

# Self-Generated Complexity and Human–Machine Interaction

Christopher M. Schlick, *Member, IEEE*, Carsten Winkelholz, Florian Motz, and Holger Luczak

**Abstract**—A novel approach to measure the self-generated complexity of human–machine interaction is presented. A complexity measure that relies solely on information-theoretic quantities is defined. An efficient estimation procedure for the complexity measure is introduced. The theoretical framework is validated by a study of human interaction with electronic map displays. Thirty experienced master mariners participated. The task was to search for multiple vessel symbols under time pressure. Samples from the visual scanpath, as well as manual responses, were acquired. The workloads due to time pressure, the number of symbol clusters on the display, and the map scenario were systematically varied. The results of an analysis of variance (ANOVA,  $\alpha = 0.05$ ) show a significant complexity decrease when the time pressure (or workload) is increased. The workload effect on the complexity of master mariners' visual scanpath was stronger than on their manual response. There was also a significant effect of the number of symbol clusters on the display: a display with two clusters showed a significantly higher search complexity for manual response than a nonclustered display.

**Index Terms**—Human factors, information theory, man–machine system, user interfaces.

## I. INTRODUCTION

THE IDEA of developing a theory of complexity has fascinated the scientific community for many decades. An early approach dates back to the great mathematician A. N. Kolmogorov. The Kolmogorov complexity of a symbol string  $s$  with respect to a universal computer is the minimum length over all programs that print  $s$  and then halt [5]. Kolmogorov preferred an algorithmic concept, and his complexity measure of patterns is maximal if the generated symbols follow a discrete stochastic process with independent and uniformly distributed random variables, e.g., fair coin tosses. For design and evaluation of human–machine systems, the algorithmic approach is an unsatisfactory basis in two respects. The first is pragmatic, because there is no algorithm to compute such a quantity for arbitrary patterns. The second reason is conceptual and most important: the algorithmic measure is closely related to the intuitive notion of randomness rather than complexity. In this sense, the Kolmogorov complexity can be considered a form of information per symbol stored in a specific pattern and for the applied cases agrees with Shannon's information entropy [5].

Manuscript received October 31, 2003; revised July 30, 2004. This paper was recommended by Associate Editor E. J. Bass.

C. M. Schlick and H. Luczak are with the Institute of Industrial Engineering and Ergonomics, Faculty of Mechanical Engineering, RWTH Aachen University of Technology, Aachen 52056, Germany (e-mail: c.schlick@iaw.rwth-aachen.de).

C. Winkelholz and F. Motz are with the German Research Establishment for Applied Science FGAN, Wachtberg 53343, Germany.

Digital Object Identifier 10.1109/TSMCA.2006.859096

Following the complexity understanding of Grassberger [10] in theoretical physics, we argue that our visual intuition of the complexity of patterns is an ensemble concept and is not related to individual patterns. Furthermore, complexity in the intuitive sense of science is not monotonic increasing with the Kolmogorov complexity or Shannon information entropy. Instead, a complexity measure should be small for completely disordered (random) or completely ordered (periodic) patterns and should have a maximum for structures with long-range correlations between symbols and therefore large mutual information. This concept was enriched by later work in basic research [2], [6], [8], which strongly stimulated the approach of this paper.

It is a characteristic of human–machine interaction that the complexity is self-generated, or emergent. Self-generated complexity arises from the ability of human cognition to make spontaneous abstractions of phenomena in the task domain. The human is able to dynamically distinguish important and unimportant features of the task and thereby acquire its “meaning” during the interaction. According to Endsley's work on designing systems for situation awareness [9], complexity is a critical issue: “Left unchecked, complexity directly works against situation awareness. It reduces people's ability to perceive needed information among the plethora of that available. It can also significantly affect the ability of users to understand and predict what will happen when working with a system (...) because complexity makes it much more difficult to form a good mental model of the system or task domain.” She distinguishes three levels [9]: the complexity of the system, the operational complexity that the user must deal with, and the apparent complexity that is caused by the system interface features. Apparent complexity is subdivided into cognitive complexity, display complexity, and task complexity. This multilevel consideration can be criticized because it is only qualitative and there is a large body of generic “complexity contributing factors,” e.g., the number of system items, the degree of interaction of the items, the system dynamics, and the predictability of changes. It is difficult to integrate these factors in an explicit complexity function for empirical human–machine studies.

On the other hand, the goal of this paper is to present a quantitative complexity measure that is based on a complexity understanding of basic research [2], [6], [8], [10]. The novel measure relies on information-theoretic variables and there is no need for factor-based assessment. In the human–computer interaction community, there are other rigorous approaches, e.g., the “cognitive complexity” of Rauterberg *et al.* [20], [21] or the “user complexity” of Kieras and Polsen [14]. Rauterberg *et al.* developed a method for constructing Petri net models of behavior with the help of log files of users solving the

same task. The idea is to use the structure of the reconstructed mental process model for a complexity assessment. The complexity measure represents the difference between the total number of flow relations in the Petri net and the total number of places and transitions. The user complexity of Kieras and Polsen is defined as the complexity of a device from the point of view of the user, which depends on the amount, content, and structure of the knowledge required to operate the device successfully. Two major components of the knowledge are the users' task and the device representation. A task modeling technique using production rules and a working memory is defined. The device behavior is modeled by generalized transition networks. In spite of a very detailed analysis, Kieras and Polsen [14] do not give a complexity equation as do Rauterberg *et al.*, but mention that possible measures should include the number of production rules in the task model, the maximum number of goals in the working memory during the performance of a given interaction function, the number of conditions and actions in a production rule, etc. In light of our complexity understanding, the main weakness of both methods is that they are algorithmic and rely on deterministic interaction models. Therefore, interindividual differences, human errors, or emergent interaction phenomena due to spontaneous mental abstractions have to be modeled explicitly, or cannot be represented. Furthermore, reconstructing an interaction model from data is time consuming and error prone because the user interface designer has to make many (and possibly wrong) assumptions about human cognition in the task situation. To the best of our knowledge, there are only two alternative information-theoretic approaches—those of Kang and Seong [12] and Karwowski [13]—that avoid the problems of algorithmic methods. However, in both references, Shannon's information entropy is directly considered to be a complexity measure and therefore the fundamental critique still holds that an adequate measure should be small for random or periodic patterns and should have a maximum for rich structures with long-range correlations.

To address the shortcoming of other methods, an extensive theoretical analysis of self-generated complexity and human–machine interaction is presented in Section II-A. Moreover, the theory is demonstrated by a simple but easy to generalize example. Based on these considerations, an estimation procedure, which exploits the predictability of the human–machine interaction process, is developed in part B of the same section. The method is designed to be used in the evaluation of up-to-date human–machine interfaces with high perceptual, cognitive, and sensory-motor demands due to real-time system operation. The validity of the theory is examined in Section III. The task of the participants in the validation study was to search for multiple targets on electronic map displays under time pressure. The results are presented in detail and the findings discussed in light of the developed theory. The paper concludes with a summary of the main issues studied and an outline of future research.

## II. THEORY

### A. Complexity Measure

First, the observable quantities for complexity measurement have to be defined. Therefore, a finite set  $\mathcal{O} \equiv \{o_1, o_2, \dots, o_K\}$

of observable interaction events (OIEs) is introduced. OIEs are generated by the user, or by the machine response, and reflect the complementary relationships between the observer and the object. User-generated OIEs may represent speech acts or sensory-motor activities such as pressing buttons, clicking a mouse on interface objects, eye fixation on display symbols, etc. Machine-related events may include display state updates and alerts. OIEs are a symbolic representation of an interaction process that is continuous in reality and the coding may result in loss of information. However, the risk of significant information loss in modern human–machine systems is low because many tasks have already shifted to a symbolic level due to their event-driven subsystems. The associated subsymbolic processes of perception and motor response can be effectively encoded using the methods of symbolic dynamics [16]; therefore, a pattern, or string, of OIE  $s \equiv o_{j(1)}, \dots, o_{j(T-1)}o_{j(T)}$  ( $o_{j(T)} \in \mathcal{O}$ ) is introduced. A pattern encodes the dialogue between human and machine. The index  $j(t)$  is simply a pointer to elements of the event set  $\mathcal{O}$ . If  $d$  dialogue situations due to repeated interaction trials were observed, they are combined in the case base  $\text{CB} \equiv (s_1, \dots, s_d)^T$ . The case base stores the complete data of the interaction process.

