DEPARTMENT OF DEFENCE
DEFENCE SCIENCE & TECHNOLOGY ORGANISATION DSTO

Protocol Analysis as a Tool in Function and Task Analysis

Catherine Lees, Jeremy Manton and Tom Triggs

DSTO-TR-0883

Protocol Analysis as a Tool in Function and Task Analysis

Catherine Lees, Jeremy Manton and Tom Triggs

Information Technology Division
Electronics and Surveillance Research Laboratory

DSTO-TR-0883

ABSTRACT

A protocol is a record of procedural steps undertaken in a process. In studying human-machine systems, observations of human operators obtained from sources such as videotapes are coded to create a descriptive protocol of behaviours, task elements, goals, etc. This record is essentially sequential in nature, but methods for analysing sequential data are relatively new. The kinds of information that protocol analysis can provide, that might be useful in function/task analysis, are examined. Methods for analysing sequences are surveyed, and recent developments in using minimum message length methods for producing probabilistic finite state automaton models of sequential behaviour are discussed.

RELEASE LIMITATION

Approved for public release

DEPARTMENT OF DEFENCE
DEFENCE SCIENCE & TECHNOLOGY ORGANISATION
DSTO

Published by

DSTO Electronics and Surveillance Research Laboratory
PO Box 1500
Salisbury South Australia 5108 Australia

Telephone: (08) 8259 5555
Fax: (08) 8259 6567

AR-011-105
October 1999

APPROVED FOR PUBLIC RELEASE

Protocol Analysis as a Tool in Function and Task Analysis

Foreword

This research paper was drafted in late 1992 while Catherine Lees was on contract to the Human Factors Group of Air Operations Division, Melbourne. Not long after the draft was submitted two authors departed for different parts of the world. The draft has now been resurrected and brought up to date as a record of work completed. The topic is still current in DSTO and the paper is a contribution to this ongoing field of research.

Executive Summary

Human factors researchers seek to understand how operators interact with their equipment in order to design and develop better and more capable military systems. Capturing and representing these transactions for existing and simulated systems can inform the design of future systems and reduce the risk that systems in development will meet operational requirements. Modern techniques enable researchers to make extensive recordings of human activities in the operation of complex systems. Coding such records and using them to develop understanding and structure for function and task analysis represents a significant challenge.

This report examines the requirement for function and task analyses in human factors evaluations of systems in some detail, considers the various types of measures that can be obtained in recording operational performance and behaviour, and pays particular attention to the sequential character of operational behaviours.

A principal goal of the report is to provide an overview of two general approaches to the study of sequential behaviours of operators. First, there are those methods that examine the immediate sequential and temporal dependencies of behaviours. Second, there is a smaller set of techniques that extends the analysis so as to take account of functional hierarchies in the analysis of behaviour, where this aspect is often essential for a detailed human factors analysis of a complex system. A recent sequential analysis technique based on minimum message length encoding that has recently been applied to human behaviour is recommended as having significant potential for such human factors analysis.

A requirement in the human factors evaluation of a system is the recording of the behaviour of system operators. Some of these records are straightforward quantitative measures that vary over time and yield a time-series. Other measures have to be provided with additional labels (or "symbolic codes") or details to provide useful time-varying data. In general, the essential quality of such records is their change over time, or 'sequentially', and a complete sequential record will include the categories of behaviours, the sequence in which they occur, and the times of transitions between behaviours.

Sequential records will differ when the same set of tasks is carried out in response to similar events on more than one occasion. However, behaviour falls well short of being random. Despite the variability, there is typically an organization or structure underlying the operational behaviour. One can address this organization of behaviour in two ways. The first is concerned with the interdependencies of behavioural activities, external variables, and time-varying factors. The second approach addresses the issue of behaviour as a hierarchical set of functions. This latter approach is of central relevance to human factors systems analysis.

Where one is concerned with the analysis of interdependencies of behaviours, a further distinction can be made between those techniques focussed on studying the sequential properties of behaviours (or their sequential association) and those aimed at understanding how behaviours tend to form groups in which several behaviours play similar roles. This latter case is concerned with the degree of embeddedness, where behaviour from the same group will be embedded in the sequence of behaviours in similar ways. The study of embeddedness is likely to be less relevant for technical, procedural environments and so this report is only concerned with the first of these, the sequential association of behaviours.

In addition to providing a means of testing hypotheses, the identification of consistency in the sequences does provide directly useful information, even if what causes the consistency is not known. Knowledge of the sequential order of component actions will provide information about what options are available to operators. Such sequence information can also provide an input to task network simulation models of human operators in systems, such as MicroSAINT.

When a record of behaviour is coded into a protocol, the continuous stream of behaviour is broken up into segments that are given names or codes. The relationships between the various codes form a coding syntax. There are two recent methods available for making the coding syntax explicit. The protocol analysis package SHAPA uses a coding syntax, which is based on the PROLOG programming language. Another package is CABER, which requires that the analyst develop a syntax for the behaviours under study. The program then parses the coded input in terms of the project-specific syntax. The syntax can be changed, and the iterative development of the syntax is part of the analysis process. CABER has potential for analyzing behaviour in real time, once an input syntax has been designed. This may be useful in system training applications and simulator training. Taken overall, CABER appears to provide several analytical tools that are likely to be of potential use in the study of behaviour sequences in operational systems.

Authors

Catherine D. Lees
University of Western Australia

After completing a BSc(Hons) and several years in the Australian Public Service, Catherine Lees was awarded her PhD in Psychology from Monash University in 1988. She undertook post-doctoral research on a 3-year National Research Fellowship at the Department of Administrative Studies at Monash, followed by a year as a visiting lecturer in the Department of Psychology, and research work on a project conducted at the then Aeronautical Research Laboratories, Melbourne. In 1992 she joined the Department of Management at the University of Western Australia, later to become the Department of Organisational and Labour Studies, where she now works as a lecturer. Current research interests are in aspects of human resources management and organisational behaviour, including, individual, group and team decision making, the analysis and design of work, and research methods.

Jeremy Manton
Land Operations Division

Jeremy Manton completed his PhD in Experimental Psychology and started working on sonar display systems at the Royal Australian Navy Research Laboratory in Sydney. He was attached to the Admiralty Marine Technology Establishment in the UK in the early 80s working on combat information system design for surface ships. On his return he took up a position in the Human Factors Group at the Aeronautical Research Laboratory and later headed the Group. He developed research in the field of aircrew-avionics interactions in areas such as aircrew heat stress in the Sea King helicopter, the Seahawk operational test, the P-3C Orion Update and the ALR 2002 Radar Warning Receiver programs. He recently returned from Washington, where he was the Attache Defence Science, and is now the Research Leader for Human Factors in Land Operations Division. He is the National Leader for the Technical Cooperation Programme in the area of Human Resources and Performance.

Tom Triggs
Monash Universitry

Professor Triggs is academic Director, Department of Psychology, Clayton campus at Monash University in Melbourne and Director of the Monash University Accident Research Centre. He is a fellow of the Human Factors society of Australia. His research interests are in the areas of human decision making, human factors in transportation and attention processes. He serves on the editorial boards of several international journals, and is a member of several sub-committees of the U.S. Transportation Research Board.

Contents

1. Introduction

The aim of the Human Factors evaluation of a system is to ensure that the physical and procedural elements of the system provide the necessary *functionality* to enable the human elements to accomplish their mission.

The mission itself is a global, over-arching function or goal, which can be broken down into component functions, and these also can be broken down successively into narrower, more specific functions until, as Meister (1985) notes, "At a certain level of detail - which is difficult to specify - the function shades almost imperceptibly into a task" (Meister, 1985, p.20). Tasks also can be further broken down successively into their component elements.

The first step in the Human Factors study of a system is the description of the hierarchy of functions and tasks to some practical degree of detail. At the finest level of detail, which is the task element, the syntax of the description is the same as it is at the highest level, the mission description. It contains a verb, an object, and usually a qualifier, which makes the content of the description specific enough to be useful. For example, at the lowest level of the hierarchy for an Air Traffic Controller (ATC) a task element might be,

> "SEARCH *Situation Display* for information pertaining to potential penetration of special use airspace" (Phillips and Melville, 1988, p.41).

The verb is 'search', the object is 'situation display', and the remainder is the qualifier, without which the terms search and situation display would be far too general to be of use.

At the highest level of the function hierarchy is the ATC's overall goal or mission which can be expressed as follows,

> Maintain a safe, orderly, and expeditious flow of air traffic.

Here the verb is maintain, the object is the flow of air traffic, and the qualifier is safe, orderly, and expeditious.

These descriptions of functions at many levels are of the same kind, but differ in scope and complexity at different levels of the hierarchy. The levels are usually given names to make it easy to refer to them, but the number of levels and the names used will differ among different authors and different projects. For example, in the comprehensive study of air traffic control reported in the U.S. Federal Aviation Administration's Air Traffic Control Operations Concepts (Ammerman, Fairhurst, Hostetler, and Jones, 1988), the following six levels of description were used: Mission, Function, Activities, Sub-activities, Tasks and Task elements.

The higher levels of description, perhaps the first three levels of the preceding list taken together, are generically called "functions". These upper levels are basically hierarchical. A diagram of the relationships between them may give the appearance of a flow chart, and it may be called a function flow diagram, but as Ammerman et al have pointed out, such a diagram is not a flow chart of operations, it represents state transition possibilities rather than strict dynamic control flow.

At some point, perhaps around sub-activities in the preceding list, we go from levels generically called "functions" to levels that can generically be called "tasks". At these lower levels relationships are less hierarchical and more sequential. For example, a particular sub-activity usually represents the operator's initial response to some event that has occurred. It is the first step among equal steps, which, together, comprise the response to the event, it is not some overall level that encompasses the subsequent steps within it.

2. Event-Based Task Analysis

It is at the lower level of generic tasks that the heavy-duty undertaking of task analysis begins. The technique used to produce descriptions of the function/task statements now takes a change of direction, because at this level tasks are goal-directed responses to events. For example, in the human factors project to define the FAA's Air Traffic Control Operations Concepts, (Change 1) (Ammerman, Fairhurst, Hostetler, and Jones, 1988) in support of the Advanced Automation System development, a fundamental assumption was,

> "Controllers, and other system users may be characterized as event-sensitive; that is, acting generally in response to or anticipation of Air Traffic events rather than initiating action independently. The term "event" here encompasses both actual occurrences such as status changes, and predicted occurrences such as a predicted aircraft conflict" (p. 1-2).

An event is an actual or predicted occurrence that impinges on the operator directly, or intersects with the operator's responsibilities. Where operator behaviour can be regarded as being event-driven, the task analysis can be approached by first defining a set of system events, which cause responses from the operator. Continuing with the Air Traffic Control example, Ammerman et al list 102 events applicable to Air Traffic Control, some examples of which are given below.

AIRCRAFT-AIRCRAFT CONFLICT: This is the most critical event in air traffic control. The controller may detect the potential conflict or may receive a system-generated message alert that two aircraft are in conflict.

BOMB THREAT: An aircraft that is under the duress of a bomb threat will convey this information to ATC.

BALLOON, GLIDER: Balloons (both manned and unmanned) and gliders represent non-controlled objects of which the controller must maintain awareness.

The event definitions simply state the events that may impinge upon the operator, and they note where these events intersect with the operator's responsibilities, but they do not specify what responses should be made to the events.

