explicitly defined for ATC displays. For instance, an unfamiliar symbol could become familiar after some practice. To apply this method to ATC displays, we may replace the concept of "unfamiliar items" with "symbolic items." A symbolic item is similar to an "unfamiliar item" in the sense that it needs to be associated with other information to be comprehended. Another working memory based metric of complexity can be derived from neuroimaging studies. Christoff (1999) described the task features that activate executive WM brain areas. For example, one of the features is the number of items to be analyzed along different dimensions. Those features can be used as complexity measures.

The relational complexity metric proposed by Halford et al. (1998) is also based on working memory. This metric is extremely useful because it is directly associated with the capacity of human cognitive processing. The problem with using this metric is the difficulty in determining the interacting variables and dimensions of interaction. Consequently, the successful applications of the method so far have been mostly limited to the areas of text comprehension and logical reasoning. In contrast, ATC displays contain a lot of graphical information, making the use of his metric problematic.

Human-computer-interface

Methods that assess the complexity of a human-computer-interface (HCI) require modeling the system in terms of states and transitions. Then the complexity measures such as McCabe's cyclomatic complexity can be computed by counting the numbers of states and transitions. Unfortunately, those indices of complexity simply would not work with ATC displays. While such displays can also demand inputs from controllers, they only use the automation tools to acquire information and do not manipulate them. Therefore, there are no clearly defined states and transitions in using ATC tools. If the use of an automation tool can be described explicitly with states and transitions, it implies that controllers are forced to manipulate the tool and be manipulated by it. In that way the tool takes control over the controllers. That would be against the philosophy of automation aid design.

One important concept embedded in the methods of HCI complexity measures is that all the methods emphasize "rules" or "connections" as the main factor contributing to complexity. By this concept, it is possible that a system composed of ten elements may have the same complexity as a system composed of 100 elements as long as the elements are independent of each other. Thus, while the "size" factor of complexity describes how complex a system appears, the "rules" factor determines how complicated the computation would be to use or interpret a system. HCI and ATC automation displays are similar in the sense that users do not have to use all pieces of information on a display at once. Controllers may use some parts of displayed information at one time and others at a different time. Therefore, the complexity measures of ATC displays, like those of HCI, have to consider the "rules" factor as well as the "size factor " when measuring complexity.

Image complexity and pattern complexity

Measures of digital image complexity compute size and variability factors. Notice that variability is computed on the basis of a given scale of a view window. Although ATC tools rarely display images, the idea of developing a scale-dependent measure of complexity might be helpful in defining elements of ATC displays. On the other hand, Klinger and Salingaros' algorithm of pattern complexity (2000) is probably more suitable for ATC displays since they are mainly composed of text and graphical patterns. The algorithm identifies some basic visual features and their relationships (variability and symmetry). It has been shown that when going from a low to high value of Klinger and Salingaros' complexity, a visual image tends to alter one's response from relaxing to distressing. Yet it is unclear how this measure may correspond to controllers' workload in using automation tools.

Display complexity and layout appropriateness

Tullis' (1984) display complexity is an easy-to-use metric that quantifies the layout of screen elements. However, this method is most suitable for text displays, while ATC displays contain graphical patterns and are color-coded. In addition, the method is not concerned with the cognitive load raised from operating an interface. Nevertheless, the method would possibly be a good start toward the complexity measures of ATC displays.

In contrast, Sears' (1993) Layout Appropriateness formula computes a user's action cost in using an interface. Sears' method counts on mouse movement as the cost of action. The method does not directly compute the complexity of information on displays; rather it is more suitable for validating the complexity measures. A shortcoming of the method is the approach used to compute the cost of an action. While excessive action demands from an ATC display are not desirable, parameters of mouse movement and keystroke do not sufficiently reveal controllers' workload. For ATC tasks such as monitoring a situation and making decisions, eye movement parameters may be better candidates in validating complexity measures.

DISCUSSION

Why objective measurement of complexity?