Second, the model class has to be defined. To adequately represent the ensemble concept, the human–machine interaction is modeled as a discrete stochastic process. Therefore, a random variable  $O_t$  for each time slice  $t$  of the process is introduced and this represents the behavior under uncertainty.  $P(O_t = o_{j(t)}) \in [0, 1]$  denotes the probability distribution of  $O_t$  with possible experimental outcomes  $o_{j(t)}$  in terms of the action alternatives. In addition, a joint probability distribution is defined as

$$P(O_1 = o_{j(1)}, \dots, O_T = o_{j(T)}) \in [0, 1]. \quad (1)$$

The joint distribution denotes the probability over patterns of  $T$  consecutive OIEs and represents a complete source of information of an observer about an interaction process without any structural assumptions. Therefore, there is no need to explicitly model the users' cognitive processes and memory structures. The joint distribution can be factorized using Bayes' theorem of conditional probability  $P(O_1, O_2) = P(O_1) \times P(O_2|O_1)$

$$P(O_1 = o_{j(1)}, \dots, O_T = o_{j(T)}) = P(O_1 = o_{j(1)}) \times \prod_{t=2}^T P(O_t = o_{j(t)} | O_1 = o_{j(1)}, \dots, O_{t-1} = o_{j(t-1)}). \quad (2)$$

To simplify the analysis, we confine the description to stationary interaction situations satisfying

$$P(O_{t+1} = o_{j(t+1)}, \dots, O_{t+T} = o_{j(t+T)}) = P(O_1 = o_{j(1)}, \dots, O_T = o_{j(T)}). \quad (3)$$

Third, a fundamental information-theoretic variable  $H$  is defined, which is a function of the joint distribution as an

information source [10]. The Shannon entropy  $H$  of length  $n$  sequences  $o_{j(1)}, \dots, o_{j(n-1)}o_{j(n)}$  with  $n > 0$  is given by

$$H(n) \equiv - \sum_{\{o_{j(1)}, \dots, o_{j(n)}\} \in \mathcal{O}^n} P(O_1 = o_{j(1)}, \dots, O_n = o_{j(n)}) \times \log_2 P(O_1 = o_{j(1)}, \dots, O_n = o_{j(n)}). \quad (4)$$

The sum in (4) is understood to run over all possible blocks of  $n$  consecutive OIEs. This entropy is not the entropy of a finite pattern with length  $n$ ; rather, it is the entropy of patterns with length  $n$  drawn from, in principle, much longer, or infinite, sequences generated by the interaction process. The variable  $n$  is the order parameter and can be interpreted as an observation window of length  $n$  over the patterns. The Shannon entropy measures the average memory, in bits, which is required to store the outcomes of the random variables representing the behavior under uncertainty. If no observations are made, there is nothing about which to be uncertain and we define  $H(0) \equiv 0$ .

Fourth, the block entropy  $h(n)$  of order  $n$  is defined as the discrete derivative of the entropy curve  $H(n)$

$$h(n) \equiv H(n) - H(n-1). \quad (5)$$

Equivalently, the block entropy can be regarded as a conditional entropy (6). Therefore,  $H(n)$  and  $H(n-1)$  in (5) are substituted by their corresponding definitions from (4) and Bayes' rule is used to factorize  $P(O_1, \dots, O_n) = P(O_1, \dots, O_{n-1}) \times P(O_n | O_1, \dots, O_{n-1})$ , i.e.,

$$h(n) = \sum_{\{o_{j(1)}, \dots, o_{j(n-1)}\} \in \mathcal{O}^{n-1}} P(O_1 = o_{j(1)}, \dots, O_{n-1} = o_{j(n-1)}) \times (-) \sum_{o_{j(n)} \in \mathcal{O}} P(O_n = o_{j(n)} | O_1 = o_{j(1)}, \dots, O_{n-1} = o_{j(n-1)}) \times \log_2 P(O_n = o_{j(n)} | O_1 = o_{j(1)}, \dots, O_{n-1} = o_{j(n-1)}). \quad (6)$$

The units of  $h(n)$  are in bits per event. If the order parameter  $n$  is 1, the block entropy  $h(1)$  is equivalent to the Shannon entropy  $H(1)$ . Consider the case where we have already seen  $o_{j(1)}, \dots, o_{j(n-1)}$  interaction events [first factor in (6)] and we want to know how much we have learned about the underlying rules in this pattern. The more we know, the more effectively we can predict the next observation  $o_{j(n)}$  [conditional probability in (6)] and therefore the fewer bits that will be required to describe the deviation of this event from our prediction [2]. In other words, knowledge of the interaction process is measured by the degree to which one can increase the average prediction accuracy of new observations. The block entropy has three important properties. First, it is monotone decreasing with the block order  $n$ . Second, it is bounded from below by 0, representing the case of perfect predictability of the interaction process. Third, it is also bounded from above by the quantity  $\log_2 K$ , which is the characteristic of completely

random processes, e.g., fair coin tosses. The zeroth-order block entropy denotes the upper bound

$$h(0) \equiv \log_2 K. \quad (7)$$

However, the case of completely randomized interaction patterns is "pathological" in human-machine studies, and it is reasonable to assume that intelligent observers are able to learn more about the interaction rules and to make better predictions the longer they observe. Clearly, the learning procedure cannot be infinite because only limited information in the observed cases can contribute to the best predictions. Therefore, in most human-machine studies, the block entropy reaches a limit

$$h_\mu \equiv \lim_{\eta \rightarrow \infty} h(n = \eta) \quad (8)$$

relatively quickly.  $h_\mu$  is equivalent to the source entropy rate [5] and can be regarded as an "informational barrier" separating predictive information—information "treasured" in the observations improving the accuracy of predictions of an intelligent observer—from "noise."

Fifth, an information-theoretic learning curve  $\Lambda(n)$  in the sense of Bialek *et al.* [2] is defined as

$$\Lambda(n) \equiv h(n) - h_\mu. \quad (9)$$

This learning curve measures the potential to improve predictions when expanding the observation window and therefore can be regarded as an  $n$ -dependent or local predictability [8] of the interaction process. Superficially,  $\Lambda(n)$  appears to be different from conventional learning curves in ergonomics (see [29]), which often show that the frequency of errors for a particular task decreases with the number of practice trials taken until a lower limit is reached. However, both learning curves follow the same principle of the general "law of diminishing returns" [2]. Here, the cost of an error is that the encoding of  $o_{j(n)}$  is longer than it would be if we had perfect knowledge of an infinite observation window.

Sixth,  $\Lambda(n)$  is used to derive the desired complexity measure  $C$  of the human-machine interaction. Therefore, the area between the curve and the abscissa is computed by the discrete integral

$$C \equiv \lim_{\eta \rightarrow \infty} \sum_{n=1}^{\eta} \Lambda(n). \quad (10)$$

$C$  measures the amount of historical information stored in the observation that is communicated to the future. In other words,  $C$  measures the mutual information between the past and the future. The unit of  $C$  is bits. A large complexity value indicates correlations within patterns, thus suggesting that an interaction process possesses a greater degree of structure—internal structure due to interaction rules that are responsible for maintaining the correlations.

We analyze  $C$  with the help of three simple examples. First, if the interaction process is completely random and there are no rules to learn, the block entropy limit is  $h_\mu = \log_2 K$ . The limit is reached at  $h(1)$  and cannot be underrun by longer observation windows. The complexity of such an obviously meaningless

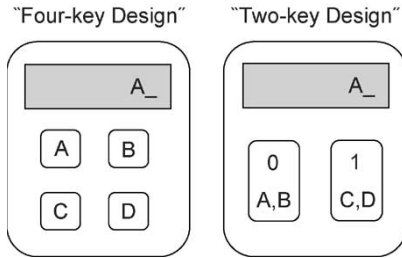


Fig. 1. Two keyboard designs for a simplified personal pager system.

human-machine interaction is minimal with  $C = 0$ . Second, if the interaction is completely monotonic and repetitive for the user, e.g., it is a periodic process with period length  $p$ , the limit is  $h_\mu = 0$ , because action and reaction are perfectly predictable after having observed the events of a full period. When the estimation procedure (Section II-B) can memorize the full period and there are no event recurrences within the period, the zero limit is reached at  $h(2)$  and there is  $C = h(1) = -p \cdot 1/p \cdot \log_2(1/p) = \log_2(p)$ . Even if there are event recurrences within the period, the complexity measure  $C$  equals  $\log_2(p)$ . The only difference is that the zero limit is not reached at  $h(2)$  but at larger block lengths  $n' \leq p$ . This is an intuitive result. Although the complexity of periodic interaction processes is rather low and the learning curve converges quickly, processes with longer periods are more complex and somehow more “meaningful” for the human than processes with shorter periods. Third, if the interaction process follows complicated rules and therefore has a large portion of long-range intersymbol correlations, which is the case for written human languages such as written English [8], then the learning curve has a long stretched tail and the complexity measure may also diverge as  $C \rightarrow \infty$ . This is the information-theoretic reason why proper writing is difficult to learn. Due to their rich hierarchical grammar and semantics, written human languages represent the most complex interaction processes. However, in most human-machine studies, the complexity is significantly lower and rapid convergence can be expected. The case  $C \rightarrow \infty$  is primarily of theoretical interest [2]. In the literature, the complexity measure  $C$  has different interpretations, such as the “effective measure complexity” [10], “predictive information” [2], or “excess entropy” [6].