Tasks are derived by stating the responses that should be made to the specified system events. All tasks are derived in this event-based approach by taking each event and working out, from all possible sources of information, what the operator has to do to respond to the event. Ammerman et al provide a good set of guidelines for task derivation in an Air Traffic Control context (Volume 1, Section 3).

Tasks themselves can be decomposed into task elements and it is possible for the task elements to occur in more than one task, and a task also can be used to respond to more than one event. It is therefore possible to draw up an event-task matrix, or an event-sub-activity matrix, depending on whatever level it is that responds directly to an event. The rows of the matrix are labelled with events, and the columns are labelled with the responding action unit, sub-activities for example.

The value of such a matrix is that it provides a way of tracing from system events to the system functionality at the operator's disposal for responding to those events, via the response action (sub-activity or task unit).The recognition of the event component, extra to the many levels of functions and tasks, is an important aid in the human factors analysis of a system. As Phillips and Melville (1988) point out, "The advantage of deriving operator tasks from system events is that the task inventory will be as complete a characterization of the job as the original event list" (Phillips and Melville, 1988, p. 38).

3. Task Characterisation

At the lowest level, the task element, a number of other types of description can be added. These include the criticality of the action, the cognitive and sensory attributes necessary for accomplishing the action, such as pattern recognition, and deductive reasoning, the identities of co-ordinatees where the action involves communication as in transmit or receive, and the required performance criteria, such as allowable accuracy and completion time limits.

The information used to construct these function and task descriptions is obtained from many sources, including system specification and requirement documents, analysis and observation of existing systems which are to be replaced or altered, and interviews with subject matter experts (SMEs) such as experienced operators or instructors.

It is possible, in some cases, to define all the functions and tasks involved in accomplishing a mission for a particular system without reference to the external operating environment of the system. For example, an aircraft might proceed along its flight path as planned, receiving all ATC clearances as requested, and weather as forecast, from take-off to landing. If this occurs, and there are no system alarms or failures on board, it would be possible to write a moderately accurate script describing the task-relevant actions of the crew, in advance. But, for most systems the events that impinge upon them from the external environment do not occur in a pre-ordained manner or at a pre-ordained time. Indeed, many systems are designed to interact with environments that are complex and somewhat unpredictable.

When an event occurs that impinges on the operator, either directly as when viewed through a window, or indirectly as through a symbol on a display, it sets in train a sequence of response actions which, in itself and in its interactions with other activities, is not entirely predictable, because neither the impinging event, nor its time of occurrence is entirely predictable.

The analysis to this stage is what Meister (1985) would call "task description", but which has been commonly called "task analysis" by more experimentally oriented authors such as Ericsson and Simon (1984), Card, Moran, and Newell (1983), and Sanderson, Verhage, and Fuld (1989). For them task analysis is an abstract, theoretical activity, that involves analysing the task set for the subject or operator, but does not involve analysing how the subject or operator performs the task. They conduct the analysis in the same way as might be done for the purpose of designing a machine to carry out the task. The importance of such task analysis will be mentioned again in a later section, here however, it is being distinguished from what Meister (1985) calls task analysis, or the study of how people actually go about doing the task, which Card, Moran, and Newell (1983) have called "methods analysis".

The various definitions of task analysis are distinguished for clarification, but the differences should not be overstated. Just as Meister acknowledges, as quoted at the beginning, that functions shade almost imperceptibly into tasks, so Card, Moran, and Newell have noted the following,

> "As analysis becomes more and more dependent on system structure, task analysis turns into method analysis. Task analysis reflects more the demands of the external environment, whereas method analysis reflects more the demands of the computer system and the ways in which the user adapts to them. There is, of course, no sharp line between task analysis and method analysis." (Card, Moran, and Newell, 1983, p.420)

4. Analysing Functionality versus Functional Analysis

Although a full description of the many levels of functions and tasks is an important step in human factors analysis it will not of itself guarantee that the system provides adequate *functionality* to support the operator in accomplishing tasks. The functional analysis approach implies that, in order to evaluate the quality of functionality of a system, you begin by looking at the functions and then look to see how those functions are used in operations. For an event-based approach, on the other hand, once having described the functions above the task level, you begin looking at system events, which are part of the operations, you define and decompose the tasks that respond to those events, and then jump across to see what functionality in the system level specification or requirements, or the existing system itself, will support the execution of those tasks. For the description and analysis of functions and tasks to be complete and practical from an operational point of view, it is necessary for actual operations to be studied, or at least expert descriptions of such operations.

To do the Human Factors evaluation of a system, in addition to the methods of task analysis referred to in preceding sections, which characterise the task in terms of what the operator is going to need to do in order to carry it out, we also must look at our video and audio records from the point of view of using them to describe what the operator actually does and how. This is the essential knowledge in a Human Factors evaluation.

5. Recording Operational Behaviour

To study the behaviour of people who are interacting with a system, or with a simulation of the system, the behaviour is usually recorded. Recorded behaviour can include video and audio records, keystroke and manual control action records (e.g. pedal presses, joystick movements), eye movements and gaze direction measures, physiological measures such as evoked potentials and heart rate variability, and subjective measures such as subjective workload ratings or "think aloud" descriptions of thoughts.

Some of these records provide straightforward quantitative measures, which, when considered over time, form a time series, eg. heart rate variability and subjective workload ratings. Others, such as gaze direction, require some symbolic code or name to be given to the behaviours over time. For example, for gaze direction, the directions themselves are usually not meaningful without reference to something in the visual field that is being gazed at. Associated with the gaze record, therefore, there must be a code with the names of objects at which the gaze is directed, before further analysis of the behaviour can be carried out.

Other records, such as "think aloud" protocols and video and audio recordings, also require an interpretative code to enable analysis of the behaviour to be carried out. The various activities of a subject must be put into categories, and the categories given names, so that the activities in each category can be enumerated, have their durations timed, and perhaps be assigned some other measures such as degree of intensity, accuracy, or skill.

The coded sequential record of behaviour is the *result* of putting together information from various measures such as those mentioned earlier, keystroke records, video and audio records, "think aloud" protocols, and comments and insights from Subject Matter Experts. A complete sequential record will include the categories of behaviours, the sequence in which they occur, and the times of transitions between behaviours.

Assigning activities to categories, or in other words, giving behaviours consistent names (codes), enables a written description to be made of the sequence of behavioural activities, as they proceed and follow each other in time.

6. Non-Sequential Measures: Task Time and Task Frequency

The two types of measures that are most straightforward to make in an analysis of a sequential record of behaviour, are task execution time, and task frequency.

6.1 Task execution time

Task execution time, or task duration, is a measure of how long it takes an operator to complete a task or a task element, or whatever level unit of the function/task hierarchy is being measured. This is the most essential measure for any time-line based approach, which might be used, for example, to model operator performance (McMillan, Beevis, Salas, Strub, Sutton, and van Breda, 1989) or to assess workload (Meister, 1985) at some point in an operational scenario.

To get a descriptive measure of task duration from a sequential record, it is necessary to have a coded record that includes the times of transitions between behaviours. These may be available as time-stamps on video or audio tapes, and in some computerised sequential behaviour analysis applications, such as CABER the transition times may be noted automatically for the sequential record from the electronic time stamps (from discussion with Jon Patrick, 14 October 1991). In some applications noting transition times may have to be done manually, by observing the tape time clock and entering the value in the protocol being coded, or by manual timing from some starting point.

It is desirable to have time of occurrence data for all activities included in the protocol. For example, where an operator is making a Subjective Workload Assessment Technique (SWAT) rating response at fixed intervals, perhaps approximately every five minutes, it would be valuable to know exactly when the response was made, so that the rating could be related to other ongoing activities at the time.

From the transition times the task duration can be worked out by subtracting the time at the beginning of task execution from the time at the end of task execution. This can be done manually but it is convenient to have an analysis application such as CABER or SHAPA (James and Sanderson, 1990) that will filter out all the instances of a particular task behaviour, from the coded protocol, collect them together, and calculate statistics such as mean execution time.

If a task is interrupted and resumed and brought to closure later in the sequence, a decision must be made about whether to include the full running time from task commencement to closure, or to subtract the period of the interruption(s). First it must be established that the interruption does not in any way contribute to the completion of the task, for example, by providing the operator with additional needed information, or by providing extra time in which to think about the task, or plan a solution to a problem. If it is clear that the interruption is not in any way related to the task being timed, then for a time-line approach the duration of the interruption should be subtracted from the task time. Without subtracting, the sum of times for all the individual tasks in a sequence may be greater than the actual duration of the whole sequence. [1]

The obvious descriptive statistics for task execution time are the mean and standard deviation, taken over all the occurrences of a particular task behaviour in a person's record, and, if available, over more than one person's record. It is unlikely in protocol analysis that there would be sufficient instances of a particular task behaviour to study the shape of the distribution of its duration. A method for modelling the factors affecting duration as a dependent variable, called event history analysis, is discussed in the section on temporal durations of behaviour.

6.2 Task frequency

The other straightforward measure that can be made from a sequential record of behaviour is a simple count of the frequency with which a particular task occurs. This frequency will depend a great deal on the frequencies of the events that happen to the operator and which trigger task execution as a response. Ancillary tasks, which can be regarded as housekeeping, such as resetting calibrations and scales on displays, tidying flight strips (in air Traffic Control), etc., and which by definition are not event-driven, need not be dependent on the frequency of impinging events, but can be, in so far as time available for ancillary tasks might diminish as event frequency increases.

Simple task frequencies have two main, and related, uses. Firstly, they can be used to elaborate the task characterisation described earlier. Elaboration of description can be added both on the side of what the operator is required to do, and on the side of what the operator actually does. In the case of ancillary tasks the frequencies provide important information necessary for any time-line approach or performance model.

In the case of event-driven tasks the frequencies in part depend on the frequencies of the events themselves. It should be noted here that in addition to frequencies of task behaviours, an observational record obtained in a "live" operational setting is an important source of information regarding the frequencies of events themselves. Events should be entered in the protocol, and they should be counted and timed so that their rate of occurrence can be calculated. In so far as tasks are triggered by events and are responses to them, event frequencies are more basic information than task frequencies.

Furthermore, for an event-based approach involving study of a single operator, live operations, or historical records of, and expert commentaries on, live operations, are the only valid source of data on event frequencies, and therefore, of data on the frequencies of tasks that are triggered by those events. In a simulation study with a single operator the events impinging on the operator are entirely determined by the scenario that has been prepared for simulation, and therefore, indirectly, so are the tasks. This is not to say that a simulator is not a useful way to collect information about the tasks an operator employs to respond to events, but only that the actual frequencies of events, and the tasks they trigger, cannot be determined in a simulation. This will also be the case where the object of study is a team or crew operating in a simulator. However, in this case those events that happen to one crew member and which are initiated as actions by another crew member can be discovered, as they are only partly determined by the events that are imposed on the simulation by the scenario.

Even for event-driven tasks, task and task element frequencies are only partly determined by event frequencies. This is because it can be possible to have more than one optional way to respond to an event. These optional tasks are analogous to Card, Moran, and Newell's (1983) "Methods" in the Goals, Operators, Methods, and Selection Rules (GOMS) model of operator performance in human-computer interaction. In the GOMS model the Methods are alternative means for carrying out an operation and the Selection Rules are used to decide which method to employ depending on the circumstances. For the text editing task used in Card, Moran, and Newell's experiments there was typically more than one available method, for example, to move the cursor to a particular point in the text. If the cursor was close to the target position, just the arrow keys might be used to re-position it. If the cursor was pages away from the target position, scrolling pages followed by the arrow keys, or issuing a search for target command, would be used.