The ultimate goal of this report was to identify objective measures of information complexity for ATC displays. Since controllers use the displays, the purpose of the complexity assessment is to make sure that the displays do not overload controllers. Traditionally, the usability and acceptability of a new tool are evaluated by having controllers use the tool and collect information from them. The information is collected through observing controllers' behavior, having them fill out questionnaires, and interviewing them. Given that such subjective techniques have been well developed, why would we want to develop additional objective measures?

While controllers' opinions about a new technology are always important sources for evaluation, several factors may bias the results obtained from subjective measures. First, subjective measures mostly reveal the degree to which people like complex interfaces. For instance, Sears (1994) reported that people usually tend to judge a tool by its perceived functionality. However, what is really needed in the evaluation is information about whether the tool helps task performance. If a tool is helpful, not only should it have great functionality, but also the benefits of using the tool should be significantly greater than the costs. Unfortunately, such benefits and costs are difficult to compute from subjective measures. One example is an ATC tool called pFAST (passive Final Approach Spacing Tool). While its functionality was highly praised by controllers, pFAST is currently not being used due to several human factors reasons (Cardosi, 2003).

Second, subjective measures are usually obtained in a simulation environment where a new tool is used stand-alone. The actual ATC environment is much more complex. The tool has to share a controller's time and attention with many other stimulus sources and tools. Thus, controllers' opinions about a new stand-alone tool can be quite different from their opinions when the tool is integrated into the operational setting. An analogy is buying a new car. One can have quite different opinions about a car when looking at it at a car dealership and by driving it under varying traffic conditions.

Third, controllers use mental models of the air traffic situation to perform their tasks and integrate information from ATC tools into their mental models. Their answers to questionnaires are based on this integration. However, the mental models are not the same for all controllers. In fact, Bressolle et al. (2000) reported that controllers adapt various strategies in using ATC tools. Therefore the answers to the same question can be drastically different among controllers.

Finally, learning to use a new tool optimally requires an extensive process of adaptation. It takes practice for controllers to achieve optimal strategies for using a new tool. Cardosi (2003) reported that the success of a tool is largely dependent on how well the system is adapted to the specific sites and its operations. Moreover, the results of the adaptation could be quite different from what was originally anticipated. Therefore, the assessment of a new tool should not simply rely on the results collected when the subjects were only briefly exposed to it. In summary, all these factors suggest that an objective, intrinsic evaluation of a new tool is an important and necessary complement to subjective measures.

What kind of complexity do we want to measure?

The complexity of a system implies some degree of computational cost. Therefore, employing an automation tool requires cognitive costs associated with integrating the tools with controllers' mental models of the situation and in programming physical actions required to use the tools. Computations such as these involve cooperative activities of many brain areas that cannot be accurately estimated using subjective means. Neuroimaging technologies may be of some benefit since they can reveal the amount of brain activity; however, we cannot put controllers under a neuroimaging machine while they perform ATC tasks. Alternatively, it may be possible to access complexity of ATC displays in terms of the visual and cognitive features and relate those features to brain activities and processing capacities.

Our goal was to develop objective measures of the complexity of ATC displays. The two running threads in this literature review were the concepts of information complexity and cognitive complexity, with various definitions for both. The basic distinction lies in that the former is typically used to describe a system while the latter is targeted at human cognitive activities. Therefore, while information complexity can have concrete, mathematically specified measures, cognitive complexity can only be estimated since the cognitive structures of human subjects are not directly observable. Unfortunately, in the end we are still faced with a basic question: What kind of complexity do we want to measure? We want to develop complexity measures that can be applied independently of controllers and yet we want the measures to be associated with controllers' task performance. The current methods of measuring cognitive complexity are user-dependent. On the other hand, generic definitions of information complexity are user-independent. More importantly, it is not evident how the measures are related to controllers' task performance.