*Example:* This subsection demonstrates the complexity theory by a didactic example. Consider two keyboard designs for a simplified personal pager system, shown in Fig. 1. The pager is always online and ready to transmit and receive text messages with a reduced alphabet of four letters, i.e., A, B, C, and D. The first keyboard design on the left side of Fig. 1 has four keys (index 4K) and therefore allows a direct user input of the four letters. The second keyboard design is cheaper to manufacture and has only two keys 0 and 1 (index 2K). Therefore, a state-dependent input mode is necessary that is similar to cellular phones: if a user wants to input an A letter, he/she has to press the 0 button once and then make a short pause P to indicate the end of the letter input. If a B letter has to be input, the 0 button has to be pressed twice and then the short pause P, and so on. The coding scheme of the whole input mode is shown in Table I. As Kieras and Polsen [14] state, the coding scheme is a device representation.

TABLE I  
CODING SCHEME OF THE STATE-DEPENDENT INPUT MODE FOR THE KEYBOARD DESIGN WITH TWO KEYS

| Letter to be input  | A  | B   | C  | D   |
|---------------------|----|-----|----|-----|
| Input event pattern | 0P | 00P | 1P | 11P |

TABLE II  
PROBABILITY DISTRIBUTION OF SINGULAR LETTERS  $S$  OF THE REDUCED GRAMMAR

| $s$    | A   | B   | C   | D   |
|--------|-----|-----|-----|-----|
| $p(s)$ | 1/4 | 1/4 | 1/4 | 1/4 |

TABLE III  
CONDITIONAL PROBABILITY DISTRIBUTION OF LETTER  $S_2$  GIVEN  $S_1$  OF THE REDUCED GRAMMAR

| $s_1$        | A   | A   | B   | B   | C   | C   | D   | D   |
|--------------|-----|-----|-----|-----|-----|-----|-----|-----|
| $s_2$        | A   | C   | A   | D   | B   | D   | B   | C   |
| $p(s_2 s_1)$ | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 |

We hypothesize that the interaction complexity is larger when using the cheaper keyboard design with two keys, because the user has to memorize the device state. Only user-generated OIEs are considered, e.g., pressing a key on the pager, and not device-generated events such as display updates. Therefore, in case of the four-key design, there is an event set  $\mathcal{O}_{4K} \equiv \{o_1 = A, o_2 = B, o_3 = C, o_4 = D\}$  with a cardinality  $|\mathcal{O}_{4K}| = 4$ . In the second case, there is  $\mathcal{O}_{2K} \equiv \{o_1 = 0, o_2 = 1, o_3 = P\}$  with  $|\mathcal{O}_{2K}| = 3$ . A user-generated pattern of OIE is simply denoted by  $s = AC$  or  $s = 0P1P$ , respectively. For the sake of simplicity, it is assumed that the pager user follows a reduced language grammar when sending text messages and not the very complex grammar of English. The reduced grammar is specified through two probability distributions  $p(s)$  and  $p(s_2|s_1)$  and can be interpreted as a stochastic task network in the sense of Kieras and Polsen [14]. The initial distribution  $p(s)$  of observing any of the four letters in input strings of arbitrary length is shown in Table II. From this table, there is a uniform probability distribution of singular letters. The conditional distribution  $p(s_2|s_1)$  is shown in Table III. All letter pairs that are not listed in Table III violate the grammatical rules and therefore are assigned zero probability values. An example of this is the probability of observing B, given A is zero, and therefore a pattern AB is grammatically incorrect. One may argue that in written English the most common letter is E with a frequency of approximately 13%, whereas the least common letters Q and Z occur with a frequency of approximately 0.1% [5], and that conditional probabilities of much higher order must be considered. However, the interested reader may choose other probability values and more grammatical rules; however, noticeable complexity differences among both keyboard designs will remain and the results can be easily generalized.

TABLE IV  
PROBABILITY DISTRIBUTION OF SINGULAR EVENTS  $S$   
IN CASE OF THE TWO-KEY DESIGN

|        |      |      |      |
|--------|------|------|------|
| $s$    | 0    | 1    | P    |
| $p(s)$ | 3/10 | 3/10 | 4/10 |

In case of the first keyboard design, the zeroth-order block entropy (7) is

$$h_{4K}(0) = \log_2 |\mathcal{O}_{4K}| = \log_2 4 = 2. \quad (11)$$

When computing higher orders of  $h$ , we use the conditional entropy definition in (6). From the previous section, it is known that  $h(1)$  equals  $H(1)$ . Therefore, we can use the distribution of Table II and compute its Shannon entropy directly. The entropy value  $h(1)$  for an observation window of length  $n = 1$  is

$$h_{4K}(1) = - \left( 4 \times \frac{1}{4} \log_2 \frac{1}{4} \right) = \log_2 4 = 2. \quad (12)$$

If the observation window is expanded to a length of  $n = 2$ , the block entropy from (6) using the values from both Tables II and III is

$$h_{4K}(2) = - \left( 4 \times \frac{1}{4} \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \right) = \log_2 2 = 1. \quad (13)$$

From the conditional distribution of the reduced grammar in Table III, there are only neighbor correlations between letters and therefore larger window lengths  $n = 3, 4, \dots$  do not further decrease the block entropy. In other words, there is a limit  $h_{\mu 4K} = h_{4K}(2) = 1$ . A very rapidly converging learning curve  $\Lambda(n)$  is found because of the simple grammar and a direct symbol input. The complexity measure  $C_{4K}$  on the basis of (9) and (10) using the above entropy values is

$$C_{4K} = h_{4K}(1) - h_{\mu 4K} = h_{4K}(1) - h_{4K}(2) = 2 - 1 = 1. \quad (14)$$

Conversely, in case of the second keyboard design, the zeroth-order block entropy is

$$h_{2K}(0) = \log_2 |\mathcal{O}_{2K}| = \log_2 3 \approx 1.59. \quad (15)$$

Obviously,  $h_{2K}(0)$  is lower than  $h_{4K}(0)$  because there are less action alternatives *a priori*, but it will turn out that the state-dependent input mode generates new correlations, influencing higher entropy orders. Because the coding scheme of Table I has to be considered when computing  $h_{2K}(1)$ , a new probability table of singular events has to be computed, which is shown in Table IV. The probability values can be derived easily because they equal the frequency of input events in the second row of Table I.

On the basis of Table IV, the calculation of  $h_{2K}(2)$  is

$$h_{2K}(1) = - \left( 2 \times \frac{3}{10} \log_2 \frac{3}{10} + \frac{4}{10} \log_2 \frac{4}{10} \right) \approx 1.57. \quad (16)$$

TABLE V  
CONDITIONAL PROBABILITY DISTRIBUTION IN CASE OF THE  
TWO-KEY DESIGN FOR A BLOCK LENGTH  $N = 2$

|              |     |     |     |     |     |     |
|--------------|-----|-----|-----|-----|-----|-----|
| $s_1$        | 0   | 0   | 1   | 1   | P   | P   |
| $s_2$        | 0   | P   | 1   | P   | 0   | 1   |
| $p(s_2 s_1)$ | 1/3 | 2/3 | 1/3 | 2/3 | 1/2 | 1/2 |

TABLE VI  
CONDITIONAL PROBABILITY DISTRIBUTION IN CASE OF THE  
TWO-KEY DESIGN FOR A BLOCK LENGTH  $N = 3$

|              |     |     |     |     |     |     |     |     |     |     |
|--------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $s_1$        | 00  | 11  | 0P  | 0P  | 1P  | 1P  | P0  | P0  | P1  | P1  |
| $s_2$        | P   | P   | 0   | 1   | 0   | 1   | 0   | P   | 1   | P   |
| $p(s_2 s_1)$ | 1   | 1   | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 |
| $p(s_1)$     | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |

The conditional probability table for the block length  $n = 2$  is shown in Table V. For instance, there is  $p(P|0) = 2/3$  in Table V because the 0 may be the first number coding B or A. Therefore, the probability that a pause follows is twice as high as a following 0.

The block entropy  $h_{2K}(2)$  according to (6) using Tables IV and V is

$$h_{2K}(2) = -2 \times \frac{3}{10} \left( \frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3} \right) - 1 \times \frac{4}{10} \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \approx 0.95. \quad (17)$$

The same procedure is used to construct the conditional probability Table VI, which is required for  $h_{2K}(3)$ . The result is shown as

$$h_{2K}(3) = -4 \times 0.2 \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) = 0.8 \log_2 2 = 0.80. \quad (18)$$

Note that conditional probability values of one do not contribute to  $h_{2K}(3)$  because  $\log_2 1 = 0$ .

To reduce the calculation, the probability values of the given string  $s_1$  are also shown in Table VI. For example,  $p(P|00) = 1$  in Table V, because the coding scheme in Table I does not allow a double null to be followed by an event other than a pause P. Moreover,  $p(00) = 0.1$ , because Bayes' rule (2) leads to the factorization  $p(00) = p(0) \times p(0|0) = 0.1$ .

The full conditional probability table for block length  $n = 4$  exceeds the space and is not given. Instead, we show a compact entropy Table VII, which only distinguishes the conditional string  $s_1$ , its probability  $p(s_1)$ , and the conditional entropy  $h$  of the strings  $p(s_2|s_1)$  calculated from (6).