In Air Traffic Control, procedures and regulations limit the availability of alternative methods. Some freedom of action does remain, however, and Air Traffic Controllers themselves have said that different controllers will produce different solutions for separating aircraft in a particular conflict situation (from discussions with Melbourne

Sector 3 Area Controllers). Thus, although in a simulator study events are determined, and so are the tasks they necessarily trigger, where there are optional tasks or task elements, the operator's use of alternative methods is free to vary, even in a simulator, and so can be studied.

The frequencies of tasks driven from events, therefore, elaborate the description of what the operator is required to do, and the frequencies of optional tasks and task elements elaborate the description of what the operator actually does. These optional frequencies can be entered in the task descriptions, and descriptions of task modules (task modules are descriptions of sets of the task elements that comprise a particular task).

The second way in which simple task frequencies can be used is when specific behaviours of interest need to be counted. For example, in studying the communications between an Air Traffic Controller and others such as pilots, other controllers and advisory positions, the frequencies of communications with each position could be extracted from the coded protocol. If only one operational position is being studied these simple frequencies could provide a useful elaboration of the description of the operator's job in terms of the intensity of communication links; flow of information, receipt of requests and issuing of instructions with other positions.

Where more than one operator is being studied, for example a crew, or a number of components of a system such as the ATCs and pilots mentioned above (Kerns, 1990), the simple frequencies of interactions between the operating positions will not provide a clear picture. For example, almost all the information received by one position might emanate from one other position suggesting a very close link, yet the issuing position might be involved in much more communication with other positions, so that communication with the receiving position only constitutes a small part of its activities. Furthermore, attending to those communications might be important for the receiving position, or it might constitute only a small part of the receiving position's activities, which might not be predominantly centred around communication. So, in a multi-operator system simple frequencies of behaviours need to be considered in relation to the amount of that behaviour in the overall system, and also in relation to the amount of total behaviours in the overall system.

7. Sequential Measures

Simple task execution times and task frequencies are fundamental forms of description in the task analysis part of a Human Factors evaluation of a system, but they are static descriptions. They do not convey or describe any of the dynamic properties of behaviour, which are supposed, somewhat tautologically, to be entailed in a sequential record of behaviour. By definition a sequential record is dynamic, simply because it is a record of behaviour as it unfolds in time. But is it importantly dynamic? Can an analysis of the essential sequentiality of the record reveal important properties of behaviour over and above those that can be discovered with the static descriptive measures described above, or those that can be discovered in traditional experimental designs with static measures such as performance, or more economically than those traditional experimental methods allow?

A traditional laboratory experiment can be used to test whether task execution time depends on many factors such as the experience of the operator, task difficulty, sleep deprivation, number of competing tasks, etc., but it would not usually be used to test whether task execution time depends on the execution time of the preceding task, on the identity of the preceding task, or the period of time elapsed since the last occurrence of the task. (It should be noted that Gregson (1983) has suggested that the traditional controlled laboratory experiment can sometimes profitably be regarded as a time series.)

7.1 Simple sequence

The simplest and most obvious form of essentially sequential information, which can be obtained from a sequential record but not from a traditional static laboratory experiment, is a description of the order in which the tasks or task elements that comprise the execution of a higher level task unit, are carried out. If these elements are carried out in invariant order with a recognised closing element that signals accomplishment of the task, they form a task module and may be relatively easy to detect, code, time, and count in a protocol even though they may occasionally be interrupted during execution and resumed later.

Note that the event-based nature of the analysis is an important aid to identifying behaviour that is aimed at satisfying a particular task. The string of task elements that together constitute the response to an event must be able to be linked directly to satisfying that event, even though the response may be delayed and temporarily put on hold due to more urgent activities. This link to events, together with the need for a specifiable closing element that signals completion of the response to an event, give a concreteness to descriptions at the generic task level. For more abstract functions that are higher in the function/task hierarchy it may be more difficult to identify behaviour that is aimed at satisfying a specific high level goal. Observable task behaviours will not necessarily be nested neatly under individual higher level functions, but may serve different functions on different occasions. The identification of behaviour satisfying abstract goals will be discussed in a later section.

7.2 Sequential analysis

Usually, the sequential record will be *slightly* different when the same set of tasks is carried out in response to similar events on more than one occasion, whether by the same person or not. The sequential record will not be *radically* different on different occasions because both the task and the available system functionality will impose constraints on how the task can be successfully accomplished.

Thus, sequential records will differ, but behaviour is not random. In the midst of this behavioural variability there is organisation or structure to be understood. Apart from the constraints imposed by the task, the sequence and timing of behaviours will have structure, some of which is ascribable to the organising activity of the human generating the behaviour, some to causal and interactive relationships with external events and context variables, and some to the available system functionality.

Techniques for studying the sequencing of behaviour in complex tasks have been reviewed by Ericsson and Simon (1984) for verbal problem solving, Gottman and Roy (1990), Bakeman and Gottman (1997), and van Hooff (1982) for social interaction in animals and people, and Sanderson, James, and Seidler (1989), Sanderson (1991), and Sanderson and Fisher (1997) for "Exploratory Sequential Data Analysis (ESDA)" of operator behaviour, among others. Sequential data has also been called "narrative data" and analysed in the context of sports by Patrick and Chong (1991), and Patrick and McKenna (1986).

Van Hooff (1982) has divided the study of the organisation of behaviour into two basic directions. One direction is concerned with the interdependencies of behavioural activities with each other, with external factors, and with time-dependent factors. The other basic direction is concerned with behaviour as an hierarchical set of functions, with activity satisfying one function being embedded within activity that is aimed at satisfying another function. In the case of human factors systems analysis the functional hierarchy approach is taken as the convention, has a formal function/task specification (which might not exactly parallel the human operator's functional hierarchy), and defines the context of further analysis. A considerable problem arises in trying to fit the analysis of sequential records conducted in the style of the first direction, concerned with the sequential and temporal dependencies of the behavioural acts, into the framework of the functional hierarchy approach of the second direction. What follows is an attempt to address this problem. Techniques for the analysis of sequences of behaviour are reviewed and discussed, both in regard to their ability to assist in the description of behaviour, and to provide assistance in building predictive models of behaviour.

8. Analysis of Sequential Records of Behaviour

Van Hooff (1982) made a further distinction between analysis techniques aimed at studying the sequential *association* of behaviours, which emphasise the notion of sequence, and techniques aimed at studying the *embeddedness* of behaviours, which show how naturally occurring behaviours tend to form groups in which several behaviours play similar roles and can be used in place of each other. The term "embeddedness" used by van Hooff (1982) means that behaviours that come from the same group and serve much the same purpose will be embedded in the ongoing sequence of behaviours in much the same way as each other: they will be similar in their patterns of transitions to and from other behaviours. The redundancy in naturally occurring behaviour such as social behaviour makes the study of embeddedness particularly appropriate. In a more formal, procedural environment, such as an aircraft cockpit, where an operator is interacting with technical tools, there may be little potential redundancy for task execution. (NB exception being the use of different 'methods' to achieve the same goal under different conditions, as in Card, Moran, and Newell's (1983) GOMS model of human-computer interaction). That is, the available methods for achieving a particular goal, or carrying out a particular operation, may be quite specific within the technology, and for a particular situation, and afford no alternative choices.

Furthermore, where some minor sequence of activities forms a group in the sense that, when all are completed they accomplish some higher-level activity or goal, it is not the case that each sub-activity will be embedded in the total sequence in a similar way. For example, let the sub-sequence of activities A, B, C, D, be a group, which, when completed, accomplishes some goal, and the sequence can be interrupted and resumed where it was left off. Each of the activities A, B, C, and D, will have a different profile of transitions to and from other behaviours, including those in the sub-sequence itself. D will often follow C immediately or shortly thereafter, but B will not follow C, except in the case of error or back-tracking to refresh memory. Thus, procedures that look for similarity in the profile of transitions to and from other behaviours would not treat these behaviours as a group. For this reason, techniques used specifically to study embeddedness, such as principal-component analysis, factor analysis, and cluster analysis, will not be considered here. As a point of departure, however, progress is being made in methods for the analysis of embedded hierarchical structures of behaviour, of which Neville-Manning and Witten (1997) provide an example.

The sequential association of behaviours can be studied in two ways. These are the analysis of the structure of the sequential record itself, and the analysis of factors affecting the structure of the sequential record. The distinction is somewhat artificially drawn for the purpose of exposition, as the two are often inextricably linked and the same techniques can be used to conduct the analysis in both cases.

One way to analyse sequences is to look just at the behaviours themselves in one long sequence, ignoring the context in which they occur, and which may be changing during the sequence. In the analysis of a behaviour protocol coded, for example, from a videotape of operator behaviour, the frequencies of the various behaviours might be counted and descriptive statistics such as mean duration of each behaviour calculated, as has been described in previous sections. These are common summarising statistics made available in computerised sequential data analysis packages (Bakeman and Quera, 1995; Hetrick, Isenhart, Taylor, and Sandman, 1991; James and Sanderson, 1990; Noldus, 1991; Patrick and McKenna, 1986). The duration times and frequencies of behaviours are useful, for example, in predicting how long it will take an operator in future to carry out similar sets of behaviours or tasks.

Transition frequencies can also be calculated and tabulated. This is the frequency with which a particular behaviour follows immediately after one or more other specified behaviours, in the sequence. A transition matrix is a special kind of contingency table, in which both the rows and the columns are lists of the same possible behaviours. The cells give the frequency with which the behaviour listed for the column follows immediately after the behaviour listed for the row, that is, the frequency of transitions from the row behaviour to the column behaviour.

Traditionally, the sequence has been examined for internal sequential association by first analysing it as a Markov chain.

8.1 Markov analysis

A Markov chain is a sequence of elements in which the probability of occurrence of a particular element at some point in the sequence depends, in part, on the identity of the element(s) that precede(s) it. That is, the probability of occurrence of an element in the sequence is not independent of the part of the sequence that immediately precedes it. The sequence has structure located in the transitions from one element to the next. This is not the only kind of internal structure a sequence can have, but the problem of detecting other structures will be discussed later.

In examining operator behaviour for the purpose of function/task analysis of a system, it is unlikely that the operator will perform actions in a random order. The completion of a mission may require that certain tasks be performed some number of times, and many of them must also be performed in a particular order, e.g. an aircraft must become airborne before it can be landed. Some constraints on the order of activities are inherent in the task, and these should be discovered as far as possible from an analysis of the task carried out before study of the human operator performing the task. Writers such as Newell and Simon (1972), Card, Moran, and Newell (1983), Ericsson and Simon (1984), and Sanderson, Verhage, and Fuld (1989) have all stressed the need for a priori analysis of what performance of a task entails for the operator. Sanderson, Verhage, and Fuld (1989), for example, advocated analysis of the system dynamics, a state-space analysis of the operator's control task, and an analysis of what knowledge of the system a human operator could possibly have or acquire. While such a thorough a priori task analysis is feasible in a closed system such as a Tower of Hanoi problem, a text-editing human-computer interaction task, or an experimental system in a laboratory, it is less feasible in an open system interacting with the external environment, such as an aircraft cockpit or an air traffic control centre. In these latter situations the situational context in which the operator is required to respond to events and carry out tasks will be slightly different on every occasion. One benefit of analysing recordings of operators behaving in actual operations is the possibility of identifying ways in which operators adapt their task performance, for example, the order in which they carry out sub-activities, to cope with this situational variability.