A series of studies by Rauterberg (1992, 1993, 1995, 1998) may shed some light on the relationship between system and cognitive complexity. Rauterberg developed a framework to estimate cognitive complexity by observing user's behavior in using computer-human-interfaces. He used the term "system complexity" to refer to the information complexity of a system. This complexity is given by the concrete system structure and is independent of users and tasks. The term cognitive complexity denotes the complexity of the user's mental model of a system. In order to perform tasks, a user's cognitive structure has to closely match the system structure. In this sense, if the cognitive structure were too simple, task performance would include errors. In his work, Rauterberg defined two other terms describing complexity: behavioral complexity, the complexity of a user's observable behavior that can be estimated by analyzing recorded concrete task performance; and task complexity, the necessary knowledge to perform a task that is user-independent. With the notion that learning to perform a task using a given system means decreasing behavior complexity and increasing cognitive complexity, Rauterberg assumed that the difference between behavioral complexity and task complexity is equal to the difference between system complexity and cognitive complexity. In the case of a "best solution," cognitive complexity is equal to the information complexity of the system.

CONCLUDING REMARKS

This report reviewed a number of definitions and measures of complexity, each providing us with some useful ideas on how to assess the complexity of ATC displays. One of the major accomplishments of the report is the identification of three basic complexity factors: numeric size, variety, and rules. All complexity definitions and measures can be described by these factors. Another accomplishment is the demonstration of the power of integration: Complexity involves the integration of the system and the observer. Through the analysis of available complexity measures, we have shown that neither information complexity that focuses on the system nor cognitive complexity that aims at observers can provide a complete description for ATC application. The great variety in complexity measures reflects the fact that the contribution of each of the three factors to overall complexity depends on how information is processed by the observer; as we cited in an earlier section of the report: "The complexity of things depends on which aspect you are concerned with" (Edmonds, 1999). Therefore, we generalized that complexity is the integration of the observer with the three basic factors, as expressed in the following formula: *complexity = integration of observer and basic factors (size, variety, rules).* To achieve our ultimate goal of developing objective complexity measures for ATC tools, we need to integrate the methods presented in this report with the specifications of ATC displays. That is our target for the next step.

REFERENCES

Adams-Webber, J. R. (1970). Actual structure and potential chaos: Relational aspects of progressive variations within a personal construct system. In Bannister, D. (Ed.), *Perspectives in Personal Construct Theory*. London: Academic Press.

Adams-Webber, J. R. (1996). Cognitive complexity. In Corsini, R. and Auerbach, A. J. (Eds.), *Encyclopaedia of Psychology*. New York: Wiley.

Bates, J. E. and Shepard, K. H. (1993). Measuring complexity using information fluctuation. *Physics Letters A*, 172, 416-425.

Bennett, C. H. (1990). How to define complexity in physics, and why. In Zurek,W. H. , Addison-Wesley (Eds.), *Complexity, Entropy and the Physics of Information*. (pp. 137-148). Redwood City, California.

Bezzi, A. (1999). What is this thing called geoscience? Epistemological dimensions elicited with the Repertory Grid and their implications for scientific literacy. *Science Education*, 83, 675-700.

Bieri, J. (1955). Cognitive complexity-simplicity and predictive behavior. *Journal of Abnormal and Social Psychology*, 51, 263-268.

Bieri, J., Atkins, A., Briar, S., Leoman, R. L., Miller, H., and Tripodi, T. (1966). *Cognitive Structure and Judgment*. New York: Wiley.

Billinghurst, M. and Weghorst, S. (1995). The use of sketch maps to measure cognitive maps of virtual environments. In *Proceedings of IEEE 1995 Virtual Reality Annual International Symposium*. (pp. 40-47). IEEE Computer Society, Los Alamitos, CA.

Bressolle, M. C., Benhacene, R., Boudes, N., and Parise, R. (2000). Advanced decision aids for air traffic controllers: Understanding different working methods from a cognitive point of view. *The 3rd International Air Traffic Management Seminar*, Napoli, Italy.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In Kennedy A. and Wilkes A. (Eds.), *Studies in Long-term Memory*. (pp. 3-18). New York: Wiley.

