Federal Aviation
Administration

DOT/FAA/AM-07/26
Office of Aerospace Medicine
Washington, DC 20591

Information Complexity in Air Traffic Control Displays

Jing Xing
Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

September 2007

Final Report

NOTICE

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:
www.faa.gov/library/reports/medical/oamtechreports/index.cfm

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-07/26	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Information Complexity in Air Traffic Control Displays		5. Report Date September 2007 6. Performing Organization Code
7. Author(s) Xing J		8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplemental Notes

Work was accomplished under approved task HRRD522

16. Abstract

Air traffic controllers typically use visual displays to interact with various automation systems. Automation tools are intended to reduce controller task load, but they may also create new tasks associated with acquiring, integrating, and utilizing information from displays. Consequently, the complexity of information displayed may reduce the efficiency and effectiveness of an automation system. Moreover, complexity could cause controllers to miss or misinterpret visual data, thereby reducing safety. Thus, information complexity in air traffic control (ATC) displays represents a potential bottleneck in ATC systems. To evaluate the cost and benefit of an automation system, it is important to understand whether the information it provides is too complex for controllers to process. The purpose of this study was to answer three basic questions: 1) What constitutes information complexity in automation displays? 2) What level of display complexity is “too complex” for controllers? 3) Can we objectively measure information complexity in ATC displays? In this study, we first developed a general framework for measuring information complexity. The framework reduces the concept of complexity into three underlying factors: *quantity*, *variety*, and the *relations* between basic information elements; each factor is evaluated at three generic stages of human information processing: *perception*, *cognition*, and *action*. By this definition, we decompose complexity into a 3x3 matrix, measuring the effects of a complexity factor on information processing at a given stage. We then take the following steps to develop complexity metrics for ATC displays: 1) Identify task requirements of using the displays in ATC; 2) Determine corresponding brain functions pertinent to the task requirements; and 3) Choose the metric that can measure the effects of the complexity factor on the brain functions. Using this approach, we developed nine metrics of ATC display complexity. These metrics provide an objective method to evaluate automation displays for acquisition evaluation and design prototypes.

17. Key Words Information Complexity, Display, Interface Design, Evaluation, Air Traffic Control		18. Distribution Statement 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) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 19	22. Price

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

ACKNOWLEDGMENTS

The author thanks Dr. Carol Manning for her consultations throughout this study. Thanks, also, to Drs. Lawrence Bailey, Ben Willems, Pam Della Roccco, and Earl Stein for their suggestions and advice. Additional thanks go to Dino Piccione, Steve Cooley, and Paul Krois for their support of the Federal Aviation Administration's requirements for information complexity research. Finally, the author expresses appreciation for the Federal Aviation Administration internal reviewers, whose comments greatly improved the quality of this paper.

CONTENTS

INFORMATION COMPLEXITY IN AIR TRAFFIC CONTROL DISPLAYS

INTRODUCTION

"*This thing is too complex to use*!" In an age of information technology, most new technology users have either said or heard this phrase at one time or another. Over the past several decades, computer technologies have achieved enormous progress in the speed at which they can process and present information. With these processing limitations overcome, the new challenge for developers of new technologies is to provide systems with more types of information to process. Such efforts result in information being presented in complex ways. Conversely, human information processing capabilities have not experienced the same level of improvement. Therefore, the complexity of information provided by new technologies can easily exceed human capacity limitations. As a result, complexity has become a big concern in the information technology industry. A survey (Economist, Oct. 28, 2004) indicated that information complexity was holding the industry back as many technologies were not implemented because of their complexity. One survey respondent stated, "66% of all IT projects either fail outright or take much longer to install than expected because of their complexity." As the complexity of technology increases, its usefulness decreases, because it becomes "too complex to use." Also, the efficiency of using these technologies decreases, because managing the use of complex systems often increases rather than mitigates workload. Furthermore, high levels of complexity can result in an operator missing or misunderstanding even critical pieces of information. These types of performance errors are of particular concern to those work domains where safety is mission-critical.

Unfortunately, as information technology has become an inescapable part of the aviation system, the aviation industry cannot avoid the problem of information complexity. In particular, air traffic control (ATC) work is performed in a dynamic environment where controllers continuously receive large volumes of information from multiple sources; they must monitor this information for changes in the environment, and they must make decisions and perform effective actions in a timely manner. As technological capabilities have developed over time, there has been a tendency to add information to ATC displays and to automate controller functions. Although these new technologies are meant to improve ATC performance, they can also increase the information complexity with which the controller has to contend. The last thing controllers need from new technologies is to be overloaded by large amounts of information they cannot use. Thus, to have a new technology operate safely, efficiently, and at its full performance potential, its complexity should be minimized within its operational scope. Before the industry will be able to reliably control the complexity of new systems, it will need an objective way to measure the complexity associated with new information technologies.

The purpose of this study was to develop metrics of information complexity that can be applied to the evaluation of ATC displays. For it to be applicable, the complexity measures need to meet the following requirements: 1) quantitatively correlated with performance measures; 2) independent of specific ATC displays; and 3) as much as possible, independent of users' experience with a display. As an initial attempt, Xing and Manning (2005) performed an extensive review of previous complexity studies and analyzed the findings of these studies for their potential application to ATC displays. It was hoped that previously developed measures could be found that would meet the purpose and requirements stated above. Instead, they found that most of the measurements of complexity focused either on the display system or on the information processing of the human operator, while seldom addressing both. A complete description of complexity for ATC technologies needs to consider the factors involved with both the human operators and the display systems. The literature analysis concluded that, for ATC display applications, a measurement that integrated both human information processing and display factors would need to be developed.

The purpose of this study raised three basic questions: What is complexity? How complex is too complex for users? Finally, is it possible to quantitatively measure the complexity of ATC displays? We organized this paper into two parts to address the above questions. In the first part, "Identifying Complexity Metrics," we summarize our understanding of information complexity from a literature analysis. Then we describe a framework that decomposes complexity into a set of factors associated with human information processing. After that, we present a set of complexity metrics developed for ATC displays. Parts of the preliminary version of these results have been published in an earlier report (Xing, 2004). In the second part, "A Preliminary Case Study: Assessing the Cognitive Complexity of the Microsoft PowerPoint Interface," we present a preliminary case study to demonstrate how the metrics can be used to evaluate the complexity of a display.

IDENTIFYING COMPLEXITY METRICS

Methods

We first derived a definition of information complexity from an analysis of the literature. We next developed a framework for measuring complexity. The framework is generally applicable to interactive visual displays. The framework specifies that complexity is constrained by task requirements. We then applied the framework to ATC displays and developed a set of complexity metrics by combining the mechanisms of human information processing with ATC task requirements.

