DOT/FAA/AM-04/17

Office of Aerospace Medicine
Washington, DC 20591

Measures of Information Complexity and the Implications for Automation Design

Jing Xing

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

October 2004

Final Report

U.S. Department
of Transportation

Federal Aviation
Administration

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-04/17	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Measures of Information Complexity and the Implications for Automation Design		5. Report Date October 2004 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 ARR-522.

16. Abstract

Information complexity associated with automation aids is a bottleneck that limits their use. While automation systems are designed to bring new functions to users and increase their capacities, automation also creates new tasks associated with acquiring and integrating information from displays. For example, a complex display increases information load to human operators and reduces usability. Thus, the efficiency of an automation system largely depends on the complexity of displayed information. To evaluate the costs and benefits of an automation aid, it is important to understand how much information is shown on the display, how users look at multiple information sources to build and maintain situation awareness, and whether the information is displayed in a compatible way so it can be integrated and understood easily without the user having to make internal conversions or calculations. In this paper, we present a set of measures to assess information complexity. The metrics count information complexity as the combination of three basic factors: numeric size, variety, and relation; each factor is evaluated by the functions at three stages of brain information processing: perception, cognition, and action. Ideally, these measures provide an objective method to evaluate automation systems for acquisition and design prototypes.

17. Key Words Information Complexity, Visual Display, Interface Design, Evaluation, Prototype	18. Distribution Statement Document is available to the public through the National Technical Information Service Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	15	

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

ACKNOWLEDGMENTS

The author thanks Dr. Carol Manning for the valuable discussions throughout this study. I also thank Drs. Scott Shappell, David Schroeder, Lawrence Bailey, and Earl Stein for their invaluable suggestions and advice about this study. I also appreciate the effort of all the FAA internal reviewers whose comments greatly improved the quality of this paper.

Measures of Information Complexity and the Implications for Automation Design

INTRODUCTION

Computer-based automation tools are everywhere today, often interacting with human operators through visual displays. From Microsoft Windows™ on your computer to more sophisticated flight management systems in today's aircraft, computer-based automation displays literally exploded on the scene. Modern air traffic control (ATC) displays are no different. Indeed, many automation tools are developed to help the controller identify potential conflicts between aircraft in an increasingly complex airspace. In addition to simple navigation, these tools provide decision-support information for air traffic control; thus, they serve a dual purpose as both automation aids and decision-support tools.

While these tools are intended to ensure safety, increase capacity, and offload tasks from controllers, they also create new tasks associated with interface management. Moreover, the use of new tools requires that controllers integrate information from displays into their own methods of managing their cognitive resources (Bressolle, Benhacene, Boudes, & Parise, 2000). Therefore, introduction of new systems can introduce additional complexity to task management. In worst cases, information provided by the tools may be too complex and overwhelm controllers' cognitive capacities. Consequently, key information could be either missed or misinterpreted by controllers and thereby increases the risk of performance errors. For these reasons, it is desirable to have an objective method to assess information complexity of automation displays and to assess the impact of complexity on operators' task performance.

Numerous studies have examined the perceptual and cognitive compatibility between the human and interface format. For instance, Wickens and Andre (1990) have demonstrated that the efficiency of an automation tool depends largely on whether the information is displayed properly. Likewise, it appears that the proximity of a display should be compatible with the proximity of the tasks (Carswell & Whickens, 1987). In addition, Carswell (1992) also pointed out that performance is attenuated with graphical display as a function of visual dimensions used to code data values. Similarly, Tullis (1985, 1986) used visual characteristics of display formats to quantify how well users can extract information from displays. Sears (1994) took a different approach by developing metrics to predict both user preference and performance in using a computer interface. Such research provides insight into the complex world of display design and evaluation.

Unfortunately, most previous human factors studies have focused on how information should be presented, not necessarily information complexity (IC), although the latter has been theoretically explored. Information theories consider a system as an automaton consisting of a series of elementary units distributed in space. From the viewpoint of information theories, the most straightforward definition of IC is the minimum description size of a system (Grassberger, 1986, 1991; Crutchfield & Young, 1989). That is to say, if the description of a system can be greatly compressed without loss of meaning, then it is considered simpler than one that cannot. However, this definition is only concerned with the storage demands of a system. In contrast, Bennett (1990) introduced the concept of logical depth as a measure of complexity. Logical depth combines resource demands and computational power into a single description of the computational resource required to calculate the results of a program of minimal length. This definition is a combination of both resource demands and computational power. Scott (1969), on the other hand, proposed a measure of information redundancy to describe complexity. Similarly, Langton (1991) suggested that complexity is associated with high levels of mutual information, which is the correlation between information at separated sites. In general, these studies focus on the difficulty of compressing a representation, with little direct connection to the practical aspects of a functioning organism. In addition, information theories define information in relation to the probabilities of all other inputs that might have been encountered. However, it is difficult to specify probabilities when applying theories like these to such realistic circumstances as ATC.