TABLE VII  
ENTROPY TABLE IN CASE OF THE TWO-KEY DESIGN FOR A BLOCK LENGTH  $N = 4$

| $s_1$        | 00P        | 11P        | P0P        | P1P        | P00 | P11 | 0P0 | 0P1        | 1P0 | 1P1        |
|--------------|------------|------------|------------|------------|-----|-----|-----|------------|-----|------------|
| $p(s_1)$     | 0.1        | 0.1        | 0.1        | 0.1        | 0.1 | 0.1 | 0.1 | 0.1        | 0.1 | 0.1        |
| $h(s_2 s_1)$ | $\log_2 2$ | $\log_2 2$ | $\log_2 2$ | $\log_2 2$ | 0   | 0   | 0   | $\log_2 2$ | 0   | $\log_2 2$ |

TABLE VIII  
ENTROPY TABLE IN CASE OF THE TWO-KEY DESIGN FOR A BLOCK LENGTH OF  $N = 5$

| $s_1$        | 00P0       | 00P1       | 11P0       | 11P1       | P0P0       | P0P1 | P1P0 | P1P1 | 0P11 | 1P00 | 1P11 |
|--------------|------------|------------|------------|------------|------------|------|------|------|------|------|------|
| $p(s_1)$     | 0.05       | 0.05       | 0.05       | 0.05       | 0.05       | 0.05 | 0.05 | 0.05 | 0.05 | 0.1  | 0.05 |
| $h(s_2 s_1)$ | 0          | 0          | 0          | 0          | 0          | 0    | 0    | 0    | 0    | 0    | 0    |
| $s_1$        | P00P       | P11P       | 0P0P       | 0P1P       | 1P1P       |      |      |      |      |      |      |
| $p(s_1)$     | 0.1        | 0.1        | 0.1        | 0.05       | 0.05       |      |      |      |      |      |      |
| $h(s_2 s_1)$ | $\log_2 2$ | $\log_2 2$ | $\log_2 2$ | $\log_2 2$ | $\log_2 2$ |      |      |      |      |      |      |

In Table VII, the conditional entropy values of P00 and P11 equal zero because the next OIE must be P to encode either A or B (Table I). The remaining zero entropy values are caused by the reduced grammar, which does not allow arbitrary pairs of letters. With the help of the compact entropy Table VII,  $h_{2K}(4)$  can be computed quickly from

$$h_{2K}(4) = 6 \times 0.1 \log_2 2 = 0.60. \quad (19)$$

The final calculation step is the determination of  $h_{2K}(5)$ . There is no need to consider larger block lengths because a full letter coding and the next input event are already known and therefore larger observation windows do not change the local predictability. Therefore, the limit of the block entropy  $h_{\mu 2K}(8)$  equals  $h_{2K}(5)$ . Using the same structure as Table VII, the values are given in Table VIII. From the third row of Table VIII, the conditional entropies of the input patterns are all zero because a coded letter A, B, C, or D and the last user-generated OIE determine the next event. For example, if the pattern 00P0 is analyzed, the prefix 00P encodes the letter B and the suffix 0 indicates that the next letter is either A or B. However, by the grammar in Table III, the letters A or D are only allowed to follow and therefore it must be A.  $h_{2K}(5)$  is derived as

$$\begin{aligned} h_{2K}(5) &= 3 \times 0.1 \log_2 2 + 2 \times 0.05 \log_2 2 \\ &= 0.4 \log_2 2 \\ &= 0.40. \end{aligned} \quad (20)$$

Finally, we are able to compute the complexity measure  $C_{2K}$  for the second keyboard design

$$\begin{aligned} C_{2K} &\approx (h_{2K}(1) - h_{2K}(5)) + (h_{2K}(2) - h_{2K}(5)) \\ &\quad + (h_{2K}(3) - h_{2K}(5)) + (h_{2K}(4) - h_{2K}(5)) \\ &\approx (1.59 - 0.40) + (0.95 - 0.40) \\ &\quad + (0.80 - 0.40) + (0.60 - 0.40) = 2.34 \text{ bits}. \end{aligned} \quad (21)$$

In conclusion, the interaction complexity when using the cheaper keyboard design with only two keys and the state-dependent input mode (21) is more than twice as large as the complexity of the four-key design (14). The considerable additional complexity of the cheaper design is due to the greater mutual information being communicated to the future (2.34 - 1 = 1.34 bits!), which has to be memorized by the pager user. This intuitive result is easy to generalize for designs with more keys, more complex grammars, and more hierarchical states for the indirect input mode.

### B. Complexity Estimation

When estimating the complexity measure  $C$  in (10) from a case base  $CB \equiv (s_1, \dots, s_d)^T$  instead of known grammars, the finite time horizon of the interaction has to be taken into account. Therefore, the block entropy limit  $h_\mu$  is approximated by the apparent block entropy  $h(n_{\max})$  of a maximum observation window length  $n_{\max}$ , which can be caused by length limitations

of observations (e.g.,  $\forall s : |s| \leq n_{\max}$ ).  $C$  is approximated by the partial sums

$$C \approx C(n_{\max}) = \sum_{n=1}^{n_{\max}-1} (h(n) - h(n_{\max})). \quad (22)$$

It would be useful from the perspective of human-machine system design if the estimation procedure not only gave an accurate estimate of  $C(n_{\max})$  but also included a prediction mechanism for OIE in future time slices. The prediction of OIE is of particular interest for the design of “anticipatory human-machine interfaces” [23], which are able to anticipate users’ actions in a task context and to provide appropriate decision support. Sheridan [26] calls a similar concept the “ultimate decision aid.” Fortunately, there is a procedure, the “Shannon guessing game” [25], which can fulfill these requirements. Shannon applied the knowledge of native speakers to put bounds on the entropy of written English, and his strategy made use of the predictability of OIE for alphabetic letters. Following the definition of the conditional entropy in (6), he showed  $M$  times randomly selected strings of length  $n$  to native speakers (as readers of written English), asked them to guess the next letter, and recorded how many guesses  $x$  were required before the answer was correct. The number of guesses can be considered as a random variable  $X_n$  with the empirical distribution  $\hat{p}_M(X_n = x)$ . The entropy of the distribution of guesses is an upper bound of the  $n$ th order block entropy [25]

$$\hat{h}(n) \equiv - \sum_x \hat{p}_M(X_n = x) \log_2 \hat{p}_M(X_n = x) \geq h(n). \quad (23)$$

Shannon also derived a lower bound; however, this paper only considers the upper bound because it is normally closer to the true value. Clearly, the design of anticipatory human-machine interfaces requires a prediction model instead of humans. Therefore, variable-length Markov chains (VLMCs, see [3]) are preferred. VLMCs generalize Markov chains of fixed order. VLMCs have a decision-tree-compatible memory structure called a context tree. The context tree is a graphical model that can be exploited for event prediction. VLMCs factorize the joint distribution of (2) as

$$P(O_1 = o_{j(1)}, \dots, O_T = o_{j(T)}) = \prod_{t=0}^{T-1} p_c(O_{t+1} = o_{j(t+1)} | f_c(O_1 = o_{j(1)}, \dots, O_t = o_{j(t)})). \quad (24)$$

In (24),  $f_c$  denotes a context function [3] mapping patterns  $s = o_{j(1)}, \dots, o_{j(t)}$  onto a prediction context  $c$ . The prediction context can be the full  $s$  if the memory depth of the model is sufficient and a shorter prediction context significantly reduces the prediction accuracy. If the memory depth of the model is insufficient, or a shorter prediction context does not reduce the accuracy, the context is a left-truncated substring of  $s$ . Each prediction context  $c$  is represented by a node in the context tree and has a predictive distribution  $p_c$ , which can be considered as the probability vector of dimension  $|\mathcal{O}|$ . The vector components simply denote the conditional probability of the next OIE

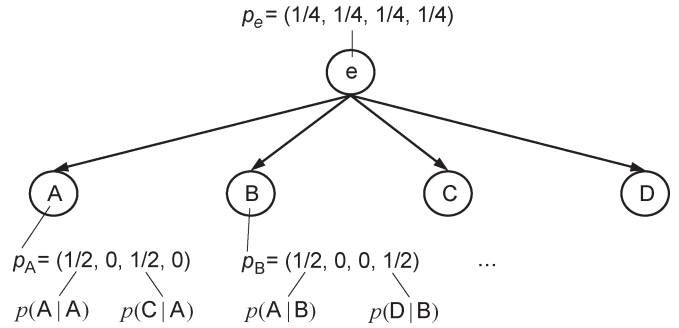


Fig. 2. Context tree example that is derived from the probability distributions in Tables II and III.

given the prediction context. The predictive distributions are the parameters of the model and the tree encodes its structure. In each time slice, the prediction procedure follows the maximum likelihood principle and the predicted OIE is the event with the maximum probability in the  $p_c$  vector component. The “e”-symbol represents the empty context  $\{\}$  and is the root node. Parent nodes in the tree denote suffix strings of their child nodes. The number of tree layers determines the maximum memory depth, e.g., if  $\hat{h}(n)$  has to be estimated on the basis of (23), the number of layers is bound to be a maximum of  $n$  layers.