RESULTS

Understanding Information Complexity

Although there are many definitions of complexity in the literature, the term has proven to be very difficult to accurately define. For instance, a simple Internet search on complexity will yield literally hundreds of definitions and measures. Xing and Manning (2005) reviewed and synthesized the major contributions to complexity associated with visual displays. In their report, they reviewed the literature from several lines of study: general concepts, information complexity, cognitive complexity, and display complexity. While each of these areas is focused on different aspects of human or machine systems, the definitions have a great deal in common. Essentially, all the reviewed definitions and measures converged on three factors associated with complexity: *quantity* of basic information elements in a system, *variety* of elements, and the *relations* between elements. Xing and Manning's analysis revealed that the concept of complexity is multi–dimensional and cannot be sufficiently described with a single measure, and they proposed that complexity is the combination of the three basic factors.

Quantity. Intuitively, the *quantity* of basic elements is related to the complexity of a system. Whether referring to minimum description size of a system (Bennett, 1990; Crutchfield & Young, 1989), number of states (McCabe, 1976), or number of "chunks" in cognition (Klemola, 2000), all studies of quantity seem to agree that a larger quantity corresponds to a higher degree of complexity. Nevertheless, quantity alone is not sufficient to define complexity in its entirety.

Variety. Indeed, the *variety* of the elements in a system is also a key component of many definitions of complexity. The concept of variety has been widely used in the literature. For instance, many researchers have used the degree of disorder or entropy in information theories as the measure for variety or complexity, even though disorder alone cannot sufficiently describe complexity. As Drozdz and other researchers have pointed out, complexity lies somewhere between order and disorder (Drozdz, Kwapien, Speth, & Wojcik , 2002). Burleson and Caplan (1998) summed up the use of variety for defining complexity when they stated, "The concept of complexity refers to diversity of forms, to emergence of coherent patterns out of randomness and also to some ability of frequent switching among such patterns."

Relations. *Relations* among the basic elements (rules of structures, interconnections, etc.) of a system also contribute to its complexity. Individual parts of a system are held together by the relations of its internal structure. An example would be a chess pattern. A chess pattern can be of great complexity to a player because the player values the relations between the elements, not just the number and the variety of them.

Xing and Manning also identified two principles that contribute to the diversity of complexity measures. The first principle is *observer dependency*. As Edmonds (1999) described, "Complexity only makes sense when considered relative to a given observer." One example would be the experiment performed by Grassberger (1991) where subjects were asked to assess the complexity of a set of images. Figure 1 shows three images that Grassberger used in the experiment. The variety of the images, measured as the disorder of image pixels, increased from left to right. Thus, the image on the right is the most complex from the perspective of computer image processing. Yet, human subjects perceived the middle one as the most complex. The reason is that the human visual system processes visual features within small local areas rather than individual pixels, per se. Those features include line orientation, spatial frequency, luminance contrast, color, etc. The image in the middle of Figure 1 contains many distinctive visual features, while the left and right images are composed of homogeneous texture with little visual features. This experiment indicates that the effects of variety on complexity depend on how observers process information.

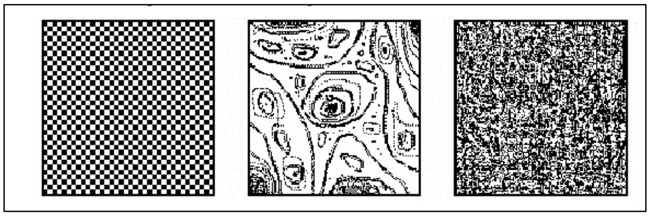

Figure 1: Variety increases from left to right, yet humans perceives the middle one the most complex.

The second principle is *task dependency*. That is, the complexity of things depends on which aspect you are concerned with. For example, if the task is to count peas in a basket, then the complexity of the peas does not vary with the number of the peas and variations in the shape or color of the peas. Therefore, it is essential to determine the task requirements of a visual display before assessing its complexity.

To summarize our understanding of information complexity, we generalized the following definition from the literature: Complexity is the combination of three basic factors: quantity, variety, and relation of basic elements; all three of which are evaluated by the mechanisms of observers' information processing and constrained by task requirements.

A Framework for Decomposing Factors of Information Complexity

Since complexity depends on how observers process information, we looked into the mechanisms of information processing in the human brain. Based on the extensive literature, we outlined a conceptual model of human information processing associated with the use of visual displays. In this simple model, information presented via visual display devices is processed in three stages: perception, cognition, and action. Through perception, a user acquires visual features of displayed information. The perceived information then feeds into the cognition stage, where one's perception is integrated with information from long-term memory, and an internal (mental) representation of the display is generated. Based on this representation, users can then make action plans to use the information or interact with the display.

Given that several hundreds of brain areas have been identified for special aspects of information processing, the above model is over-simplified and we only use it for application purposes. The definitions of perception and cognition have been controversial; however, opinions about the basic distinction between perception and cognition are more consistent among researchers: Perception does not necessarily involve working memory while cognitive activities are based on working memory. At the perception stage, the brain responds to a stimulus with neuronal activities, and the activities end with the offset of the stimulus. However, after the stimulus ends, the neuronal activity for cognitive processing continues until an action or a plan for actions is made and executed. In this sense, our model is similar to Wickens' model for interface evaluation, where information is processed at four sequential stages: perception, working memory, decision, and action (1991).

While our three-stage classification of brain functions is coarse, those stages have intrinsically different mechanisms of information processing. One example would be the neuronal response patterns over time, as illustrated in Figure 2. Perceptual neuronal responses (left panel) start following the onset of a stimulus and ends after the stimulus is no longer present, while the cognitive responses (middle panel) remain for an extended period without the stimulus, such sustained activity is the substrate of working memory. On the other hand, neuronal responses in the cortical pre-motor areas become activated before an action and end after the action plan is executed. Another example is the way in which information is encoded. Perceptual information processing is initially performed in a parallel manner. Thousands of visual neurons are activated by a visual image and they simultaneously encode many pieces of the image. Thus, the perceptual system offers a relatively broad information bandwidth. On the other hand, working memory, as the basis of cognitive activities, has a highly limited capacity. That is, only a few pieces of information can be simultaneously encoded (Cowan, 2001). Consequently, the information bandwidth of the cognition stage is much less than that for perception. Finally, the cortical areas that encode action plans are characterized with "one plan at a time," yielding an even narrower bandwidth (Georgopoulos, Schwartz, & Kettner, 1986; Pouget, Zemel, & Dayan, 2000; Xing & Andersen, 2000).