Burleson, B. R. and Caplan, S. E. (2002) Cognitive complexity. In McCroskey et al. (Eds.), *Communication and Personality: Trait Perspectives*. Cresskill, NJ: Hampton Press.

Card, S. K., Moran, T. P., and Newell, A. L. (1983). *The Psychology of Human Computer Interaction*. Hillsdale, NJ: Erlbaum.

Cardosi, K. (2003). Human factor integration Challenges in the terminal radar approach control (TRACON) environment. Washington DC: Federal Aviation Administration; No: DOT/FAA/AR-02/127.

Casti, J. L. (1979). *Connectivity, Complexity and Catastrophe.* New York: Wiley.

Cha, S., Chung, I. S., and Kwon, Y. R. (1993). Complexity measures for concurrent programs based on information-theoretic metrics. *Information Processing Letters*, 46, 43-50.

Christoff, K. (1999). Complexity and working memory resources: Task characteristics necessitating the executive control of attention. In Kokinov, B. (Ed.), *Perspectives on Cognitive Science*, *Vol. 5*. Sofia: New Bulgarian University.

Cowan, N. (2001). The Magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*. 24, 87-114.

Crockett, W. H. (1965). Cognitive complexity and impression formation. In Haher, B. A. (Ed.) *Progress in Expremental Personality Research* (Vol. 2, pp. 47-90). New York: Academic Press.

Crutchfield, J. P. and Young K. (1989). Inferring statistical complexity. *Physics Review Letters*, 63, 105.

Delahaye, D. and Puechmorel, S. (2000). Air traffic complexity: Towards intrinsic metrics. *Proceedings of the 3rd USA/Europe Air Traffic Management R&D Seminar*. (pp. 13-16). Napoli, Italy.

Downs, R. J. and Stea, D. (1973). Cognitive maps and spatial behavior. In Downs, R.J. and Stea, D. (Eds.), *Image and Environment* (pp. 8-26). Chicago: Aldine Publishing Company.

Downs, R. J. (1976). Cognitive mapping and information processing: A commentary. In Moore, G. and Golledge, R. (Eds),. *Environmental Knowing: Theories, Research and Methods*. (pp. 67-70). Dowden: Hutchinson and Ross, Inc.

Drozdz, S., Kwapien, J., Speth, J., and Wojcik, M. (2002). Identifying complexity by means of matrices. *Physica A*, 314, 355-361.

Edmonds, B. (1999). What is complexity? - The philosophy of complexity per se with application to some examples in evolution. In Heylighen, F. and Aerts, D. (Eds.): *The Evolution of Complexity*. (pp. 1-18). Kluwer, Dordrecht.

Fisher, D. L. (1984). Central capacity limits in consistent mapping, visual search tasks: Four channels or more? *Cognitive Psychology*, 16, 449-484.

Grassberger, P. (1986). Towards a quantitative theory of self-generated complexity, *International Journal of Theoretical Physics*, 25, 907-938.

Grassberger, P. (1991). Information and complexity measures in dynamical systems. In Atmanspacher, H. and Scheingraber, H. (Eds.), *Information Dynamics.* (pp. 15-33). New York: Plenum Press.

Gronlund, S. D., Dougherty, M. R. P., Ohrt, D. D., Thomson, G. L., Bleckley, M. K., Bain, D. L., Arnell, F., and Manning, C. A. (1997). The Role of memory in air traffic control. Washington DC: Federal Aviation Administration; No: DOT/FAA/AM-97/22.[1]

Gronlund, S. D., Ohrt, D. D., Dougherty, M. R. P., Perry, J. L., and Manning, C. A. (1998). Aircraft importance and its relevance to situation awareness. Washington DC: Federal Aviation Administration; No: DOT/FAA/AM-98/16.[1]

Halford, G. S., Wilson, W. H., and Phillips, W. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. *Behavioral Brain Sciences*, 21, 803-831.