Thus, while an operator does not perform tasks in random order, due at least to the constraints of the task, nor does successful accomplishment of the mission necessarily entail that the operator will complete the component tasks in a fixed order. Some of the Markov structure in a behaviour sequence may be due to task constraints and some may be due to an operator's strategy. It may be difficult to make this distinction for an open system. For this reason Markov analysis of behaviour sequences may be best suited to those aspects of operator behaviour that are largely self-generated and self-controlled, and which are not closely constrained by task requirements. Moray and Rotenberg (1989), for example, used a Markov approach to analyse eye movements as indicators of shifts of attention between four sub-systems in an experimental water bath control task. Moray and Rotenberg (1989) employed a particular aspect of Markov analysis, the mean first passage time, which is the mean number of sequence elements or steps traversed before reaching a particular state, from some other state. If both starting and stopping states are the same, the mean first passage time is called the recurrence time (Kemeny, Mirkil, Snell, and Thompson, 1959). Analysing the sequence of gaze directions, Moray and Rotenberg (1989) used mean recurrence times to indicate the return of the subjects' attention to a particular sub-system, the shorter the mean recurrence time, the more attention was being paid to that sub-system.

Mean first passage time is not the aspect of Markov chain analysis most usually employed in studying sequences of behaviour. Where Markov analysis has been used most extensively, in the study of social behaviour (Gottman and Roy, 1990), it has been employed to look for internal dependencies within the sequence, and to note how these dependencies differ in the presence of different external context factors (i.e. experimental treatment conditions). As a simple example, the behaviour of one individual interacting with another may depend on whether the immediately preceding behaviour of the other has been friendly or aggressive. The extent and nature of this dependency may differ with factors such as education, gender, whether one person is an adult and the other a child, where one is a therapist and the other is a client in some stage of treatment, where one is a supervisor and the other a subordinate, etc.. In these examples the behaviours of the participants under study are free to vary at will, and where they have sequential structure it is self-imposed. This is also the case with the gaze directions of Moray and Rotenberg's (1989) subjects, who were free to look at whichever sub-system they chose. The same applies to control room tasks that are mainly of a monitoring or vigilance kind. In an aircraft cockpit or an air traffic control centre, the gaze direction of the pilot, or air traffic controller, carrying out certain tasks would also be free to move around in order to monitor various sources of information. Yet, other tasks such as selecting radar targets with a track-ball, or keying-in course information, would be necessary rather than discretionary components of task execution. For these kinds of tasks, which are the main concern of this report, Markov dependency in the sequence of behaviours may be as much due to the constraints of the task as to the strategy of the operator. The question of the need to distinguish sequential structure that is inherent in the task, from that which is not, will be discussed later.

Gottman and Roy (1990) have described two basic steps in Markov analysis of sequential behavioural data. The first step they refer to as "fitting the timetable", and the second step involves seeing how the "timetable" changes as an effect of experimental factors.

"Fitting the timetable" involves determining whether there is any Markov dependency in the sequence, and finding the order of that dependency. It is a first-order dependency if the probability of a particular element depends in part on the previous element, it is second order if that probability depends on the identities of the previous two elements, and so on.

Just knowing the order of the sequential dependency in a sequence is not very informative. As Gottman and Roy (1990) have pointed out, in research there is usually an experimental design, and in sequential analysis we want to see if the likelihood of occurrence of a specific sequence varies with the experimental factors. If there is Markov dependency in the sequence, and if it is "stationary", which means that it is fairly stable throughout the sequence, log-linear analysis can be used to study how factors in the experimental design affect the contents of the "timetable", the Markov sequential structure of the behaviours.

Log-linear analysis involves testing the fit of models, similar to the structural models of analysis of variance, to the data in contingency tables, where the cell entries are frequency counts. Log-linear analysis is used extensively in fields of study where frequentistic data are common, such as political science, sociology, and market research. It has developed in recent years because the extensive computations required can now be carried out by computer (Wickens, 1989).

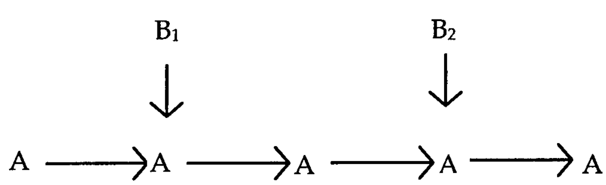
Log-linear analysis is a statistical method for contingency tables of frequency data in general, and has nothing in particular to do with sequential data. Because the transition matrices used to describe sequential data are contingency tables (albeit of a special kind) log-linear methods have been applied to them. Gottman and Roy (1990) and Bakeman and Gottman (1997 and 1986) give the history of this development. It should be noted, however, that, unlike the standard contingency tables in which entries in one cell are sampled in such a way as to be independent of those in another cell (apart from factor effects), for transition matrices there is no such independence. The entries in the cells are the frequencies of digrams of behavioural elements, or tri-grams in the case of a three-way transition matrix representing second order dependency. These digrams share behavioural elements in common, so their frequencies are not independent. Gottman and Roy refer to Monte Carlo studies by Bakeman and Dorval (1988) (cited in Gottman and Roy, 1990, p.109) carried out to test the impact of violating this assumption of sampling independence. They concluded that the effect of the violation of this important theoretical assumption is inconsequential in practice, and the application of contingency table statistical methods to transition matrices was still recommended.

Gottman and Roy (1990) provide a good text on the application of both log-linear analysis, and the related logistic regression, to transition matrices of sequential data. Log-linear analysis models the frequency of observations in cells to study the association between variables that define the contingency table. Logistic regression, on the other hand, can be used to treat one of the variables as a dependent variable, while the others are regarded as independent variables. These methods are aimed at accounting for variation in frequency data and are analogous to the analysis of variance or multiple regression methods used for continuous data. In the specific context of sequential analysis the frequency data are cell entries in a transition matrix giving the frequency with which a particular, consequent behaviour, follows one or more preceding behaviours. This "timetable" is embedded as a cell itself in a larger contextual design (of treatment variables), and we can examine how the transition frequencies depend significantly on the values of variables in the larger design. For example, we might show that the extent to which one task behaviour follows contingently upon another (a dependency within the Markov timetable), depends significantly on whether the operator is an expert or a novice (a dependency in the contextual design), and that this effect interacts with the cumulative work done in the test session.

8.2 Autocontingency

The use of log-linear and logistic regression methods to analyse sequential data seems appealing, and is strongly advocated by writers such as Gottman and Roy. However, it is necessary to mention, as they do, some concerns regarding the valid interpretation of the results of these analytic methods. The violation of the assumption of independent sampling has already been mentioned and discounted as unimportant. A further obstacle exists, however, in the problem of autocontingency in sequences of behaviour.

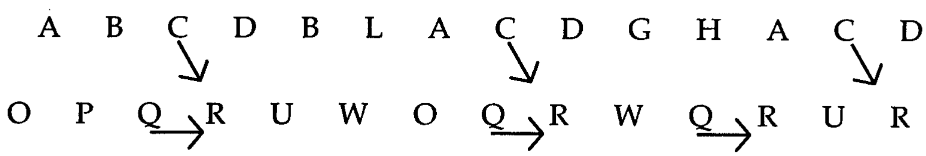
As the name suggests, autocontingency refers to the dependence or contingency of an individual's behaviour on that same individual's preceding behaviour. Quite apart from the events that are impinging upon the individual operator, or the situational context, an individual's purposive behaviour will include many internal dependencies. For example, we might want to know whether the identity of variable A, perhaps the method chosen to carry out a particular task, depends on the identity of variable B, but behaviour A forms a time series, and is itself enmeshed in a sequential record of other behaviours on which it may be dependent.



Variable B might be an independently controlled context variable, or a sampled context variable also measured over time, or it might be part of the same sequential record, perhaps another behaviour by the same operator. Because behaviour A is a sequence in time, its successive values cannot be assumed to be independent, and indeed, they will usually be related to each other. Because A is influencing itself, it makes it difficult to draw conclusions about the effect that B is having on A.

In some research the effect of one individual's behaviour on that of another is studied. This is common in social research, for example, where the question of interest might be the dependence of a child's behaviour on the behaviour of its mother, or the dependence of the behaviour of a political leader on the behaviour of an opponent. In human factors, group communication is often an object of study. Even where communication is not explicit, in team work, the activities of one team member may be thought to depend on the activities of another. In these situations it is of interest to see how the behaviours in a sequence for one individual depend on actions which are themselves part of another person's (or group's) behaviour sequence. Autocontingency, which almost certainly exists within the respective behaviour sequences, can invalidate conclusions about contingency between the sequences.

For example, let the two following sequences of letters represent coded behaviours for two people talking to each other and taking turns to speak. The top row represents one person's behaviour and the bottom row represents the other person's behaviour.



In the first part of the sequence it appears that the second person's R behaviour may depend on the first person's C behaviour (a cross contingency) but it also depends on the second person's Q behaviour (an auto contingency). Later on in the sequences it appears that autocontingency may be the better explanation because R occurs following Q in the second person's sequence but in the absence of C in the first person's sequence. Later again in the sequence it appears that dependence on C in the first person's sequence may account for at least some of the occurrences of R in the second person's sequence.

There are numerous methods proposed to deal with the problem of autocontingency, and these have been discussed by Gottman and Roy (1990). They generally involve regarding dependence within the individual's sequence as the more fundamental form of dependence and the more parsimonious explanation of dependence. This is partialled out in some way before considering the contribution of cross-contingent dependence. This procedure would become complicated if we wished to consider a team of people interacting with each other.

Furthermore, in the case of two or more sequences that are not obviously related, such as the monthly population of kangaroos in Australia over the last five years and the monthly public popularity rating of the Prime Minister over the same period, it makes sense to regard autocontingent dependence as the more fundamental explanation of predictability in a sequence, rather than cross-contingency. However, where individuals are interacting with each other directly, and taking part together in a larger evolving context, such as an aircraft in flight, it is not obvious why dependence that could be regarded as either autocontingent or cross-contingent, should necessarily be regarded as the former rather than the latter. The behaviours of both people may be part of a sequence of responses the group must make to an aspect of the external situation.

8.3 Amount of data required

An additional obstacle to using a Markov based analysis of sequential dependency on coded records of operational behaviour is the amount of data required by these methods. Because these methods are based on the analysis of contingency tables, an average expected cell frequency of at least some figure between 5 and 10 is recommended (Gottman and Roy, 1990, p.170; Wickens, 1989, p.30). Thus, if behaviour is to be coded into R different types, for a first order transition matrix with R columns and R rows, there are R^2 cells, and at least $5R^2$ observations are required. For a second order matrix there are R^3 cells, and at least $5R^3$ observations are required. So, if there are 10 behavioural categories, R=10, which would not be an unusually large number in protocol analysis, a second order matrix would require at least $5(10^3) = 5000$ observations, or recorded instances of behaviours taking place.

Overall insufficiency of data is a separate problem to the issue of dealing with empty cells that occur either as structural zeros, where a particular combination of categories cannot logically occur, or as sampling zeros, by chance. Gottman and Roy (1990, p.220) point out that modern logit and log-linear methods can handle the occurrence of these zeros. However, this does not mean that an inadequacy of data in the table overall can be ignored.