Given the inherent differences in the three stages, the three complexity factors should be separately evaluated at each stage. This results in a 3x3 matrix, as shown in Table

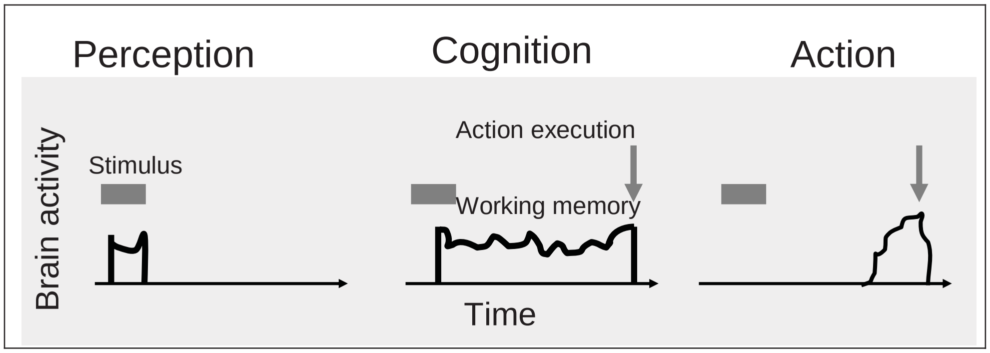

Figure 2: Activity patterns over time for information processing in the three stages.

Table 1: Metrics of information complexity

	Perception	Cognition	Action
Quantity			
Variety			
Relation			

1, with the rows being the three complexity factors and columns being the three stages. Therefore, we decompose complexity into nine metrics. Each metric describes the effect of the complexity factor on information processing at a given stage, so it is associated with the operator's task performance. For example, one of the metrics that we proposed is the degree of clutter (described in detail later). Clutter occurs when the perception of a central target is masked by the presence of overlaying and immediately surrounding stimuli. Clutter directly affects the speed and accuracy of text reading and target detection. With the known capacity limits of brain information processing, these metrics can elucidate why a visual display can be too complex for human operators. Given the principle that complexity is constrained by task requirements, the format of each metric may vary with different applications. Next, we used this framework to develop metrics specifically for ATC displays.

Metrics of Information Complexity for ATC Displays

Since complexity is constrained by task requirements, developing metrics for ATC displays needs to consider the generic tasks associated with using the displays. Therefore, we used the following steps to develop the metrics:

- Identify task requirements;
- Determine corresponding brain functions pertinent to the task requirements; and
- Choose the metric that can measure the effects of the complexity factor on the brain functions.

This procedure requires understanding the nature of the tasks associated with using displays. Since the purpose of this study is to develop complexity measures for generic ATC displays, we extracted some typical characteristics of those displays:

- The displays contain mainly text, icons, and other binary graphical patterns (symbol, charts, etc.);
- Many categories or types of data are presented in intermingled and sometimes overlapped patterns;
- Displays are often dynamic; information is continuously updated; and
- Displays are interactive; users actively access to and update information.

Perceptual Complexity

We derived the following basic perceptual tasks for using generic ATC displays. The generic tasks include: 1) Detect critical messages; 2) Search for data of a given category; and 3) Scan / rapidly read graphic patterns and text. The perceptual functions involved in these tasks include target pop-out (*Pop-out* means that a visual target can be effortlessly detected irrespective of the amount of surrounding visual materials), detection, search, segmentation, text reading, etc. To drive measurements of the effects of the complexity factors on the performance of these tasks, we need to understand the underlying mechanisms of these functions.

The mechanisms of the above visual tasks can be described with the well-known two-step model of visual information processing. Figure 3 illustrates the model in which the visual system processes information in two steps. In the first step, a visual image is segmented into distinctive visual objects, and salient targets pop out of the image. This step involves parallel processing. Based on the results of the parallel processing, the second step involves the visual system serially focusing on the salient targets or selected objects so that information can be analyzed in detail. Next we evaluated the three complexity factors in this perceptual model to derive the metrics of perceptual complexity.

Quantity evaluated by perception. According to the perceptual model, the quantity factor does not affect image segmentation and target pop-out due to their parallel processing. However, it does affect the serial processing of visual details. Processing time increases with the number of visual elements in a display. Since serial processing is limited to the information available within retinal fovea where the eyes are fixated, the basic visual element in serial processing is the fixation. Therefore, we propose that the metric of *Quantity evaluated by perception* is the number of fixation groups. A fixation group is defined as a set of visual stimuli that can be perceived within a foveal fixation for detailed analysis. Typically, a foveal fixation spans a viewing angle of about 2-4 degrees. The average time to search for a particular target on a display increases with the number of fixation groups. While there is no physiological limit on how many fixations one can make on

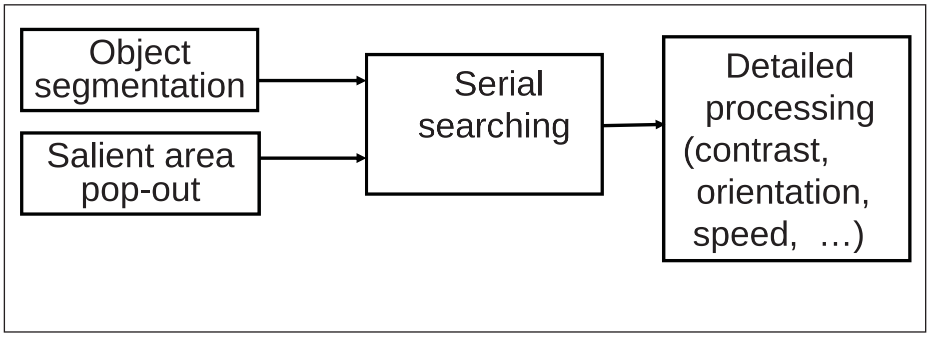

Figure 3: Diagram of the two-step perceptual model.

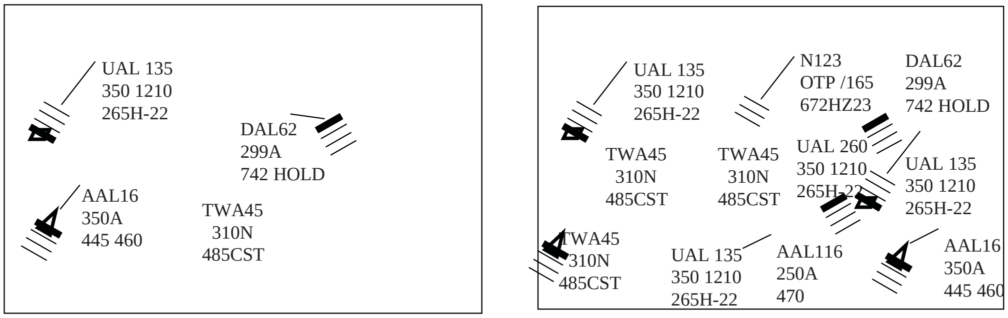

Figure 4: Graphic illustration of the effects of the quantity factor on perception.

a display, visual experiments have demonstrated that it takes 600-700ms for an observer to perceive the information in one fixation. Therefore, the capacity limit of this metric is determined by the time available for users to monitor a display. For example, if an air traffic controller has ten seconds maximally to scan all the information on a display, then the number of fixation groups included in the display should be less than 14.