The objective of this report was to develop observable metrics of IC for automation displays. This objective raises three basic questions: What is complexity? Why can information be too complex for the human brain? Finally, how do we quantify the complexity of visual displays? This paper is organized into three sections to address the above questions. We will first summarize our understanding of information complexity from an analysis of the literature. Then we will demonstrate that information is processed at three distinctive stages in the human brain, and complexity should be evaluated by the functions at each stage. In the last section, we will describe a set of IC metrics that we propose to use for automation displays in ATC.

RESULTS

Understanding information complexity

Although there are many definitions of complexity in the literature, the term has proven to be very difficult to define accurately. 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 studies: 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: numeric size of basic elements in a system, variety of elements, and relation 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.

Intuitively, the numeric size of basic elements is related to the complexity of a system. Whether referring to minimum description size of a system (Grassberger, 1986, 1991; Crutchfield & Young, 1989), number of states (McCabe, 1976), number of "chunks" in cognition (Cant, Jeffery, & Henderson-Sellers, l995; Klemola, 2000), all studies of size seem to agree that a larger size corresponds to a higher degree of complexity. Nevertheless, numeric size in itself is not sufficient to define complexity in its entirety.

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 et al. and other researchers have pointed out, complexity lies somewhere between order and disorder (Drozdz, Kwapien, Speth, & Wojcik, 2002). Perhaps 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 among the basic elements (such as rules of structures, interconnections, etc.) of a system contribute to complexity. Individual parts of a system are held together through 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 counts on the relations between the elements, not just the number and the variety of the elements.

In addition to generalizing from the literature that complexity is the combination of three basic factors, we also identified two principles that contribute to the great diversity of complexity measures. One 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 experiments. The variety of the images, measured as the disorder of image pixels, increased from left to right, yet subjects perceived the middle one as the most complex. One explanation for this result is that the human visual system processes features such as orientation and spatial frequency rather than individual pixels per se. This experiment indicates that the impact of the variety factor on complexity depends on how observers process information.

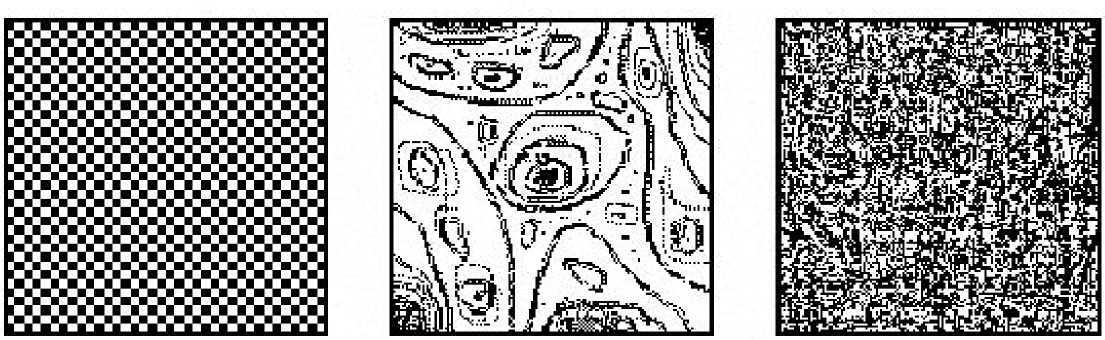

Figure 1. Variety increases from left to right, yet the human perceives the middle one the most complex (Grassberger 1991)

Another principle is task dependency — 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 the variation 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 of complexity: Complexity is the combination of three basic factors: numeric size, variety, and relation; all three of which must be considered within the constraints of task requirements.

In the following two sections, we will describe how the human brain evaluates the factors and how we determined complexity metrics of automation displays within the constraints of ATC tasks.

Human visual information processing

Given that complexity depends on how observers process information, we looked into the mechanisms of information processing in the human brain. Figure 2 outlines a conceptual diagram of human visual information processing associated with the use of visual displays. In this simple representation, information presented via visual display devices is processed by three stages in the brain: perception, cognition, and action. Through perception, an observer acquires information about the current status of the world. Visual attention to specific regions of interest is needed to modulate perception; otherwise, we would have no way of filtering out unwanted information. The perceived information then feeds into the cognition stage, where one's perceptions are integrated with information from the observer's experience and memory. An internal (mental) representation of what was observed can then be generated. Based on this representation and personal strategies, the observer can then make decisions and convert them into actions. The actions allow interaction between the observer and the system.

