# TASK CLASSIFICATION AND DECISION PROCESSES IN MONITORING BEHAVIOUR

THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

by

Raja Parasuraman, B.Sc., M.Sc.



Applied Psychology Department,

University of Aston in Birmingham.

April, 1976.

SUMMARY

Task classification is introduced as a method for the evaluation of monitoring behaviour in different task situations. On the basis of an analysis of different monitoring tasks, a task classification system comprising four task 'dimensions' is proposed. The perceptual speed and flexibility of closure categories, which are identified with signal discrimination type, comprise the principal dimension in this taxonomy, the others being sense modality, the time course of events, and source complexity.

It is also proposed that decision theory provides the most complete method for the analysis of performance in monitoring tasks. Several different aspects of decision theory in relation to monitoring behaviour are described. A method is also outlined whereby both accuracy and latency measures of performance may be analysed within the same decision theory framework.

Eight experiments and an organizational study are reported. The results show that a distinction can be made between the perceptual efficiency (sensitivity) of a monitor and his criterial level of response, and that in most monitoring situations, there is no decrement in efficiency over the work period, but an increase in the strictness of the response criterion. The range of tasks exhibiting either or both of these performance trends can be specified within the task classification system. In particular, it is shown that a sensitivity decrement is only obtained for 'speed' tasks with a high stimulation rate. A distinctive feature of 'speed' tasks is that target detection requires the discrimination of a change in a stimulus relative to preceding stimuli, whereas in 'closure' tasks, the information required for the discrimination of targets is presented at the same point in time. In the final study, the

specification of tasks yielding sensitivity decrements is shown to be consistent with a task classification analysis of the monitoring literature.

It is also demonstrated that the signal type dimension has a major influence on the consistency of individual differences in performance in different tasks. The results provide an empirical validation for the 'speed' and 'closure' categories, and suggest that individual differences are not completely task specific but are dependent on the demands common to different tasks. Task classification is therefore shown to enable improved generalizations to be made of the factors affecting 1) performance trends over time, and 2) the consistency of performance in different tasks.

A decision theory analysis of response latencies is shown to support the view that criterion shifts are obtained in some tasks, while sensitivity shifts are obtained in others. The results of a psychophysiological study also suggest that evoked potential latency measures may provide temporal correlates of criterion shifts in monitoring tasks. Among other results, the finding that the latencies of negative responses do not increase over time is taken to invalidate arousal-based theories of performance trends over a work period. An interpretation in terms of expectancy, however, provides a more reliable explanation of criterion shifts. Although the mechanisms underlying the sensitivity decrement are not completely clear, the results rule out 'unitary' theories such as observing response and coupling theory. It is suggested that an interpretation in terms of the memory data limitations on information processing provides the most parsimonious explanation of all the results in the literature relating to sensitivity decrement.

Task classification therefore enables the refinement and selection of theories of monitoring behaviour in terms of their reliability in generalizing predictions to a wide range of tasks. It is thus concluded that task classification and decision theory provide a reliable basis for the assessment and analysis of monitoring behaviour in different task situations.

## ACKNOWLEDGEMENTS

I would like to thank Dr Roy Davies, Reader in Experimental Psychology in the Applied Psychology Department, for his supervision and his valuable advice and encouragement at every stage in the development of this thesis. I have benefitted greatly from discussions with Dr Davies and other members of the Applied Psychology Department.

I am indebted to Christine Maddison for her efficiency and speed in typing the manuscript, and to the technical skills of Peter Bailey, Les Bagnall and Paul Bernard, and also John Lindley, now at the Department of Engineering Production in the University of Birmingham.

I would also like to thank Professor Edwin A. Fleishman of the Graduate School of Administration, University of California, Irvine, and Dr D.E. Broadbent of the Department of Experimental Psychology, University of Oxford, for their valuable comments on previous drafts of some of the material presented in this thesis, and Dr S. Milosevic, of the Institute of Psychology, Belgrade, for providing me with unpublished data.

Finally, I would like to thank all the people who kindly participated in the experimental studies, and all friends and acquaintances for their encouragement and support.

# CONTENTS

# C H A P T E R   O N E

## INTRODUCTION

The experimental study of human behaviour stems from the need to understand and develop the principles underlying human functioning in different situations. The formulation of such principles and their embodiment within a theory usually follows from investigations of the effects of certain independent variables on the performance of a task sampling the particular type of behaviour under study. By assessing the relative influence of these variables, it is assumed that a fuller understanding of the principles governing behaviour may be gained.

In empirical studies in human performance research, it is often the case that investigators use a number of different tasks in the study of similar types of behaviour. The use of a wide range of tasks should ideally provide a rich source of data whereby the generality of a theory and its ability to explain performance trends in different tasks may be evaluated. Quite often, however, a given theory may be able to make only very general, and sometimes trivial, predictions about performance. In consequence, several difficulties may arise in attempting to evaluate the effects of independent variables on performance in a range of tasks which sample a similar class of behaviour.

Such difficulties may also arise in attempting to predict individual or group performances in task situations other than the ones for which empirical data are available. However, despite the wide range of tasks for which performance data are available, the development of a system for organizing tasks within a limited and specified number of categories ought to improve significantly the ability to make reliable predictions of performance. The rationale behind the use of such a <u>task classification</u> system is that it may help not only to organize performance data so that more reliable generalizations about factors affecting human performance may be made, but also serve as a framework for guiding further research. Many prominent psychologists have, over the years, recognized the need for the development of such classification systems (Alluisi, 1967; Fitts, 1962; Gagné, 1962), but efforts to organize task classification schemes have been initiated only recently, the principal developments stemming from the work of Edwin A. Fleishman and his associates (Fleishman, 1972, 1975a, b; Theologus and Fleishman, 1971). Fleishman has also put forward similar arguments to the ones presented here for the need to consider task classification systems in human performance research. Such arguments can be formulated with reference to the general field of human performance, but they apply equally forcefully to smaller areas of behaviour, in particular to monitoring behaviour, which is the subject under investigation in this thesis.

Monitoring behaviour is an area of human performance research where a number of experimental studies have been reported (Davies and Tune, 1970). The tasks used in these studies involve the detection or discrimination of relatively infrequent 'signals' over a prolonged work period. The 'signals' in monitoring tasks may be defined in a number of ways, and several types of stimuli have been used. In most laboratory investigations, 'sensory' stimuli have been employed, although more 'cognitive' tasks employing verbal and symbolic stimuli have also been investigated (see, for example, Bakan, 1959; Sipos, 1970). The diversity of signal types and stimuli used in monitoring tasks emphasizes the need for taxonomic systems whereby such tasks may be described. However, although some attempts have been made to identify the common features of such tasks (Bergum, 1966; McGrath, 1963), very little attention has been paid to the problem of the development and evaluation of a task classification system for monitoring tasks.

The historical antecedents of research into monitoring behaviour may be traced to early field studies of the effects of monotony and fatigue on different industrial inspection tasks (Wyatt and Langdon, 1932). Present interest in the subject stems mainly from a number of laboratory experiments carried out since the pioneering work of N.H. Mackworth (1948, 1950) on the performance of radar operators. Although most of these experiments have been concerned with the investigation of performance trends over a work period, monitoring tasks have also been used as standard tasks in the assessment of environmental stress effects (see, for example, Lewis, Baddeley, Bonham and Lovett, 1970; Poulton, 1970; Wilkinson, 1968), and in the 'diagnosis' of 'attentional dysfunctions' in children, old people and certain categories of mental patients (Alexander, 1973; Anderson, Halcomb and Doyle, 1973; Dardano, 1969). Furthermore, although arguments against the relevance of laboratory research in monitoring behaviour have been put forward (Smith and Lucaccini, 1969), applied interest in the subject is also maintained, not only in the previously mentioned 'diagnostic' applications, but also in certain industrial situations such as inspection and quality control, (Drury and Fox, 1975; Poulton, 1973; Wiener, 1975b).

Research into monitoring behaviour has been the subject of periodic reviews (Deese, 1955; Frankmann and Adams, 1962; Craig and Colquhoun, 1975), and the experimental studies reported up to 1969-70 have been reviewed in books by Broadbent (1971), Davies and Tune (1970) and J.F. Mackworth (1969, 1970). As has been noted previously, however, very little of this work has been carried out within a taxonomic framework. A major aim of this thesis is thus the development and evaluation of a task classification system which will enable more reliable generalizations to be made of the effects of independent variables on monitoring performance. At the same time, however, it is recognized that the potential benefits of systematization techniques such as task classification may not be realized if appropriate methods of performance measurement are not available. This thesis is thus also concerned with the further development of more reliable methods of analysing monitoring performance. More specifically, it is proposed that this requirement leads to an analysis of the decision processes involved in monitoring tasks. A characteristic of the literature on monitoring behaviour is that such analyses are not frequently reported. In this thesis however, it is argued that decision theory provides the most complete method for the analysis of performance in a wide range of monitoring tasks.

An analysis of monitoring behaviour from the point of view of decision theory was first reported by Egan (Egan, Greenberg and Schulman, 1961a), Broadbent (Broadbent and Gregory, 1963a) and J.F. Mackworth (Mackworth and Taylor, 1963). While this type of analysis has proved to be successful in interpreting a number of different aspects of performance in monitoring tasks, some investigators have disputed its validity (Jerison, 1967a; Wiener, 1973). However, their objections have been directed, in the main, at the use of certain 'parametric' statistics (such as d' and $\beta$) of the theory of signal detectability (TSD). As will be apparent in the following chapters, a TSD analysis of monitoring behaviour does not necessitate the sole use of these statistics, but, when taken within the framework of statistical decision theory, implies a broader consideration of the decision and response processes involved in monitoring tasks than do the more 'traditional' (Broadbent, 1971) types of analysis. It is also shown in this thesis that decision theory enables, within limits, the interpretation of both 'discrete' measures of performance accuracy, as well as of 'continuous' measures of response latency. Furthermore, it is demonstrated that, under certain conditions, physiological correlates of behaviour, such as evoked potentials, may also be analysed within a decision theory framework.

Two broad approaches to performance assessment in monitoring tasks have thus been identified. The first, the task classification approach, has not been considered in relation to monitoring behaviour in any consistent manner. The second, the decision theory approach, has been applied with increasing success in a number of diverse research areas (see Broadbent, 1971; Swets, 1973), although there are a number of areas where the theory is difficult to apply in its present form, such as in continuous performance tasks, in multiple-choice situations, and other more complex and operational tasks (but see Ingleby, 1973; Luce and Green, 1972). However, one aim of this thesis is also to establish the limits of applicability of both the task classification and the decision theory approach to performance assessment.

The work reported in this thesis thus represents an investigation into the influence of, and the interaction between, task classification and decision processes in relation to performance in different monitoring situations. The first part of the thesis, comprising Chapters 1 to 6, outlines the theory and method behind the task classification and decision theory approaches, and the background to the experimental work, which is reported in Chapters 7 to 12. The literature review and outline of various theoretical considerations begins with the following chapter, in which task classification is introduced.

# C H A P T E R T W O

# TASK CLASSIFICATION

## 2.1 Task Classification and Performance

Task classification can be viewed as a method of systematization whereby a task or a group of tasks may be described on the basis of a limited number of classification categories. The classification categories may be defined by the common physical features of different tasks, or by the 'processes' which are assumed to mediate performance, or in a number of other ways. Common to all such taxonomic approaches, however, is the assumption that classification increases the ability to evaluate or predict performance as a function of selected independent variables.

### 2.1.1 Approaches to task classification

Over the years, a number of investigators have stressed the need for the development of reliable task classification systems (Cotterman, 1959; Melton and Briggs, 1960; Fleishman, 1967a, b, 1972, 1975a, b). Progress in this field has, however, been fairly slow. Many of the task classification systems which have been proposed have tended to be very specific in their application and limited in their objectives (Farina, 1969). On the other hand, some taxonomic approaches have been based on such broad categories of behaviour as to be of little use in the description of task performance within any of the categories. For example, Alluisi (1967) proposed a task classification system based on such categories as memory, vigilance, communication functions, and so on. Both intuitive reasoning and factor analytic research (e.g. Fleishman, 1967a) suggest a greater specificity of performance than that represented by these categories. Indeed, this thesis is concerned with the development of a task classification system <u>within</u> Alluisi's classification category of 'vigilance'.

A major research programme aimed at developing reliable taxonomies of human performance has been initiated by Edwin A. Fleishman and associates at the American Institutes for Research (AIR) in Washington D.C. Although still in their development phase, the task classification systems proposed by the AIR group have proved fairly successful in meeting the objectives of improving predictive and organizational capacity in the analysis of human performance in both laboratory and operational tasks. The probable reason for this is that the AIR group have gone beyond purely descriptive systems, and have attempted to establish an empirical basis for the various classification categories by using a correlational-factor analytic approach. Several classification schemes have been developed, including classifications based on human abilities (Theologus and Fleishman, 1971), task characteristics (Farina and Wheaton, 1971), information measures (Levine and Teichner, 1971), and task strategies (Miller, 1971). A number of reliability and validation studies for the first two classification systems have been reported (Fleishman, 1975a).

The reliability of a task classification system may be examined by investigating its efficiency and consistency in describing performance in different tasks. Ideally, a task classification system would be applied so that a more precise description of task performance may be achieved on the basis of the classification categories or 'dimensions' comprising each task, rather than on the basis of existing broad categories of behaviour such as memory, discrimination, vigilance, etc.

Somewhat inevitably, however, one may be faced with the problem of what to identify as a 'task dimension' (to be included in the task classification system), and what to identify as an 'independent variable'. For instance, the sense modality of stimulus presentation has often been identified as an important factor in monitoring behaviour (Davies and Tune, 1970). The problem emerges of deciding whether to include sense modality in the classification scheme, or to treat it as an independent variable. The problem cannot be dealt with in advance of an empirical evaluation of the taxonomy, and this depends on the specification of the classification categories in the first place, so that one is caught in a kind of vicious circle. This need not present any major difficulties however, if, on the basis of familiarity with the existing literature, one can identify some of the major task dimensions, so they may be included in the taxonomy, while other, seemingly less important ones may be included in the set of independent variables. Of course, some of these latter variables may be included in the taxonomy subsequently. A task taxonomy should not be regarded as a fixed system, but rather as a tool which may be progressively refined as further evidence is gathered. As Fleishman has pointed out, "..... individuals who attempt classification do not view the development of such a system, in and of itself, as an end. Rather, they view a system of classification as a tool to increase their ability to interpret, predict, or control some facet of performance.... This goal is to be achieved by seeking relationships between that which is classified (e.g. tasks, processes mediating performance, etc.) and selected variables of interest to a particular investigator" (Fleishman, 1975b, p. 50).

### 2.1.2 A brief survey of some classification schemes

Wheaton (1968) has reviewed some of the major taxonomic approaches in several areas within the behavioural sciences. He identified two major types of task taxonomy: specific or 'utilitarian' systems, and 'theoretical' systems. Utilitarian systems were defined as those which are designed to have utility for a specific, limited application area[1]. Wheaton noted that most of the taxonomies in this category are concerned with classification schemes for the evaluation of training methods (e.g. Gagné, 1962). He labelled these taxonomies 'specific' because they are not easily applicable to the analysis of performance as influenced by other independent variables. Wheaton also noted that such systems viewed the unit 'task' as comprising a larger and more complex set of operations than 'tasks' as referred to in the experimental literature on human performance. (See Fleishman, 1975b, for a discussion of the definition of a task).

The other type of task taxonomy identified by Wheaton (1968) was the so-called 'theoretical' classification systems. Such systems are assumed to be 'theoretically based' and 'autonomous', and capable of being applied to a wide range of task situations. The specific application does not, in principle, influence the form of the task classification system. These systems may vary according to the number, complexity and other features of the categories comprising each system; examples include the classification systems of Farina and Wheaton (1971), Guilford (1967), Miller (1967), Teichner (Teichner, 1972; Teichner and Olsen, 1969) and Fleishman (Fleishman, 1967a; Theologus and Fleishman, 1971). These classification schemes appear to be of greatest relevance to monitoring tasks, and, on the whole, are the only ones which have been empirically tested. We shall also see that this approach treats classification as an integral part of the development of a theory of monitoring performance.