When the required number of observations is not available from one sequential record, data can be pooled from a number of records. For example, data can be pooled from different subjects carrying out the same tasks, or from different records or sections of records in which the same subject (or unit such as a team) has carried out the tasks of interest more than once. It is necessary, however, to test these records for homogeneity before pooling them. Records are homogeneous if they exhibit the same kind of Markov dependency. This will not necessarily be the case, just as a record from one subject will not necessarily exhibit stationarity, that is, have the same Markov structure throughout its length.

In summary, the Markov approach to sequential structure, and the log-linear and logistic regression approaches to determining how that Markovian structure depends on external factors, while being appropriate forms of analysis for sequential records of behaviour, have a number of associated difficulties.

- It is necessary to test the stationarity of the Markov structure. The analysis can only be interpreted meaningfully for a section of the sequential record over which it is stationary.

- If sequential records are pooled it is necessary to test for homogeneity. Again, interpretation is not meaningful if the analysis is applied to a set of records that are not homogeneous.

- If more than 4 or 5 behavioural categories are used in coding the behavioural record, large numbers of observations (i.e. instances of a behavioural category occurring in the sequential record) are required.

9. Information Theoretic Approach

Another approach to determining what influences the frequencies of behaviours can be found in information statistics, which have also been applied to sequential behavioural records, particularly in studying communication (Van den Bercken and Cools, 1980). Krippendorff (1986) has presented a description of how information theory can be used for structural modelling of qualitative data. Krippendorff claims that the information theory approach is more elegant than the log-linear approach (Krippendorff, 1986, p.92), and provides greater analytic power. This approach has not been considered here because it deals only with qualitative data, and does not include the capability of dealing with continuous predictor variables, which is available through logistic regression.

10. Lag Sequential Analysis

Gottman and Roy (1990, p.100) have said, "Lag sequence analysis is a trick to get around the problem of not having enough data for a complete Markov analysis of second or third order". It does not provide the complete analysis of sequential relationships afforded by Markov analysis, but it is more practical and has been incorporated into software packages for observational data such as SHAPA (Version 2.0) (James and Sanderson, 1990) and SATS (Yoder and Tapp, 1990). Faraone and Dorfman (1987) state that this method is a form of exploratory data analysis.

Lag sequential analysis involves testing the significance of dependence of the occurrence of a target behaviour at some specified number of observations removed from some other key behaviour. Various statistics for measuring the dependence have been suggested and these have been discussed by James and Sanderson (1990), Yoder and Tapp (1990), Gottman and Roy (1990) and by Faraone and Dorfman (1987), who concentrate on the problem of distinguishing cross-dependence from auto-dependence.

It is possible to test the significance of dependence between any two behaviours at any number of steps separation, and thus to build up profiles of how several behaviours follow after some other specific behaviour. It is important to note, however, as Gottman and Roy (1990, p.100) point out, that this is not a complete picture of the dependence relationships in the sequence, of the kind that a Markov analysis would be, because only dependence on single events is taken into account, not dependence on pairs of events or triples, etc.

When there are numerous behavioural categories in the code the number of key and target behaviour pairs that might be examined at various lagged positions with respect to each other becomes considerable. For example, if there are 10 behavioural categories there are 90 ordered pairs of behavioural categories that might form the key and target behaviours, just for lag position 1. This must then be multiplied by the number of lag steps that are to be investigated. If say, 4 lag steps are all considered relevant, and all 360 analyses are carried out, interpreting the pattern of results even for those which turn out to involve significant dependence could be difficult. Again the number of categories that would typically be used in a behavioural record for Human Factors purposes poses a problem.

Thus, although lag sequential analysis has been specifically recommended as a tool for exploratory data analysis (Faraone and Dorfman, 1987, p.312), the exploration could be treacherous and should be confirmed by subsequent prediction and hypothesis testing. The method could be used effectively for testing a priori hypothesised dependencies, as few tests would be required.

The techniques described above all deal with the frequencies of behavioural categories, attempting to identify the sequential structure (Markov, log-linear, and lag sequential analyses), how that sequential structure depends on contextual factors (log-linear analysis), and how the frequency of a particular behaviour may depend on contextual factors and on other behaviours at various points in the sequence (logistic regression). The question of how the results of these analyses might be used to assist in the function/task analysis will be considered later.

The frequencies of behavioural categories are usually tabulated in task analyses and form an important part of the analysis, as well as an essential input to operator performance models, such as might be prepared using MicroSAINT (Laughery, 1989).

11. Temporal Durations of Behaviours

In addition to the frequencies of behaviours, the temporal durations of behaviours form the other main input to practical task analysis. While, in theory, the frequencies of behaviours entailed by a task might be calculated without actually recording an operator carrying out the task, for temporal duration, or task execution time, it is more practical to obtain measures through observation. Although it should be mentioned that performance models such as those presented by Card, Moran, and Newell (1983) are aimed at making such predictions of task execution time.

The temporal duration of some episode, such as a task behaviour, is a special measure, which has its own history of study. The study of episodes such as survival times in medicine, biology, and insurance, and product failure times in manufacturing (Kalbfleisch and Prentice, 1980) has provided the statistical techniques for analysing duration data. These techniques have been applied to the events that occur during people's lives, such as education and employment episodes (Blossfeld, Hamerle, and Mayer, 1989), and to sequential records of people interacting in conversations (Gardner and Griffin, 1989; Griffin and Gardner, 1989; Gardner, 1990).

Griffin and Gardner (1989, p.497) state that the shapes of distributions of duration data are typically highly asymmetric and do not satisfy the assumptions of ordinary regression techniques. It is easy to see that this is the case for a duration such as human lifetime. A frequency distribution of human life duration will be bimodal, with a higher probability that lifetimes will end in infancy or after about 60 years, and a lower probability that lifetimes will end in the intermediate years (Blossfeld, Hamerle, and Mayer, 1989). However, Griffin and Gardner (1989) do not present evidence that behavioural durations as short as half a minute, for example, for a speaking turn in a conversation, or one second for gaze duration in eye movement records, or 200 milliseconds for a reaction time, will also exhibit highly asymmetric and non-normal distributions and would benefit from an event history approach to analysis.

Event history analysis is a form of regression analysis with different distributional assumptions. Griffin (1995) provides an overview of event history analysis. It is distinct from other sequential regression techniques such as those discussed by Gregson (1983) in that it analyses data recorded in continuous time and does not require a series of discrete intervals of time. Episode duration can be regressed on a number of variables, which may be continuous or discrete, or some of each kind. The predictor variables may be internal to the behaviour sequence, such as identity of the previous behaviour, duration of the previous behaviour, or accumulated number of occurrences of a particular behaviour to date in the sequence. The predictor variables may also be contextual factors such as whether the sequence has been collected from an experienced or an inexperienced operator, or the conditions under which the sequence is collected. Whether these variables are controlled experimental factors, or uncontrolled variables measured as they arise does not matter.

As with log-linear analysis, event history analysis works by testing the fit of structural models to find one that fits the data best. Many of the problems that are associated with log-linear analysis of the frequencies of sequential behaviours also apply to the analysis of the durations of behaviours using event history analysis. These are the problems of insufficient data, autocontingency, and heterogeneity of sequential records.

11.1 Amount of data required

In its historical origins event history analysis employs hundreds or thousands of subjects. It is not exactly clear how much data is required to carry out event history analysis of sequential behavioural records. Kerlinger and Pedhazur (1973, p.446-447) state that for any multiple regression analysis there should be at least 100 subjects, preferably 200, especially if there are to be many independent variables. In Griffin and Gardner's (1989) study of mother-son verbal interaction, 206 families were used, and over 30,000 behavioural events were recorded. This included 1,581 instances of negative statements by mothers to the boys, which was the behaviour of theoretical interest in that study. In Gardner and Griffin's (1989) study of one married couple interacting for a 20-minute conversation, behaviour was coded into only four categories, husband and wife, each looking at the other or away from the other. These gazes lasted around four seconds, so that over the 20-minute period there were approximately 300 occurrences of each behavioural category. Such numbers of occurrences of individual categories of behaviour are not likely to occur in the human factors analysis of records (such as videos) of operational behaviour. The exception perhaps would be gaze fixations on a limited number of targets, such as monitoring displayed instruments.

Autocontingency is again a problem in that if duration of behaviour depends, in part, on the duration of the previous occurrence of the same behaviour, which will often be the case, and if some independent variable is also a variable that depends in part on its own past history in the sequence, such as the duration of another behaviour, then contingency that appears between the variables may be spurious and due to the fact that both variables have their own sequential dependence structure.

The autocontingency problem is a difficult one for event history analysis (Griffin and Gardner , 1989) and its presence, if unaccounted for, may seriously bias the results of the analysis. If there are sufficient data, the duration of the previous occurrence of the behaviour of interest can be taken account of by including it as an independent variable (Gardner, personal communication, 12 September, 1991).

Griffin and Gardner (1989) have also warned of the dangers of unmeasured heterogeneity when the records of different subjects are pooled, which would often be necessary in human factors to gain generalisability and sufficient data. Apparently the developers of event history analysis have had difficulty identifying an appropriate distribution for an error term, which would absorb any variability due to variables which were not able to be specified in the model under test.

In summary, event history analysis for the study of durations of behaviours is a promising technique, but its application to sequences of behaviours is still under development. Like the analyses described earlier, large amounts of data are required for event history analysis and it is unlikely that such amounts of data would be available from the kinds of records of observations used in function/task analysis. Indeed, in a modern technological system, any single task component that might need to be carried out 200 times in say, two hours of video record, is likely to have been automated out of the task, as for example, repeatedly pressing cursor keys in a word processor has been eliminated by the introduction of the mouse and drag bars on screens.

In some highly proceduralised environments such large amounts of data may be obtainable for some behaviours. For example, in air traffic control a controller may handle 30 aircraft per hour. There are some tasks, such as acceptance and later hand-off to another controller, which must be carried out for every aircraft. For these tasks numbers of observations would soon reach an acceptable level after only a few hours of recording. Other tasks such as resolution of some aircraft conflicts may be considerably less frequent, yet take much longer when they do occur, and have more serious safety implications both from the point of view of the conflict concerned and the prolonged distraction of the ATC's attention from other ongoing events. Accumulating sufficient data to carry out event history analysis on these kinds of behaviours would require long periods of recording.

12. Other Approaches to Sequences

The techniques described above fall within that division of the study of behaviour, identified by van Hooff (1982), which looks at the immediate sequential and temporal dependencies of behaviours, they do not address the alternative framework of a functional hierarchy, an approach that is more consistent with traditional Human Factors function/task analysis. Methods such as log-linear analysis for task frequencies and event history analysis for task durations would be very useful in the detailed analysis of behaviour that can be carried out in experimental laboratory studies, or in studies involving behaviour in simulators, where events, displays, etc., can be manipulated. They are, first and foremost, techniques for testing hypotheses. Because data is always limited it is not possible to use these techniques in a wildly exploratory manner, setting up and testing for every conceivable source of influence. Indeed, it is difficult to collect enough data to support testing of specific hypotheses of interest using these techniques.

A Markov analysis will reveal whether a sequence is essentially random, with transition frequencies depending only on the base rates of the component behaviours, or whether it has a form of sequential structure in which the frequency of a behaviour depends on the identity of the immediately preceding behaviour(s). Except in extreme cases it is usually not possible to make this assessment intuitively by examining the transition matrix oneself. In Human Factors analysis we want to know both whether there is some consistent order in which the operator carries out component actions, and what determines that sequence.