[1]This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications Web site: http://www.cami.jccbi.gov/aam-400A/index.html

Holm, H. J. (1993). Complexity in Economic Theory. *University of Lund: Lund Economic Studies*, Lund, Sweden.

Humphreys, M. S., Bain, J. D., and Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic and procedural tasks. *Psychological Review*, 96, 208-233.

John, B. E. (1995). Why GOMS? *Interactions*, *2*, 80-89.

Jonassen, D. H., Tessmer, M., and Hannum, W. H. (1999). *Task Analysis Methods for Instructional Design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kauffman, S. A. (1993). *The Origin of Order*. New York: Oxford University Press.

Kelly, G. A. (1955). *The Psychology of Personal Constructs*. New York: W. W. Norton and Company.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*, London: Cambridge University Press.

Klemola, T. (2000a). A Cognitive model for complexity metrics. *4th International ECOOP Workshop on Quantitative Approaches in Object-Oriented Software Engineering (*ECOOP 2000)*,* Sophia Antipolis, France.

Klemola, T. (2000b). Cognitive complexity metrics and requirements comprehension, *Australian Conference on Software Measurement (ACOSM 2000)*, Sydney.

Klinger, A. and Salingaros, N. A. (2000). A pattern measure. *Environment and Planning B: Planning and Design*, 27, 537-547.

Kornwachs, K. (1987). A quantitative measure for the complexity of man-machine interaction process. In Bullinger, H-J. and Shackel, B. (Eds), *Human-Computer-lnteraction INTERACT'87*. (pp. 109-116). North-Holland: Elsevier Science.

Kuipers, B. (1983). The cognitive map: Could it have been any other way? In *Spatial Orientation: Research, Theory, and Application*. New York: Plenum Press.

Landsberg, P. T. and Shiner, J. S. (1998). Disorder and complexity in an ideal non-equilibrium Fermi gas. *Physical Letter A,* 245, 228-32.

Langton, C. (1991). Life at the edge of chaos. In Langton, C., Taylor C., Farmer J., and Rasmussen, S. (Eds.). *Proceedings Artificial Life II*. (pp. 41–91). Addison-Wesley.

Laudeman, I. V., Shelden, S. G., Branstrom, R., and Brasil, C. L. (1998). Dynamic density: An air traffic management metric. *NASA-TM*, 1998-112226.

Linville, P. W. (1985). Self-complexity and affective extremity: Don't put all of your eggs in one cognitive basket. *Social Cognition*, 3, 94-120.

Linville, P. W. (1987). Self-complexity as a cognitive buffer against stress-related illness and depression. *Journal of Personality and Social Psychology*, 52, 663-676.

Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.

McCabe (1976). A complexity measure. *IEEE Transactions on Software Engineering*, 2, 308-320.

Moertl, P. M., Canning, J. M., Gronlund, S. D., Dougherty, M. R., and Johansson, J. (2002). Aiding planning in air traffic control: an experimental investigation of the effects of perceptual information integration. Washington DC: Federal Aviation Administration; No: DOT/FAA/AM-02/22.[1]

Mogford, R. H., Guttman, J. A., Morrow, S. L., and Kopardekar, P. (1995). The complexity construct in air traffic control: A review and synthesis of the literature. Washington DC: Federal Aviation Administration; No: DOT/FAA/CT-TN95/22.[1]

Morçöl, G. (2002). *A New Mind for Policy Aanalysis: Toward a Post-Newtonian and Postpositivist Epistemology and Methodology.* Westport, CT: Praeger.

Morçöl, G. and Asche, M. (1993). *The repertory grid in problem structuring: A case illustration*. The International Journal of Personal Construct Psychology, 6, 371-390.