Teichner and Olsen (1969) developed a classification system consisting of performance categories common to various tasks. Each performance category was defined by a dependent measure: the performance classes of 'searching', 'switching', 'coding' and 'tracking' were defined by the measures detection probability, reaction time, percentage of correct responses, and percentage time-on-target, respectively. The results of an evaluation study by Teichner and Whitehead (1971) indicated that these categories were useful in predicting performances on tasks incorporating these performance classes. It would appear however, that these categories are too broad to be capable of generating any but the most general and simple predictions. Furthermore, this taxonomy does not take into account a number of other dependent measures which might be employed in the description of these behaviours. Teichner has, however, claimed some success for his approach for the problem of deriving reliable empirical relationships from existing literature (Teichner, 1972, 1974). He points out that "a frequent complaint is that the scientific literature is neither reliable nor relevant enough.... I shall try to show that this complaint is not justified. Rather, the problem seems to be simply that the effort to use the literature properly has not been made by those who complain." (Teichner, 1972, p. 420).

The abilities classification system has been developed and described by Fleishman (1975 a, b). The rationale behind this approach is that certain basic abilities may be identified through a determination of the performance consistencies in different tasks within a correlational or factor analytic framework. Several such correlational studies have been carried out, and a comprehensive set of ability categories has been developed, covering perceptual, motor, comprehension and physical proficiency tasks (Theologus and Fleishman, 1971).

The ability categories developed by Fleishman have undergone considerable refinement in reliability and validity tests (Theologus and Fleishman, 1971). In particular the AIR group have shown that by the use of a refined, anchored rating scale, the ability category ratings could be used to predict performances on a number of (laboratory) psychomotor tasks. A similar type of result was obtained in a predictive study of operational performance, using tasks performed by Navy sonar operators (Wheaton, Shaffer, Mirabella and Fleishman, 1973). Integration studies of existing human performance data have also been facilitated with this approach. Two examples concern the areas of monitoring performance (Levine, Romashko and Fleishman, 1971) and the effects of alchohol on performance (Levine, Greenbaum and Notkin, 1973). In both studies, improved generalizations could be made regarding the effects of independent variables, and previously obscured effects could be identified. Furthermore, throughout an integrated series of studies carried out by the AIR group the same ability categories have consistently appeared and accounted for performance in a wide range of tasks.

Finally, a task characteristics classification system has been developed by Farina and Wheaton (1971). This taxonomic approach treats the task as incorporating a set of conditions which elicit performance, and the classification is based on the (objective) properties of the task itself. It is assumed that certain common characteristics can be identified across tasks. The properties may include the goal or purpose of the task, the relevant task stimuli, instructions, etc. Farina and Wheaton noted that this taxonomy was relatively free of the subjective and indistinct descriptions found in other task-oriented classifications. They concluded that although the taxonomy was 'explorative', the results of a 'post-diction' study were encouraging and suggested that further development might lead to a refined tool.

## 2.2 Monitoring Tasks: The Major Task Dimensions

Having briefly discussed some of the major task classification systems, we now turn to a consideration of taxonomic systems for monitoring tasks in this and the following section. In this section we shall examine some of the features of monitoring tasks and consider in further detail the classification categories proposed by Levine, Romashko and Fleishman (1971).

Since the original use of the Clock Test by Mackworth (1950), a number of different tasks have been used in the investigation of monitoring or vigilance behaviour. Despite the wide range of tasks used, very little attention has been paid to the problem of the development of taxonomic tools that would facilitate the evaluation of performance in these tasks. This is in keeping, as we have seen, with similar neglect in other human performance areas. It has been left to researchers outside the monitoring field to point to the need for developing such task classification systems (Fleishman, 1972).

The development of a taxonomy of monitoring tasks also assumes importance from the point of view of the evaluation of different theories of behaviour. In using two different tasks in studying the same psychological mechanisms, two investigators may obtain different empirical results supporting two conflicting theories; for instance, theory A may predict that monitoring performance over time is characterized by an increasing strictness in the criterion adopted by the subject, while theory B may predict that performance decrement is primarily associated with a decrease in perceptual sensitivity over time. Each investigator may assume his task to be fairly representative of monitoring tasks, and his theory to be supported since the data fit the respective predictions. In considering both studies, we are faced with the problem of reconciling the conflicting results, and deciding which theory is correct, or if both theories are correct, for certain non-overlapping situations only. There is thus a clear need for some sort of organizing taxonomic framework. In this regard, Davies and Tune (1970) have pointed out that "the theories advocated by different experimenters to explain their data appear, at least in part, to depend on the type of task they have used. Investigators who have used stimulating tasks tend to favour theories concerned with division of attention, while those who have used unstimulating tasks tend to favour theories based on the concept of arousal" (Davies and Tune, 1970, p. 11).

While comparatively little work has been done towards the development of a taxonomy of monitoring tasks, a number of investigators have specified different task features as important determinants of monitoring performance. Some of these have been included into theoretical approaches to monitoring behaviour. Others, such as the 'coupling' concept of Elliott (1960) have been proposed as features which are present in differing degrees in different monitoring tasks. Some of these 'dimensions' are listed in Table 2.1. Most of these task dimensions have not been evaluated within a taxonomic framework. Because of the diversity of the supporting literature, a synthesis is difficult, but a closer examination of the range of task dimensions may prove fruitful.

The first three dimensions listed in Table 2.1 might be termed 'intuitive', in that they are common to a large number of tasks, and cannot be attributed to a single author. On an intuitive basis, accordingly, the first item in Table 2.1, the sense modality of stimulus presentation, might be included in any general task classification system, since modality specific effects have been reported in a number of different behavioural areas, including monitoring behaviour. In general, it has been noted that auditory monitoring performance is superior to visual monitoring performance[2]. A comparison of performances across modalities has been considered to imply a critical test of the hypothesis that monitoring performance is mediated by a 'common central process' (see Davies and Tune, 1970, pp. 30 - 37). It has been commonly assumed, from the results of some early studies, that the 'centrality' hypothesis does not hold since these studies found that monitoring performances are not correlated across sense modes (Buckner and McGrath, 1963b; Pope and McKechnie, 1963). However, it has also been reported that performances are uncorrelated <u>within</u> modalities (Baker, 1963). Moreover, some more recent studies have found that if certain task factors are controlled, a pattern of consistency in the individual differences in different monitoring tasks emerges (these studies are discussed in greater detail in Chapter 3).

Four studies (Gunn and Loeb, 1967; Hatfield and Loeb, 1968; Hatfield and Soderquist, 1970; Loeb and Binford, 1971) have investigated whether one of these task factors might be related to 'coupling', but the results were generally inconclusive (see 3.2). Elliott defined coupling as "an arrangement of the task so as to ensure that the signal put out by the experimenter gets into the appropriate sensory input of the vigilant subject" (Elliott, 1960, p. 360). It is intimately related to sense modality; in general, visual displays may be said to be poorly coupled, since it cannot be ascertained that the observer receives a representation of the signal on each occasion (e.g. he may look away). With an auditory display, the signal is always peripherally received (e.g. at the cochlea; see Picton, Hillyard, Galambos and Schiff, 1971).

The second item in Table 2.1, stimulus source complexity, may also be included on an intuitive basis in a task classification system. Although the relationship between performance trends over time in single and multi-source tasks remains to be fully explored, a number of authors have stressed that this dimension effectively dichotomizes the range of monitoring tasks (Howell, Johnston and Goldstein, 1966; Jerison and Wallis, 1957; Johnston, Howell and Williges, 1969), and that 'simple' and 'complex' monitoring performance need to be considered separately.

The third item in Table 2.1 refers in particular to the experimental methods of the theory of signal detectability (TSD). Conventionally, only simple 'signal present' or 'Yes' responses have been required in monitoring tasks, but in some studies responses to each stimulus event have been required, in a manner similar to the Yes/No paradigm of TSD (Davies, Lang and Shackleton, 1973; Parasuraman and Davies, 1975; Whittenburgh, Ross and Andrews, 1956). The rating method has also occasionally been

| | DIMENSION | EXAMPLES | SOURCE |
|---|---|---|---|
| 1. | Sense modality | Visual/Auditory/Vibrotactile/etc. | |
| 2. | Source complexity | Single/Multi-source | |
| 3. | Response type | Single/Binary/Rating/etc. | |
| 4. | Degree of coupling | Loosely/Closely coupled | Elliott (1960) |
| 5. | Signal duration | Limited/Unlimited hold | Broadbent (1958) |
| 6. | Degree of pacing | Paced/Unpaced | (e.g. Wilkinson, 1961) |
| 7. | Time course of events | Discrete-slow/Discrete-fast/ Continuous | (e.g. Davies and Tune, 1970) |
| 8. | Attention requirement | Intermittent/continuous | Simpson (1967) |
| 9. | Stimulation value | (A taxonomy for all continuous performance tasks) | Bergum (1966) |
| 10. | Perceptual load | Sub-optimal/Optimal/Super-optimal | Poulton (1960) |
| 11. | Required abilities | Perceptual Speed/Flexibility of closure | Levine, Romashko and Fleishman (1971) |

TABLE 2.1 Some proposed task 'dimensions' of relevance to a taxonomy of monitoring tasks (arrows connected by unbroken and broken lines indicate that dimensions are related either closely or indirectly, respectively).

used in monitoring studies (Broadbent and Gregory, 1963a; Loeb and Binford, 1964; Milosevic, 1975). The evidence from detection and recognition tasks is not entirely conclusive that these various procedures are equivalent; surprisingly, however, the differences across procedures are somewhat smaller for monitoring tasks, as we shall see in Chapter 4 (see also Appendix D).

In some early studies on monitoring behaviour, some tasks were used which were such that the signals for detection remained present until detected. Examples of such 'unlimited-hold' tasks are the twenty-dials and twenty-lights tasks (Broadbent, 1950, 1951, 1954). A distinction between tasks in which signals are presented only briefly (transient or limited-hold tasks) and unlimited-hold tasks was made in the context of Broadbent's (1958) filter theory. Broadbent proposed that the latter type of task did not suffer from the adverse effects of deviations in attention over a monitoring period[3].

Perhaps a more compelling basis for the distinction between transient and unlimited-hold tasks lies in the available information processing strategies in either case. Laming (1973) has distinguished tasks in which the sensory information is limited and unreliable (in the statistical sense), from those where it is 'unlimited'. This distinction appears to include Broadbent's classification of tasks by signal duration. In transient tasks, it is reasonable to assume that the observer's information processing strategies follow those assumed by fixed-sample theories such as TSD (see Chapter 4). On the other hand, in unlimited-hold tasks, it is likely that the observer would use a multi-sample detection strategy that would enable, in principle, 'perfect' performance to be achieved. Various sequential-decision strategies may be proposed for this case (see Laming, 1973).

It is therefore important to distinguish between tasks on the basis of which type of analysis, e.g. TSD or one based on sequential-decision models, is appropriate. In Chapter 4, we shall see that TSD, as a specific form of the more general statistical decision theory, is capable of being applied in many cases, but finds its greatest relevance for tasks where the availability sensory information is limited, that is, in the so-called 'data-limited' tasks (Norman and Bobrow, 1975).

A task dimension which has important implications for theories of monitoring, in particular to Broadbent's (1958) filter theory and other 'attentional' theories, is the time course of events. This may vary in a continuum from slow, to fast, to continuous presentation. With tasks having 'discrete' events, filter theory predicts that performance will not suffer from the effects of deviations in attention when the presentation rate is low, but will do so if the rate is high or continuous. Essentially the same distinction applies to the classification proposed by Simpson (1967), on the basis of the 'attention requirement' of different tasks; he proposed that performance decrement is observed only for tasks requiring 'continuous' attention, that is, in high or continuous rate tasks. Despite some inadequacies in the filter theory approach to monitoring performance, we shall see that this task dimension is in fact an important one, and that in certain cases the rate of stimulus presentation is an important determinant of performance.

A general taxonomy for the analysis of performance on continuous work tasks has been proposed by Bergum (1966). This was based on the general conceptual framework of arousal or activation theory (e.g. see Duffy, 1962; Welford, 1962), and distinguishes tasks on the basis of their 'total stimulation value'. Bergum applied his classification scheme to a wide range of continuous performance tasks, including production-line, assembly and other monotonous tasks. As such, it is of limited value for the purposes of a taxonomic analysis of monitoring tasks, since the task dimensions are too broad. The 'stimulation value' dimension is also rather imprecisely defined, and it is difficult to provide an independent measure of it. Nevertheless, Bergum's analysis is important for providing one of the first attempts at task classification for monitoring and other prolonged tasks, and his analysis did allow some measure of predictive capacity across classification categories. His concept of 'stimulation value', and the proposal of an optimal value for efficient performance is related to similar conceptualizations such as the inverted-U relation (Corcoran, 1965), the Yerkes-Dodson Law (e.g. Broadbent, 1965) and Poulton's (1960) concept of 'optimum perceptual load'.

## 2.3 Classification Schemes for Monitoring Tasks

We have now examined a number of the task 'dimensions' listed in Table 2.1 as possible candidates for inclusion in a taxonomy of monitoring tasks. However, the last item listed in Table 2.1 has not been discussed thus far. This is the abilities classification system of Fleishman (1972) and Levine et al. (1971).

### 2.3.1 Abilities classification

The abilities classification approach is one of an inter-related set of taxonomic approaches to the evaluation of human performance which were briefly mentioned in 2.1. The abilities classification system has been described in detail elsewhere (Fleishman, 1972, 1975a, b; Theologus and Fleishman, 1971). Briefly, this system proposes that certain basic abilities can be identified as the major determinants of performance in a variety of tasks. The abilities are inferred from factor-analytic analyses of performance consistencies in different tasks. Theologus and Fleishman (1971) identified four main ability 'domains': cognitive, perceptual-sensory, physical proficiency, and psychomotor. Within each of these domains, several ability categories were identified; this forms the major point of departure from other taxonomies, which have often only specified a few, broad categories. In the 1971 version of the abilities taxonomy, 37 basic abilities were postulated. Each was derived from factor loadings on tasks sampling the ability, and refined rating scales were developed so that the 'ability requirements' of different tasks could be quantitatively expressed. Each of these scales were 'anchored' with empirically determined scale values corresponding to the ability requirements of different tasks. Figure 2.1 illustrates the assessment scale for the rate control ability, which is relevant to performance in different tracking situations.

In their application of the abilities classification to the organization of a portion of the literature on monitoring performance, Levine, Romashko and Fleishman (1971, 1973) considered two ability domains, the perceptual-sensory and cognitive domains, to be of relevance to monitoring tasks. The former domain was accorded greater importance, and two 'primary abilities' were selected from this domain: perceptual speed and flexibility of closure. Perceptual speed refers to the ability to rapidly compare successively presented patterns or stimulus configurations for identity or degree or similarity. The sensory patterns to be compared occur within the same sense modality and not between modalities. Flexibility of closure refers to the ability to detect or identify a previously specified stimulus configuration which is part of a more complex sensory field. Both the relevant stimulus configuration and the 'noise field' occur within the same sense modality (Levine et al., 1971, p. 9).

Levine et al. also considered two 'secondary abilities', selective attention, and time sharing. Definitions of these abilities are as follows:

RATE CONTROL This is the ability to make timed, anticipatory motor adjustments relative to changes in the speed and/or direction of a continuously moving object. The purpose of the motor adjustments is to intercept or follow a continuously moving stimulus whose speed and/or direction may vary in an unpredictable manner.