It is important to emphasize that the three stages of information processing correspond to different functional areas of the brain. Neurophysiological studies have mapped specific areas of the brain. Each area responds to specific functions of perception, cognition, or action. For instance, neurons in the primary visual cortex appear to respond to orientation, color, and other basic visual features of a stimulus, whereas neurons in the temporal cortex respond to face recognition and neurons in the frontal cortex respond to planning of eye movements. Recently, using brain-imaging techniques, scientists have been able to associate brain activities to specific cognitive functions. While a human subject performs some cognitive tasks, a functional MRI (Magnetic Resonance Imaging) instrument can register the amount of neuronal activity to functioning areas of the brain over time. For example, Carpenter, Just, and Reichle (1999) have found that the activities of the prefrontal cortex increase

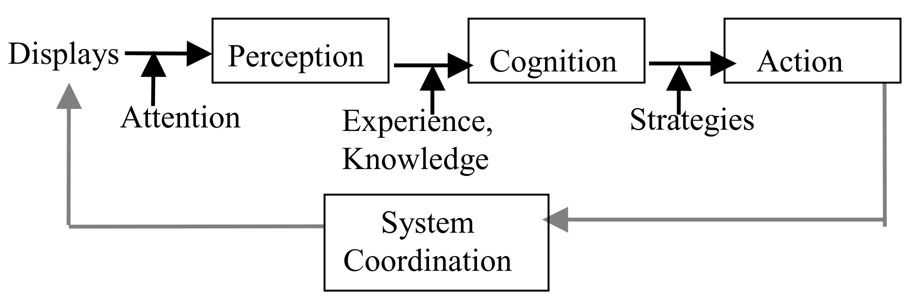

Figure 2. A diagram of information processing in the human brain

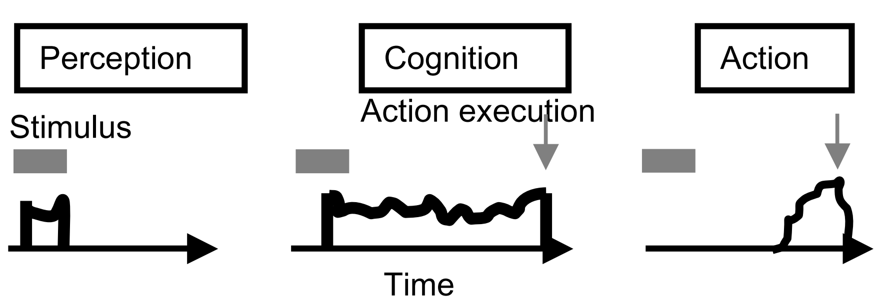

Figure 3. The three stages, perception, cognition, and action, are distinctive in mechanisms of information processing

with working memory load while the activities of the posterior cortex are highly correlated to mental rotation of a visual pattern.

Moreover, the three stages result in distinct information processing mechanisms. One way of visualizing this is by plotting brain activity over time, as illustrated in Figure 3. The vertical axis in the figure represents neuronal activity, and the horizontal axis represents time. The three panels (from left to right) represent perception, cognition, and action. The presence of a visual stimulus is indicated with a horizontal bar, and the time when an action is made is indicated with a vertical arrow. At the perception stage, neurons in the perceptual areas of the brain respond to a stimulus, and the activity goes off with the offset of the stimulus. However, neuronal activity for cognitive tasks is associated with working memory – after the stimulus goes off, neuronal responses to a stimulus continue until an action or action plan is made. Such memory activity can last for several seconds. At the action stage, neurons in the pre-motor and motor areas of the brain become active right before an action signal is issued and the activity dies away after the action is taken.

Perhaps most important to this discussion is that the three stages are distinctive in function. Many neurophysiological studies have quantified the relationship between localized brain activities and behavioral functions of task performance. Listed below are some behavioral functions at each stage that are related to the use of visual displays.

Perception. The human visual cortex is specialized to perform many kinds of perceptual functions including target searching, text reading, color discrimination, texture segmentation, motion detection, and many others. Perception processes information serially and in parallel. Thousands of visual neurons can be activated simultaneously by a visual image. Thus, they extract information rapidly in parallel. Based on the result of such parallel processing, the visual system then serially focuses the fovea on salient spots so that information can be analyzed in detail.