Firstly, if there is a consistent order in which component tasks and task elements are carried out it is important to know that order, even if we do not know what causes it. Knowing the order of component actions will provide more detailed knowledge of the use of optional methods by operators as discussed earlier and described by Card, Moran, and Newell (1983). This detail can be used in making a MicroSAINT model of operator performance. Also, if we know something of the order in which tasks are carried out, and the circumstances that affect that order, it is possible to make some prediction of what an operator is likely to be doing at any particular point in a scenario. This is potentially helpful in predicting bottlenecks and overloads. For example, in the Air Traffic Control domain Bisseret (1971) claimed to have found two kinds of reasoning reflected in the order in which controllers checked attributes of conflict situations in developing a solution, and to have shown by experiment that one of these was more economical than the other. However, details of the study were not provided. As another example, Kerns (1990) reported important changes to the order in which pilots and air traffic controllers carry out certain procedures when data link is provided. Both pilots and air traffic controllers were found occasionally to act on information received via datalink, before acknowledging to the other party that the information had been received. This is potentially hazardous and does not happen with voice communications except perhaps if there is some technical fault that prevents acknowledgment.

Secondly, it is important to know what determines the sequence of actions, not only in terms of case specific behaviours that lead to other specific behaviours, or affect their durations, but also in terms of what kinds of factors can possibly have an influence on sequence;

- environmentally imposed task and situation context?
- available system functionality?
- humanness of the operator?

The categories listed above are intended to describe the possible sources of constraints on the sequence of behaviours. The first category derives from traditional task analysis as used by Card, Moran, and Newell (1983), Ericsson and Simon (1984), and Sanderson, Verhage, and Fuld (1989). It refers to sequentiality that is inherent in the task set for the operator, and which would also be present in the performance of a machine required to conduct the task. The second category refers to sequentiality that derives directly from the technical and procedural functionality available to the operator in the system that must be used to carry out the task. This category covers part of what Card, Moran, and Newell (1983, p.420) called 'methods analysis', and deals with sequentiality due to what, in their case, was, "...the demands of the computer system", but which in general would be the entire technical and procedural dimensions of the system. The third category refers to sequentiality that is entirely due to the human attributes that are embodied in the operator. The latter are sources of sequentiality such as cognitive limitations such as attention and memory, training, practice and expertise.

12.1 Task-entailed sequentiality

It is clear that in the case of the first category, task-entailed sequential constraints, there will be many sequential dependencies in an operator's behaviour that are due to this type of constraint. As a trivial example, it is not possible to land a (real) aircraft until it has taken off. If we have a number of protocols of flights, in every one the aircraft will have taken off at some point earlier in the sequence than the point in time at which it lands. This type of sequentiality does not derive from operator behaviour, but from the task itself. Many protocols will be riddled with such task-derived sequentiality. Indeed, for many tasks, such as landing an aircraft, successful performance may depend on careful adherence to the task-entailed sequential constraints, and many errors may consist of violations of the required sequence (there can, of course, be other sources of error such as errors of timing which would also be critical in landing an aircraft).

If sequential analysis is to become a regular part of Human Factors function/task analysis it will be necessary to find a means of identifying task-entailed sequential constraints and separating their effects on the sequence of behaviours from the effects of other sources of sequential constraint such as system functionality and the human attributes of the operator. A state-space analysis of the dynamics of the system being controlled by the operator (e.g. a chemical plant or an aircraft in flight), recommended as an essential part of task analysis by Sanderson, Verhage, and Fuld (1989), certainly will go some way towards specifying the task-entailed constraints on sequential orders of behaviours. However, while a state-space analysis of system dynamics specifies the paths that various system variables may take, and therefore how those variables would respond to control actions, it does not make any explicit statement about the possible sequential order of those control actions. Neither does an analysis of the control task set for the operator, another aspect of task analysis recommended by Sanderson, Verhage, and Fuld (1989), make sequential constraints clear, although a specification of sequential constraints could be incorporated into this part of task analysis. A possible means for separating task-entailed sequential constraints is discussed in a later section.

12.2 System-entailed sequentiality

Another example of a sequential constraint, but this time imposed by system functionality, is that an operator such as an Air Traffic Controller might not be able to punch-in co-ordinates to a computer until someone else, such as a pilot, who may have to tell the controller the co-ordinates, has communicated them to the controller. This is not as strict a constraint as the previous example in which the aircraft could not land until after it had taken off. That is a logical impossibility. But the present example, while effectively impossible, may be violated in error. For example, the controller may punch-in co-ordinates which the controller believes to have been received and to be correct, but which in fact comprise information that the transmitting party was sending about another matter. Many system constraints will be of this kind. Others will be of the more rigid, former kind, in which the technical configuration of the system makes it logically impossible to do one thing until something else has been done. Indeed, 'bugs' in computer systems often result from an assumed rigid sequential input not being rigid enough, allowing the user to make inputs which the user believes to be meaningful, but which do not have the intended effect.

12.3 Human-entailed sequentiality

Over and above the sequentiality entailed by the task, there will be some sequentiality due to the available system functionality, and there *may* be some sequentiality due to the nature of the human operator. The Human Factors analyst would like to be able to identify and separate out particularly these two sources of sequentiality in behaviour, assuming that there *is* some sequentiality due to the human attributes of the operator. If the sequential structure entailed in the task and the system can be accounted for, (a possible method will be discussed in a later section), so that the operator is free to arrange the remaining unconstrained tasks on hand in any order, is the basic null hypothesis of random sequential order reasonable for purposive behaviour?

A similar issue has been considered by Ericsson and Simon (1984) for problem solving, and by Card, Moran, and Newell (1983) for human-computer interaction. Ericsson and Simon (1984) referred to work by Haines on consumer decision making (1974, cited in Ericsson and Simon , 1989) which purported to show that the coding of subject's protocols was unreliable. Ericsson and Simon claim instead that the coding is fairly reliable in terms of inter-rater reliability, but that the protocols in that study were different for every subject, so that there was no underlying and generalisable sequential pattern to the particular process being examined.

Card, Moran, and Newell (1983), examining human-computer interaction, found that their ability to predict the sequence of behaviours from their GOMS model decreased as the "grain of analysis" (p.171) became finer. They were actually interested in the total execution time of a higher level task, an 'operator' in their nomenclature, which might be composed of a number of lower level tasks or operators. To predict the total execution time they had to predict the operators that the subject would employ, in the case of optional operators or methods, and they also investigated the effect of the sequential order in which subjects used the operators. Not surprisingly it was found that the actual order did not make much difference to the total execution time. However, the finer the grain of analysis, that is, the more components a task is broken down into, the more opportunities there are for different sequences to emerge. Card, Moran, and Newell 1983, p.161) state that it is not possible, a priori, to know which grain size is appropriate. This is an empirical question. It is necessary to try different grain sizes and see at what level the ability to predict something about the sequence can be obtained.

Ericsson and Simon (1984) also reported that for problems in which the path to the correct solution was not unique, subjects did not use the same sequences of processes.

> "This makes it difficult to use a single computer model to predict or account for the detail of numbers of different protocols. This doesn't mean that these models are incorrect, but rather, that at the level of detail they capture, we cannot always generalize across individuals". (Ericsson and Simon , 1984, p. 196).

In referring to Ericsson's work on the Eight Puzzle, in which subjects move tiles around a board, Ericsson and Simon further commented:

> "The similarity of move sequences among subjects starting from the same puzzle configuration was no greater than would be predicted by chance. Hence, no single model could be expected to predict the exact sequences. However, when the solutions were analyzed at a more abstract level (in terms of the attainment of certain specified configurations of tiles), most subjects followed the same orderly and predictable sequence. The same special configurations (sub-goals) were attained by most subjects for most of the different problems in the same order". (Ericsson and Simon, 1984, p.197).

Note that the special configurations of tiles, the sub-goals, that most subjects went through on their way to solution, were not observable behaviours, they were observable states of the tile board. The behaviours are moves of the tiles and were unpredictable, the tile-board also, therefore, goes through unpredictable intermediate states.

While at an abstract level it might be possible to say that working towards a particular sub-goal is a behaviour, in its own right, you cannot actually *observe* it as a behaviour. It is only possible to infer the behaviour of working towards a particular sub-goal, either by working backwards in the protocol from the attainment of the sub-goal, or by having the subject assert in debriefing that this is what the subject was doing, or by having it asserted by a Subject Matter Expert (SME), or by yourself asserting, as an analyst, that you know that the subject was working towards a particular sub-goal, because you know enough about the subject matter domain to make this interpretation with confidence. Sebillotte (1988), for example, relied entirely on subject's verbalizations to infer a hierarchically organised subjective goal structure. The point is that behaviour at this level of abstraction cannot be observed.

Yet, the work Ericsson and Simon (1984) referred to has shown that some consistent pattern in sequential behaviour can be expected at this abstract level of goals and sub-goals, even if it cannot be found at the more fine-grained level of the component elements that are used to achieve the sub-goals. Sub-goals should be objectively defined, like the specific configurations of tiles on the tile board described by Ericsson and Simon (1984), rather than being notions of individual sub-goals generated by subjects. If sub-goals can be defined objectively, so that they will be consistent from one subject to another, or from one sub-sequence of behaviour to another within the same subject's record, then it might be expected that consistent patterns of sequential behaviour could be found, at least for this level of abstraction. The sequences of activity at a finer grain of analysis, between these sub-goals, will not necessarily have any consistent pattern, and may, indeed, represent the level at which sequential order is neither constrained, nor of any important value in terms of efficiency or economy of effort.

However, to discover that sequences of sub-goals are consistent from one sequential record to another it is necessary to compare them. Card, Moran, and Newell noted in 1983:

> "There is no standard statistical technique for indexing how well one sequence matches another." (Card, Moran, and Newell, 1983, p.157)

They used a method that inserts dummy symbols into the sequences to bring two sequences into correspondence as far as possible, and then counts the matching symbols and expresses the result as a percentage of overall sequence length (Card, Moran, and Newell, 1983, appendix to chapter 5, p 190). Card et al. were not able to say what significance this percentage of symbol matches has in terms of any assumed probabilistic process generating the sequence differences.

It is important to be able to make confident statements about whether sequences of behaviours are similar or not, both for sub-sequences drawn from one subject's longer record, and for sets of sequences from two groups of subjects. Such statements have been based on appearance, or on measures such as that described above, which do not have an accompanying test of significance.

Hierarchical approaches to describing sequential structure, such as grammars (Schiele and Green, 1990), have also suffered from the lack of a technique for making comparisons. The "goodness" of the grammar has been determined aesthetically, not on the basis of tests (van Hooff, 1982). However, with the introduction of new techniques it now appears to be possible both to compare sequences themselves for similarity, and to compare the goodness of fit of representations of their structure. Such a technique is discussed in the next section.

13. Using the Wallace Information Measure to Compare Sequences

Patrick and Chong (1991) have introduced the use of a measure they refer to as the Wallace Information Measure to compare structural representations of sequences of behaviour. The Wallace Information Measure is a statistic based on minimum message length encodings of the sequences. A sequence can be represented by its structure. If a grouping of symbols in the sequence, such as a sub-sequence, occurs more than once, it can be given a code of its own, and then that code can be used in place of the sub-sequence when describing the structure, thus shortening the description needed to represent the whole sequence. To attain minimum message length an optimal way of representing the structure of the original sequence must be found. For example, the most frequently occurring sub-sequence can be given the shortest representational code, the next most frequent can be given the next shortest code and so on. There is an information cost associated both with describing the group with its code, which is only incurred once for each grouping in a sequence, and another information cost associated with referencing that group with its code, which is a cost incurred once for every time the group appears in the overall sequence. The Wallace Information Measure takes both of these costs into account. As there are many ways to chop up a sequence into recurring sub-groups an optimal dissection must be sought. The Wallace Information Measure provides a decision statistic for comparing two sequences by comparing their minimum message length encodings, and making a confident selection of the shortest alternative.