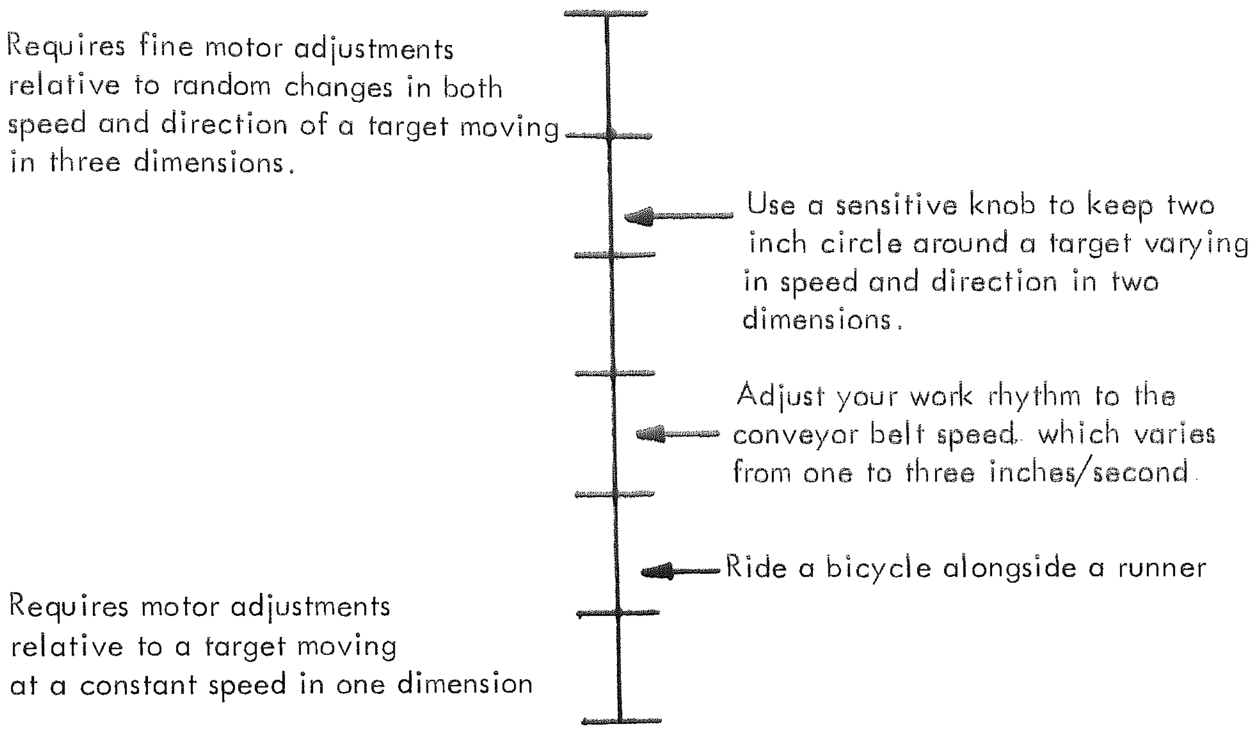


FIGURE 2.1 Anchored rating scale to assess requirement for Rate Control ability (after Theologus, Romashko and Fleishman, 1970, p. 192).

Selective attention

> The ability to perform a task in the pres ence of distracting stimulation or under monotonous conditions without loss of efficiency. When distracting stimulation is present in the task situation, it is not an integral part of the task being performed, but rather is extraneous to the task and imposed on it. The task and the irrelevant stimulation can occur either within the same sense or across senses. Under conditions of distracting stimulation, the ability involves concentration on the task being performed, and filtering out of distracting information. When the task is being performed under monotonous conditions, only concentration on the task being performed is involved.

Time sharing

> The ability to utilize information obtained by shifting between two or more channels of information. The information obtained from these sources is either integrated and used as a whole or retained and used separately.
>
> (Levine et al., 1971, pp. 9-10)

Levine et al. classified 53 monitoring tasks used in the literature on each of the four abilities; each task was categorized by the predominant ability required for efficient performance. Most of the tasks (50 out of 53) fell into either the perceptual speed or flexibility of closure ability categories. Very few tasks could be assumed to require the selective attention or time sharing abilities predominantly (hence their connotation as 'secondary'). It would therefore appear that these two abilities do not entirely satisfy Fleishman's (1967a) criteria for a reliable classification system, namely that the classification categories should not be too generalized or too specific. The selective attention category appears to fail to satisfy the first of these requirements, in that demands on selective processing may be imposed not only in monitoring tasks, but also in many other tasks; and one would expect to find performance specificity within each of these categories, as for the vigilance and memory categories of Alluisi (1967).

However, the two primary abilities, perceptual speed and flexibility of closure appear to be representative of important task features, and may provide a neat dichotomous dimension for the classification of monitoring tasks. Levine et al., (1971) found that there were some differences between tasks requiring these abilities in both the mean performance trends over time and in the effects of certain independent variables. One result was that for tasks requiring flexibility of closure, the percentage of correct detections declined with time on task up to a certain point of time (about $\frac{3}{4}$ hour into the vigil), and then increased, while for tasks requiring perceptual speed performance did not reverse, but levelled off. This is illustrated in Figure 2.2, which suggests that tasks requiring perceptual speed may be more susceptable to performance decrement than tasks requiring flexibility of closure. Furthermore, the impact of independent variables such as signal rate, sensory modality and knowledge of results, were found to be a function of the ability requirements of tasks.

While the theoretical import of these results is not immediately clear [4], they do point to the feasibility of the abilities classification approach as a useful tool for the integration and generalization of research findings. Further research appears to be needed to clarify the nature of the perceptual speed and flexibility of closure categories in relation to monitoring performance. Some further insight may be gained by examining the tasks given by Levine et al., (1973) as examples of tasks requiring these abilities. For perceptual speed, the task used by Eason, Beardshall and Jaffee (1965) was described. In this task subjects were required to press a switch when they detected a flash of light appearing in a 1-inch circular hole for 0.8 seconds. The light was normally flashed for 0.5 seconds, the interval between flashes being 3 seconds. For flexibility of closure the task of Adams (1956) was described. In this task subjects were required to detect a 2 millimetre blip of light

Aston University

Content has been removed for copyright reasons

FIGURE 2.2 Median percentage correct detections as a function of time on task for tasks requiring perceptual speed and flexibility of closure (from Levine, Romashko & Fleishman, 1971).

appearing in the centre of a 5-inch white screen.

Examination of these tasks reveals that the major feature of the task requiring perceptual speed is that a detection of a change in the duration of a stimulus with respect to preceding stimuli is required, while for the task used by Adams (1956), detection of a previously specified stimulus configuration is required. The perceptual speed and flexibility of closure categories may thus be easily inferred from the signal and task characteristics of each monitoring task.

### 2.3.2 Classification by task characteristics

The task dimensions of perceptual speed and flexibility of closure emerge as a possible starting point for a taxonomy of monitoring tasks. These two categories may be used to classify a number of monitoring tasks, which may subsequently be examined for performance consistencies and differences across categories. Table 2.2 lists some tasks which may be classified in this manner. Examples of both visual and auditory monitoring tasks are given.

It will be noted that of the tasks listed in Table 2.2, some are 'discrete' tasks, and others are 'continuous' tasks. This distinction was earlier pointed out to be an important one. We have also considered the modality of monitoring displays as an important task dimension; and in the following Chapter, its importance is underscored as a result of a consideration of the relationships between performance consistency and modality in monitoring tasks.

In considering the abilities classification system, we have rejected the 'selective attention' category because of its generality and imprecise definition. What of the time sharing category? This category is also

| | PERCEPTUAL SPEED | FLEXIBILITY OF CLOSURE |
|---|---|---|
| VISUAL TASKS | 1. Detect a change in the intensity of intermittent light flashes (Broadbent and Gregory, 1963a; Hatfield and Loeb, 1968).<br>2. Detect increase in deflection of meter needle (Baker, 1963).<br>3. Detect increase in duration of intermittent light flashes (Williges, 1973). | 1. Detect specified configuration in complex pattern of letters (Adams, Humes and Stenson, 1962).<br>2. Detect a disc of specified hue in display of six discs (Colquhoun, 1961).<br>3. Detect a blip of light appearing occasionally on a screen (Adams, 1956). |
| AUDITORY TASKS | 1. Detect a change in the intensity of intermittent noise bursts (Hatfield and Soderquist, 1970).<br>2. Detect a sequence of digits in a series of auditorily presented numbers (Bakan, 1959).<br>3. Detect decrease in duration of intermittent tone (Deaton, Tobias and Wilkinson, 1971). | 1. Detect a tone embedded in noise bursts (Hartley, Olsson and Ingleby, 1973).<br>2. Detect a brief interruption in continuous white noise (Ware, Sipowicz and Baker, 1961).<br>3. Detect occasional tone in a background of continuous white noise (Colquhoun, Blake and Edwards, 1968). |

TABLE 2.2 Examples of visual and auditory monitoring tasks classified on the perceptual speed and flexibility of closure task categories.

fairly broad; and it may include time sharing both within a task (e.g. between different sources of a multi-source task), and between a main task and a secondary task, such as tracking (Wiener, 1975a), or memory and encoding tasks (Tyler and Halcomb, 1974). To limit our classification to more specific features of the monitoring situation, we need only consider a task dimension based on source complexity; that is, whether the task has a single or multiple number of stimulus sources.

We may now propose a task classification system based on the important task dimensions we have considered. Although two of the classification categories form part of Fleishman's (1972) abilities classification, in

| DIMENSION | EXAMPLES |
|---|---|
| Type of signal discrimination | 'Speed'/'Closure' |
| Sense modality | Visual/Auditory |
| Source complexity | Single/Multi-source |
| Time course of events | Discrete-slow/Discrete-fast/ Continuous |

TABLE 2.3 The components of a proposed task classification system for monitoring tasks.

the absence of a correlational or factor analytic base for these categories, we will refer mainly to the 'task dimensions' or 'task characteristics'. Conceptually, the relation between task features and hypothetical abilities may be retained, but in practice, we shall talk in terms of a task characteristics classification system.

Table 2.3 displays some of the components of such a system, which will be further examined in the experimental studies reported in this thesis. It is proposed that this system is sufficiently comprehensive to cover a wide range of tasks, while at the same time being small enough to ease empirical investigation. It will be noted that the ability categories of perceptual speed and flexibility of closure have been taken as part of the task dimension 'type of signal discrimination', since they may be inferred from such an examination of the task features. In forthcoming chapters, these categories will often be referred to as the 'speed' and 'closure' categories. Reference will also be made to 'speed tasks' and 'closure tasks'. Unless otherwise stated, this should be taken to mean that the category represents a dominant feature or 'ability requirement' of each task.

While a number of other task dimensions may be identified, this task classification system may be taken to represent a first approximation to a monitoring task taxonomy, so as to limit the scope of the initial enquiry. One may also include the dimensions of pacing and signal duration in subsequent modifications. We believe, however that the components of Table 2.3 represent the most important of the many task dimensions we have considered. The one exception is the dimension of visual search, which we have not considered before. This dimension obviously represents a strong candidate for future research into classification systems for monitoring tasks, since it is an important component of many inspection and other industrial tasks. Again, it has been neglected to limit the breadth of the empirical investigations.

The task classification system presented here has arisen primarily out of considerations of the previous literature, and its evaluation in terms of the reliability of various proposed task categories. The taxonomy is intended to be a preliminary, exploratory tool, and one whose reliability is yet to be empirically determined. It does not exclude further extension and revision in the light of subsequent empirical evidence.

## 2.4 Notes to Chapter Two

1. Utilitarian task classifications include the task taxonomic approach known as 'task analysis' (e.g. Annett and Duncan, 1967).
2. This generalization does not always hold (for exceptions, see Colquhoun, 1975; Kennedy, 1971). As Elliott (1960) has pointed out, comparisons across modalities are only valid if the inputs to the different sensory channels are 'equalized', so that the observed performance differences can be attributed to modality factors alone.

3. The empirical data available in 1958 generally supported the filter theory approach to monitoring performance. More recent evidence, however, has shown this approach to be untenable, except for certain task situations (see Chapter 5).

4. A major difficulty in the interpretation of these results is the lack of an analysis in terms of independent sensitivity and bias measures (this was not done presumably because very few studies in the literature, at the time, reported false positive error rates).

# CHAPTER THREE

## INDIVIDUAL DIFFERENCES IN MONITORING PERFORMANCE:
## A TASK CLASSIFICATION ANALYSIS

## 3.1 General Features of Individual Differences

In our discussion of the relation between task classification and performance, we have been generally concerned with the mean performance associated with groups of individuals. It is also of interest, however, to examine aspects of task classification from the point of view of individual performances, or the variability in performance between individuals. This is especially true for monitoring tasks, since performance in such tasks is usually found to vary quite considerably between subjects. Marked individual differences in performance is one of the commonest findings in research on monitoring behaviour (Davies and Tune, 1970).

J.F. Mackworth (1969) has identified four major ways in which individuals may differ in performing monitoring tasks. She proposed that subjects may differ 1) in their ability to detect a signal under alerted conditions, or in their basic reaction time, 2) in their change in performance over time, 3) in their chosen criterial level for responding, and 4) in their level of arousal or activation. Studies of individual differences in monitoring performance have generally sought to examine the inter-relationships of these factors (e.g., the relationship between individual differences in arousal level and detection rate), and in obtaining their psychological and physiological correlates.

### 3.1.1 Reliability of individual differences

N.H. Mackworth (1950) was one of the first investigators to report the existence of large individual differences in monitoring performance. He noted that these differences were fairly reliable, in that they were maintained consistently both between successive periods of a watch, and between watches. The variability in performance in a monitoring task is generally of about the same order as the mean level of performance. Figure 3.1 illustrates typical values of the mean and standard deviation

FIGURE 3.1 Mean and standard deviation of correct detections as a function of time on watch in a monitoring task (from Buckner, Harabedian and McGrath, 1960).

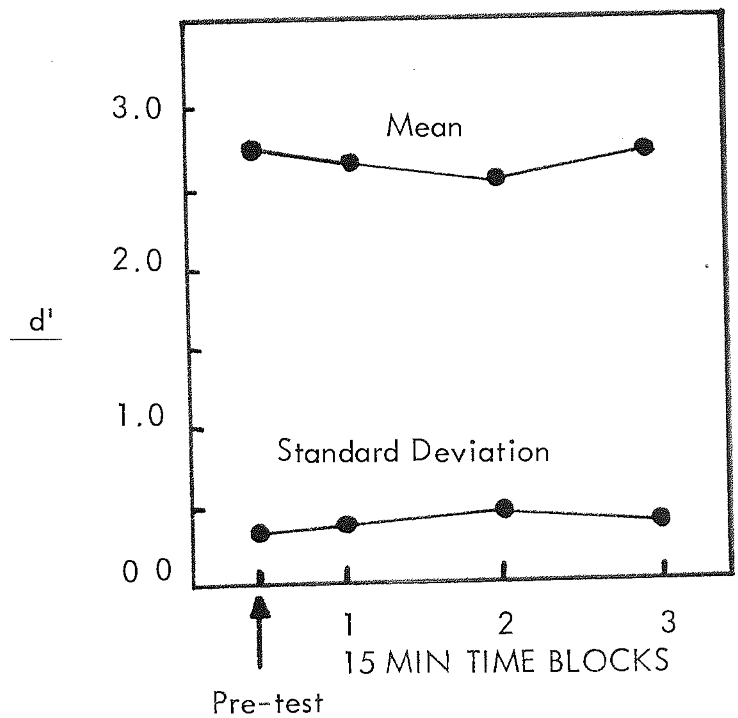


FIGURE 3.2 Mean and standard deviation of values of d' as a function of time on task (from Experiment 1, Chapter 7).

of the percentage of correct detections for performance in a task used by Buckner, Harabedian and McGrath (1960). A similarly consistent but somewhat smaller variability in performance is apparent in Figure 3.2, in which values for the d' measure of performance have been plotted from an experiment reported in Chapter 7.

The detection rate is usually found to be fairly reliably related to the initial detection level. Mackworth (1950) reported, for instance, that the rate of decrement in correct detections over a vigil was dependent upon the initial level of detection. Subjects who detected a high number of signals at the start of the vigil exhibited less decrement than those subjects who had a poor initial performance level. In one of several experiments using the Clock Test, Mackworth (1950, Clock Test 10) also found that subjects who had already been monitoring the Clock for half an hour had a greater decrement subsequently than subjects who began the vigil 'fresh'.

The finding that the rate of decrement in correct detections is dependent on the initial mean level of performance has also been reported in a number of experiments by Buckner (Buckner, 1963; Buckner et al., 1960). In these experiments subjects performed on 'alerted' short-duration pre- and post-monitoring session tests; and it was found that pre-test and main session performances were highly correlated. Teichner (1974) has discussed the results of studies on simple visual detection tasks in a similar manner. By collating data from existing studies, he was able to show a relationship between the rate of decrement in correct detections and the initial level of detections.