Cognition. The high-level modules of brain cortical areas, called associational cortex, integrate inputs from the perceptual cortex with information stored in the brain's long-term memory. The associational cortex performs cognitive functions such as working memory, text comprehension, planning, selecting, etc. A common feature of cognitive functions is their limited capacity. That is, only a few pieces of information can be processed simultaneously in the associational cortex. Consequently, the bandwidth of information processing in the cognition stage is much less than that in the perceptual stage.

Action. The premotor and motor cortex of the brain are responsible for encoding various manual actions such as eye, head, hand, and arm movements. Those brain areas are also able to encode sequential movements. The motor cortex, unlike other cognitive and perceptual areas of the brain, is believed to work in a serial manner, i.e., all the neurons in the motor cortical area work together to encode a single movement, and only after the movement command is executed do they begin to encode the next movement. Consequently, with such a narrow bandwidth of information processing, an effective automation tool should impose only very limited action requirements for human operators.

Given the differences inherent among the three information-processing stages, the three complexity factors should be evaluated separately at each stage. This results in a 3x3 matrix as shown in Table 1, with rows being the three complexity factors and columns being the three information-processing stages. Each box in the matrix corresponds to one IC metric. For each box, the complexity factor should be evaluated by the functions that occur at that stage, and each complexity metric should be associated with the operator's task performance. For

Table 1. Metrics of information complexity for automation displays in air traffic control

	Perception	Cognition	Action
Numeric Size	No. of fixation groups	No. of functional units (tasks)	Amount of keystrokes and mouse movements
Variety	Variety of groups	Dynamic change in the units	No. of transitions
Relation	Degree of clutter (Text readability)	No. of variables in a unit	Action depth (steps) of a functional unit

example, one of the metrics that we proposed (details in the next section) is the clutter effect, which means that the perception of a central target is masked by the presence of its immediate surrounding stimuli. Clutter directly affects the speed and accuracy of text reading; therefore, it affects task performance. With the known capacity limits of brain functions, the metrics can elucidate why a visual display can be too complex for human operators. Table 1 lists the metrics we have proposed to assess complexity associated with automation displays in ATC. Each metric will be explained in the next section.

Metrics of information complexity for automation displays in air traffic control

Steps for developing complexity metrics

If we evaluate every complexity factor against every function of the three stages of brain information processing, the resulting metrics would have too many dimensions. Fortunately, complexity metrics are constrained by task requirements. Only the functions critical to the given tasks are relevant. The following steps were used to develop the metrics:

1. Identify task requirements;
2. Determine corresponding brain functions pertinent to the task requirements;
3. Choose the metric that can reflect the impact of the complexity factor on the brain functions.

Task requirements of using automation displays in air traffic control

ATC systems have unique features that differentiate them from other applications. Below are some typical characteristics of ATC automation displays:

1. Displays contain mainly text, icons, and other binary graphical patterns (symbol, charts, etc.). Spatially continuous digital images are very rare;
2. Controllers look for particular information on displays to assist in decision-making;
3. Displays are dynamic: Information is continuously updated with the evolution of the traffic situation;
4. Unlike most human-computer-interaction systems, many ATC automation tools are presented as aids, not the objects that controllers have to operate on. Controllers use aids only when they are helpful (i.e., the benefit is greater than the cost) and ignore them when they are not.

Given these characteristics, we derived some basic requirements for using ATC automation. ATC displays must allow: 1) searching for information in a timely manner; 2) reading text reliably; 3) facilitating rather than disturbing decision-making; and 4) minimizing time-costing actions. The complexity metrics described below were developed to measure these requirements. For each metric, we will first introduce its definition and how it relates to ATC task requirements. We will then describe its impact on ATC performance and the capacity limit to address the question of "why information can be too complex for users."

Metric-1: Perceptibility

Size factor evaluated by perception

The proposed metric is the number of fixation groups. The basic element for searching and reading tasks is eye fixation. A fixation group is defined as a set of visual stimuli that can be grabbed with one eye fixation. Typically, a foveal fixation spans a view angle of about 2-4 degrees. The average time to search for a particular target on a visual display increases with the number of fixation groups. While there is no physiological limit on how many fixations one can make on a display, visual experiments have demonstrated that it takes 600-700ms for an observer to perceive the information in one fixation (Joseph, Chun, & Nakayama, 1997). Therefore, the capacity limit of this metric is determined by the time that a user has available to spend on an automation aid. For example, if a controller has 5s maximally to acquire the information from an automation aid, then the number of fixation groups included in the display should be less than 14 (5000/700).