Nielsen, M. E. (1996). H-comp: A program to calculate information complexity. *Behavior Research Methods, Instruments & Computers*. 28, 483-485

Orland, B., Weidemann, E., Larsen, L., and Radja, P. (1994). *The effects of visual variety on perceived human preference*. The Fifth International Symposium on Society and Resource Management. Fort Collins, CO.

Patel, L.N. and Holt, P. (2000). Testing a computational model of visual complexity in background images. In *Advanced Concepts for Intelligent Systems*. (pp. 119-123). Baden.

Perlman, G. (1987). An axiomatic model of information presentation. *Proceedings of the 31st Human Factors Society Annual Meeting* (pp. 1229-1233). Santa Monica, CA.

Piasecki, R., Martin, M. T., and Plastino, A. (2002). Inhomogeneity and complexity measures for spatial patterns. *Physica A*, 307, 157-171.

Pressing, J. (1999). Cognitive complexity and the structure of musical patterns. *Proceedings of the 4th Conference of the Australasian Cognitive Science Society*. Newcastle.

Rauterberg, M. (1992). A method of a quantitative measurement of cognitive complexity. In G. van der Veer, M. Tauber, S. Bagnara, and M. Antalovits (Eds.), *Human-Computer Interaction: Tasks and Organisation—ECCE'92* (pp. 295-307). Roma: CUD.

Rauterberg, M. (1993). AMME: an "Automatic Mental Model Evaluator" to measure complexity of user behaviour recorded on logfiles. In: *Abridged Proceedings HCI International '93* (School of Industrial Engineering, Purdue University (pp. 191). West Lafayette, IN.

Rauterberg, M. (1995). About a framework for information and information processing of learning systems. In: Falkenberg, E., Hesse , W. and Olive, A. (Eds.), *Information System Concepts—Towards a consolidation of views*. (pp. 54-69). London: Chapman and Hall.

Rauterberg, M., Schluep, S., and Field, M. (1998). Modelling of cognitive complexity with Petri nets: An action theoretical approach. In Trappl, R. (Ed.), *Proceedings of Cybernetics and Systems EMCSR'98*, 2, 842-847. Vienna, Austria.

Schwartz, D. R. (1988). The Impact of task characteristics on display format effects. *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 352-356). Santa Monica, CA.

Scott, W. A. (1969). The Structure of natural cognition. *Journal of Personality and Social Psychology*, 12, 261-278.

Sears, A. (1993). A layout appropriateness: *A metric for evaluating user interface Widget layout. IEEE Trans Software Engineering.* 19, 1-13.

Sears, A. (1994). Automated metrics for user interface design and evaluation. *International Journal of Bio-Medical Computing,* 34, 149-157.

Sirdhar, B., Seth, K. S., and Grabbe, S. (1998). Airspace complexity and its application in air traffic management. *Proceedings of the 2nd USA/EUROPE ATM R&D seminar,* Orlando, FL.

Stevens W. P., Myers G. J., and Constantine, L. L. (1974). Structured design. *IBM System Journal,* 13, 115-139.

Tullis, T. S. (1984). A computer-based tool for evaluating alphanumeric displays. In Shackel, B. (Ed.), *Human-Computer Interaction: INTERACT '84*. London, England.

Tullis, T. S. (1985). Designing a menu-based interface to an operating system. *Proceedings of CHI'85 Conference on Human Factors in Computing Systems*. San Francisco, CA.

Tullis, T. S. (1986). Optimizing the usability of alphanumeric displays. *Proceedings of HCI'86 Conference on People and Computers*. York, England.

Vikal, S. (2000). *Analysis of complexity evolution management and the human performance issues on commercial aircraft automation systems.* Ph.D thesis, MIT-ICAT. Cambridge, MA.

Wilson, W. H. and Halford, G. S. (1994). Robustness of tensor product networks using distributed representations. *Proceedings of the Ninth Australian Conference on Neural Networks.* (pp. 47-51). Brisbane, Australia.

Woehr, D. J., Miller, M. J., and Lane, J. A. S. (1998). The development and evaluation of computer-administered measure of cognitive complexity. *Personality and Individual Differences,* 25, 1037-1049.