An example context tree can be easily derived from the grammatical probability distributions of Tables II and III for the previous pager example. This tree is shown in Fig. 2. The example tree has five prediction contexts  $\{e, A, \dots, D\}$  on two layers. The prediction contexts on the first layer are derived from Table II. If the prediction context is initially empty, e.g., no OIE occurred, the probability of observing ‘A’ in the next time slice is 1/4 according to the first component of the probability vector of the root node, and so forth. If a one-digit context  $c = A$  is considered, the probability  $p(C|A)$  of observing C in the next time slice is 1/2 (Fig. 2), and so forth. These conditional probability values stem from Table III. The detailed context tree parameterization method is given in [3].

Now an estimate  $\hat{C}$  of the measure  $C$  can be derived. The estimation procedure is as follows.

- 1) The case base CB is split into parameterization cases and evaluation cases.  $M$  denotes the number of OIE in the evaluation cases.
- 2) The parameterization cases are used to estimate the structure and parameters of the VLMC prediction model (24) according to the parameterization method [3] with a maximum context tree depth of  $n = 1$  layer.
- 3) The parameterized VLMC model is used to predict  $M$  times OIE in the evaluation cases. For each predicted event, the number of guesses to find the correct answer is recorded. Because the context tree gives not only an event prediction but also a predictive distribution  $p_c$ , the number of guesses is equivalent to the rank of the next OIE in a list of probability mass values of descending order.
- 4) The empirical distribution of guesses  $\hat{p}_M(X_n = x)$  is computed and  $\hat{h}(n = 1)$  is estimated using (23). The result is pushed onto a stack of estimated block entropy values.

This procedure is repeated with incremented order  $n' = n + 1$  in step 3 until the estimated block entropy increases or the maximum length  $n_{\max}$  is reached. An increasing block entropy contradicts the postulate of a monotonic decreasing quantity (Section II-A) and the maximum order can be estimated as  $n_{\max} = n' - 1$ . Finally, the estimated values of  $\hat{h}(n)$  on the stack of block entropy values are inserted in (22), respectively. The final estimation formula for the complexity measure is simply

$$\hat{C}(n_{\max}) = (1 - n_{\max})\hat{h}(n_{\max}) + \sum_{n=1}^{n_{\max}-1} \hat{h}(n). \quad (25)$$

A computer algebraic analysis of the estimation procedure based on a Hidden-Markov-Model showed that the complexity estimates are conservative and always lower than the true values. Concerning interaction processes of medium complexity generating only short patterns of OIE (see Section III-F), the estimates were on average 35% lower. However, the complexity differences caused by systematic variations of interaction process dynamics were accurately estimated with average errors below 7%.

### III. VALIDATION STUDY

#### A. Methods

The goal of the validation study was to investigate the internal validity of the complexity measure  $C$ . Therefore, a laboratory experiment with electronic map displays was chosen because their human-centered design and evaluation are important for safe and effective work in many situations. The task was to search for multiple vessel symbols on the electronic map display within a limited time. In the literature on human factors, this task type is called a “search for multiple targets” [7], [29]. Although there is a large body of publications concerning the prediction of search times [7] and ergonomic design of display symbolism [18], complexity issues have not previously been studied. The display format was a computer-generated radar picture. Twenty valid vessel symbols without course and speed vectors were used as search targets and 17 vessel symbols with vectors were distractors. All 37 symbols were displayed concurrently. The ship representing the observer’s base was always at the center of the display. According to the standards for visual search experiments, after the start, successful recognitions of valid vessel symbols on the radar picture were indicated to the experimenter with a point-and-click on the display positions with the computer mouse. In addition, an eye movement analysis based on the cornea infrared light reflex method was used to acquire the visual scanpaths of the participants. The visual scanpath was measured by eye fixation in a narrow radial area around the vessel symbols.

#### B. Participants

Thirty vessel master mariners whose ages ranged from 35 to 65 ( $\bar{x} = 44.8$  years,  $SD = 9.3$  years) participated in the validation study. All the master mariners were male and had

more than 10 years of professional experience. They were all very familiar with reading electronic maps.

#### C. Apparatus

The electronic map display used the standardized green-on-black color scheme of a commercial off-the-shelf 21-in CRT monitor. The software was developed specifically for the test. The manually indicated search patterns were automatically written into the first case base  $CB_{\text{man}}$  by the display PC. The eye fixation patterns were measured by a head-mounted ASL 501 eye movement recorder [28]. The fixations were computed by the standard recorder algorithm and were written into the second case base  $CB_{\text{eye}}$ . The “compressed” fixation-recording mode [4] was used, which integrates multiple fixations on the same object. Due to calibration problems with certain eyeglasses, the scanpaths of three master mariners had to be excluded from  $CB_{\text{eye}}$ . Nevertheless, their manually indicated search patterns were satisfactory and were kept in  $CB_{\text{man}}$ .

#### D. Experimental Design

There were two dependent measures  $C_{\text{eye}}$  and  $C_{\text{man}}$ , representing complexity contributions of different functional stages of human information processing. We distinguished three stages of perception, cognition, and motor response [29]. In the given task, these stages are embedded in repetitive cycles of perception of a symbol on the map, cognition of the symbol meaning, and motor response on the map display with the mouse.  $C_{\text{eye}}$  represents the complexity contribution of the stage of visual perception and can be estimated using the eye fixation data in  $CB_{\text{eye}}$ . Conversely,  $C_{\text{man}}$  measures the interaction complexity of the joint three stages of human information processing and is represented by the correlations between the events in  $CB_{\text{man}}$ . Consistency of the theory requires that the complexity of visual perception in the given task setting is not greater than the complexity of the joint stages, because the human cognitive system memorizes previously recognized valid vessel symbols and therefore adds mutual information (see example in Section II-A).

There were three independent variables (factors) defining the experimental conditions. Because the participants were difficult to recruit, a repeated-measures design was used and each participant was measured under all conditions (factor levels). First, the effect of time pressure was studied, and the maximum time-on-task was varied to be either 10 or 20 s. The 20-s time limit represents a “natural” search frequency at an acceptable level of mental workload, whereas the time limit of half this length implies a high mental workload. Second, the effect of spontaneous mental abstractions was investigated. According to Hacker’s theory of hierarchical action regulation [11], humans are much better than machines when it comes to identifying hierarchical relations among features of observed objects on self-generated layers of abstraction. Three basic hierarchy-generating relations can be distinguished: spatial, temporal, and functional. Hierarchical relations in mental representations of a task generate medium and long-range intersymbol correlations and increase the interaction complexity. In the given search

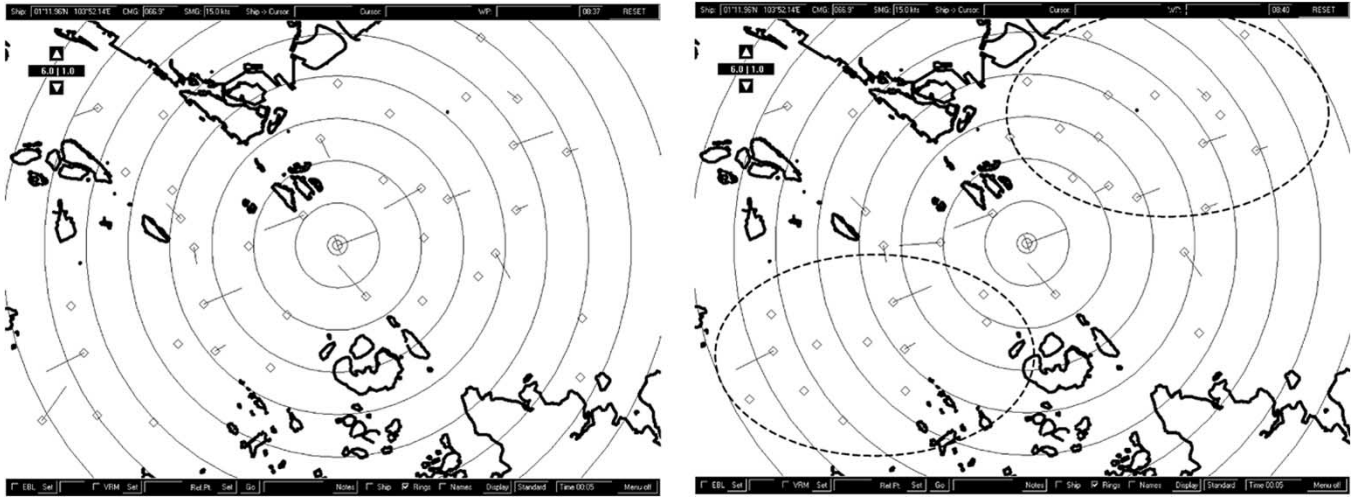


Fig. 3. Electronic radar map displays with zero (left) and two clusters (right) of vessel symbols (inverted images for greater print clarity).

task, spatial hierarchical relations were considered on two layers of abstraction. On the macrolevel, there are spatial clusters of vessel symbols on the map that often have a meaning of their own, e.g., a formation of ships. On the microlevel, a cluster can be mentally decomposed and single vessel features considered. In the validation study, display layouts with zero, one, and two clusters were investigated. Map examples with zero and two clusters are shown in Fig. 3. Third, the complexity influence of the map scenario was studied. The first map scenario covered the “Strait of Singapore” and the second scenario the Hong Kong area. The scenarios were, by design, of comparable complexity for experienced master mariners.