In many applications, displays are very busy and it takes many fixations to view all the information. One strategy to reduce perceptual complexity is the use of color-coding, because information can be segregated into several categories with color-coding. Consequently, visual searching can be limited to the visual targets illustrated with a particular color. By doing so, the number of fixations can be greatly reduced.

Variety factor evaluated by perception

This proposed metric is the variety of fixation groups. Variety is defined as the differences in visual features such as size, texture, luminance, contrast, and colors of the groups. Increasing the variety of visual features increases complexity. Visual studies have found that switching between visual features such as color and luminance contrast increases searching time. This effect is called "cost of switching." In addition, switches may also reduce the reliability of reading text and increase visual fatigue. Consider, for example, two figures (A and B) that contain the same text. The text in Figure A has the same format while the text in Figure B is manipulated in font, letter size, and luminance contrast to increase visual variety. As a result, Figure B will appear to be more complex than Figure A due to its increased variety.

Relation factor evaluated by perception

The proposed metric is the degree of clutter. Clutter is the effect of masking the visual perception of a stimulus with the presence of other stimuli. Consequently, clutter can increase search time and reduce text readability. The effect is apparent when background visual stimuli are spatially superimposed on the text. Moreover, the perceived contrast of a visual target can also be largely suppressed by the presence of neighboring stimuli. The reduced luminance contrast results in deterioration of text readability and a corresponding increase in search time. Xing and Heeger (2001) examined this effect in a series of experiments. They 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 suggest two methods that reduce the clutter effect: 1) reducing the amount of text in a display and, 2) reducing the continuity of graphics so that targets do not have immediate surrounds.

Metrics-2: Cognitive capacity

Air traffic control is cognitively demanding. Basic ATC tasks include monitoring, controlling, checking, diagnosing, and decision-making. Many cognitive models of ATC have been proposed (Kallus, Barbarino, & Van Damme, 1997). The kernel of those models contains two components: mental representation (or "mental model," as it is called by some in the literature) and memory. Cognitive processing is based on a mental representation of the task environment. Mental representations of a given situation are built by organizing information into many independent entities that are kept on-line for awareness. On the other hand, working memory enables us to hold in our mind's eye the content of our conscious awareness, even in the absence of sensory inputs. In a sense then, working memory manipulates entities in one's mental representation. It links pieces of information that are simultaneously required for a particular task. Therefore, measures of cognitive complexity should quantify how much a task imposes demands on both mental representations and working memory.

Size factor evaluated by cognition

Given that a mental representation is the platform for cognitive processes, the size factor corresponds to the number of basic, independent elements in a given mental representation. The challenge is to define these entities with respect to the use of automation displays. These elements represent the essential characteristics of information provided by a display. A common strategy used to support cognitive processing is categorizing pieces of information, where categories represent independent dimensions that an operator comprehends. In this way, categories correspond to the entities of a mental representation. It makes intuitive sense, then, that complexity would be greater when an operator views a display as having many categories and must make fine distinctions among those categories.

While categorization can be based on perceptual features, a number of studies have demonstrated that the categorization process in ATC task performance is mostly goal-oriented. "Goal-oriented" refers to any feature that is an important objective of the task. Therefore, the basic elements of a mental representation can be specified as the fundamental functional units of a display. Each of the units represents a distinctive objective of the tasks. The units are independent of each other and cannot be combined to a chunk. Hence, we defined the number of functional units as the metric of size factor evaluated by cognition. A display may have many functional units; each unit achieves specific functional goals. To use the display fully, a user stores the functional units in the mental representation of the situation. Complexity therefore would logically increase with the number of units in a given display. As the number gets larger, the memory load could impair task performance; the user may either misinterpret the information or choose to ignore it. Conway and Engle (1996) reported that normal adults could actively maintain 9-16 independent items in their memory during the operation of a task. This limit is potentially related to the capacity of a mental representation.

Variety evaluated by cognition

The proposed metric can be specified as dynamic complexity, measured as the rate of information change over time. Information changes in a display impose cognitive loads in two ways: 1) increasing working memory load. Psychophysical experiments have demonstrated that a sudden onset of visual targets or even changes in luminance of visual patterns automatically takes working memory (Schmidt, Vogel, Woodman, & Luck, 2002); 2) reducing the stability of mental representation. To build a mental representation takes time. For example, it takes several minutes for air traffic controllers to "warm up" before their visual scan patterns become regular and they can reliably perform their tasks (Stein, 1992). As a result, if too many entities are updated at a high rate, the mental representation tends to deteriorate. That corresponds to controllers "losing the picture" (Hopkin, 1995).