Figure 4 illustrates the concept of the metric. Both pictures contain aircraft symbols and datablocks, mimicking those on controllers' radar displays. Controllers typically scan the datablocks and fixate on important ones for detailed processing. Thus, each datablock can be counted as one fixation group. The picture on the left panel has four datablocks and to scan them all would take at least 4 × 600ms = 2. 4 sec. In contrast, the picture on the right panel has 11 datablocks so it would take at least 11× 600ms = 6.6 sec to scan all of them.

In many applications, displays are crowded and it takes many fixations to view all the information. One strategy to reduce perceptual complexity is to use color or other cues to aid visual segmentation, so information can be segregated into several objects. Consequently, the number of fixations required to complete the visual search is reduced.

Variety evaluated by perception. Variety is related to image segmentation and pop-out, both based on the uniformity and distinction of the visual features. Therefore, we proposed that this metric involved the number of different visual features, including distinctive colors, luminance contrast, spatial frequency or size, texture, and motion signals in a display. Increasing the variety of visual features leads to difficulties in visual segmentation and target pop-out; as a result, a complex display cannot be efficiently organized, and salient targets cannot be instantly detected without being searched. In addition, visual studies have demonstrated that switching between visual features such as color and luminance contrast increases search time. This effect is called "cost of switching." Switching also reduces the reliability of text reading and target detection. Consider, for example, that the two pictures in Figure 5 contain the same materials. The text in the left panel has the same font and uses three colors. The red letters "LA" indicate an alert that controllers need to instantly detect and pay attention to. The blue and black text represents two categories of datablocks. This picture uses three visual features (colors) to segment the displayed information, and the segmentation is very effective. However, the figure in the right panel uses many visual features (colors, font, sizes, etc) and segmentation becomes impossible. As a result, the right figure appears to be more complex than the one on the left due to its variety.

Relation evaluated by perception. The relationship of visual elements affects the processing of detailed visual information. In particular, the perceived contrast of a visual stimulus depends on the physical contrast of the stimulus and other visual stimuli in its surrounding area. Contrast is important as the difficulty of text reading and graphic target detection are primarily determined by the perceived contrast. Therefore, we proposed that the metric for *relation evaluated by perception* was the degree of clutter, defined as the effect of the visual perception of a stimulus being masked by the presence of other stimuli. Clutter can increase search time and reduce target detection as well as text readability. The effect is apparent when background visual stimuli are spatially superimposed on the target. Moreover, the perceived luminance contrast of a visual target can also be largely suppressed by the presence of surrounding stimuli. Reduction in perceived contrast results in significant deterioration of text readability. Xing and Heeger (2001) examined this effect and found that the perceived contrast of a sine-grating patch embedded in a large patch of the same kind of gratings was about half the contrast perceived when the central patch was presented alone. However, when a blank gap was introduced between the central and surrounding patch, the suppression effect became much weaker. These experimental results implicitly suggested two methods to reduce the clutter effect: 1) reducing the amount of text in a display and, 2) keeping blank surrounds for targets to be quickly read or detected.

Figure 6 is an illustration of the clutter effect. The picture on the left shows the datablocks in a baseline conflict alert display to alert pilots of potential conflicts with other aircraft. The picture on the right shows the datablocks on a prototype of the display improvement. Since pilots primarily need only the aircraft heading and altitude information, the prototype displays altitude in the datablock and uses a simple triangle to indicate the current heading direction of the aircraft. The additional information is hidden and displayed only upon the user's request. The clutter is thus greatly reduced, and the information can be more easily read. Notice that this declutter strategy works for pilot but not controllers who need to see the hidden information most of the time. Controllers typically declutter their radar displays by showing part of the datablock (called *limited-datablock*) when too many aircraft make the displays crowded.

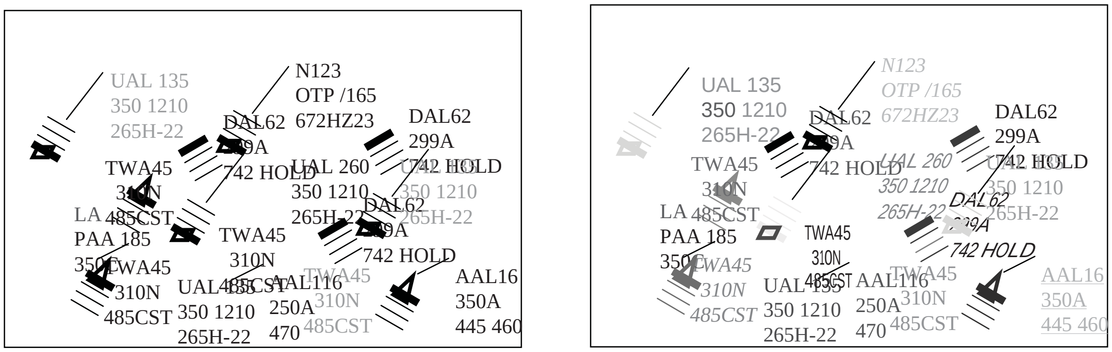

Figure 5: Illustration of the effects of the variety factor on perception.

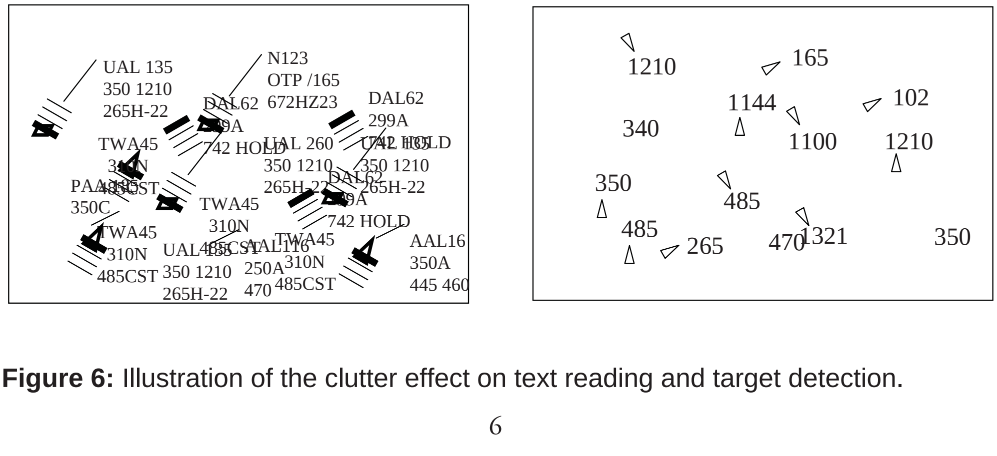

Figure 6: Illustration of the clutter effect on text reading and target detection.