FIGURE

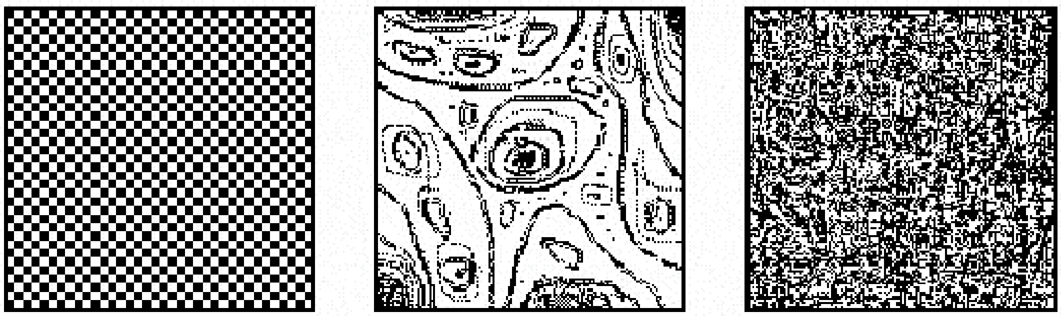

Figure 1: Images with increasing variation from left to right. (Grassberger, 1991)

TABLES

Table 1. Definitions of complexity

Source	Definition
General understanding	Combination of size, variety and rules.
Complexity by Drozdz (2002)	A trinity of comprising coherence, chaos and a gap between them
Kolmogorov complexity (Casti, 1979)	Minimum description size
Effective Measure Complexity (Grassberger, 1986)	The amount of information that must be stored in order to make an optimal prediction about the next symbol to the level of granularity
Topological complexity (Crutchfield &Young, 1989)	The minimal size of the automaton that can statistically reproduce the observed data within a specified tolerance
Complexity by Langton (1991)	Level of mutual information, which measures the correlation between information at sites separated by time and space.
Bennett logical depth (Bennett, 1990)	Computational cost (time and memory) taken to calculate the shortest process that can reproduce a given object.
Hieratical complexity (Bates & Shepard, 1993)	Number of local states, dimensionality and rule-range.
Cyclomatic complexity (McCabe, 1976)	Difference of the total number of transitions and the total number of states.
Edmonds complexity (Edmonds, 1999)	The difficulty to formulate an overall behavior with given atomic components and their inter-relations
Cognitive complexity (Crokett, 1965)	The entities of differentiation, articulation and hierarchic integration
Bieri's index of cognitive complexity Bieri, 1955)	Number of constructs and matches between the constructs
Relational complexity (Halford et al., 1998)	The number of interacting variables that must be presented in parallel to perform a process entailed in a task.
Kauffman complexity (Kauffman, 1993).	Number of conflicting constraints

Table 2. Methods of complexity measures

	Brief description of the method
Etropy	Map the system to discrete elements and determine the probability of each element relative to others.
Random matrix theory	Determine the elements and specify the relationship between elements.
Kelly's grid technique	Define the elements; describe the properties of elements or compare the similarity between pairs of elements.
Sketch map	Reveal one's mental representation of a system by having subjects sketch the structure and details of the system.
Working-memory metrics	Determine the items that need to be associated with other items for task performance, and determine the level of relations by which the items are interacted.
Human-computer-interface complexity	Model the system into an automaton composed of elements and their interconnections, then determine the complexity from the numbers of elements and interconnections.
Pattern complexity	Determine visual features of the pattern such as size, density, line curvature, color, symmetry, similarity of shapes, etc, and then compute the harmony and variations of those features.
Image complexity	Compute the variations and inhomogeneity of image pixels with a given size of window of view.
Display complexity	Specify text density, text blocks and relative positions of text blocks then compute complexity using Tullis's metrics (1984).
Human-to-computer complexity	Determine the actions needed to use the interface and compute the cost of the actions.