Relation evaluated by cognition

A task can become more complex as the number of interacting factors increase. Thus complexity can be measured by the dimensionality of the relation or number of variables that are related in a task. 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 (Halford, Wilson, & Phillips, 1998). Relational complexity is defined as the number of independent elements or variables that must be simultaneously considered to solve a problem. Many cognitive processes, such as selection of actions, manipulation of goal hierarchies, reasoning, and planning actions, are examples of processing at high levels of relational complexity. Halford et al. argued that relational complexity reflects the cognitive resources required to perform a task. The more interacting variables that have to be processed in parallel, the higher both the cognitive demand and computational cost will be. For example, an equation $a = 3 * b$ is a binary relation while an equation $a/b = c/d$ is a quaternary relation and therefore more complex. Hence, relational complexity is suitable to measure the affect of relation on cognitive load.

Because working memory links pieces of information that are needed simultaneously for task performance, relational complexity turns out to be a straightforward measure of the working memory load of a task. Halford et al. further demonstrated that the processing capacity of working memory for normal adults is limited to quaternary relations: Adults can 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 working memory at about four items (Cowan, 2001).

Metric –3: Action feasibility

The purpose of an ATC automation aid is to increase capability and decrease workload. Therefore, it is desirable that a display provides information without demanding too much action from users. This is especially important for time-critical tasks such as air traffic control. If an automation aid requires too many inputs from controllers, it shunts controllers' attention away from the main tasks and may increase the risk of operational errors. However, given that today's automation systems are designed to provide large volumes of information, they inevitably require controllers to interact with them. Specific actions may include 1) eye/head movements to search for specific information; 2) keystrokes to update information and make inquires and 3) mouse movements to select specific information on a display. The following metrics of action complexity were determined by quantifying how feasible it is to perform those movements in a timely manner.

Size evaluated by action

The proposed metric is the number of keystrokes and mouse movements. Compared with keystrokes and mouse movements, the time needed for eye and head movements is negligible. Therefore, only keystrokes and mouse movements are considered here. Mouse movements are typically made to select information in a region of interest (ROI). That is, the larger the area of an ROI, the less time that is needed to perform a selection action with the mouse. Thus, the moving distance and the ROI size both contribute to the cost of mouse movements.

Variety evaluated by action

The proposed metric is the number of action transitions required by a functional unit. An action transition is a change of action modes, such as from keystrokes to mouse movements or vice versa. Those transitions take time and require the brain to coordinate different action modes.

Relation evaluated by action

The proposed metric is the degree of action depth needed to achieve the goal specified by a functional unit. Action depth is the number of serial steps needed to achieve the task goal of a functional unit. An example of action depth is the number of layers of pop-up windows needed to accomplish a given task. Complex systems are usually characterized by a multi-level structure. Theoretically, a two-level structure is desirable to maintain low complexity. With a two-level structure, the information hierarchy required by task goals is achieved by a number of parallel, independent subgoals. However, following the need to increase the variety of actions, today's automation aids 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. The subgoals are determined by interactions between the sub-structure of the original task and the details of the system interface. Researchers have used the number of serializable subgoals as a measure of complexity for a system with a multi-level structure (Heylighen, 1998).

DISCUSSION

Complexity measures in the literature

Perceptual complexity and Tullis' display complexity

Many algorithms have been developed to address image or pattern complexity (Landsberg & Shiner, 1998; Grassberger, 1991; Klinger & Salingaros, 2000; Orland, Radja, Larsen, & Weidemann, 1994). However, most are based on information theories and have low correlations with human judgment. In contrast, Tullis (1985,1986) developed a set of metrics to measure display complexity based on human performance of visual search tasks. The metrics used 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 of groups and average group size, and d) layout complexity, which describes how well the arrangement of items on a display follows a predictable visual scheme. Using these display characteristics, Tullis was able to obtain correlation coefficients of .71 for predicting search times.

One drawback associated with Tullis' metrics is that the measures were specified for text-dominant displays but not for graphical ones. It is unclear how to identify Tullis' groups in spatially continuous 2-D graphical patterns. In addition, Tullis' metrics did not take into account the effects of color-coding. Color is routinely used in today's displays, given that it is a very efficient way to segregate information visually. Color segregation means that targets can be searched "effortlessly" by colors. In terms of the complexity measures proposed in this report, color-coding reduces the number of fixation groups because one only needs to make eye fixations to targets of a given color.