Cognitive Complexity

We theorized that the following cognitive tasks are performed when operators interact with ATC displays. The generic cognitive tasks include 1) constructing, maintaining, and updating the mental representation of the information contained in the display; 2) comprehending text and graphic information; and 3) binding (or associating) items of information to plan an action or make a decision. All of these cognitive tasks require working memory. Therefore, measures of cognitive complexity should be based on quantifying the demand that using a display imposes on working memory.

Traditionally, working memory has been considered as the limited-capacity storage system involved in the *maintenance* and *manipulation* of information over short periods of time. However, recent findings suggest that working memory for *maintenance* is different from that for *manipulation*. Memory for *maintenance* is about chunks, or elements that are in our conscious awareness in the absence of sensory inputs, while memory for *manipulation* is about the independent elements or variables that must be simultaneously considered to plan an action or make a decision (Cowan, 2001; Halford, Wilson, & Phillips, 1998). Based on recent findings from psychophysical and neurophysiological studies, we generalized a working memory model that incorporates *maintenance* and *manipulation* memory, and we used the model to evaluate the complexity factors.

Figure 7 shows a diagram of the working memory model. It consists of input mechanisms (long-term memory and inputs from sensory systems) and two working memory buffers: *processing* memory and *maintenance* memory. We used the term "*processing*" rather than the traditional term "*manipulation*" to be consistent with the recent literature and to emphasize that this type of memory is for processing information, not for maintaining it. The circles in each buffer represent the elements of information. Open circles represent items of available information; filled circles represents information selected for an action. The arrows represent information flow between the buffers.

The model performs tasks through interactions between processing memory and maintenance memory. Below are the basic ways that the model processes a complex task:

- Maintenance memory keeps items of information "on-line" without being attended to; such items form a "to-do" list for completing a task; they can be quickly retrieved and are subject to decay if not attended to over a period of time.
- Processing memory binds pieces of information that are simultaneously needed for planning an action or making a decision.
- Processing memory retrieves information from sensory systems, maintenance memory, and / or long-term memory. Information needed for an action or decision is selected from those sources and associated in processing memory.
- Depending on the task, processing memory can discard information that is no longer needed or register new information into maintenance memory for later use.

The limit for representational complexity. According to our model, "too complex to use" is when the memory demand for processing the displayed information exceeds the capacity of working memory. We generalized the

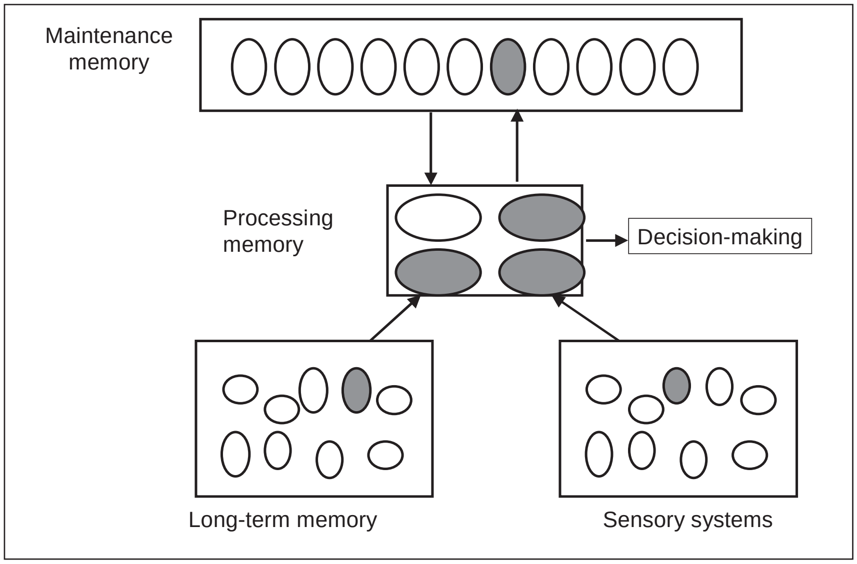

Figure 7: The diagram of a working-memory model for cognitive information processing.

results of working capacity limits from the literature and used them as the upper-limits of cognitive complexity for displays. In the model, the items in maintenance memory do not need to be retrieved simultaneously but should be maintained over a period during which the user is attending to other information. Thus, representational complexity, defined as the number of functional units, is in accordance with the operation span of working memory. Conway and Engle (1996) proposed that the operation span is about the limitation in the ability to use controlled processing to maintain information in an active, quickly retrievable state. They found that the number of items one can reliably recall is typically in a range of 5~16, with a mean value of 9~11. Other studies hinted at similar limitations. For example, Willems, Allen, and Stein (1999) reported that a controller could correctly recall up to 11 aircraft that he or she operated.

Variety evaluated by cognition. Once information is organized into independent units or chunks and held in maintenance memory, the variety of information chunks does not affect cognitive task performance. Task performance is not affected because the chunks are meant to be different. However, the variety factor in the temporal domain (i.e., dynamic changes of information) affects information flow between processing and maintenance memory. Therefore, we proposed that the metric was the amount or frequency of unpredictable information onset that demands a change in the contents of maintenance memory. We refer to this metric as dynamic complexity. Information changes in a display impose cognitive loads in several ways: 1) increasing maintenance memory load; sudden onset of visual targets or even changes in luminance of visual patterns (Schmidt, Vogel, Woodman, & Luck, 2002); 2) increasing information flow between maintenance and processing memory; and 3) reducing the stability of mental representation. To build a mental representation takes time. As a result, if too many entities in maintenance memory are updated at a high rate, the mental representation tends to deteriorate or even collapse. This corresponds to users' "losing the picture" (Hopkin, 1995).

The limit of dynamic complexity. While many studies have demonstrated that information can be lost if it is rapidly presented or immediately disturbed by the onset of new stimuli, there are no conclusive results in the current literature about the temporal capacity of working memory.

Relation evaluated by cognition. According to the memory model, the relation between elements of information is handled by processing memory that binds pieces of information simultaneously needed to plan an action or make a decision. Thus, this metric measures the demand for processing memory. We used the definition of relational complexity proposed by Halford et al. as the metric to describe how the relation factor of complexity affects cognition. The metric is defined as the number of independent elements or dimensions of information that must be simultaneously bonded to use information. Many cognitive processes, such as selection, manipulation of goal hierarchies, reasoning, and planning, are examples of processing at high levels of relational complexity. Halford argued that the more interacting variables that have to be processed in parallel, the higher both the cognitive demand and computational cost will be. Since processing memory links pieces of information that are simultaneously needed for task performance, relational complexity is a straightforward measure of the processing memory load of a task.