While individual differences are reliable within a monitoring session, a few studies have also found that individual performances are highly correlated over several sessions spent working at the same task (Buckner et al., 1960, obtained reliability coefficients in the range .72 to .91 for monitoring sessions separated by up to six intervening sessions). It may thus be taken as established that individual differences in monitoring performance are maintained consistently within a monitoring session, between pre-test and main session, and between different sessions, at least as far as detection performance is concerned (see also Appendix B).

### 3.1.2 Correlates of individual differences

Given the reliability of individual differences in monitoring performance, a number of investigators have sought to identify their correlates. On the whole, however, such efforts have been fairly unsuccessful. Mackworth (1950) originally reported that neither intelligence nor visual acuity was related to monitoring performance. A number of similarly unsuccessful studies have since been reported (see Mackworth, 1969, pp. 111-120; Wiener, 1975b, pp. 105-107).

Despite this generally negative picture, however, some studies have found that the introversion-extraversion dimension of temperament might provide a possible correlate of monitoring performance; the general finding has been that introverts detect a greater number of signals and have a smaller decrement than extraverts (Bakan, 1959; Davies, Hockey and Taylor, 1969; Keister and McLaughlin, 1972), for temperament as assessed, respectively, by the Heron Personality Inventory (Heron, 1956), the Maudsley Personality Inventory (Eysenck, 1959) and the Eysenck Personality Inventory (Eysenck and Eysenck, 1964). These studies have generally interpreted monitoring performance within an arousal framework and related personality, autonomic arousability or reactivity, and performance (see Broadbent, 1963). The exact form of the relationship, however, remains to be demonstrated. It should be noted that the previously mentioned studies have not generally reported the relationship between temperament and <u>bias-free</u> performance measures; hence it is not clear whether the apparent superiority of introverts (higher detection rate) represents a genuine superior perceptual sensitivity (d'). Davies and Hockey (1966) and Davies et al. (1969) found that while introverts detected a greater number of signals than extraverts, they also tended to make more false alarms. In a related context, Parasuraman (1975a) showed that subjects with a high level of electrodermal reactivity made a significantly greater number of detections (at all confidence levels) on a short-duration tone discrimination task than low-reactive subjects; however this apparent superiority was not due to a superior ability to discriminate signals, but a result of significantly 'riskier' criterion levels adopted by high-reactive subjects. It thus remains to be demonstrated whether introverts and extraverts differ in their sensitivity or their chosen criterial level, or both.

The results included in this section do not, then, allow for any firm conclusions to be made regarding the relation between temperament and monitoring performance. It therefore appears that while individual differences are consistently present in monitoring performance, it is difficult to identify their correlates. It has been suggested, furthermore, that it is relatively futile to seek such correlates since individual differences are highly specific from task to task (Baker, 1963). This view was mainly based on the results of some early studies which found that inter-task correlations in monitoring performance were low and not significantly different to that to be expected by chance (Baker, 1963; Buckner et al., 1960; Pope and McKechnie, 1963). However there are some reasons why this is probably too facile a view. We shall discuss these in some detail in the following section, in which we shall examine the implications of task classification in relation to performance consistency in different monitoring tasks.

## 3.2 Inter-Task Consistency of Individual Differences

We have noted that one of the commonest findings in monitoring performance is that of large individual differences. In some early studies, it was reported that while performance differences between subjects are maintained consistently both within and over sessions spent working on the same task, individual differences are not consistent when subjects work with different tasks (Baker, 1963; Buckner, Harabedian and McGrath, 1960). This finding has generally been taken to be a characteristic of monitoring performance, the conclusion being that individual differences in monitoring performance are highly task specific (Buckner and McGrath, 1963b; Mackworth, 1969). However, this view may be challenged on the basis of the findings of more recent studies, in which generally high and significant correlations have been obtained between performances in different monitoring tasks, at least for some performance measures (Gunn and Loeb, 1967; Hatfield and Loeb, 1968; Hatfield and Soderquist, 1970; Loeb and Binford, 1971; Sverko, 1968; Tyler, Waag and Halcomb, 1972).

These more recent studies have generally controlled task factors more closely by equating tasks for the type and difficulty of signal discrimination. It is also noteworthy that in a study in which significant inter-task correlations were obtained for all the performance measures taken (Sverko, 1968), the monitoring tasks used were such that the dominant ability required for efficient performance ('numerosity', or the ability to detect a specified number of temporally spaced discrete stimuli) was different to that normally encountered in monitoring situations but was <u>common to all</u> the tasks used. It thus appears possible that tasks which make similar demands on the subject are more likely to share common performance variance than tasks which do not, although none of the existing studies provide a direct confirmation of this view. In later chapters, we shall be concerned with how the similarity of demands of different tasks may be assessed.

These studies may therefore be divided into two general categories: the earlier, or Type A studies, which report little or no correlation in performance between different monitoring tasks, and the more recent, or Type B studies, which report generally high and significant inter-modal correlations in monitoring performance. We have seen that the feature distinguishing these studies is the closer control of task factors in the Type B studies. These studies are also the only ones reporting correlations for the d' measure of performance. We shall accordingly consider the Type A and B studies separately. The correlation coefficients reported in both types of studies are tabulated in Table 3.1.

### 3.2.1 Type A studies

In one of the earliest studies on inter-task correlations in monitoring performance, Buckner et al., (1960) reported low, positive correlations in performance between visual, auditory and combined audio-visual tasks (see Table 3.1). In the visual task, Navy personnel were required to detect slight increments in the brightness of an intermittent light, while in the auditory task detection of small increments in the intensity of intermittent 750 Hz pure tones were required. Three audio-visual tasks combining the visual and auditory tasks under three degrees of signal redundancy were also employed. Buckner et al reported correlations for the correct detections measure only, so we cannot determine whether the low correlations were indicative of inconsistencies in subjects response criteria, or of a genuine inconsistency in detection sensitivity.

Pope and McKechnie (1963) also found no correlation in performance between visual and auditory monitoring tasks. In the visual task subjects were required to detect a low-intensity spot of light on a frosted glass

| | Performance Measures | | | | |
|---|---|---|---|---|---|
| | Hits | FAs | d' | Log β | DL |
| Type A Studies | | | | | |
| Buckner et al. (1960) | .24 | | | | |
| Buckner and McGrath (1963a) | | | | | |
| Visual task | .20 | | | | |
| Audio-visual tasks: | | | | | |
| Redundant | .22 | | | | |
| Partially redundant | .54 | | | | |
| Non-redundant | .23 | | | | |
| Gruber (1964) | .18 | .15 | | | .42 |
| Pope and McKechnie (1963) | -.11 | | | | |
| Type B Studies | | | | | |
| Gunn and Loeb (1967) | | | | | |
| Experiment I | .11 | .57* | .48* | .52* | |
| Experiment II | .21 | .79* | .68* | .78* | |
| Hatfield and Loeb (1968) | | | | | |
| CC Visual Task | .65* | .15 | .34 | .33 | .76* |
| LC Visual Task | .48* | .21 | .27 | .32 | .76* |
| Hatfield and Soderquist (1970) | | | | | |
| CC Visual Task | -.08 | .52* | -.09 | .50* | .78* |
| LC Visual Task | .47* | .93* | .14 | .96* | .71* |
| Loeb and Binford (1971) | | | | | |
| CC Visual Task | .65* | .15 | .36 | .56* | |
| LC Visual Task | .67* | .29 | .88* | .72* | |
| Sverko (1968) | | | | | |
| Light Flashes | .77* | .57* | | | |
| Pointer Deflections | .66* | .58* | | | |
| Tyler et al. (1972) | .80* | .77* | .87* | .61* | |

TABLE 3.1 Visual-auditory correlation coefficients in Type A and B studies (* = $p < .05$ or better; CC = Closely coupled; LC = Loosely Coupled; Hits = Correct Detections; FAs = False Alarms; DL = Detection Latency; rank order correlations for data of Gruber computed by Davies and Tune, 1970; partial correlations reported by Tyler et al., 1972).

display, while the auditory task involved the detection of a low intensity (55dB) pure tone (1350 Hz) in a continuous steady 90dB, 500 Hz tone. Both the product-moment and rank-order correlations were low and negative, each being - .11.

Gruber (1964) tested subjects on a visual task in which the signal was the disappearance of a faint horizontal line on a CRT display, and an auditory task requiring the detection of .1 second interruptions of a 1000 Hz pure tone in noise. Gruber did not compute correlation coefficients for his data, but this was done by D.R. Davies, who found that low visual-auditory correlations were obtained for three measures, correct detections, false positives and detection latency (Davies and Tune, 1970, p. 35).

A low correlation in performance was also reported by Baker (1963), who, however, unlike earlier investigators, used two visual tasks. Baker found that a low correlation was obtained between two dissimilar tasks, the Continuous Clock and Dial Tasks (for descriptions, see Mackworth, 1970, p. 158); however he also reported that if two tasks differed only in the signal duration, then a high correlation was obtained. Baker suggested that the low correlation was obtained because the tasks were either differentially difficult, or dissimilar (since the Clock Task requires visual search movements while the Dial Task requires continuous visual fixation), and concluded that individual differences in monitoring performance are almost wholly task specific.

### 3.2.2 Type B studies

The study reported by Gunn and Loeb (1967) can be taken to be the first study belonging to this category. They tested the same subjects on a visual and auditory task in two fairly short (15 min) monitoring sessions. Detection of intensity increments in a background of either light or white noise pulses were required in the visual and auditory tasks, respectively. Unlike earlier investigators, Gunn and Loeb equated the two tasks for difficulty by matching d' values obtained in short, 'alerted' pre-tests. Both d' and $\beta$, which are measures of perceptual sensitivity and the response criterion respectively in the TSD model, were found to be significantly correlated across tasks; however the correlations for the percentage of correct detections were not significant in either experiment (see Table 3.1). This latter result, when compared to those of the Type A studies, indicates the importance of using a bias-free index of detection performance such as d'.

A further series of experiments by M. Loeb and his colleagues have given results which are, on the whole, similar to those of Gunn and Loeb (1967), although some apparently inconsistent results have also been obtained (Hatfield and Loeb, 1968; Hatfield and Soderquist, 1970; Loeb and Binford, 1971). These experiments used visual and auditory tasks very similar to those used by Gunn and Loeb (1967). In addition, all three studies also employed a 'closely coupled' visual task (see 2.2), which was identical to the 'loosely coupled' task except that subjects had their eyes taped to reduce the contribution of 'observing responses'. Significant correlations in performance between visual and auditory tasks were obtained for subsets of the five performance measures, but the pattern of correlations is not consistent between studies (see Table 3.1). Hits and false positives were correlated across tasks in 8 and 7 cases out of 11, respectively; and d' and log $\beta$ were correlated significantly in 4 and 7 cases out of 9, respectively.

Task difficulty was controlled in the Type B studies by matching group d' values between tasks, for the experiments reported thus far. Loeb and Binford (1971) and Tyler at al. (1972) however, equated task difficulty by matching <u>individual</u> d' values. Loeb and Binford did not obtain high correlations for each performance measure they took, and their results do not resolve the inconsistencies in the pattern of correlations in the previous Type B studies. Tyler et al. (1972) tested subjects on three monitoring tasks on the same day, separated by 3-min. rest periods. They used a visual task in which the signal was an increase in the duration of an intermittent red light; in their auditory task a similar discrimination for intermittent 1 KHz tone stimuli was required. Uniformly high and significant correlations in performance between the visual and auditory tasks were obtained, for each performance measure taken (see Table 3.1). On the basis of their results, Tyler et al. concluded, in contrast to Baker (1963), that individual differences in monitoring performance are not task specific, but are mediated by a 'common vigilance factor'.

Finally, Sverko (1968) tested subjects on three rather unusual visual, auditory and electrocutaneous tasks. In each task, subjects were asked to monitor a series of pulses in trains of two to seven pulses, and to indicate 'signal present' when two successive trains contained the same number of pulses. Pointer deflections and light flashes were the stimuli used for the two visual tasks, while click and weak electrical shock stimuli were used for the auditory and cutaneous tasks, respectively. Inter-task correlations in the range .57 to .84 were obtained, all correlations being significantly different from zero.

### 3.2.3 Discussion of type A and B studies

From an examination of the reports in the literature on inter-task correlations in monitoring, it is clear that these fall into two main categories, Type A and Type B studies. These studies may be distinguished by the fact that the Type B studies have generally controlled task difficulty and compatibility factors more closely than the Type A studies. The overall finding that uniformly low and nonsignificant performance correlations are reported in the Type A studies, while generally high and significant correlations are reported in the Type B studies emphasizes the importance of these distinguishing factors.

While high correlations were reported in all the Type B studies, low and nonsignificant correlations were also obtained for some performance measures, and overall, the pattern of correlations shows little consistency between studies. The most consistent result was that log $\beta$ was highly correlated across tasks in all the Type B studies, thus presumably reflecting a consistency in subjects' mode of response in different tasks (although there are some difficulties in comparing response criterion levels on the basis of $\beta$ values alone, see 4.4.2). A high correlation was also obtained for detection latency in the two experiments where it was recorded (see Table 3.1). Finally, Table 3.1 also reveals that 'coupling' has no reliable effect on inter-modal performance consistency. Indeed, if anything, the results appear to indicate that the inter-modal correlations are <u>higher</u> for <u>loosely</u> coupled tasks than for closely coupled tasks.

A methodological point which should be noted is that most of the experiments we have discussed have not included a control for the possible effects of the sequence of testing. Particularly severe sequence effects were obtained in the experiment of Tyler et al. (1972), who tested subjects on three successive half hour monitoring sessions separated by 3-min. intervening rest periods.

Given these drawbacks, the results nevertheless lend themselves to the general interpretation that individual differences in performance are consistent between tasks when they are equated for the type and difficulty of signal discrimination, as in the Type B studies. The evidence also points, albeit less forcefully, to the importance of task factors (which might be identified within a task classification system) for determining performance consistency in monitoring behaviour. An evaluation of the correlation studies in terms of the speed-closure dimension is not possible since all the Type B studies used very compatible, similarly classified tasks. Such an evaluation is reported in experiments described in Chapters 8 and 10. These experiments apply a task classification approach to the problem of task specificity in monitoring performance.

The need for a task classification approach was stressed some time ago by H.J. Jerison at the First International Symposium on Vigilance (Buckner and McGrath, 1963a). A glance at Table 3.1 reveals that in the Type A studies, the reported correlations, with one exception (Pope and McKechnie, 1963), are all positive and fall in the range .15 to .54. The implication of this fact was recognized by Jerison, as is apparent in the following transcript of a part of the discussion following the paper by Buckner and McGrath (1963b);

Buckner: It is an additional illustration of the task and mode specificity of individual differences in vigilance. In every study we have conducted, the between-modes correlation has been a low positive value, usually between .2 and .3.

Jerison: What strikes me is that one finds consistently low, but consistently positive correlations. This itself has implications; and the matter is not quite as discouraging as it would be if we were obtaining a distribution of correlations around zero. The implication is that, *we are not dealing with task specificity purely and simply, but with a multidimensional problem*. It is clear that some common factor is present in vigilance performance, *but changing the task adds many factors to it*.

(From Buckner and McGrath, 1963b, p. 69; italics added)

It is apparent that in order to cope with the 'multi-dimensional problem', some sort of task classification system which specifies the important task factors is required. We are faced with the problem of not only identifying the 'common factor', but also those factors which are 'added' when the task is changed. As we pointed out in Chapter 2, it is also clear that these factors have to be related to the detection and decision aspects of monitoring tasks rather than to 'broad' factors common to the monotonous and prolonged aspects of monitoring and other continuous performance tasks. This point is further emphasized by the results of a study by Baker and Ware (1966), who examined inter-task correlations for four continuous tasks, a simple vigilance task, a bean-sorting task, an addition task, and a simple assembly task. There were no correlations in performance between the vigilance task and the other tasks.

We may therefore conclude that a task classification approach provides a departure point for a further investigation of the problem of apparent task specificity in monitoring performance. Naturally, the expectations of this approach are that individual performances on two different tasks will be correlated to a degree dependent on the degree of 'similarity' of the tasks. As yet, we have no firm empirical evidence whereby we can specify the task factors by which the degree of 'similarity' may be ascertained. This problem is essentially the problem of determining which factors might comprise a task classification system for monitoring tasks, which we encountered in Chapter 2. A part of the experimental work reported in this thesis is devoted to an investigation of this problem.