Cognitive complexity

Previous studies of cognitive complexity have focused on text comprehension, creativity, social phenomena, etc. In comparison, little research has been devoted to the complexity of visual displays. Crokett (1965) used the concept of "level of hierarchic integration of constructs" to define complexity. With this definition, cognitive complexity is associated with increasing differentiation (containing greater number of constructs), articulation (consisting of more refined and abstract elements) and hierarchic integration (organized and interconnected). Kelly (1955) described cognition as a construct system, composed of constructs and elements, where constructs are transparent templates that a person uses to understand the world. Bieri (1955) developed the first index of cognitive complexity. The intent of the index was to measure differentiation. Two measures were used: number of constructs and matches (i.e., similarity) between the constructs. A key issue in applying the data to Bieri's index is determining the independent constructs. Several numeric computational methods, such as principal components analysis and factor-analysis, were used to reveal the independence of the elicited constructs (Bezzi, 1999; Woehr, Miller, & Lane, 1998). Morçöl and Asche (1993) used this index 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, suggesting that a person is more creative if he or she perceives things in more complex ways.

We identified three measures of cognitive complexity: number of functional units, rate of information change over time, and relational complexity. The number of functional units is similar to the measure of constructs in the literature. We proposed relational complexity as a metric to quantify how the relation factor of complexity affects cognition. This measure corresponds to the interconnected hierarchic integration proposed by Crokett (1965). We also proposed to use rate of information change to measure the variety factor of complexity; the dynamic aspect of cognitive complexity has been ignored in previous studies.

Essentially all the measures of cognitive complexity in the literature deal with static subjects. Yet information presented by an ATC automation display is constantly changing. Visual psychophysical experiments have demonstrated that changes of visual information taxed viewers' working memory. We proposed to use the rate of information change as the measure of such dynamic complexity. Unfortunately, this measure is not well defined in this report. With that being said, we are still exploring various possibilities about how to compute information change. A critical, unsolved problem is: What kinds of changes in displayed materials affect a controller's cognitive load?

Action complexity

Many methods have been developed to assess the complexity of human-computer interfaces (McCabe 1976; Kornwachs, 1987; Rauterberg, 1994). Those methods require modeling a system's states and transitions between states. Unfortunately, it is implausible to apply such methods to ATC displays directly. Bressolle et al. (2000) reported that controllers use automation tools adaptively and no common procedure can be specified. Thus, 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 following a fixed procedure. That would be against the philosophy of automation aid design.

One measure related to action complexity is Sears' Layout Appropriateness metric (1994). Sears proposed a layout complexity metric to assess users' performance on a display. The metric was the summed product of the frequency of action transitions and the cost of transitions. The two factors in our metrics, amount of manual movements and transitions between the movements, are essentially the same as Sears' metric. Sears used the distance that users must move and the size of the objects to be moved as the cost of a transition. Such a metric of layout complexity can be used to evaluate the efficiency of a user-interface layout. It can also be used to compute the extent to which a display demands action. Therefore, within the ATC domain, the number of keystrokes, the distance of mouse movements, and head/eye movement patterns can all be measured as the "cost" of action. Likewise, the number of switches between action modalities can be quantified and added to the complexity of ATC displays.

Challenges to the validation of the metrics

The metrics presented in this report are based on theoretical studies and field observation of ATC automation displays. They are preliminary and need to be experimentally validated before being applied to display evaluations. One criterion for a good evaluation method is that the entities of the metrics should be maximally orthogonal to each other so that a compact, independent set of pertinent factors can be elucidated. The metrics presented in this report were developed under a theoretical framework: Complexity lies in the interaction between the system and the observer with the constraints of task requirements. This framework views the complexity as an entity of three orthogonal dimensions: numerical size, variety, and relation. It results in a 3x3 table containing metrics of complexity: three complexity factors evaluated by the three stages of brain information processing.

While there is evidence that the mechanisms of information processing are distinctive at these three stages, we do not yet have strong evidence that the functions from different stages are orthogonal. For example, the metric "action depth" might correlate with the metric "relational complexity" because they both affect working memory load. This problem brings challenges to the validation process. Analytic methods such as factor analysis should be employed to validate the orthogonally of the metrics. In addition, we also need to determine whether the chosen metrics are appropriate for ATC.

One potential approach to metric validation is to use eye trackers to monitor controllers' visual scan patterns while they use automation displays. For example, Stein (1992) observed that the scan pattern for controllers usually involves dividing the primary screen into separate segments and allowing the eyes to clear each segment momentarily before moving to the next segment. That study further showed that regular visual scan patterns are essential to task performance. It is possible that using an automation aid may interrupt a controller's regular scan patterns to some degree since a controller needs to move his or her eyes to the screens/windows associated with automation aids. Such movements will logically increase the variation of the scan patterns on the primary display and may thus decrease task performance. Willems, Allen, and Stein (1999) used a Markov matrix to compute the regularity of visual scan patterns. The method can be used to quantify the disruption of scan patterns due to the use of automation aids.