The limit of relational complexity. Halford et al. further demonstrated that normal adults could reliably integrate up to four relations in parallel while children can only integrate one or two relations. This quaternary limitation appears to be consistent with other studies that demonstrated the capacity limit of processing memory as 3~5 items (Cowan, 2001). Cowan reviewed the literature on processing memory and concluded that the human capacity limit is three to four items on average. Many visual experiments have revealed that the capacity of visual processing memory is about four items. As another example, in a single sentence the maximal number of concepts the sentence can contain and still be reliably interpreted is also four.

Action Complexity

We derived the following basic action-related tasks for using ATC displays from ATC task analyses found in the literature. The generic tasks include: 1) responding to onset of alert/warning messages and other requests from a display; 2) locating and acquiring specific information; and 3) entering data into the display system. The action functions involved in these tasks include planning eye/head/hand/body movements, sequential movements, oral communication, etc.

Figure 8 is a simplified diagram of action information processing. According to the literature, the parietal and frontal cortices are involved in planning actions, and the motor cortices are responsible for executing planned actions (Andersen, Snyder, Bradley, & Xing, 1997). A common feature of these cortices is population coding. That is, all the neurons in a cortical area work together to encode a single action plan, and only after the plan is executed do they begin to encode the next (Georgopoulos, Schwartz, & Kettner, 1986; Pouget, Zemel, & Dayan, 2000; Xing & Andersen, 2000). Therefore, the brain can only execute one action plan at a time. Using a display to perform a specific task may require multiple

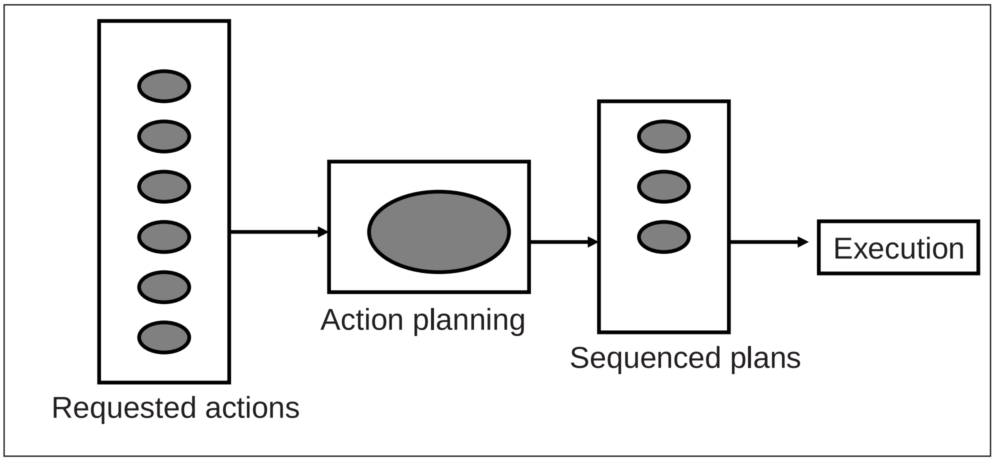

Figure 8: Diagram of action information processing.

actions. Figure 8 provides a framework for multi-step action planning. The framework is based on the fact that the brain is able to hold several action plans and execute them sequentially. When performing multiple tasks with a display, users actually switch action planning back and forth. The switch can be so quick that they feel that they are doing multiple things at the same time. However, information can be misinterpreted or lost during these quick switches. Next, we use this framework to evaluate the complexity factors and derive the metrics of action complexity.

Quantity evaluated by action. The quantity factor affects the first part of the action model, the requested actions. Performing physical interactions with a display costs time and takes users away from other perceptual and cognitive tasks. Therefore, we proposed that the metric was the minimal amount of keystrokes, mouse movements, and transitions of action modes required to use displayed information. Compared with keystrokes and mouse movements, the time needed for eye and head movements is negligible. Therefore, we only considered keystrokes and mouse movements. An action transition is a change of action modes, such as from keystrokes to mouse movements or vice versa. Those transitions also take time and require the brain to coordinate different action modes. Sears (1994) proposed a layout complexity metric as the summed product of the frequency of action transitions and the cost of transitions. The two factors in our metrics, the amount of manual movements and the transitions between the movements, are essentially the same as Sears' metric.

Variety evaluated by action. The variety factor affects storage of action plans. Therefore, we propose that the metric is action depth, defined as the number of serial steps needed to plan (or select from a number of action options) to acquire information. Following the need to increase the variety of actions, today's display systems tend to use multi-level structures to cope with more diverse environmental perturbations and reduce the difficulty of decision-making. In such systems, a task of any complexity can be decomposed into a series of subtasks, each represented by a subgoal. Some researchers have used the number of serializable subgoals as a measure of complexity for a system with a multi-level structure (Heylighen, 1989).

Relation evaluated by action. The relation factor affects action planning. The brain cortices related to motor planning can only program one action plan at a time. Therefore, we propose that this metric should be the number of simultaneous action goals required to use displayed information. Since the brain can only reliably program one action plan at a time, ideally each action should result in only one action goal for the next step. In the case where more than one action goal needs to be planned simultaneously, the brain has to switch back and forth between the goals. Errors may occur when switches of planning occur at a fast pace.

Table 2 summarizes the 3x3 metrics we developed for generic ATC displays. The metrics describe the objective aspects of complexity. So far, we have developed the metric definitions, yet the algorithms or methods of computing each metric remain to be developed or implemented from the literature. Next we describe a preliminary case study to explore the methods of applying our metrics to computing cognitive complexity of a human-computer interface.

A PRELIMINARY CASE STUDY: ASSESSING THE COGNITIVE COMPLEXITY OF THE MICROSOFT POWERPOINT INTERFACE

The purpose of this case study was to explore how to apply the proposed metrics to evaluate display complexity. Microsoft PowerPoint™ is one of the most popular software applications for making presentations. Using our complexity metrics, we assessed the cognitive complexity of the PowerPoint interface.

Table 2: Metrics of information complexity

	Perception	Cognition	Action
Quantity	No. of fixation groups	No. of functional units	Amount of action cost
Variety	No. of visual features	Dynamic complexity	Action depth
Relation	Degree of clutter	Relational complexity	No. of action goals

Methods

We performed the study in the following steps:

1) Assessing representational complexity: 24 PowerPoint users were asked to visualize their PowerPoint interface and list the functions that are most essential and useful (or frequently used). The average number of listed functions was used as the index of representational complexity.

2) Assessing processing complexity: We calculated the elements of information needed to access each basic function and took the average number of the elements as the index of processing complexity.