## 3.3 Interrelationships between Motivation, Task Factors and Abilities

Thus far in our discussion of individual and group differences in monitoring performance, we have concentrated mainly on task factors, and only briefly on organismic factors. However, in doing so, we have neglected an 'underlying' variable which should be taken into account in any evaluation of human performance, namely, motivation. Accordingly, in this section we shall briefly consider certain influences of motivational variables on aspects of monitoring performance, and their implications for the task classification approach.

### 3.3.1 Motivation and monitoring performance

Motivation has been considered to be an important variable in a number of theoretical approaches to monitoring behaviour. However, at least one investigator has considered it to be the principal determinant of performance (Smith, 1966). Other investigators have also reported that monitoring performance is significantly affected by a number of independent variables thought to influence motivation level, such as the provision of true or false knowledge of results, artificial signals, financial incentives, etc. (for reviews, see Davies and Tune, 1970, pp. 95-114; Mackworth, 1970, pp. 109-130). Broadbent (1971) has recently also discussed the effects of such variables on motivation and monitoring performance, within a decision theory framework.

The theory of monitoring behaviour put forward by Smith (1966) postulates that the major determinant of performance is the interaction between monotony and motivation. On this theory, the motivation level of the individual, and his reaction to the monotony of the monitoring situation largely determines both his mean performance level and his performance over time. In support of his theory, Smith argued that "typical experimental subjects differ not so much in their ability to maintain attention as in their willingness to do so" (Smith, 1966, p. 2).

While this theory may hold for certain simple tasks and subject populations, it does not appear to be sophisticated enough to explain the empirical data available on the effects of various task and environmental variables on performance. We have seen that subjects do differ in their basic ability to detect signals, even under 'alerted' pre-test conditions, where the effects of the interaction of monotony and motivation would appear to be minimal. Moreover, Smith's theory is somewhat embarrassed by the results of Baker and Ware (1966), who, as we have previously noted, could find no relation between monitoring performance and other 'monotonous' assembly and addition tasks.

Davies and Tune (1970, pp. 210-212) have nicely outlined some further limitations of Smith's theory, which need not concern us here. The important point is that Smith's (1966) approach to motivation does not provide a basis for the interpretation of the effects of a number of task and environmental variables on monitoring performance. A more fruitful approach may be one based on decision theory, with which a relationship may be established between motivation and variations in the response criterion; we have already noted that Broadbent (1971) has discussed monitoring performance in this way. A somewhat broader discussion of the relationship between motivation and decision theory is to be found in an excellent new review by Galanter (1974).

### 3.3.2 Motivation and the interpretation of performance consistencies

An important consideration of the effects of motivation on performance arises out of the type of interpretation to be placed on the results of correlational and factor analytic studies, in which certain response consistencies and differences are observed. The usual method of interpretation of a correlation in performance between two tasks is to postulate the existence of a common 'ability' or 'factor'. Fleishman has pointed out that "abilities are defined by empirically determined relationships among observed separate performances. For example, that individuals who do well on Task A also do well on Tasks B and C, but not on Tasks D, E and F, implies a common process is involved in performance on the first three tasks, distinct from that involved in the last three tasks. To account for these consistencies and distinctions, an ability is postulated" (Fleishman, 1972, p. 1018).

However, if like Smith (1966), we believe motivational factors to be of the greatest importance, we can assert that the 'common process' is nothing other than motivation level. We can then account for the results by assuming that the response consistencies for Tasks A, B and C are related to instances when the subjects were consistently motivated, while the distinction for Tasks D, E and F arose out of the subjects being differentially motivated when performing these tasks.

Such an interpretation is plausible, but rather unlikely to reflect the true situation. The process of assigning abilities is not entirely arbitrary. A common ability is usually attributed to two tasks sharing certain features. The motivational interpretation must assume that motivation levels are maintained consistently only for similar tasks, and not for different tasks, but there appears to be little justification for this assumption. Moreover, if we repeat our correlational study or carry out another study in which the Tasks A to E are included as a sub-set among other tasks, and observe the same response consistencies, the motivational interpretation cannot be reliably supported. Fleishman (1967a, b) has reported a number of such cross-validation studies.

Nevertheless, the possible intervening effects of motivation should always be considered, especially when a correlation in performance between only two tasks is to be interpreted. If a high correlation coefficient is observed, the following interpretations are possible:

1) Performance is a function of motivation level, which is maintained consistently between subjects on both tasks, and is not influenced by task factors or individual differences in ability.

2) Performance is a function of a 'general monitoring ability'. Motivation may influence performance, but not differentially for the two tasks. Task specific abilities are unimportant.

3) Performance consistency is due to the existence of a common ability

which is related to certain features shared by the two tasks. Motivation does not influence performance differentially.

The first of these interpretations may be rejected on the basis of previous considerations, but with somewhat lesser confidence. If, however, a cross-validation study confirms the observed correlation, the motivational interpretation can be more confidently rejected. Interpretations 2 and 3 are difficult to distinguish on the basis of a single result, but interpretation 3 may be chosen on the basis of our review of the correlation studies in 3.2; we saw there that the evidence points to monitoring performance being neither wholly task specific nor wholly nonspecific, the viewpoints of Buckner and McGrath (1963b) and Tyler et al. (1972), respectively. Instead we can assert a third view, in the middle ground as it were, by pointing to the influence of both a general factor or ability relating to the monotonous aspect of a monitoring situation, as well as to task factors or abilities which may be identified within a task classification system.

This section concludes the discussion of various aspects of task classification and monitoring performance. The second of our two approaches to performance assessment, the decision theory approach, is outlined in the following two chapters. Decision theory in the analysis of detection and discrimination performance is introduced in the next chapter.

# CHAPTER FOUR

## DECISION THEORY AND THE ANALYSIS OF DETECTION BEHAVIOUR

## 4.1. Statistical Decision Theory and Detection Performance

The preceding chapters have reviewed a number of different aspects of task classification and monitoring performance. We now turn to the second of our two approaches to performance assessment, the decision theory approach. This chapter is concerned with methods of analysis of detection and discrimination behaviour within the framework of statistical decision theory. There is a very large body of research literature in this area, and thus the treatment here is selective, and considered for its particular relevance to decision processes in monitoring behaviour.

### 4.1.1 Statistical representation of stimuli

Theoretical approaches to the analysis of detection performance commonly assume the existence of 'internal states' in the observer. These are assumed to be activated in the presence of sensory stimulation. Decisions between alternative responses are made contingent upon some utilization of these states. The differences between the various theories of detection arise principally from considerations of the nature and representation of these states. For example, they may be assumed to be 'discrete' or 'continuous', and within these broad categories further assumptions regarding the exact probabilistic representation of states may be made.

Theories of detection and discrimination which can be identified along these lines may be treated as specific forms of statistical decision theory which is a general theory for the analysis of decision making under conditions of uncertainty[1] (e.g., Edwards, Lindman and Phillips, 1965). In particular, the theory of signal detectability (TSD) of Tanner and Swets (1954), which is probably the best known theory of perceptual discrimination, follows naturally from statistical decision theory. The specific nature of TSD arises from the type of the assumptions made regarding the representation of stimuli and the choice of responses; in TSD the internal states are assumed to form a unidimensional continuum, activated by each source of stimulation, and which can be represented probabilistically by a random variable on a Gaussian distribution.

In the simplest form of TSD, only two sources of stimulation are considered. Thus, in the language of decision theory, TSD narrows the general form of the statistical decision theory representation by restricting the number of 'states of the world' to two, and by constraining the form of the 'evidence' to a continuous Gaussian probabilistic representation. These assumptions make tractable the isolation of parameters describing the sensory and decision aspects of detection performance.

While a restriction of the possible forms of stimulus representation to a narrow set enables the computation of desired performance metrics, it also prompts the question as to whether such restrictions can be justified in a given experimental situation. If they cannot, there is clearly a case for making less restrictive assumptions, or of abandoning them altogether, as in the so-called 'nonparametric' models (Richardson, 1972). However, even in shedding many of the assumptions of TSD, the basis of the theory in statistical decision theory remains, and a generalized TSD analysis can thus be applied to a number of decision making situations. Any assumptions regarding the 'underlying' processes may then be made post-hoc, if necessary. Such an approach thus preserves the most general tenets of TSD, which are based in statistical decision theory, without first having to accept the 'restrictive' assumptions which have influenced some researchers to deny the usefulness of TSD (Jerison, 1967a; Parducci and Sandusky, 1970; Wiener, 1973).

### 4.1.2. Operating characteristics

In the general decision theory approach to detection performance, the observer is assumed to respond on the basis of a decision rule which maps the set of stimuli onto the set of responses. The pertinent analogy to this concerns the method of choosing between alternative statistical hypotheses on the basis of a sample of fixed size. In testing statistical hypotheses we are usually concerned with minimizing the probability of making either Type 1 or 11 errors, and use some decision rule for achieving this. The consequences of using different decision criteria can be examined by plotting the points (p(S/n), p(S/s)), where p(S/n) is either the probability of a Type 1 error in statistical terminology, or of a false alarm in detection terms, and p(S/s) is either the probability of a Type 11 error not occurring, or of a correct detection. Laming (1973) has shown that all such pairs of points form a convex set within the unit square (1,1), and represent the totality of decision rules for deciding between responses. The upper bound of the convex set, which is associated with the operating characteristic (OC), represents the only admissible decision rules which cannot be improved upon in a given experimental situation.[2] This is illustrated in Figure 4.1.

The OC represents the limit of an observer's discrimination performance, and completely specifies it. Laming (1973) and Thomas (1973) have described several general properties of the OC and its relation to different theoretical probability distributions which may be used in the representation of stimuli. Two general properties include the monotonicity of the OC (given certain classes of distribution; see Thomas,1973) and the invariance of the OC under any monotonic transformation of the evidence continuum. This latter property also indicates that a TSD type analysis can be applied without any a priori assumptions of the form of the probability distributions. The analysis of detection performance in

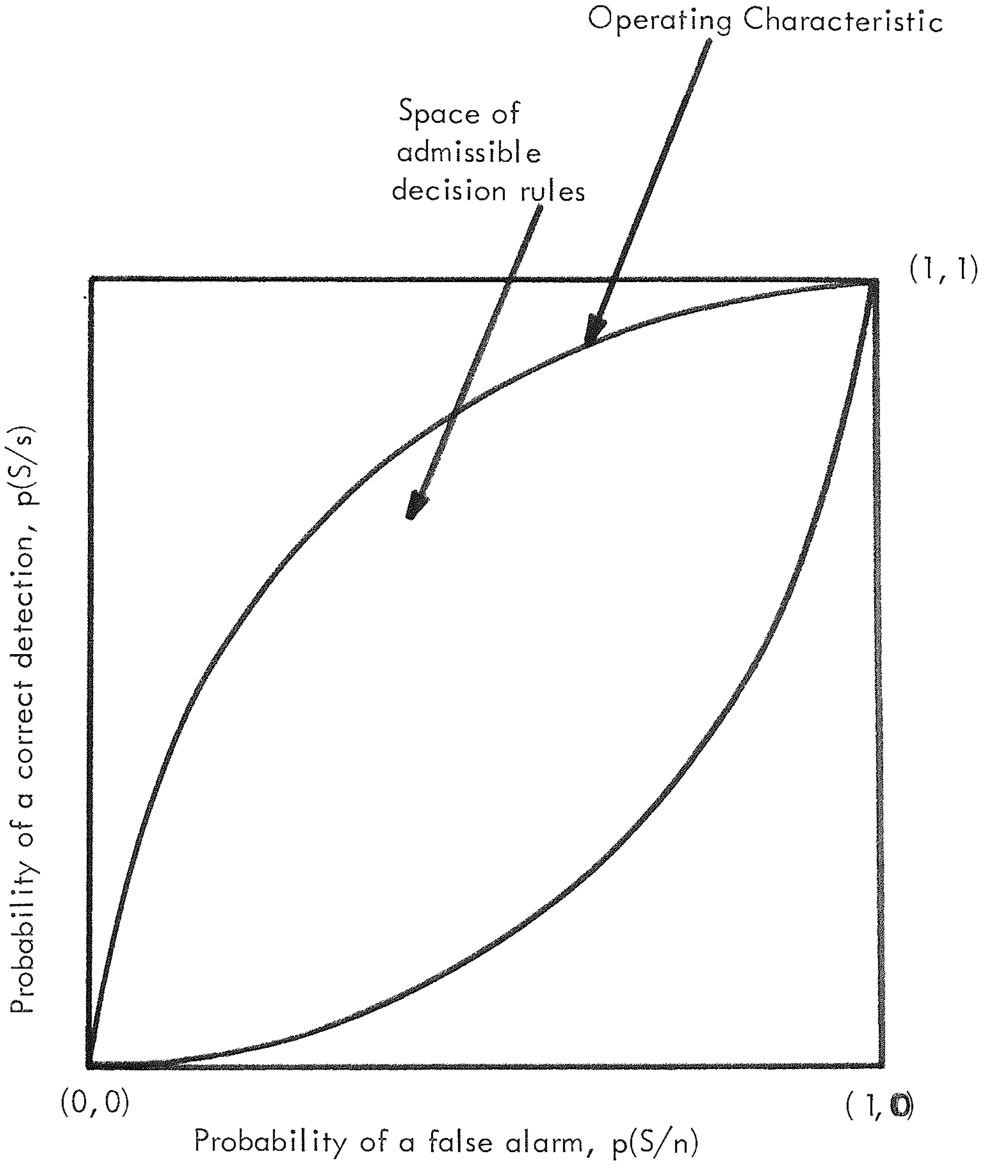


FIGURE 4.1 Representation of the totality of decision rules for the making of statistical decisions. The only rules which cannot be improved upon are those lying on the upper bound of the convex set, or operating characteristic (after Laming, 1973, p.70).

terms of statistical decision theory and the OC thus represents a general type of analysis for detection, discrimination and recognition situations in which the observer is required to make decisions on the basis of unreliable sensory information.

## 4.2. Elements of the Theory of Signal Detectability

Having discussed detection behaviour in the general terms of statistical decision theory and the operating characteristic, we may now turn to a more specific and detailed consideration of the theory of signal detectability (TSD). As we have seen, TSD may be treated as a part of statistical decision theory, in that it makes certain assumptions regarding the nature of the stimulus representations and the choice of responses, which are not made in decision theory. The fundamentals of TSD have been described in detail in several recent books (Egan, 1975; Green and Swets, 1973; McNicol, 1972), and thus only the elements of the theory are sketched here.

### 4.2.1. Gaussian distributions of signal and noise

In the Gaussian model of TSD, the sensory stimulation received by the observer (the 'evidence') is assumed to be represented by a random variable X which is normally distributed. In the most general case, a probability distribution for each type of stimulus may be assumed. More commonly, however, two stimulus conditions are considered, 'signal' and 'noise'; these are represented by Gaussian probability density functions of the evidence favouring either signal or noise, as shown in Figure 4.2.

The axis in Figure 4.2. represents the continuum along which the evidence variable x has its range of variation. This is also known as the decision axis, for reasons which will become clear shortly. Any value of

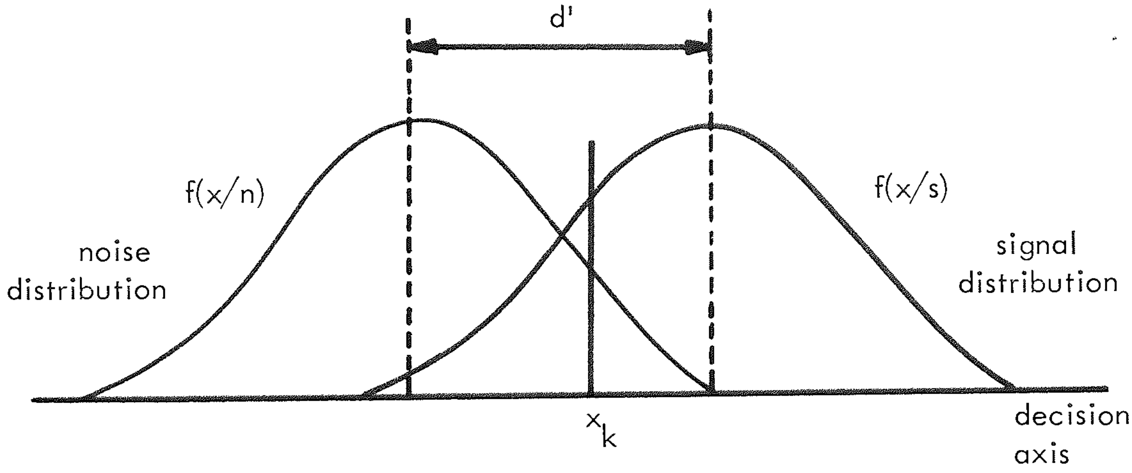


FIGURE 4.2 Probability density functions of the evidence given signal and noise in the Gaussian model of the theory of signal detectability (TSD). The choice of criterion at $x_k$ determines the hit and false alarm probabilities. The signal and and noise distributions have equal variance in the simple TSD model.