The inferential statistics are based on a linear model [22]

$$\hat{C}_{ijkl} = \mu + \beta_i + \gamma_j + \delta_k + \beta\gamma_{ij} + \beta\delta_{ik} + \gamma\delta_{jk} + \beta\gamma\delta_{ijk} + \varepsilon_{ijkl}. \quad (26)$$

In (26),  $\mu$  denotes the overall complexity mean and  $\varepsilon$  represents normal distributed random effects due to uncontrollable experimental conditions. The variables  $\beta_i$ ,  $\gamma_j$ , and  $\delta_k$  represent complexity effects of the levels of the introduced within-subjects factors “time-on-task” ( $i = 1, 2$ ), “clusters” ( $j = 0, 1, 2$ ), and “map scenario” ( $k = 1, 2$ ). The remaining four variables denote two-way and three-way interaction effects.

There were three hypotheses to be tested.

- 1) A significantly lower interaction complexity for the increased workload level due to shorter time-on-task was postulated because mental strain destroys event correlations and the learning curve  $\Lambda(n)$  is flattened. The null hypothesis  $H_0^1$  was  $\beta_1 = \beta_2$  [see (26)].
- 2) An increasing interaction complexity for an increasing number of clusters to be searched by the master mariners was postulated and  $H_0^2$  was  $\gamma_0 = \gamma_1 = \gamma_2$  (26).
- 3) An equal interaction complexity for both scenarios was postulated and  $H_0^3$  was  $\delta_1 = \delta_2$  (26).

For both dependent measures  $C_{eye}$  and  $C_{man}$ , repeated-measures analyses of variance (ANOVA) and the introduced three within-subject factors were the testing procedures. Wilks’ test statistic was used. The significance levels were  $\alpha = 0.05$ .

To obtain a better statistic, a fivefold cross-validation scheme with six replications was applied. This scheme divides the case bases into five partitions per replication; four partitions were selected when parameterizing the VLMC model and the fifth partition was used to evaluate the predictions. Then, four different partitions were selected, and so on, until all five combinations were computed and the next replication commenced.

### E. Procedure

The data collection procedure had three main steps per participant.

- 1) The master mariner had to read written instructions describing the search task.
- 2) Training trials covering the experimental conditions using a different map area were carried out.
- 3) In the measurement trials, the visual scanpaths and manually indicated search patterns were collected under all conditions ( $2 \times 3 \times 2$ ) and written into  $CB_{eye}$  and  $CB_{man}$ , respectively.

The presentation order of experimental conditions was balanced to avoid intervening effects.

### F. Results and Discussion

The complexity results for visual perception as well as joint stages of human information processing are shown as grouped boxplots in Fig. 4. In a boxplot, the bottom is the minimum, the bottom of the box is the estimated 25% percentile, the middle of the box represents the median, the top of the box is the estimated 75% percentile, and the top is the maximum.

As shown in Fig. 4, the complexity median of visual perception in both map scenarios is approximately 85% of the median of the overall complexity. The estimated 25% and 75% percentiles for visual perception show a similar behavior and are approximately 15% below the overall complexity percentiles. The large complexity contribution of the perceptual stage is consistent and plausible because the search task is clearly dominated by the perceptual abilities of the master

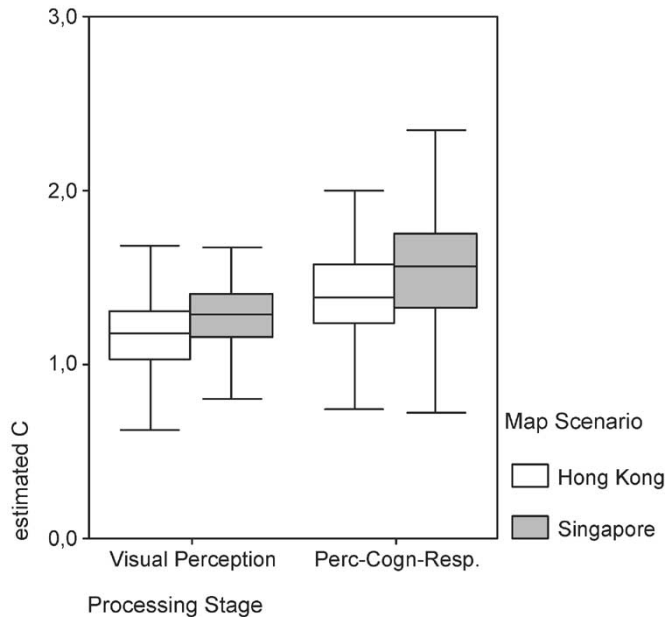


Fig. 4. Complexity measure for visual perception based on eye fixations and for joint stages of perception, cognition, and motor response ( $n_{\text{man}} = 30$ ,  $n_{\text{eye}} = 27$  subjects).

mariners involved. In the cognitive stage, the participants only have to memorize already discovered symbols and have to distinguish between valid symbols and distractors. Furthermore, the distance between the 25% and 75% percentiles indicates a considerably higher standard deviation for the joint stages of human information processing, which is also plausible because of a larger dispersion of cognitive abilities and search styles. Obviously, inconsistencies of the new complexity measure did not occur.

The repeated-measures ANOVA of the complexity  $C_{\text{man}}$  of the joint processing stages shows a three-way interaction effect  $\beta\delta\gamma_{ijk}$  between cluster (CLUST), time-on-task (TOT), and scenario (SCEN) that does not significantly differ from zero ( $F(2, 28) = 0.459$ ,  $p_{\text{SCEN} \cdot \text{CLUST} \cdot \text{TOT}} \leq 0.636$ ). Therefore, from (26), the summation of the variables is a reasonable assumption. Furthermore, the two-way interaction effects involving the time-on-task ( $\beta\gamma_{ij}$  and  $\beta\delta_{ik}$ ) are not significant ( $F(2, 28) = 0.108$ ,  $p_{\text{CLUST} \cdot \text{TOT}} \leq 0.898$ ;  $F(1, 29) = 0.002$ ,  $p_{\text{SCEN} \cdot \text{TOT}} \leq 0.967$ ) and the complexity difference, caused by a systematic variation of time pressure ( $F(1, 29) = 45.370$ ,  $p_{\text{TOT}} < 0.0005$ ), is not biased. Obviously,  $p_{\text{TOT}}$  is below the 5% level of significance and the first null hypothesis  $H_0^1$  is rejected. The lack of significant factor interactions concerning the time-on-task is also demonstrated in the profile diagram of Fig. 5.

From the ANOVA results and Fig. 5, it is concluded that the mental strain due to time pressure destroys intersymbol correlations in human information processing and the interaction complexity under high workload is significantly lower than for normal workload. In other words, the data demonstrate a significant workload-complexity tradeoff: the higher the mental workload, the lower the complexity generated by visual perception, cognition, and motor response. There is an additional significant complexity effect of the number of

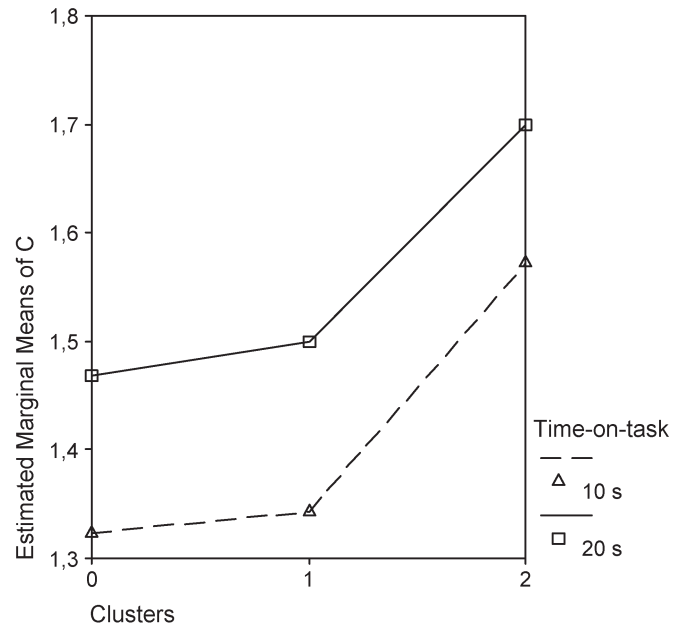


Fig. 5. Profile diagram of the overall search complexity with perception, cognition, and motor response concerning the time-on-task and the number of clusters in the electronic map ( $n_{\text{man}} = 30$  subjects).

clusters on the map display ( $F(2, 28) = 21.794$ ,  $p_{\text{CLUST}} < 0.0005$ ) and the second null hypothesis  $H_0^2$  is rejected. The trend of the estimated marginal means in Fig. 5 is in accord with the second hypothesis and shows increasing complexity means for an increasing number of clusters. Therefore, the validation data showed the complexity increasing effects of spontaneous mental abstractions due to spatial hierarchical relations between recognized vessel symbols.