3) Assessing dynamic complexity: We counted the number of unexpected changes in the interface and used it as the index of dynamic complexity.

RESULTS

Representational Complexity

Figure 9 shows the indices of estimated representational complexity. The horizontal axis represents the number of frequently used functions the users listed. The vertical axis represents the number of users responding. All but one user listed 5~13 basic functions. The mean number of listed basic functions was 8.1.

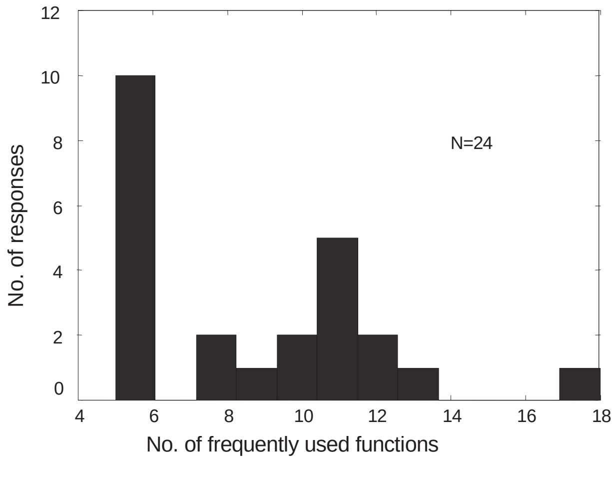

Figure 9: Estimated representational complexity.

Processing Complexity

Table 3 indicates how we estimated processing complexity. The table presents the top 11 most frequently listed functions, the elements of information required to access the function, and the indices of processing complexity counted as the number of the elements. The "*" associated with an index indicates the situation where a function can be accessed through a drop-down menu or directly through an icon on the top layer of the interface, depending on individual customization. The results show that the indices of processing complexity for these individual functions are either zero or one. We used the average of the indices as a rough estimation of the processing complexity of the interface. The average index is 0.66.

In addition, dynamic complexity for the PowerPoint interface is zero because the interface does not have any unpredicted changes of information or functions.

Table 4 presents the summarized results of the complexity evaluation of the PowerPoint interface. The upper row displays the three metrics; the middle row indicates the estimated indices of the metrics; and the bottom row compares the indices to the corresponding capacity limits of working memory. The results indicate that the estimated indices of complexity for the PowerPoint interface are well below the capacity limits.

DISCUSSION

This paper presents a framework to decompose information complexity and 3x3 complexity metrics for ATC displays. Previous work has reported measures similar to the individual metrics we proposed (see Xing & Manning for a review). Value can be gained from comparing our metrics with similar ones found in the literature and by integrating existing complexity measures into our framework.

Perceptual Complexity in the Literature

Many algorithms have been developed to address image or pattern complexity, but most are based on information theories and have low correlations with human judgment. In contrast, Tullis (1985) developed a set of metrics to

Table 3: Estimated processing complexity for PowerPoint interface

Basic function	Elements needed to access the function	Processing complexity of the function
Copy/paste	Edit	1
Insert picture	Insert	1
Slide layout	Format	1 or 0*
Save file	File	1
Text box	Insert	1
Format text	Format	1 or 0*
Open PowerPoint	File	1
Slide show		0
Set font characteristics	Format	1 or 0*
Insert new slide	Insert	1
Draw shape / line	Draw (Autoshapes)	1 or 0*

Table 4: Summary of the complexity evaluation of PowerPoint

Metric	Representational complexity	Processing complexity	Dynamic complexity
Index	8.1	0-1	0
Relative to capacity limits	8.1/(9~11)	(0~1)/(3~4)	0

measure display complexity from the perspective of human performance on visual search tasks. The metrics were comprised of four basic characteristics of display formats to describe how well users can extract information from displays. They included a) overall density of displayed items, b) local density of characters, c) number and average group size, and d) layout complexity, which describes how well the arrangement of items on a display follows a predictable visual scheme. Tullis showed that these metrics are highly correlated with subjects' performance time in visual search tasks. The overall and local density characteristics, together, can be a measure of our clutter metric; the number of groups corresponds to our metric of fixation groups; and the layout complexity is somewhat related to the variety of visual features. Therefore, for limited applications, we can use Tullis' metrics as the quantitative measurements for the perceptual complexity we proposed.

One drawback of Tullis' metrics is that those measurements were specified for text-dominant displays but not for graphical ones. It is hard to define Tullis' groups in spatially continuous two-dimensional images with varying colors and luminance contrast. On the other hand, Rosenholtz, Li, Mansfield, and Jin (2005) proposed a *feature congestion measure* of display clutter in which clutter is considered as the degradation of task performance caused by the density of visual features such as color or luminance contrast. This measure is related to our clutter metric and can be applied to graphic visual displays. Unfortunately, the algorithm requires displayed materials to be converted to digitized images to compute the feature congestion. That is often not practical for dynamic displays in which visual images evolve rapidly.

Comparisons With Other Measures of Cognitive Complexity in the Literature

Previous studies of cognitive complexity have focused on text comprehension, creativity, social phenomena, etc. For example, Crokett (1965) used the concept of "level of hierarchic integration of constructs" to define cognitive complexity. 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). We identified three metrics of cognitive complexity based on working memory: number of functional units, number or frequency of unpredictable changes, and number of relations. Our metric, the number of functional units, is similar to the measure of constructs in the literature. We also proposed relational complexity as a metric to quantify how the relation factor of complexity affects cognition. This measure corresponds to the interconnected hierarchical integration proposed by Crokett. In addition, we proposed to use the frequency of unpredictable information changes to measure the variety factor of complexity. However, this dynamic aspect of cognitive complexity has been seldom studied.

Measures of cognitive complexity, explicitly or implicitly, depend on cognitive task analyses that reveal the cognitive aspects of tasks and knowledge needed for situation awareness, decision-making, planning, etc. One popular cognitive task analysis method is GOMS: Goal, Operator, Methods, and Selection (Card, Moran, & Newell, 1983). This method seeks to analyze and model the knowledge and skills a user must develop to perform tasks on a device or system. The result is a description of the Goals, Operators, Methods, and Selection rules for any task. Currently we are exploring how to calculate the three cognitive metrics based on GOMS and other similar analyses.

Relevant Measures of Action Complexity in the Literature

Many methods have been developed to assess the complexity of human-computer interfaces (McCabe, 1976; Rauterberg, 1992). Those methods require modeling a system's states and transitions between states. Unfortunately, it is implausible to directly apply such methods to ATC displays. Controllers use displays adaptively, and no standardized procedure has been specified. Thus, there are no clearly defined states and transitions in their interactions with ATC displays. If the use of a display can be described explicitly with states and transitions, it implies that controllers are forced to manipulate the display following a fixed procedure. That would be contrary to the design philosophy.