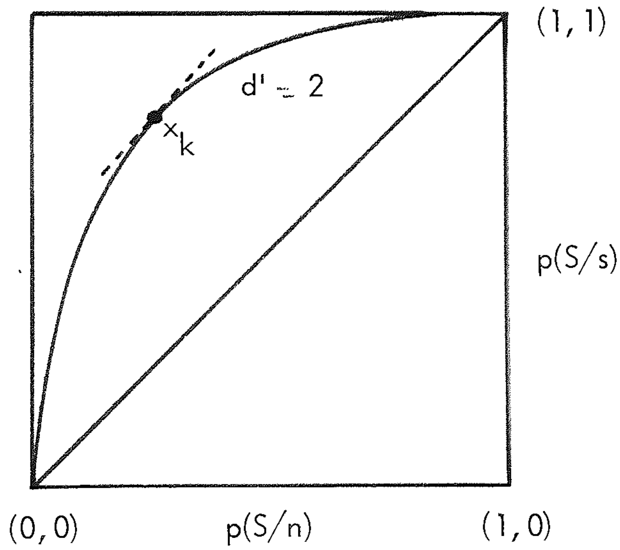


FIGURE 4.3 Operating characteristic (OC) for the equal-variance model of TSD. The slope of the OC at $x_k$ is equal to the likelihood ratio at the criterion.

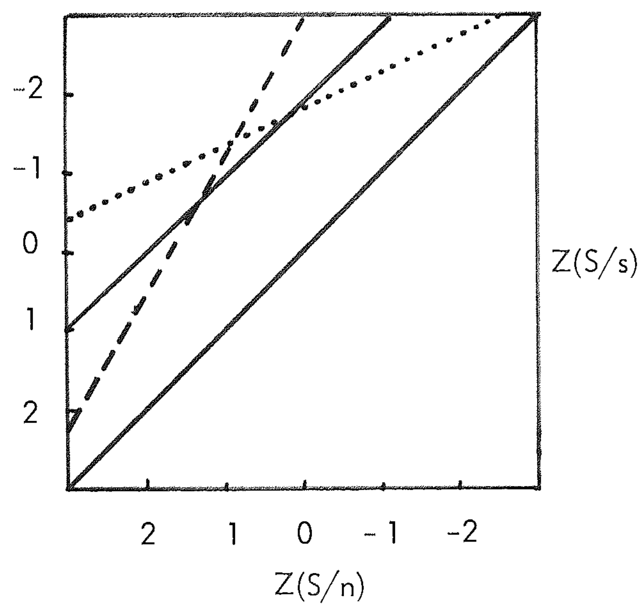


FIGURE 4.4 Operating characteristic (solid line) replotted in z space (normal deviate axes). The dotted and dashed OCs are theoretical predictions of the cases when signal variance is greater than noise variance or vice versa. The convention adopted for the z transformation is $0 \geq z \geq 0$ as $\frac{1}{2} \leq p \leq \frac{1}{2}$.

x along this axis, say $x_k$, may be associated with a <u>likelihood ratio</u> $\left| f(x/s)/f(x/n) \right|_k$ , where $f(x/s)$ and $f(x/n)$ are the probability density functions conditional upon signal and noise respectively. The likelihood ratio represents the odds favouring the signal given the sample observation $x_k$. It is a central concept in decision theory, and in TSD is assumed to provide the basis for making decisions [3]; the observer is assumed to use a decision rule based on likelihood ratio which partitions the x-axis (evidence or decision axis) into two nonoverlapping regions such that a negative response (or a less confident positive response) is elicited when the observed value of x is less than a criterion value $x_k$, and a positive response follows if x is greater than $x_k$.

The choice of criterion thus determines the probability of a correct or incorrect response, and thus correct detection (hit) and false alarm probabilities may be defined:

$$p(S/s) = \int_{x_k}^{\infty} f(x/s)\,dx \quad ; \quad p(S/n) = \int_{x_k}^{\infty} f(x/n)\,dx$$

Responses are thus determined by the placement of $x_i$ criteria along the decision axis. The TSD model assumes furthermore, that

$$\int_{x_0}^{x_1} f(x/s)\,dx = \int_{-\infty}^{x_1} f(x/s)\,dx \quad ; \quad \int_{x_0}^{x_1} f(x/n)\,dx = \int_{-\infty}^{x_1} f/x/n)\,dx$$

$$\int_{x_{n-1}}^{x_n} f(x/s)\,dx = \int_{x_{n-1}}^{+\infty} f(x/s)\,dx \quad ; \quad \int_{x_{n-1}}^{x_n} f(x/n)\,dx \quad \int_{x_{n-1}}^{+\infty} f(x/n)\,dx$$

As $x_i$ varies over its possible range $(-\infty, +\infty)$, pairs of probabilities $(p(S/s),\ p(S/n))$ are generated; $x_i$ may be varied by varying instructions or signal probability in binary choice or 'Yes/No' tasks, or several values of $x_i$ may be created by giving the observer a rating task, in which responses along a scale of confidence are required (see 4.3.). As we have seen, if the hit and false alarm probabilities are plotted against each other, the operating characteristic (OC) is derived. The OC thus represents a transformation of the decision axis. It is invariant under any monotonic change in the axis; it is, however, dependent on the form of the underlying distributions. The TSD model assumes normal (Gaussian) distributions, and if, additionally, the distributions are assumed to have the same variance, it can be shown that the form of the OC is as displayed in Figures 4.3 and 4.4. When plotted in the unit square, the OC describes a bow-shaped function symmetric about the negative diagonal, and a straight line when plotted on double-probability axes, or in z space.

Some simple but important properties can be derived from a consideration of the OC in z space and the probability density functions for signal and noise (see Green and Swets, 1973, 58 ff.). In particular, independent measures of sensitivity[4] and response bias may be derived under this model (see also Figures 4.2, 4.3 and 4.4). The sensitivity index d' is the difference between the means of the two distributions, scaled in units of the standard deviation of the noise distribution. It can be shown that

$$d' = z(S/n) - z(S/s)$$

where z(S/s) and z(S/n) are the normal deviates (or z transformed values) corresponding to the operating probabilities p(S/s) and p(S/n) respectively. This equation also defines the OC in z space.

The measure of bias $\beta$, is defined as the likelihood ratio at the operating point (criterion), or

$$\beta = y(S/s)/y(S/n)$$

where y is the ordinate of the corresponding distribution at the appropriate z value. By substituting for y in the above equation, a useful relation between $\beta$, d' and the evidence variable x may be derived:

$$\ln \beta = d'x - \tfrac{1}{2}d'^2$$

This equation is important because it illustrates that if the criterion is fixed at a given point on the decision axis, then $\beta$ is not constant if d' varies; on the other hand, $\beta$ may be held constant if d' varies and if x changes appropriately. This equation is also important for some other considerations which we shall touch on in later sections, such as the provision of a reliable measure of bias, and whether a subject chooses to fix $\beta$ or x when there is a change in d'.

The above formulations have been presented for the case where the underlying distributions have equal variance. More often than not, however, this assumption does not hold. For the unequal variance case, sensitivity and bias measures cannot be estimated from a single pair of operating probabilities, since there are four free parameters (the two means and the two variances) compared to only two in the equal variance case.

The OCs in the unequal variance case are somewhat different to those for the equal variance case. These differences can be better appreciated for OCs plotted on double probability axes. An important property of such plots is that the OC has a slope which is equal to the reciprocal of the signal to noise standard deviation ratio. Two corollaries follow: if signal variance ($\sigma_s^2$) is greater than noise variance ($\sigma_n^2$), then the OC slope will be less than 1, and vice-versa. This also follows from the following equation:

$$d = (\sigma_s/\sigma_n)z(S/s) - z(S/n)$$

where d is the distance between the two unequal variance distributions; d reduces to d' when $\sigma_s = \sigma_n$. OCs for $\sigma_s \neq \sigma_n$ are shown in Figure 4.4. Thus the slope of an empirical OC can be measured to test the equal variance assumption; if an empirical OC with slope equal to 1 is obtained, then the assumption can be assumed to hold.

The measurement of sensitivity and bias in the unequal variance case is not as straightforward as when the variances are equal. Maximum-likelihood estimation techniques giving values of d and $\sigma_s/\sigma_n$ have been proposed (e.g., Abrahamson and Levitt, 1969; Grey and Morgan, 1972; see also 4.4), but there is no commonly accepted single index of sensitivity. Some of the indices which have been suggested are discussed in 4.4. Measuring bias is even more difficult, and no completely satisfactory index has yet been proposed. The relationship between $\beta$, d and x for the unequal variance case is given by

$$\ln \beta = \frac{x^2}{2}(1 - 1/\sigma_s^{\,2}) + dx - d^2/_2 - \ln \sigma_s \qquad ; (\sigma_n \text{ scaled to } 1)$$

which indicates the difficulty of defining a measure of bias independently of sensitivity.

### 4.2.2. The choice of responses

In the usual TSD Gaussian model, the observer is assumed to decide between responses on the basis of a comparison between likelihood ratio (or some monotonic function of likelihood ratio) and a criterion. In other words, the decision rule is to respond as if the signal were present if the likelihood ratio exceeds the criterion, and to respond as if noise alone were present if it does not exceed the criterion. The criterion may be chosen so as to optimise a number of different decision goals; Green and Swets (1966) have shown that a number of such goals, such as maximizing expected value, minimizing the probability of an error, and so on, can each be expressed in terms of likelihood ratio. The decision goal which has been most commonly considered is the one in which expected value is maximized; expected value is assumed to be dependent on both the a priori signal probability and the costs and values attached to the various responses.

For a given set of experimental conditions, therefore, the optimum value of the likelihood ratio at criterion can be calculated:

$$\beta_{Opt} = \frac{1 - p}{p} \cdot \frac{V_{N/n} - V_{S/n}}{V_{S/s} - V_{N/s}}$$

where p is the a priori signal probability and the V's refer to the costs associated with the different types of response. For a symmetrical payoff, that is, where $V_{S/s} = V_{N/n}$ and $V_{S/n} = V_{N/s}$, the optimum $\beta$ value reduces to:

$$(1 - p)/p$$

Hence an ideal observer wishing to maximize expected value would choose a criterion according to the above equations. It has usually been found, however, that the actual detection behaviour of subjects does not match that of the ideal observer; subjects may be said to be conservative, in that they choose criteria which are less extreme than that of the ideal observer.

Some investigators have utilized this fact to suggest that detection behaviour is often not compatible with an all-or-none division of the decision axis as assumed by TSD, but appears rather to reflect a probability matching strategy (Parks, 1966; Thomas and Legge, 1970). Under this strategy, observers are assumed to match the frequency of their positive responses to the a priori signal probability when the payoff matrix is symmetrical:

$$p(Yes) = p(p(S/s)) + (1-p)(p(S/n) = p$$

Thomas and Legge, (1970) have reviewed some studies whose results indicated that the matching hypothesis is obeyed, as far as group mean response frequencies are concerned, for experienced subjects run with trial-to-trial feedback. They showed that the empirical finding of the non-optimality of detection behaviour is a specific prediction of the matching hypothesis; or that, if $\beta$ is the likelihood ratio for responses chosen on the basis of probability matching.

$$\beta \lesseqgtr \beta_{Opt} \quad , \text{ as } \quad \beta \gtreqless 1 \quad .$$

However, while the matching hypothesis appears to fit group data fairly well (but see Creelman and Donaldson, 1968), Dusoir (1974) has shown that there are large deviations from the hypothesis for the data of individual subjects, and that response frequency is not constant over different discriminability levels, as implied by the matching hypothesis. Thomas (1975) has proposed some modifications to the model of Thomas and Legge (1970) which can account for deviations from probability matching, but the elegance and simplicity of the original model is then lost.

## 4.3. The Derivation of Empirical Operating Characteristics

There are a number of different ways in which changes in an observer's detection behaviour may be induced so as to generate an empirical operating characteristic (OC). These methods may be considered in relation to the two major experimental paradigms of the theory of signal detectability (TSD), the Yes/No procedure, and the Rating procedure. A third procedure, the forced choice procedure, is used to provide a relatively pure measure of sensitivity only, there being little concern with the observer's decision criterion (at least for psychophysical tasks). This procedure is thus of limited relevance to the analysis of monitoring behaviour, since, as we shall see, variations in the decision criterion play an important part in monitoring performance. However, it is of value in establishing appropriate sensitivity levels either between subjects or between conditions, prior to a monitoring session.

### 4.3.1. The Yes/No procedure

In a Yes/No detection task, the observer is presented with one of two stimulus alternatives on each trial. The stimuli correspond to the presentation of 'noise alone' or 'signal plus noise' in the TSD model. An observation interval is clearly marked out with warning signals. A simple 'Yes' or 'No' response is required on every trial. Feedback may or may not be provided. Observers are encouraged to respond even if they are doubtful and cannot decide which stimulus was presented.

The Yes/No method thus provides data from which a pair of operating probabilities may be estimated from the relative frequencies of correct and incorrect detections over a block of trials. The OC is generated by inducing observers to adopt different degrees of response strictness in different sessions. For example, they might be instructed to be 'lax' in one session, that is, to respond to any event they suspect to be a signal, and to be 'strict' in another, that is, to respond positively only when they are absolutely sure they have perceived a signal. Other ways of obtaining an OC include 1) varying the a priori signal probability and requiring the observer to maximize expected value 2) varying the costs and values attached to response outcomes and requiring a maximization of expected gain, and 3) asking the observer to adopt the Neymann-Pearson criterion (Swets, Tanner and Birdsall, 1961). Atkinson and Kinchla (1965) have also obtained Yes/No OCs by varying the amount of (misinformative) feedback in different blocks of trials.

These methods have their relative advantages, but in general, the Yes/No procedure suffers from the drawback that the experimental effort required to obtain a reliable OC may be large. For instance, Green and Swets (1973, p.393) suggest that to estimate a single pair of operating probabilities, a total of about 500 trials should be used; hence if an OC of say, 5 points is desired, 2500 trials have to be run for each subject and condition. For many experiments these demands are high; they are virtually impossible to meet in situations where the number of signal and noise trials is set by other considerations, as in monitoring tasks.

### 4.3.2 The rating procedure

The rating procedure provides an attractive and economic alternative to the Yes/No procedure. It is identical to the Yes/No procedure, except that observers are required to register their confidence on a rating scale. A Yes/No task can thus be treated as a special case of a rating task with the number of confidence categories collapsed to two. By allowing responses of differing confidence level, it is assumed that the corresponding criteria held by the observer may be sampled. Presenting a n-category rating scale ranging from 'certain signal' and 'quite certain signal' through to 'certain no signal', provides n-1 potential operating probabilities ('potential', since all the categories may not be used).

### 4.3.3 Comparison of procedures

The rating method is thus a more efficient way of determining the empirical OC. For an m-point OC, only 500 trials are needed with a m+1 rating scale, while 500m trials are needed with the Yes/No method. However, while the rating procedure is undoubtedly efficient, there is some doubt about whether it is equivalent to the Yes/No procedure, that is, whether it yields similar OCs and values of sensitivity and bias. The evidence on this point is conflicting.