The ANOVA also shows a significant factor interaction effect between the number of clusters and the map scenario ( $F(2, 28) = 8.831$ ,  $p_{\text{SCEN} \cdot \text{CLUST}} \leq 0.001$ ). Fortunately, the factor interaction effect is consistent because the ranking of the estimated marginal means of  $C$  is independent of the map scenarios and there is merely a complexity leverage effect from the first (Hong Kong) to the second (Singapore) scenario. As described in Section III-D, the complexity differences of both map scenarios were not anticipated and the null hypothesis  $H_0^3$  is rejected ( $F(1, 29) = 32.319$ ,  $p_{\text{SCEN}} < 0.0005$ ). Complexity in the semantic sense of vessel navigation seems to differ from the syntactical assessment. Nevertheless, the mean relative complexity difference between both scenarios is below 15%, and this small discrepancy may be considered as semantically indistinguishable.

In the repeated-measures ANOVA of the interaction complexity  $C_{\text{eye}}$  of visual perception, there is a significant three-way interaction effect ( $F(2, 28) = 9.781$ ,  $p_{\text{SCEN} \cdot \text{CLUST} \cdot \text{TOT}} < 0.001$ ) and the simple summation of variables (26) is not a reasonable assumption. This strong factor interaction is mainly due to complexity differences of the scenario and must be expected when using eye fixation measurements because fixation patterns are strongly guided by local features of the electronic map, such as luminance gradients, contrast differences, and suchlike [1]. The further analysis strategy [22] was to split the sample data by the map scenarios and to compute

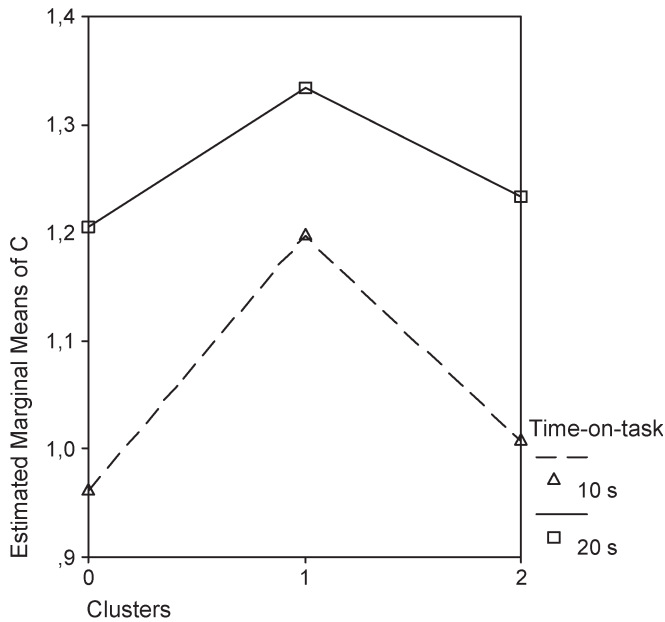


Fig. 6. Profile diagram of complexity of visual perception based on eye fixation measurements ( $n_{eye} = 27$  subjects).

separate two-way ANOVAs with the remaining factors. These separate analyses showed a two-way factor interaction effect for the Hong Kong scenario significantly different from zero ( $F(2, 28) = 17.191$ ,  $p_{CLUST \cdot TOT} < 0.0005$ ); instead, in the Singapore scenario, significant factor interaction effects did not occur ( $F(2, 28) = 1.402$ ,  $p_{CLUST \cdot TOT} \leq 0.263$ ). Therefore, only the sample data of the Singapore scenario were considered. After data splitting, the complexity differences of the time-on-task ( $F(1, 29) = 65.576$ ,  $p_{TOT} < 0.0005$ ) as well as the number of clusters ( $F(2, 28) = 20.626$ ,  $p_{CLUST} < 0.0005$ ) are both significant at the required 5% level. The null hypotheses  $H_0^1$  and  $H_0^2$  can be rejected as above. The effect is stronger for the time-on-task variation. The profile diagram is shown in Fig. 6. With regard to this figure, the significant complexity difference of visual perception is again in the postulated direction: the higher the mental workload, the lower the complexity of eye fixation patterns.

It is interesting to see a stronger workload effect in the perceptual stage than in joint stages (Fig. 5 compared with Fig. 6). Obviously, the task-induced mental strain affecting the perceptual stage is smoothed in the cognition and motor response stages. This workload effect on visual perception was also validated using the mean saccade length as a classic psychophysiological variable. From May *et al.* [17] and Nakayama and Takahashi [19], it must be expected that a higher mental workload level correlates with lower mean saccade lengths. Clearly, the null hypothesis to be tested postulated equal mean saccade lengths for the 10-s and 20-s levels of the time-on-task. This hypothesis was tested by the well-known  $t$ -test for dependent samples.

The eye fixation measurements for the 10-s time-on-task had a mean saccade length of 3.89 cm (SD = 0.79 cm). On the other hand, for the 20-s condition, the mean was 4.06 cm (SD = 0.52 cm). The computed test statistic is above the

critical  $t$ -value and the null hypothesis is rejected. The mean saccade length also indicates a significantly higher workload level for the halved time-on-task.

It is also interesting to find the inverse V-shaped complexity curve of clusters in the display layout (Fig. 6). Under both workload conditions, a layout with one cluster, rather than two clusters, shows a maximum complexity of visual perception. The visual appearance of the single cluster layout with a local concentration of 50% of the search targets in one map sector leads to a higher perceptual attention than the zero and two cluster layouts with a more spread distribution of targets across the whole display. It can be concluded that spontaneous mental abstractions due to spatial hierarchical relations are a phenomenon of cognition (cf. Fig. 5) because the most strongly (bipartite) cluster display layout shows a maximum interaction complexity in the experimental setting with a cognitive processing stage only. From the experimental data, the visual perception stage does not significantly contribute to a bipartite hierarchical action regulation.

#### IV. CONCLUSION AND FUTURE RESEARCH

This paper presented a novel information-theoretic approach to measuring the self-generated complexity of human-machine interaction. In the validation study of the theory, the task was to search for multiple vessel symbols in electronic radar map displays, and 30 experienced master mariners participated. The task allowed a straightforward measurement of manually indicated search patterns. In addition, perceptive search patterns were measured by eye fixations. The mental workload due to time pressure, the number of clusters in the display layout, and the map scenario were varied systematically. The results demonstrated a significantly decreased complexity under high time pressure at a significance level of 5% for visual perception as well as for the joint processing stages. In both cases, the number of clusters had a significant effect at the same significance level. However, the highest average complexity was measured for two clusters only in case of joint stages of human information processing. From Hacker's theory of hierarchical action regulation [11], it can be concluded that this effect was not due to the novel measure but is a well-known phenomenon of spontaneous spatial mental abstraction. Finally, the map scenario had significant complexity effects that were not anticipated in the design, but the mean complexity differences were below a level of 15% and therefore are difficult to distinguish semantically.

Due to length limitations, we were unable to present the results of a second validation study [24] in the same application domain with 30 additional participants. The second study was carried out in a motion-based simulator. The platform motion and the ambient light conditions were varied systematically. The results showed a significantly lower interaction complexity when the simulator was put in sea state characteristics and the participants were facing considerable motion forces. Furthermore, significantly lower complexity values were measured if ambient light intensity was reduced from daylight to twilight condition. Under the daylight condition, the average complexity reduction due to platform motion was 10%, and under the

twilight condition, there was a stronger average reduction of 17%. In summary, the novel measure seems to be a valid and consistent quantity for complexity assessment. The complexity measure and the estimation procedure have three main advantages compared with the previous information-theoretic approaches. First, OIEs are the only information source and subjective (and possibly unreliable) estimates of the channel capacity of human cognition under different display or workload conditions are not required. Second, the information difficulty of the machine-supported task in terms of memory demands and hierarchical mental representations is considered. Third, a prediction model of users' behavior is an integral part of the estimation procedure, which can be readily exploited for in-depth behavioral analysis, or for the design of anticipatory user interfaces.

Future research will focus on additional validation studies and will pay special attention to differences between experts and novices. The experiments conducted were of relatively low complexity and only meaningful for the expert's domain of visual search tasks with electronic map displays where mental workloads due to time pressure, display layout, motion forces, and ambient light conditions are critical issues. In the future, we will also study complexity effects in concurrent task settings with novice operators, for example, the supervisory control of semiautonomous multirobot systems [27] or swarms of self-organized autonomous agents [15]. These inherently concurrent systems are very difficult for a human to control and it is desirable to develop multimodal user interfaces that may achieve a better balance between visual, auditory, and tactile information channels. One can hypothesize that multimodal user interfaces reduce the workload of a human's visual system and therefore allow significantly increasing interaction complexity.