Minimum message length encoding techniques originated in computer science. Once a statistical theory had been developed for minimal message length it was possible to use this technique to make inferences about sequences of many kinds. For example, minimum message length methods have been used to compare DNA sequences (Allison, Wallace, and Yee, 1990), and in Patrick and Chong's (1991) Capture and Analysis of Behavioural Events in Real Time (CABER) application this technique is used to select a Probabilistic Finite State Automaton that will account for the transitions between behaviours in a behavioural sequence.

13.1 Comparison with other sequential statistics

Markov analysis compares sequences in terms of the structure of transition probabilities. This is a particular kind of sequential structure but not the only kind. It cannot, for example, take into account sequential structure in which a behaviour depends on what has occurred a number of steps back in the sequence, but does not depend at all on what intervenes, as when there is a delayed reaction to an event. Lag sequential analysis, on the other hand, can examine only this latter form of sequential dependency. Minimum message length encoding considers sequential structure from another point of view. It uses intelligent methods to assign sections of the sequence to categories and the Wallace Information Measure can then be used to compare the informational economy of different categorisations. The most economical categorisation that can be found is accepted. As Patrick (1991) has put it,

> "This approach may also be regarded as an operational form of Occam's Razor, that is, to prefer that theory which yields the shortest "explanation" of the available data." (Patrick, 1991, p. 2).

Further research would be required to make a detailed comparison of the minimum message length approach to analysing sequences and the Markov method of analysing sequences. It is not immediately apparent what differences in results the two methods would produce.

However, there are aspects of the minimum message length approach that may have potential benefits beyond the immediate one of providing a way of testing the similarity of sequences, and these are discussed below.

13.2 Assistance in discovering a categorisation structure

In order to calculate the message length, a categorisation structure is imposed on the sequence, and it may be possible to use this structure to assist in the process of building a coding scheme for the behaviour sequence. Both the coding scheme and the task taxonomy itself may benefit from this automatic method of producing a categorisation structure. Another possible area for further research would be to compare the categorisations produced by the minimum message length method to those produced by other categorisation methods that have been applied to behaviour sequences; principal component and factor analysis, multidimensional scaling, and cluster analysis (Van Hooff, 1982), and more recently, SEQUITUR by Neville-Manning and Witten (1997).

14. Syntactic Coding of Behaviour Sequences

When a record of behaviour, for example on a videotape, is coded into a protocol, the continuous stream of behaviour is segmented, broken up into pieces with beginnings and endings, and these are given names or codes. Usually, the code that is assigned to each piece of behaviour is not unique to that piece of videotape, the same behaviour may occur again in another part of the tape and be given the same name. Also, as discussed earlier, several behaviours may be seen as related, perhaps forming part of a higher level activity, as when a number of task elements form a typical sequence that completes a task. These relationships between the various codes form a coding syntax.

The coding syntax may be implicit in that it is not specified, but is retained in the analyst's mind and used in the process of coding the protocol. Alternatively, the coding syntax may be made explicit, perhaps by writing down the relationships between the codes. The protocol analysis package SHAPA (Sanderson, James, and Seidler, 1989), uses a coding syntax which is based on the prolog programming language (Bratko, 1986), and which in its application to behaviour, is reminiscent of Dawkins's (1976) definition of a hierarchy. The package acts as a database with the syntax specifying the relationships between categories of behaviour in the database. Of course it is up to the analyst to decide which behaviours are related to each other when assigning codes to the behaviour.

Another package for analysing records of behaviour, CABER (Patrick and Chong, 1991), requires, as part of the analysis, that the analyst develop a syntax for the particular behaviours being studied. The package then parses coded input in terms of this particular project specific syntax. The syntax can be changed, and the iterative development of the syntax is part of the analysis process. For example, CABER has been used to analyse player and team behaviour in various sports, such as Australian rules football (Patrick and McKenna, 1986b), rugby, and water polo. In these sports the rules of the game provide a certain amount of sequential structure. For example, an Australian rules football match is made up of four 25 minute quarters. Each quarter begins with the umpire bouncing the ball high into the air in the centre of the football ground. Thus, the first player action cannot be 'taking a mark', which is catching the ball on the volley from the kick of another player, since no player has yet kicked the ball. This inherent sequentiality in the game is like the task-entailed sequential constraints discussed in an earlier section.

The way of coding protocols in CABER appears to differ from that used in SHAPA in that in CABER the coding syntax is specifically concerned with the sequential constraints on behaviour. It is hierarchical, but it is a hierarchy lying on its side on a horizontal time dimension. From a hierarchical perspective, the high level category of a quarter of the match branches out into finer categories of player events and team events which are contained within it, and which are brought to a close in time before the next quarter of the match begins. In SHAPA, on the other hand, the relationships between codes form a hierarchy which is strictly vertical, and which does not have any sequential structure.

The ability, in the CABER package, to construct a coding syntax which is both specific to the questions being asked in the study, and of an inherently sequential nature, could be of particular use in Human Factors. It would be possible to use the development of the coding syntax to specify the task-entailed sequentiality discussed earlier, and to separate that from other sources of sequentiality. Different syntaxes can be developed to code the same behavioural record, addressing different questions of interest. So, another version of the coding syntax might include both task-entailed sequentiality and what is believed to be sources of system-entailed sequentiality. Any consistent sequentiality remaining in the behaviour, that is, not incorporated in the input syntax, would then be identifiable as due to some properties of the human operator. This form of coding syntax is suggested as a means of separating sources of sequentiality, and thus, simplifying the analysis and interpretation of sequentiality.

CABER includes the ability to build a Probabilistic Finite State Automaton (PFSA) as a model of the sequences of behaviour that have been analysed. It uses minimum message length techniques and the Wallace Information Measure to progressively adjust the PFSA to provide a representation of the set of input sequences. The PFSA can be presented in diagrammatic form with nodes representing the behaviours and arrows to indicate sequential order. It should be noted that CABER can accommodate parallel behaviour, which is necessary of course, in its use in analysing team sports, where several team members and opposition players may be carrying out significant actions at the same time.

The difference between PFSAs generated under different codings can be measured using the Wallace Information Measure. This is useful in the gradual refinement of coding syntax, and also provides the ability to test competing theories regarding the appropriate structure for the task taxonomy. Furthermore, the PFSA's generated from using the same coding syntax, but on records from different individuals or experimental groups, such as experts and novices, or records obtained under different conditions, can be compared, providing a way of testing whether these manipulations affect the sequence of behaviours.

Once an input syntax has been designed, CABER can also be used to analyse behaviour in real time. This would not usually be necessary in Human Factors, as videos and other records can be analysed in the laboratory at will. However, there may be some circumstances in which real-time analysis is helpful. For example, in training situations it would be possible to provide feedback to the operator while performance is continuing.

In summary, CABER appears to provide several analytical and statistical tools, which have already been developed with the analysis of human behaviour as the aim, would be of potential use for the study of protocols in Human Factors, and which do not appear to be available elsewhere.

Hingston and Lees (1994) have also used the minimum message length technique for inductive inference of probabilistic finite state automata to model sequential observational data, and have modified and refined Patrick and Chong's (1991) search method.

15. Conclusion

Modern technology enables Human Factors researchers to make extensive recordings of many dimensions of the behaviour of the human operator, but these recordings encode behaviour at the very lowest, 'cinematic' level of representation (Gregson, 1983). Coding these records as protocols gives a higher level of representation but it is still very like Gregson's second lowest level, verbal description. It is difficult to summarise and represent the structure of these protocols at a higher level, in a manner that will provide additional characterisation to the function/task analysis. Several statistical tools that can be used to assist in this endeavour have been reviewed, but they generally require a large amount of data. A sequential analysis technique which has only recently been applied to human behaviour, based on minimum message length encoding, is recommended as a potentially useful tool for this purpose.

The extent to which the sum of task times, when calculated as full running times, exceeds the actual duration of the whole sequence, perhaps expressed as a ratio, might be of some use as an index of the extent to which activities are held *on hand*, or carried out in parallel in a particular record. This might be compared for records taken under different circumstances, or for operators with different levels of experience, etc., and would be a measure related to overall workload. For individual tasks, the ratio of the full running time including interruptions, to the duration with interruptions subtracted, could provide an index of whether tasks represent long-term or short -term goals. This measure, however, could also be related to the priority or importance the operator assigns to the task, with less important tasks being allowed to lie on hand longer while more important tasks are attended to. To take priority differences into account it would be necessary to note the frequency with which the operator returns to the task during its full running time. This frequency can be used as an index of the attention the operator is giving the task in the same way that frequency of gaze fixations at a particular sub-system is interpreted as an index of attention being paid to that sub-system in the research reported by Moray and Rotenberg (1989). It is also important to distinguish between meaningful short-term goals, meaningful long-term goals, and goals that are neither of these, in that it does not matter very much when they are completed within some medium-term limit. Short-term goals have some urgency, as other activities depend on their timely completion. Long-term goals, by definition cannot be brought to satisfactory completion in the short term, as they consist of some achievement that depends on other intervening activities occurring. In practice it may be difficult to distinguish long-term goals from the remainder, for which completion time is not important. One way to separate them might be to look at the variability of their full running time, that is, task duration from beginning to end, including interruptions. This may reveal priority relative to other tasks on hand at the time, but not absolute priority of the task.

16. References

Allison, L., Wallace, C. S., and Yee, C. N. (1990). *Inductive inference over macro-molecules*, Dept of Computer Science Technical Report No: 90/148, Monash University, Clayton, VIC. Australia. 3168.

Bakeman, R., and Gottman, J. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed). New York: Cambridge University Press.

Bakeman, R., and Gottman, J. (1986). *Observing interaction: An introduction to sequential analysis*. Cambridge: Cambridge University Press.

Bakeman, R., and Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS & GSEQ*. Cambridge: Cambridge University Press.

Bisseret, A. (1971). Analysis of mental processes involved in air traffic control. *Ergonomics*, 14(5), 565-570.

Blossfeld, H. P., Hamerle, A., and Mayer, K. U. (1989). *Event history analysis: Statistical theory and application in the social sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Bratko, I. (1986). *Prolog programming for artificial intelligence*. Wokingham, England: Addison-Wesley Publishing Company.

Card, S. K., Moran, T. M., and Newell, A. (1983). *The psychology of human-computer interaction*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Dawkins, R. (1976). Hierarchical organisation: a candidate principle for ethology. In P. P. G. Bateson and R. A. Hinde (Eds), *Growing points in ethology*. Cambridge: Cambridge University Press. p. 7-54.

Ericsson, K. A., and Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, Massachusetts: The MIT Press.

Faraone, S. V., and Dorfman, D. D. (1987). Lag sequential analysis: Robust statistical methods. *Psychological Bulletin*, 101(2), 312-323.

Gardner, W. (1990). Analyzing sequential categorical data: Individual variation in Markov chains. *Psychometrika*, 55(2), 263-275.

Gardner, W., and Griffin, W. A. (1989). Methods for the analysis of parallel streams of continuously recorded social behaviors. *Psychological Bulletin*, 105(3), 446-455.