The early promise of the rating procedure was matched with data indicating that there were no substantial differences in the form of OCs, or in sensitivity values, obtained by the rating and Yes/No procedures (Egan, Schulman and Greenberg, 1959; Swets, 1959). More recent studies by Emmerich (1968) and Nachmias (1968) have also supported these findings, while the consistency of the rating method was demonstrated by Weintraub and Hake (1962), who obtained the same values of sensitivity using 2, 3 or 4 category rating scales. On the other hand, studies by Markowitz and Swets (1967), Clarke and Mehl (1973) and Weitzel and Dobson (1974) have revealed systematic differences between Yes/No and rating OCs. Leshowitz (1969) compared the Yes/No and two-alternative forced choise procedures, and also found differences in the obtained values of sensitivity.

Markowitz and Swets (1967) found that in general, Yes/No OCs had unit slopes while rating OCs had slopes less than 1. They explained their results by suggesting that Yes/No OCs are contaminated by variations in signal probability, while in the rating method, with signal probability fixed (usually at 0.5), no such effects are present. Some other evidence for this view is provided by Schulman and Greenberg (1970), who found a systematic co-variation between OC slope and signal probability, However, it should be noted that the sensitivity index $d'_e$ was not found to vary in this study, as well as in the Markowitz and Swets experiment. This index is further described in 4.4.1.

There are thus some doubts as to the equivalence of the Yes/No and rating methods. This therefore somewhat reduces the force of our earlier statement that the rating method provides the most efficient method of obtaining the OC. A further drawback of the rating method is that the OC is constrained to describe a smooth monotonic function, especially if the number of categories is large (Watson, Rilling and Bourbon, 1964), because of the dependence of the OC points on each other. Moreover, while it has been demonstrated that observers are capable of simultaneously maintaining several decision criteria (for example, 9 in a study by Gaussin, 1972), this has not been shown to be true in many of the experimental situations in which TSD has been applied.

At present therefore, a choice between the Yes/No and rating procedures can only be made by weighing their relative advantages and disadvantages in a particular experimental application. Until further evidence relating to the similarities and differences between the two procedures has been provided a choice between them is largely a matter of personal preference.

## 4.4 The Measurement of Sensitivity and Bias

Two aspects of the analysis of detection behaviour in terms of TSD have been considered thus far. The first concerns the treatment of TSD within the framework of decision theory, whereby empirical OCs are obtained and subsequently examined to see whether performance variation is due to a change in sensitivity or in response bias. The second aspect discusses different methods of deriving the empirical OC. A third aspect arises from the need to compare OCs between subjects and conditions, and this section is therefore concerned with the quantification of the OC.

### 4.4.1 Sensitivity indices

If the equal-variance Gaussian model of TSD is found to hold, then a reliable measure of sensitivity is readily available in d'. If, however, the equal-variance assumption does not hold, other indices have to be sought. We may also not wish to make any assumptions about the underlying detection model, and therefore might wish to consider a 'nonparametric' measure. The best such measure is provided by the area under the OC, but a number of other indices have also been proposed. Table 4.1 lists a number of parametric and nonparametric indices, as well as two indices derived from Luce's (1959) Choice Theory.

The first index listed in Table 4.1 is d', which is commonly used in the literature. As we have seen, its use depends on the assumption of equal variances of signal and noise. Dissatisfaction with this restrictive assumption has led to a number of other indices being proposed for the unequal variance case, where, as we noted in 4.2.1,

$$z(S/s) = (\sigma n/\sigma s)(z(S/n) + d).$$

The separation of the means of the signal and noise distributions, d, can be used with the signal to noise variance ratio to provide a conjoint measure $D(d, \sigma n/\sigma s)$, as Green and Swets (1966) suggest. Both of these can be obtained from the double probability OC. While D is an informative index since it completely describes a straight-line OC in z space, it does not provide the desired single parameter of sensitivity.

An alternative to both d' and D was suggested by Egan and Clarke (1966). This is $d'_e$ or $d'_s$, which is twice the ordinate of the intersection of the OC with the negative diagonal (in z space). The justification for this measure (see also Ogilvie and Creelman, 1968) derives from the association of the negative diagonal with unbiased (chance) performance; thus the negative diagonal provides a reference from which sensitivity may be measured.

Another sensitivity index which may be related to the OC in z space was suggested by Schulman and Mitchell (1966). Since any line through the origin in z space represents zero detectability, they proposed the index $D_{YN}$, which is the orthonormal distance from the origin to the OC.

Simpson and Fitter (1973) have recently reviewed some of the parametric measures of sensitivity, and concluded that $D_{YN}$ provides the basis for the 'best' measure of detectability for the case when Gaussian distributions with unequal variance are assumed. Simpson and Fitter appear to have considered three criteria for a good measure of sensitivity:

| Index | Source | Formulae and Notes |
|---|---|---|
| **A. Parametric indices assuming Gaussian distributions** | | |
| $d'$ | Green & Swets (1966) | $z(S/n) - z(S/s)$; assumes $\sigma_s = \sigma_n$. |
| $D(d, \sigma_n/\sigma_s)$ | Green & Swets (1966) | $d$ and $\sigma_n/\sigma_s$ are the horizontal intercept and slope of the OC, respectively. |
| $d_e$, $d_s$ | Egan & Clarke (1966) | $2d/(\sigma_s + \sigma_n)$; twice ordinate of OC at equal bias point (negative diagonal). |
| $D_{YN}$, $D_{FC}$ | Schulman & Mitchell (1966) | $d/(\sigma_s^2 + \sigma_n^2)^{\frac{1}{2}}$; orthormal distance to OC. $D_{FC} = \sqrt{2}D_{YN}$. |
| $d_a$ | Simpson & Fitter (1973) | $\sqrt{2}D_{YN}$; (see text). |
| $(d')$ | Grey & Morgan (1972) | $d/(\sigma_s\sigma_n)^{\frac{1}{2}}$. |
| **B. Distribution free indices** | | |
| $P(A)$ | Green & Swets (1966) | Area under the unit square OC. |
| $z(A)$ | Simpson & Fitter (1972) | Normal deviate of $P(A)$. |
| $P(C)$ | Green (1964) | (see text). |
| $P(\bar{A})$ | Pollack & Norman (1964) | Area estimate from single OC point (see text). Can be computed from formula given by Grier (1971): $\frac{1}{2} + (y - x)(1 + y - x)/(4y(1 - x))$, $x = p(S/n)$, $y = p(S/s)$. |
| $C, d(A,B)$ | Hammerton & Altham (1971) | $C = (n - N)/(r - 1)$, where $N$ and $n$ are the number of ratings of signal and noise, and $r$ is the total number. $d(A,B)$ reduces to $p(S/s) - p(S/n)$ when $r = 2$. |
| $D_{AB}$ | Sakitt (1973) | $(\bar{i}_n - \bar{i}_s)/(\sigma_s\sigma_n)^{\frac{1}{2}}$; $\bar{i}_s$ and $\bar{i}_n$ are the mean ratings of signal and noise. Equivalent to $(d')$. |
| $E$ | Simpson & Fitter (1973) | $\sqrt{2}(\bar{i}_n - \bar{i}_s)/(\sigma_s^2 + \sigma_n^2)^{\frac{1}{2}}$; equivalent to $z(A)$ and $P(A)$. |
| **C. Other indices** | | |
| $\alpha$ | Luce (1959) | $(p(S/s)p(N/n)/(p(S/n)p(N/s)))^{\frac{1}{2}}$; equivalent to $z(A)$ and $P(A)$. |
| $d'_M$ | McNicol (1972) | $2\log\alpha$ |

TABLE 4.1 Indices of sensitivity ($\sigma_s = \sigma_s$, $\sigma_n = \sigma_n$; OC = operating

1) it should be equivalent to the best nonparametric index, namely P(A), or the area under the OC; 2) it should be related monotonically to P(C), the bias-free index of sensitivity available from the forced-choice procedure; 3) it should reduce to d' when the variances are equal. By restating Green's (1964) theorem, which proves the equality of P(A) and P(C), and drawing on Schulman and Mitchell's (1966) finding that $D_{YN}$ differs from its forced-choice equivalent $D_{FC}$ by only the constant $\sqrt{2}$, Simpson and Fitter proved the equality of $D_{YN}$ and the normal deviate of P(A). Thus they argued that $D_{YN}$ is the only index which meets their criteria. In order to maintain comparisons with the two alternative forced-choice task, Simpson and Fitter accordingly suggested the use of alternative $\sqrt{2}D_{YN}$, or $d_a$.

The $d_a$ measure thus appears to be the best available index of sensitivity. It is equivalent to a measurement of the mean distance between the signal and noise distributions scaled to the root mean square of the variances. It should be noted, however, that its stability has not been put to empirical test. We noted in 4.3.3 that one advantage of the $d'_e$ index is that it is relatively invariant with a change in OC slope, as seen in the data of Schulman and Greenberg (1970). One minor problem in the use of $d_a$ (and $d'_e$) which may arise concerns empirical OCs which lie in regions of z space well away from the negative diagonal. Such OCs, which may be obtained from 'noisy' data in typical monitoring situations, may have to be fitted by eye; if so, any error in the fit will be magnified in $d_a$ or $d'_e$, since the OC will have to be extrapolated to obtain these indices (alternatively, errors in obtaining an estimate of signal to noise variance ratio from an OC fitted by eye will be reflected in $d'_e$ or $d_a$ if the computing formulae given in Table 4.1 are used to calculate these indices).

The sensitivity indices described thus far may be obtained after the OC has been fitted to the available data. Since there is an error in the estimation of both operating probabilities, normal curve fitting techniques, which consider errors in only one variable, cannot be used. Recently, however, a number of procedures borrowed from functional analysis have been suggested for the fitting of empirical OCs. To be capable of solution, such procedures must assume, a priori, a form for the underlying distributions, subsequent to which a maximum-likelihood estimation process can be applied to extract the detection parameters. A number of recent papers have outlined these procedures, for both Yes/No and rating data, and assuming both Gaussian and logistic distributions. (Abramson and Levitt, 1969; Dorfman and Alf, 1968, 1969; Grey and Morgan, 1972; Ogilvie and Creelman, 1968). Abramson and Levitt's paper discusses the most general case, while Grey and Morgan have discussed the use of an estimation technique based on the minimum chi-squared estimates of the logistic function (MLC), showing that this method can provide convergent estimates when maximum likelihood methods fail to do so. Grey and Morgan provided detection estimates using the MLC method for data from a vigilance task; these parameters had a large variance due to the inherent unreliability of probability data obtained from low signal probability tasks. For this reason, it may not be necessary to use the relatively sophisticated likelihood estimation techniques for the 'noisy' data from such tasks, where typically, false alarm probabilities may be estimated from a frequency of say 2 or 3 from among 200 or 300 noise trials. This point is further discussed in later sections. The maximum likelihood and MLC methods do however provide an excellent means of deriving the detection parameters d and $\sigma_s$ when reliable data are available. The program provided by Grey and Morgan (1972) is probably the most readily available and easiest to use of the different techniques.

Of the distribution free or nonparametric indices listed in part B of Table 4.1, the best measure is P(A), the area under the OC. Since the OC is invariant under a monotonic transformation of the decision axis, the area under the OC is not dependent on the underlying distributions; hence the term 'distribution-free'. However, its computation may not always be easy, especially if only a few operating points are available, or, as in the case with monitoring performance, the operating points are restricted to a small portion of the unit square, or if only one operating point is obtained.

Pollack and Norman (1964) and Pollack, Norman and Galanter (1964) have, however, showed that P(A) can still be roughly estimated if only one operating point is available. The OC passing through a single operating point is constrained to pass through certain defined regions. This is illustrated in Figure 4.5 for the operating point X. Lines to the origin and (1,1) from X represent the locus of operating probabilities if the observer were more biased toward making No and Yes responses, respectively. Points in the regions B and W thus represent either better or worse detectability than at X, and hence the OC must pass through X in the regions U1 and U2. A rough measure of detectability can then be taken as the mean area under the OC, or $P(\bar{A})$. While this type of analysis is only approximate, and rests on the crucial assumption that OCs may not cross (skewed OCs may do so), it does illustrate that some information regarding sensitivity can be gained from even a single operating point, more so than the consideration of hit probability alone. This point will be raised again in relation to 'traditional' (Broadbent, 1971) and decision theory analyses of monitoring behaviour.

Further nonparametric indices have been suggested by Hammerton and Altham (1971) and Altham (1973), whose indices C and d (A,B) are based on the

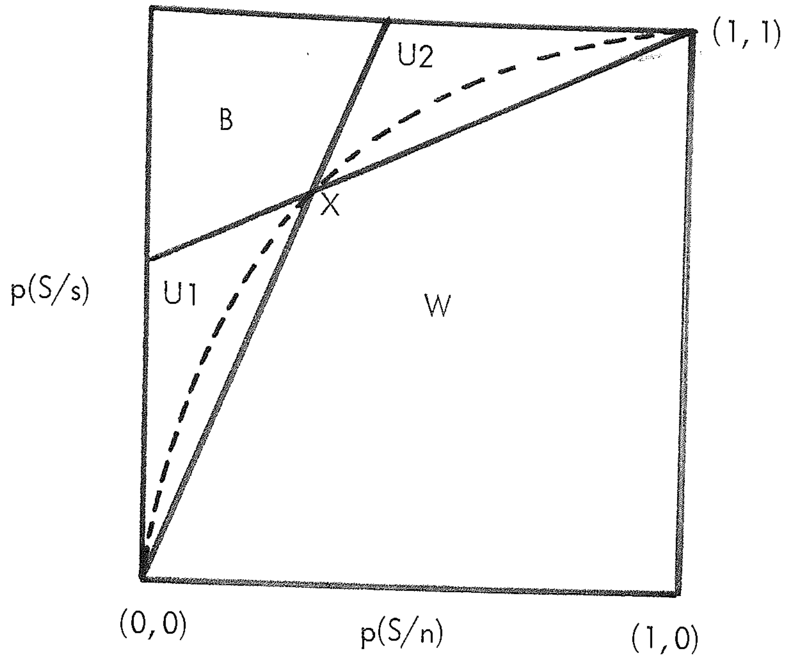


FIGURE 4.5 The regions to which a symmetric operating characteristic (OC) passing through X is constrained. Regions B and W represent the space of OCs with better or worse detectability, respectively. The dotted curve indicates the possible shape of the OC, assuming the validity of the equal-variance TSD model.