## REFERENCES

- [1] J. L. Barbur, P. M. Forsyth, and D. S. Wooding, "Eye movements and search performance," in *Visual Search*, vol. 1. New York: Taylor & Francis, 1993, pp. 253–264.
- [2] W. Bialek, I. Nemenman, and N. Tishby, "Predictability, complexity and learning," *Neural Comput.*, vol. 13, no. 3, pp. 2409–2463, 2001.
- [3] P. Bühlmann and A. J. Wyner, "Variable length Markov chains," *Ann. Stat.*, vol. 27, no. 1, pp. 480–513, 1999.
- [4] Y. S. Choi, A. D. Mosley, and L. W. Stark, "String editing analysis of human visual search," *Optom. Vis. Sci.*, vol. 72, no. 7, pp. 439–451, Jul. 1995.
- [5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [6] J. P. Crutchfield and D. P. Feldman. (2001). *Regularities Unseen, Randomness Observed: Levels of Entropy Convergence*. Santa Fe Institute Working Paper 01-02-012 2001. [Online]. Available: <http://arXiv.org/abs/cond-mat/0102181>
- [7] C. G. Drury and S. K. Hong, "Generalizing from single target search to multiple target search," *Theor. Issues Ergon. Sci.*, vol. 1, no. 4, pp. 303–314, Oct. 2000.
- [8] W. Ebeling, J. Freund, and F. Schweitzer, *Komplexe Strukturen: Entropie und Information* (in German). Stuttgart, Germany: B.G. Teubner, 1998.
- [9] M. R. Endsley, B. Bolte, and D. G. Jones, *Designing for Situation Awareness*. New York: Taylor & Francis, 2003.
- [10] P. Grassberger, "Toward a quantitative theory of self-generated complexity," *Int. J. Theor. Phys.*, vol. 25, no. 9, pp. 907–938, 1986.
- [11] W. Hacker, *Allgemeine Arbeitspsychologie—Psychische Regulation von Arbeitstätigkeiten*, 2nd ed. (in German). Bern, Switzerland: Verlag Hans Huber, 1998.
- [12] H. G. Kang and P. H. Seong, "Information theoretic approach to man-machine interface complexity evaluation," *IEEE Trans. Syst., Man Cybern. A, Syst., Humans*, vol. 31, no. 3, pp. 163–171, May 2001.
- [13] W. Karwowski, "Achieving compatibility in design and evaluation," in *The Human-Computer Interaction Handbook*, J. A. Jacko and A. Sears, Eds. Mahwah, NJ: Lawrence Erlbaum, 2003, pp. 1227–1238.
- [14] D. Kieras and P. G. Polsen, "An approach to the formal analysis of user complexity," *Int. J. Human-Comput. Stud.*, vol. 51, no. 2, pp. 405–434, Aug. 1999.
- [15] J. D. Lee, "Emerging challenges in cognitive ergonomics: Managing swarms of self-organized agent-based automation," *Theor. Issues Ergon. Sci.*, vol. 2, no. 3, pp. 238–250, 2002.
- [16] M. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 1995.
- [17] J. G. May, R. S. Kennedy, M. C. Williams, W. P. Dunlap, and J. R. Brannan, "Eye movement indices of mental workload," *Acta Psychologica*, vol. 75, no. 1, pp. 75–89, Oct. 1990.
- [18] F. Motz, H. Widdel, P. Oei, S. MacKinnon, and L. Alexander, "Investigations for ergonomic presentation of AIS symbols on ECDIS," *Int. Hydrogr. Rev.*, vol. 5, no. 2, pp. 26–36, 2004.
- [19] N. Nakayama and K. Takahashi, "The act of task difficulty and eye-movement frequency for the 'oculo-motor indices'," in *Proc. Eye Tracking Research & Applications Symp.*, New Orleans, LA, 2002, pp. 37–42.
- [20] M. Rauterberg and R. Aeppli, "How to measure the behavioral and cognitive complexity of learning processes in man-machine systems," in *Proc. World Conf. Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA)*, Boston, MA, 1996, pp. 581–586.
- [21] M. Rauterberg, S. Schlupe, and M. Fjeld, "Modeling of cognitive complexity with Petri nets: An action theoretical approach," in *Proc. Cybernetics and Systems (EMCSR)*, Vienna, Austria, 1998, pp. 842–847.
- [22] L. Sachs, *Angewandte Statistik*, 9th ed. (in German). Berlin, Germany: Springer-Verlag, 1999.
- [23] C. M. Schlick, C. Winkelholz, F. Motz, A. Künzer, and H. Luczak, "Stochastic operator models for multiple target search tasks," in *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, Hammamet, Tunisia, 2002, pp. 121–126.
- [24] C. M. Schlick, C. Winkelholz, F. Motz, S. MacKinnon, and A. Patterson, "A complexity study of human-machine interaction on motion platforms," in *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, The Hague, The Netherlands, 2004, pp. 82–87.
- [25] C. E. Shannon, "Prediction and entropy of printed English," *Bell Syst. Tech. J.*, vol. 30, no. 1, pp. 50–64, 1951.
- [26] T. B. Sheridan, *Humans and Automation: System Design and Research Issues*. New York: Wiley, 2002.
- [27] B. Trouvain and C. M. Schlick, "A study of audio and visual context switch indicators in a multirobot navigation task," in *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, The Hague, The Netherlands, 2004, pp. 2821–2826.
- [28] A. E. Welchman and J. M. Harris, "Task demands and binocular eye movements," *J. Vis.*, vol. 3, no. 11, pp. 817–830, 2003.
- [29] C. D. Wickens and J. G. Hollands, *Engineering Psychology and Human Performance*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2000.



**Christopher M. Schlick** (M'05) received the Dipl.-Ing. degree from the Berlin University of Technology, Berlin, Germany, in 1992, and the Dr.-Ing. and Habilitation degrees from the RWTH Aachen University of Technology, Aachen, Germany, in 1999 and 2004, respectively.

From 1992 to 1993, he was a Design Engineer at Krone Group. From 1994 to 1997, he was a Research Assistant and from 1998 to 2000 an Assistant Professor (Oberingenieur) at the Institute of Industrial Engineering and Ergonomics, RWTH Aachen. From 2001 to 2004, he was the Head of the Human Factors Engineering Department at the German Research Establishment for Applied Science FGAN. He is currently a Full Professor at the RWTH Aachen Faculty of Mechanical Engineering, where he is the Director of the Institute of Industrial Engineering and Ergonomics. His research interests include the design and simulation of work and business systems, human-machine systems, human resource management, and ergonomics.

Dr. Schlick received merits of honor from the German Human Factors Society GfA, RWTH (Borchers insignia), and the Holste Foundation (Holste Price 2004).



**Carsten Winkelholz** received the Dipl.-Phys. degree in physics from TH Karlsruhe University of Technology, Germany, in 1996, and is currently working toward the Dr.-Ing. degree at RWTH Aachen University of Technology, Aachen, Germany.

From 1996 to 1998, he was a Research Assistant at the Institute of Applied Computer Science, Kassel University, Kassel, Germany. Since 1998, he has been with the German Research Establishment for Applied Science FGAN, Germany, where he is a Research Scientist. His research interests include

cognitive modeling and human factors in virtual environments.



**Florian Motz** received the Dipl.-Ing. degree in mechanical engineering from RWTH Aachen University of Technology, Aachen, Germany, in 1992.

Since 1992, he has been with the German Research Establishment for Applied Science FGAN, Germany, where he is a Senior Research Scientist. He is responsible for projects related to the display and integration of navigational information on ship bridges.

Mr. Motz is a member of IEC TC 80 working group 13 "Presentation of navigation related Information" and serves as Coordinator of an IMO correspondence group dealing with the harmonization of navigation related information on ship bridges.



**Holger Luczak** received the Dipl.-Ing, Dr.-Ing., and Habilitation degrees from Darmstadt University of Technology, Darmstadt, Germany, in 1969, 1974 and 1977, respectively.

From 1969 to 1974, he was a Research Assistant and from 1974 to 1977 an Assistant Professor (Oberingenieur) at the Institute of Ergonomics, Darmstadt University. From 1977 to 1983, he was a Full Professor of Production Engineering at Bremen University of Technology, Bremen, Germany. He founded the Faculty of Production Engineering and

the Bremen Institute of Industrial Engineering and Applied Ergonomics. From 1983 to 1992, he was a Full Professor of Industrial Engineering and Ergonomics at the Faculty of Mechanical Engineering, Berlin University of Technology, Berlin, Germany. From 1992 to 2005, he was a Full Professor at the Faculty of Mechanical Engineering, RWTH Aachen University of Technology, where he was the Director of the Institute of Industrial Engineering and Ergonomics as well as the Managing Director of the Research Institute for Operations Management.

Dr. Luczak is a member of editorial boards of national and international journals in ergonomics and is an advisory and supervisory council member on several national and international committees.