In conclusion, this report presents a framework for developing metrics of information complexity in automation displays. The framework is described as follows: 1) information complexity is the combination of three basic factors: numeric size, variety, and relation; 2) complexity factors are associated with the functions at three stages of brain 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 metrics to assess the complexity of automation displays in air traffic control. We expect that these metrics will not only be used for evaluation of new systems but will also serve as a guideline for interface design.

REFERENCES

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.

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; 51:263-8.

Bressolle MC, Benhacene R., Boudes N, Parise R (2000). Advanced decision aids for air traffic controllers: Understanding different working methods from a cognitive point of view. NLR, EU. 39, A2.

Burleson BR, Caplan SE (1998). Cognitive complexity. In: McCroskey et al., eds. Communication and personality: Trait perspectives. Cresskill, NJ: Hampton Press.

Cant S, Jeffery DR, Henderson-Sellers B (1995) A conceptual model of cognitive complexity of elements of the programming process. Information Software Technology; 37(7):351- 62.

Carpenter PA, Just MA, Reichle ED (2000). Working memory and executive function: Evidence from neuroimaging. Current Opinion in Neurobiology. 10:195-9.

Carswell CM (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. Human Factors; 34(5):535-54.

Carswell CM, Wickens CD (1987). Information integration and the object display: An interaction of task demands and display superiority. Ergonomics; 30: 511-28.

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: Haher 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.

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.

Grassberger P (1986). Towards a quantative theory of self-generated complexity. International Journal of Theoretical Physics; 25(9):907-38.

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.

Joseph JS, Chun M and Nakayama K (1997). Attentional requirements in a "preattentive" feature search task. Nature; 387:805-7.

Kallus KW, Barbarino M, Van Damme D (1997). Model of the cognitive aspects of air traffic control. Brüssel, Brüssel: EUROCONTROL; Ref.No. HUM.ET1.ST01.1000-Rep-02, (55 S.).

Kelly GA, ed. (1955). The psychology of personal constructs. New York: W. W. Norton and Company.

Klinger A, Salingaros NA (2000). A pattern measure. Environment and Planning; B27:537-47.

Klemola T (2000). Cognitive complexity metrics and requirements comprehension. Australian Conference on Software Measurement (ACOSM 2000); 2000 Nov 1-3; Sydney, Australia.

Komwachs, K (1987). A quantitative measure for the complexity of man- machine interaction process. In: Bullinger H-J, Shackel B, eds. Proceedings of Human-Computer-Interaction. INTERACT87, Elsevier Science, North-Holland; 109-16.

Landsberg PT, Shiner JS (1998). Disorder and complexity in an ideal non-equilibrium Fermi gas. Physical Letters; A245:228-32.

Langton C (1991). Life at the edge of chaos. In: Langton C, Taylor C, Farmer J, Rasmussen S, eds. Proceedings Artificial Life II; Redwood City, California: Addison-Wesley; 41–91.

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

Morçöl G, Asche M (1993). The repertory grid in problem structuring: A case illustration. The International Journal of Personal Construct Psychology; 6:371-90.

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

Rauterberg M (1994). About the relationship between incongruity, complexity and information: design implications for man-machine systems. In: Rauch W, Strohmeier F, Hiller H, Schlögl C, eds. Mehrwert von Information — Professionalisierung der Informationsarbeit; Konstanz: Universitätsverlag; 122-32.

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

Scott WA (1969). The structure of natural cognition. Journal of Personality and Social Psychology; 12: 261-78.

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

Stein ES (1992). Air traffic control visual scanning. Washington DC: Federal Aviation Administration; Report No: DOT/FAA/CT-TN92/16.

Tullis TS (1985). A computer-based tool for evaluating alphanumeric displays. Human-Computer Interaction: INTERACT; 1984 Sep; London, England.

Tullis TS (1986). Optimizing the usability of alphanumeric displays. Proceedings of HCI'86 Conference on People and Computers; 1986 Sep; York, England.

Wickens CD, Andre AD (1990). Proximity compatibility and information display: effects of color, space, and object display on information integration. Human Factors; 32(1):61-77.

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

Woehr DJ, Miller MJ, Lane JAS (1998). The development and evaluation of computer-administered measure of cognitive complexity. Personality and Individual Differences; 25:1037-49.

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; In press.