One measure related to action complexity is Sears' *layout appropriateness* metric (Sears, 1994). Sears proposed this metric to assess users' performance when using a computer interface. The metric was the summed product of the frequency of action transitions and the cost of these transitions. The two factors comprising our metric of action cost, number of manual movements, and transitions between the movements are essentially the same as Sears' metric. Sears used the distance that users must move the computer mouse and the size of the objects to be moved as the cost of a transition. This metric can be used to evaluate the efficiency of a user-interface layout and compute the extent to which a display demands action. However, it does not apply to ATC displays well because it primarily emphasizes the effects of mouse movements, while many other kinds of actions are involved in using ATC displays (Allendoerfer, Zingale, Pai, & Willems, 2006) .

Perhaps the *Keystroke Level Model* (KLM) proposed by Card, Noran, and Newell (1980) is more applicable to measure action complexity within the ATC domain. The model measures the sum of the execution time of sequenced operations, including key strokes, mouse movement, switches, and mental preparation for a physical action. While most of these operations correspond to our metric of action cost, the last operation — mental preparation — is somewhat related to the other two action metrics: action depth and simultaneous action goals. Compared to Sears's metric, the KLM model considered the effects of key strokes and mental preparation. Both play crucial roles in controllers' interacting with ATC displays. Allendoerfer et al. documented the frequency of use of en route controller commands using radar displays and measured the time and number of keystrokes or mouse clicks required to perform those commands. They found that controllers entered information on the radar display more frequently than they moved things around. They also proposed that the time spent looking at the keyboard or screen while entering commands should be included in the assessment of display usage characteristics. These results suggested that the KLM model is a potential candidate to objectively measure action complexity for ATC displays. Next, we need to explore how to calculate action complexity based on the measurements like those in the KLM model.

CONCLUSIONS

This paper presents a framework for decomposing factors of information complexity and a set of metrics to measure ATC display complexity. The framework is described as follows: 1) information complexity is the combination of three basic factors: quantity, variety, and relation; 2) complexity factors need to be evaluated with the mechanisms of brain information processing at three stages of information processing: perception, cognition, and action; and 3) the metrics of complexity can be derived by associating task requirements to brain functions. The framework incorporates many human factors studies involving interface evaluation. Within this framework, we identified a set of complexity metrics for ATC displays. Future work will focus on testing the metrics in a real or simulated ATC work environment and converting the metrics into easy-to-use products for the design and human factors evaluation of new ATC technologies.

REFERENCES

Allendoerfer KP, Zingale C, Pai S, Willems B (2006). En route air traffic controller commands: Frequency of use during routine operations. Federal Aviation Administration, William J. Hughes Technical Center; Technical Report No: DOT/FAA/TC-TN06/04.

Andersen RA, Snyder LH, Bradley DC, Xing J (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*; 20: 303-30.

Bennett CH (1990). How to define complexity in physics, and why. In: Zurek WH (Ed.), Complexity, entropy and the physics of information. Redwood City, California: Addison-Wesley; 137-48.

Burleson BR, Caplan SE (1998). Cognitive complexity. In: McCroskey JC, Daly JA, Martin MM. & Beatty MJ (Eds.), Communication and personality: Trait perspectives. Cresskill, NJ: Hampton Press.

Card SK, Moran TP, Newell A (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM;* 23(7): 396-410.

Card SK, Moran TP, Newell, AL (1983). The psychology of human computer interaction. Hillsdale, NJ: Erlbaum.

Conway AR, Engle RW (1996). Individual differences in working memory capacity: more evidence for a general capacity theory. *Memory*; 4(6): 577-90.

Cowan N (2001). The Magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*; 24(1): 87-114.

Crockett WH (1965). Cognitve complexity and impression formation. In: Maher BA (Ed.), *Progress in experimental personality research*. New York: Academic Press; 2: 47-90.

Crutchfield JP, Young K (1989). Inferring statistical complexity, *Physics Review Letters*; 63: 105-8.

Drozdz S, Kwapien J, Speth J, Wojcik M (2002). Identifying complexity by means of matrices. *Physica*; A314: 355-61.

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

Georgopoulos AP, Schwartz AB, Kettner RE (1986). Neuronal population coding of movement direction. *Science*; 233: 1416-9.

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

Halford GS, Wilson WH, Phillips W (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. *Behavioral Brain Sciences*; 21(6): 803-31.

Heylighen F (1989). Self-organization, emergence and the architecture of complexity. *Proceedings of the 1st European Conference on System Science; Paris: AFCET;* 1989: 23-32.

Hopkin VD (1995). Human factors in air traffic control. Bristol, PA: Taylor & Francis.

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

McCabe (1976). A complexity measure. *IEEE Transactions on Software Engineering*; SE-2: 308-20.

Pouget A, Zemel RS, Dayan P (2000). Information processing with population codes. *Nature Review of Neuroscience*; 1: 125-32.

Rauterberg M (1992). A method of a quantitative measurement of cognitive complexity. *Human-Computer Interaction: Tasks and Organization - ECCE'92*; Roma, Italy. 295-307.

Rosenholtz R, Li Y, Mansfield J, Jin Z (2005). Feature congestion: A measure of display clutter. *Proceedings of the SIGCHI 2005;* 761-70.

Schmidt BK, Vogel EK, Woodman GF, Luck SJ (2002). Voluntary and automatic attentional control of visual working memory. *Perception and Psychophysics;* 64: 754-63.

Sears AL (1994). Automated metrics for user interface design and evaluation. *International Journal of Biomedical Computation;* 34: 149-57.

Tullis TS (1985). A computer-based tool for evaluating alphanumeric displays. In Shackel B (Ed.), Human-computer interaction: INTERACT ‘84. London, England.

Wickens CD (1991). Engineering psychology and human performance. NY: Harper Collins.

Willems B, Allen RC, Stein ES (1999). Air traffic control specialist visual scanning II: Task load, visual noise, and intrusions into controlled airspace. Federal Aviation Administration, William J. Hughes Technical Center; Technical Report No: DOT/FAA/CT-TN99/23.

Xing J, Andersen RA (2000). Memory activity of LIP neurons for sequential eye movements simulated with neural networks. *Journal of Neurophysiology;* 84(2): 651-65.

Xing J, Heeger D (2001). Quantification of contrast-dependent center-surround interaction. *Vision Research;* 41: 571-83.

Xing J, Manning C (2005). Complexity and automation displays of air traffic control: Literature review and analysis. Washington DC: Federal Aviation Administration; Report No: DOT/FAA/AM-05/4.

Xing J (2004). Measures of information complexity and the implications for automation design. Washington DC: Federal Aviation Administration; Report No: DOT/FAA/AM-04/17.