Gottman, J. M., and Roy, A. K. (1990). *Sequential analysis: A guide for behavioral researchers*. Cambridge: Cambridge University Press.

Gregson, R. A .M. (1983). *Time series in psychology*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Griffin, W. A. (1995). Assessing state changes in microsocial interaction: An introduction to Event-History Analysis. In J. M. Gottman (Ed.), *The analysis of* *change*. Mahwah, New Jersey, Lawrence Erlbaum Associates, Publishers, (pp. 309-337).

Griffin, W. A., and Gardner, W. (1989). Analysis of behavioral durations in observational studies of social interaction. *Psychological Bulletin*, 106(3), 497-502.

Hetrick, W. P., Isenhart, R. C., Taylor, D. V., and Sandman, C. A. (1991). ODAP: A stand-alone program for observational data acquisition. *Behavior Research Methods, Instrumentation, and Computers*, 23(1), 66-71.

Hingston, P. F., and Lees, C. D. (1994). *Sequential analysis with probabilistic finite state automata*. (Tech. Rep. No. 94/12). Perth, Western Australia: University of Western Australia, Department of Computer Science.

James, J. M., and Sanderson, P. M. (1990) *Heuristic and statistical support for protocol analysis with SHAPA version 2.0*. Engineering Psychology Research Laboratory report EPRL-90-16, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

Kalbfleisch, J. D., and Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley.

Kemeny, J. G., Mirkil, H., Snell, J. L., and Thompson, G. L. (1959). *Finite mathematical structures*. Englewood Cliffs, NJ: Prentice-Hall.

Kerlinger, F. N., and Pedhazur, E. J. (1973).*Multiple regression in behavioral research*. New York: Holt, Rinehart, and Winston, INC.

Kerns, K. (1990). *Data link communication between controllers and pilots: A review and synthesis of the simulation literature*. Report MP-90W00027, The MITRE Corporation., Centre for Advanced Aviation System Development, 7525 Colshire Drive, McLean, Virginia 22102-3481.

Krippendorff, K. (1986). *Information theory: Structural models for qualitative data*. Sage University Paper series on quantitative applications in the social sciences, 07-062. Beverly Hills: Sage Pubns.

Laughery, K. R. (1989). MicroSAINT - a tool for modeling human performance in systems. In G. R. McMillan, D. Beevis, E. Salas, M. H. Strub, R. Sutton, and L. van Breda (Eds), *Applications of human performance models to system design*. New York: Plenum Press, p.219-230.

McMillan, G. R., Beevis, D., Salas, E., Strub, M. H., Sutton, R., and Van Breda, L. (Eds) (1989). *Applications of human performance models to system design*. New York: Plenum Press.

Meister, D. (1985). *Behavioral analysis and measurement methods*. New York: John Wiley and Sons.

Moray, N., and Rotenberg, I. (1989). Fault management in process control: eye movements and action. *Ergonomics, 32(11)*, 1319-1342.

Neville-Manning, C. G., and Witten, I. H. (1997). Identifying hierarchical structure in sequences: A linear-time algorithm. *Journal of Artificial Intelligence Research, 7*, 67-82.

Newell, A., and Simon, H. A. (1972), *Human problem solving*. Englewood Cliffs, New Jersey: Prentice Hall.

Noldus, L. P. J. J. (1991). The Observer: A software system for collection and analysis of observational data. *Behavior Research Methods, Instruments, and Computers*, 23(3), 415-429.

Patrick, J. D. (1991). *Snob: A program for discriminating between classes*. Dept of Computer Science Technical Report No. 91/151, Monash University, Clayton, VIC. Australia, 3168.

Patrick, J. D., and Chong, K. E. (1991). Real-time inductive inference for analysing human behaviour. Paper presented at the International Joint Conference on AI (IJCAI'91), Workshop number 6 on Integrating AI into databases. Sydney, 1991.

Patrick, J. D., and McKenna, M. J. (1986a). A generalised computer system for sports analysis. *The Australian Journal of Science and Medicine in Sport*. 18(3), 19-23.

Patrick, J. D., and McKenna, M. J. (1986b). Computerised analysis of handball in Australian football. *The Australian Journal of Science and Medicine in Sport*, 18(3), 24-26.

Phillips, M. D., and Melville, B. E. (1988) Analyzing controller tasks to define air traffic control system automation requirements. In *Human error avoidance techniques conference proceedings*. Washington, D. C. December 1-3, 1987. SAE-P-204. Warrendale, PA: The Society of Automotive Engineers. p.37-44.

Sanderson, P. M., and Fisher, C. (1997). Exploratory Sequential Data Analysis: Qualitative and quantitative handling of continuous observational data. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1471-1513). New York: John wiley & Sons.

Sanderson, P. M. (1991). *ESDA: Exploratory sequential data analysis*. Engineering Psychology Research Laboratory report EPRL-91-04, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

Sanderson, P. M., James, J. M., and Seidler, K. S. (1989). SHAPA: an interactive software environment for protocol analysis. *Ergonomics*, 32(11), 1271-1302.

Sanderson, P. M., Verhage, A. G., and Fuld, R. B. (1989).State-space and verbal protocol methods for studying the human operator in process control. *Ergonomics*, 32(11), 1343-1372.

Schiele, F., and Green, T. (1990). HCI formalisms and cognitive psychology: The case of task-action grammar. In M. Harrison and H. Thimbleby, *Formal methods in human-computer interaction*. Cambridge: Cambridge University Press, (pp. 9-62).

Sebillotte, S. (1988). Hierarchical planning as method for task analysis: the example of office task analysis. *Behaviour and Information Technology*, 7(3), 275-293.

Van den Bercken, J. H. L., and Cools, A. R. (1980). Information-statistical analysis of social interaction and communication: An analysis-of-variance approach. *Animal Behavior*, 28, 172-188.

Van Hooff, J. A. R. A. M. (1982). Categories and sequences of behaviour: methods of description and analysis. In K. R. Scherer and P. Ekman (Eds), *Handbook of methods in nonverbal behavior research*. Cambridge: Cambridge University Press, (pp. 362-439).

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Yoder, P. J., and Tapp, J. T. Sr. (1990). SATS: Sequential analysis of transcripts system. *Behavior Research Methods, Instruments, and Computers*, 22(3), 339-343.

DISTRIBUTION LIST

Protocol Analysis as a Tool in Function and Task Analysis

Catherine Lees, Jeremy Manton and Tom Triggs

AUSTRALIA

DEFENCE ORGANISATION

S&T Program

Chief Defence Scientist
FAS Science Policy
AS Science Corporate Management
} shared copy

Director General Science Policy Development
Counsellor Defence Science, London (Doc Data Sheet)
Counsellor Defence Science, Washington (Doc Data Sheet)
Scientific Adviser to MRDC Thailand (Doc Data Sheet)
Scientific Adviser Policy and Command
Navy Scientific Adviser (Doc Data Sheet and distribution list only)
Scientific Adviser - Army (Doc Data Sheet and distribution list only)
Air Force Scientific Adviser
Director Trials

Aeronautical and Maritime Research Laboratory
Director
Chief of Air Operations Division
Research Leader Simulation and Human Factors

Electronics and Surveillance Research Laboratory
Director (Doc Data Sheet and Distribution List)
Chief of Land Operations Division
Head Human Factors Discipline
Author: Dr J Manton

DSTO Library and Archives
Library Fishermens Bend
Library Maribyrnong
Library Salisbury (2 copies)
Australian Archives
Library, MOD, Pyrmont (Doc Data sheet only)
*US Defense Technical Information Center, 2 copies
*UK Defence Research Information Centre, 2 copies
*Canada Defence Scientific Information Service, 1 copy
*NZ Defence Information Centre, 1 copy
National Library of Australia, 1 copy

Capability Development Division

Director General Maritime Development (Doc Data Sheet only)
Director General Land Development (Doc Data Sheet only)
Director General C3I Development (Doc Data Sheet only)
Director General Aerospace Development (Doc Data Sheet only)

Army

ABCA Standardisation Officer, Puckapunyal (4 copies)
SO (Science), DJFHQ(L), MILPO Enoggera, Queensland 4051 (Doc Data Sheet only)
NAPOC QWG Engineer NBCD c/- DENGRS-A, HQ Engineer Centre Liverpool Military Area, NSW 2174 (Doc Data Sheet only)

Intelligence Program

DGSTA Defence Intelligence Organisation

Corporate Support Program

OIC TRS, Defence Regional Library, Canberra

UNIVERSITIES AND COLLEGES

Australian Defence Force Academy
Library
Head of Aerospace and Mechanical Engineering
Serials Section (M list), Deakin University Library, Geelong, 3217
Senior Librarian, Hargrave Library, Monash University
Librarian, Flinders University

OTHER ORGANISATIONS

NASA (Canberra)
AGPS
State Library of South Australia
Parliamentary Library, South Australia

External Authors.

Dr Catherine Lees, Dept of Organisational and Labour Studies, The University of Western Australia (5 copies)
Dr Tom Triggs, School of Psychology, Monash University (5 copies)

OUTSIDE AUSTRALIA

ABSTRACTING AND INFORMATION ORGANISATIONS

Library, Chemical Abstracts Reference Service
Engineering Societies Library, US
Materials Information, Cambridge Scientific Abstracts, US
Documents Librarian, The Center for Research Libraries, US

INFORMATION EXCHANGE AGREEMENT PARTNERS

Acquisitions Unit, Science Reference and Information Service, UK

Library - Exchange Desk, National Institute of Standards and Technology, US

National Aerospace Laboratory, Netherlands

SPARES (5 copies)

Total number of copies: **60**

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA					1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)
2. TITLE Protocol Analysis as a Tool in Function and Task Analysis			3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)		
4. AUTHOR(S) Catherine Lees, Jeremy Manton and Tom Triggs			5. CORPORATE AUTHOR Electronics and Surveillance Research Laboratory PO Box 1500 Salisbury SA 5108 Australia		
6a. DSTO NUMBER DSTO-TR-0883	6b. AR NUMBER AR-011-105		6c. TYPE OF REPORT Technical Report		7. DOCUMENT DATE October 1999
8. FILE NUMBER D9505-15-185	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES 37		12. NO. OF REFERENCES 47
13. DOWNGRADING/DELIMITING INSTRUCTIONS http://www.dsto.defence.gov.au/corporate/reports/DSTO-TR-0883.pdf			14. RELEASE AUTHORITY Chief, Land Operations Division		
15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT *Approved for public release* OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE CENTRE, DIS NETWORK OFFICE, DEPT OF DEFENCE, CAMPBELL PARK OFFICES, CANBERRA ACT 2600					
16. DELIBERATE ANNOUNCEMENT No Limitations					
17. CASUAL ANNOUNCEMENT Yes					
18. DEFTEST DESCRIPTORS Task analysis Human factors Complex systems Sequential analysis Human behavior					

19. ABSTRACT

A protocol is a record of procedural steps undertaken in a process. In studying human-machine systems, observations of human operators obtained from sources such as videotapes are coded to create a descriptive protocol of behaviours, task elements, goals, etc. This record is essentially sequential in nature, but methods for analysing sequential data are relatively new. The kinds of information that protocol analysis can provide, that might be useful in function/task analysis, are examined. Methods for analysing sequences are surveyed, and recent developments in using minimum message length methods for producing probabilistic finite state automaton models of sequential behaviour are discussed.