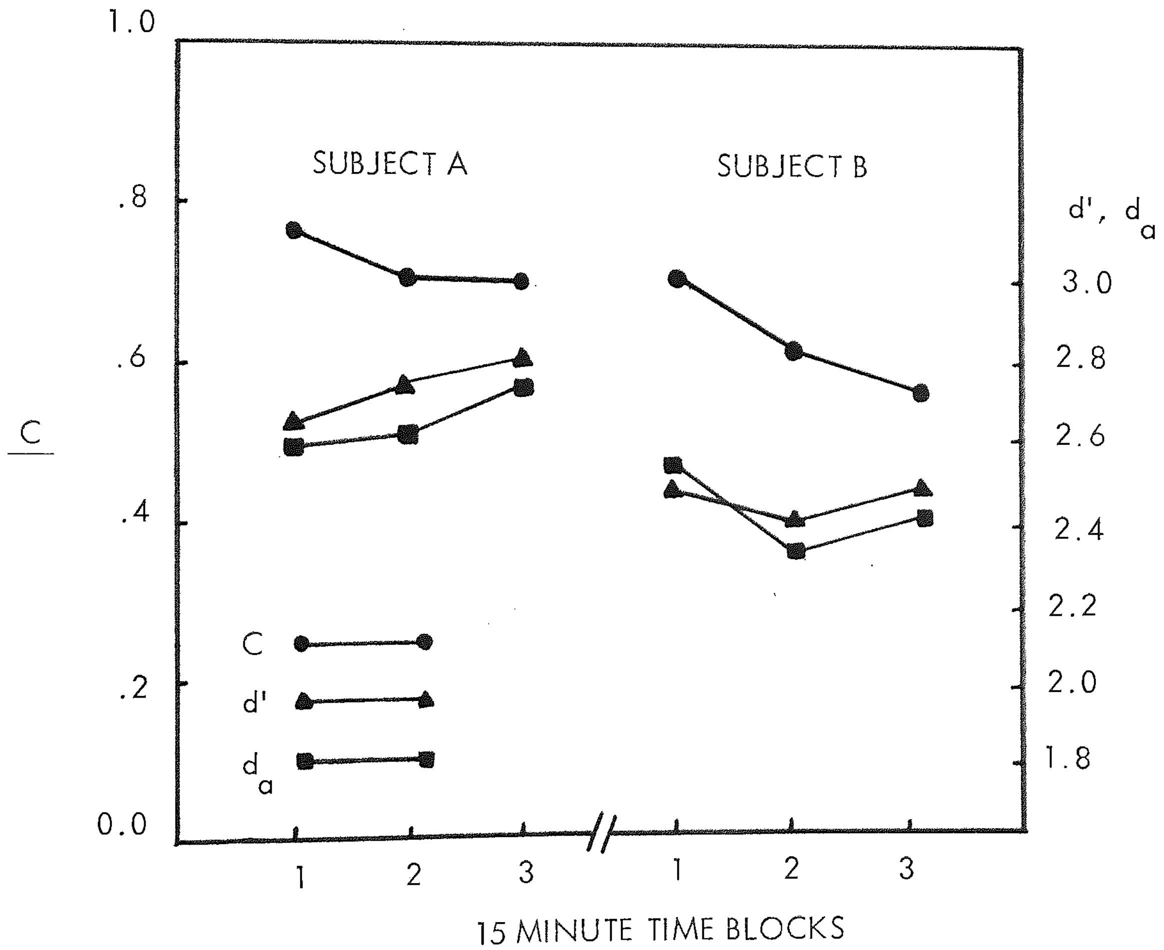


FIGURE 4.6 Three measures of sensitivity, C, d' and $d_a$, as a function of time spent on a monitoring task (data from Experiment 4).

average rating given over a block of trials to signal and to noise. The indices $D_{AB}$ and E, proposed by Sakitt (1973) and Simpson and Fitter (1973) respectively, are similarly defined, except that they are scaled to the geometric mean and root mean square of the signal and noise variances, respectively. (see Table 4.1). Both E and $D_{AB}$ are superior to C and d(A,B), with E being the most acceptable since it is directly related to both P(A) and $d_a$. The problem with the C and d(A,B) indices are that each reduces to p(S/s) - p(S/n) when the number of ratings is two, as in a Yes/No task. Hence a corresponding OC, which is an iso-sensitivity curve, would describe a straight line in the unit square; as we shall see in 4.5, most of the available data is inconsistent with this prediction. Altham (1973) presented data from a vigilance experiment showing that values of C were monotonically related to the parametric index d'. However, C does not measure the same thing as d', since C will vary over conditions even if d' is constant across conditions; this is usually the case across successive time blocks in a monitoring task. The use of C would thus indicate a loss in detectability over time blocks in a monitoring task, whereas other measures of detectability might not do so. Figure 4.6 shows performance data from a monitoring task for three indices of detectability, d', $d_a$ and C, and for two subjects. The trends over time in the indices d' and $d_a$ are very similar for both subjects, indicating that there is no decrease in detectability with time on task, while there is a steady decrement in C in both subjects. The use of C cannot therefore be recommended.

Finally, two indices of detectability derived from Luce's Choice Theory are listed in part C of Table 4.1. McNicol (1972) has shown that these indices are equivalent to those obtained if a TSD type model with underlying logistic signal and noise distributions are assumed. The predictions of Choice Theory are, in many instances, very similar to TSD, but these indices are of especial value in the analysis of so-called open-ended tasks, to which TSD methods cannot be easily applied. Choice Theory is further discussed in section 4.5.3.

### 4.4.2 Indices of response bias

We have now reviewed a number of indices of perceptual sensitivity which may be used in the quantification of detection behaviour. While several such indices have been proposed, considerably less effort has been devoted to the development of indices of response bias. This is probably due to the fact that the OC is an iso-sensitivity curve which can be readily derived using one of the methods we have discussed in 4.3; the subsequent isolation of a reliable sensitivity index is thus a relatively straightforward matter. However, no similar methods exist for the derivation of iso-bias curves, and hence the identification of a bias parameter is less straightforward. Nevertheless, some proposals for a bias index have been put forward, and we shall review these in this section. It will emerge, however, that a single, satisfactory index cannot be identified.

The available indices of response bias are listed in Table 4.2. The index $\beta$, which is the value of the likelihood ratio at the criterion C is probably the most widely used index of bias. When the equal-variance TSD model is shown to hold, this provides a reliable measure of bias, although, there is surprisingly little evidence showing it to be invariant over different levels of discriminability. We shall return to a consideration of the reliability of $\beta$ and the relationship between $\beta$, c and d' later, after first discussing the other indices reported in Table 4.2.

For rating data, McNicol (1972) suggests correcting $\beta$ at each criterial

| Index | Source | Computing Formula | Notes |
|---|---|---|---|
| $\beta$ | Green & Swets (1966) | $\left\| f(x/s)/f(x/n) \right\|_c$ | f is the normal distribution |
| c | - | | Criterion cut-off z(S/n) |
| $\beta'$ | McNicol (1972) | $\beta/\sigma_s$ | $\sigma_s$ taken as inverse of OC slope |
| B | McNicol (1972) | Rating point at which p(S/s) + p(S/n) = 1 | |
| $B_I$ | Ingham (1970) | z(S/s) + z(S/n) | |
| $B_H$ | Hodos (1970) | $1 - x(1 - x)/y(1 - y)$ , $B_H < 1$ | Computing Formulae |
| | | $y(1 - y)/x(1 - x) - 1$ , $B_H > 1$ | after Grier (1971) |
| | | x = p(S/n), y = p(S/s) | |
| log b | Luce (1963b) | $\frac{1}{2}\log(p(N/s)p(N/n)/p(S/s)p(S/n))$ | |

TABLE 4.2 Indices of response bias.

level by dividing the likelihood ratio by the signal standard deviation $\sigma_s$ since this occurs in the denominator of the expression for likelihood ratio when the variances are unequal. This index ($\beta'$) only provides a very rough estimate of bias, and since it is corrected by a measure of slope which is dependent on the operating points, none of which need lie on the OC, it can be very sensitive to errors in measurement. Its independence of sensitivity has also not been empirically tested.

Another index B, suggested by McNicol (1972) also suffers from drawbacks. This is the mean category on the rating scale corresponding to a point of equal bias towards positive and negative responses, that is, the category rating for which $p(S/s) + p(S/n) = 1$. While B has the advantage of being a nonparametric index, it provides only one measure of bias in the rating scale task, whereas one at each criterion (confidence) level is desirable. Furthermore, comparisons of B across different rating tasks is difficult since the absolute value of B depends on the number of rating categories; there will also be a tendency for the estimate of B to improve as the number of categories increase. A final drawback of B is that it is difficult to compute when bias is such that the median value of B falls outside the rating range (that is, when $p(S/s) + p(S/n) > 1$ at category 1, or $< 1$ at category n).

Having considered the index B, let us digress slightly to note that henceforth we will sometimes refer to the likelihood ratio measure $\beta$ as 'B' for typing convenience. This should not be confused with the B measure of McNicol (1972), which we have just considered and rejected. In the experimental part of the thesis, we shall almost exclusively refer to the likelihood ratio as 'B'.

Ingham (1970) proposed the index $B_I = z(S/s) + z(S/n)$ for the measurement of response bias. The rationale behind this measure is that it is orthogonal to the index d', which is z(S/s - z(S/n). However, Ingham's own experimental work did not provide full support for the independence of d' and $B_I$, although $B_I$ was less susceptible to variations in d' than $\beta$. Furthermore, while this index may be orthogonal to d', it may not be orthogonal to other parametric measures of sensitivity which are required for the unequal variance TSD model.

Hodos (1970) has reported an index purporting to be a nonparametric measure of bias. In point of fact, Hodos merely provides an arbitrary measure of bias based on a similar analysis of the unit square as carried out by Pollack et al. (1964). Taking the negative diagonal as the line of zero bias, Hodos' measure is based on the relative displacement of a single operating point from this line. For points to the left of the diagonal, the index $B_H$ is defined as the ratio of areas (U2 - U1)/(U2 + B), and the ratio (U2 - U1)/(U1 + B) for points to the right of the diagonal (see Figure 4.5 for the definitions of these areas). The index $B_H$ cannot be recommended for the following reasons: 1) the index makes no reference to a sensitivity parameter to which it is orthogonal, 2) the index is not nonparametric since it yields isobias contours which are specific to the type of detection model assumed (the most obvious divergence is between TSD and simple threshold theory), 3) for operating points <u>below</u> the leading diagonal, $B_H$ cannot be used although such points are acceptable within the unequal-variance TSD model.

The index log b is the equivalent of the criterion cut off c of TSD for logistic distributions of signal and noise. The two indices are very similar due to the similarities of the normal and logistic distributions, and we would expect the isobias contours associated with these two indices to be also very similar. We shall, therefore, include the discussion of log b within that of c.

Having considered and, for the most part, rejected the majority of the indices in Table 4.2, we are left with the indices $\beta$ and c. The problem of deciding between these two indices comes down to the problem of deciding which of these two quantities are held constant (if at all) when there is a change in detectability. The evidence on this point is not completely clear. One the one hand, in selective (dichotic) listening experiments, Broadbent and Gregory (1963b), Moray and O'Brien (1967) and Treisman and Geffen (1967) reported that the unattended ear was associated with a reduced d' with no change in $\beta$, while c increased. These results therefore suggest that subjects hold the likelihood ratio $\beta$ constant with varying discriminability, although it should be noted that only the data of Broadbent and Gregory (1963b) are consistent with the assumption of equal variances, upon which this interpretation depends. On the other hand, Hardy and Legge (1968), in an investigation of the effect of 'emotional' words on auditory discriminability, suggested that c rather than $\beta$ is held constant across discriminability levels. McNicol (1972) has similarly argued in favour of c.

It emerges therefore, that we cannot confidently use any of the proposed bias indices as a reliable measure of response bias, independent of sensitivity. Apart from the relative neglect of isobias functions in the literature, one reason for this is that much of the theoretical and and empirical effort invested in developing a reliable index of sensitivity has not been similarly transferred to the problem of finding a reliable index of bias. It may also be the case that with a change in discriminability, a given bias index may be affected in different ways according to the variable effecting the detectability change. This might therefore account for the difference between the previously mentioned results of Broadbent and Gregory (1963b), Moray and O'Brien (1967) and Treisman and Geffen (1967) on the one hand, and Hardy and Legge (1968) and McNicol (1972) on the other hand.

It seems wisest, therefore, to use both $\beta$ <u>and</u> c, although this may cause some difficulties of interpretation if these vary disproportionately.

However, until there are some further developments in this area, we will have to put up with these difficulties. In a later section we shall return to this point and investigate whether the use of decision latencies can shed some light on this problem.

## 4.5. Detection Models

In this section we shall consider some of the alternatives to the theory of signal detectability (TSD). A few of these alternative theories consist of modifications of the basic TSD model, but a number of other theories proceeding from different starting points have also been proposed. The number of extant models is fairly large, as indicated in recent reviews by, among others, Luce and Green (1974) and Nachmias (1972). We will only briefly sketch some of these models, and consider their main features with particular reference to the analysis of monitoring behaviour.

### 4.5.1 Classification of detection models

The complete range of detection models may differ along a number of dimensions, but only one major factor is considered here. This concerns the nature of the representation of the stimuli in the observer, and the relation of the responses to these representations, or 'states'. These states may be treated as the 'input' to the perceptual-decision stage within an information-flow model of the observer, and whose physical realization can, in principle, be obtained. Within this framework, three types of model can be distinguished, 1) 'finite' models, in which in the number of states is limited, 2) 'continuous' models, in which the states form a continum, and 3) stochastic models, which are related to continuous models, but which consider the stimuli to be represented by (or to 'activate') some stochastic process over time. We will consider these three types of model separately.

### 4.5.2 Finite state models

Finite state detection models have in common the feature that stimuli are represented by a limited number of discrete states. In the simple 'threshold' models, only two states are considered. Blackwell's (1963) high threshold model (2-HT), which derives directly from classical threshold theory, assumes that a threshold or limen must be exceeded before a stimulus may be perceived. This is assumed to occur on a proportion p of all signal trials.

The 2-HT model also assumes that the threshold is never exceeded on the presentation of noise alone, with the result that 'true' false alarms are never made; false alarms are treated as wrong guesses made when the threshold is not exceeded. The observed hit rate is thus assumed to be the sum of the 'true' hit rate and the guessed hit rate, which is the proportion of sub-threshold signals which are correctly guessed as signals. Since false alarms can only be wrong guesses, the guessing factor is the observed false alarm proportion, and hence the observed hit rate p(S/s) becomes

$$P(S/s) = p + (1 - p)P(S/n)$$

where p is the 'true' hit rate, P(S/n) is the guessing factor and 1 - p is the 'true' omission rate. Rearranging terms gives

$$p = \frac{P(S/s) - P(S/n)}{1 - P(S/n)}$$

which represents the traditional correction for guessing ($p \rightarrow P(S/s)$ as $P(S/n) \rightarrow 0$). Thus p is a measure of sensitivity in the 2-HT model, and the above equations can be used to obtain the theoretical OCs. The predicted OC is a straight line as shown in Figure 4.7. The OC in z space is shown in Figure 4.8.

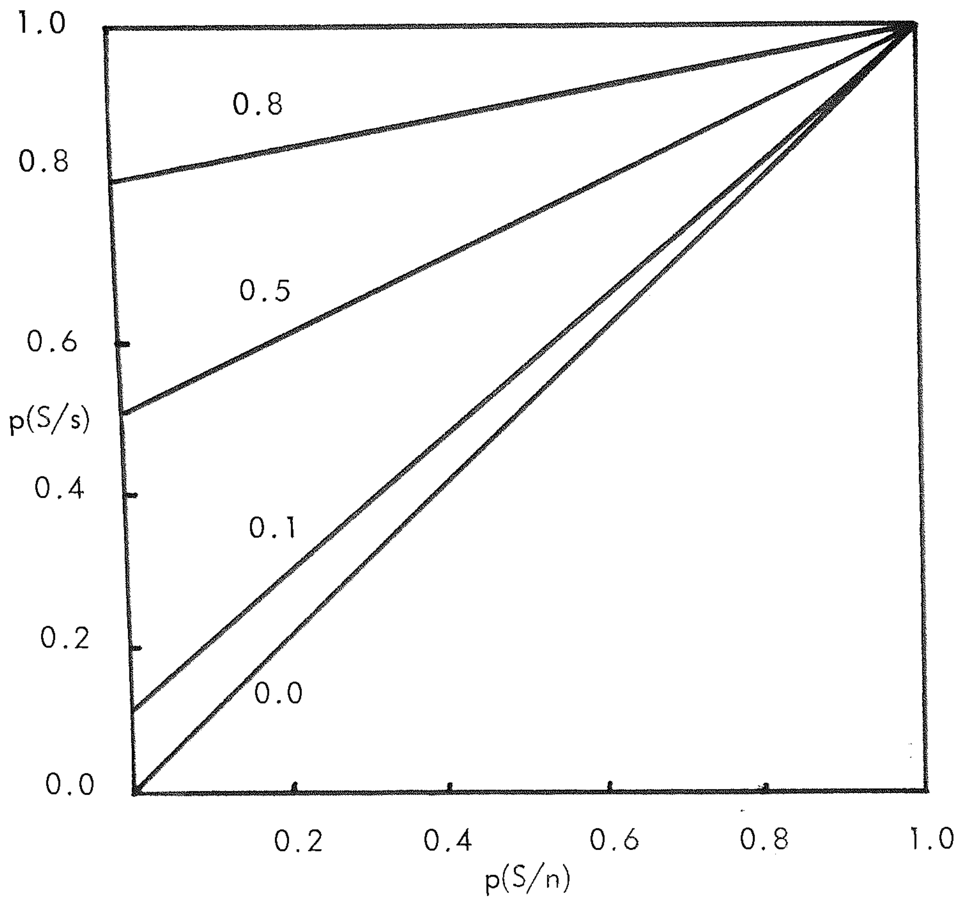


FIGURE 4.7 Theoretical operating characteristics (OCs) predicted by High Threshold Theory (the parameter numbers indicate values of p, the proportion of signal trials on which the threshold is exceeded).

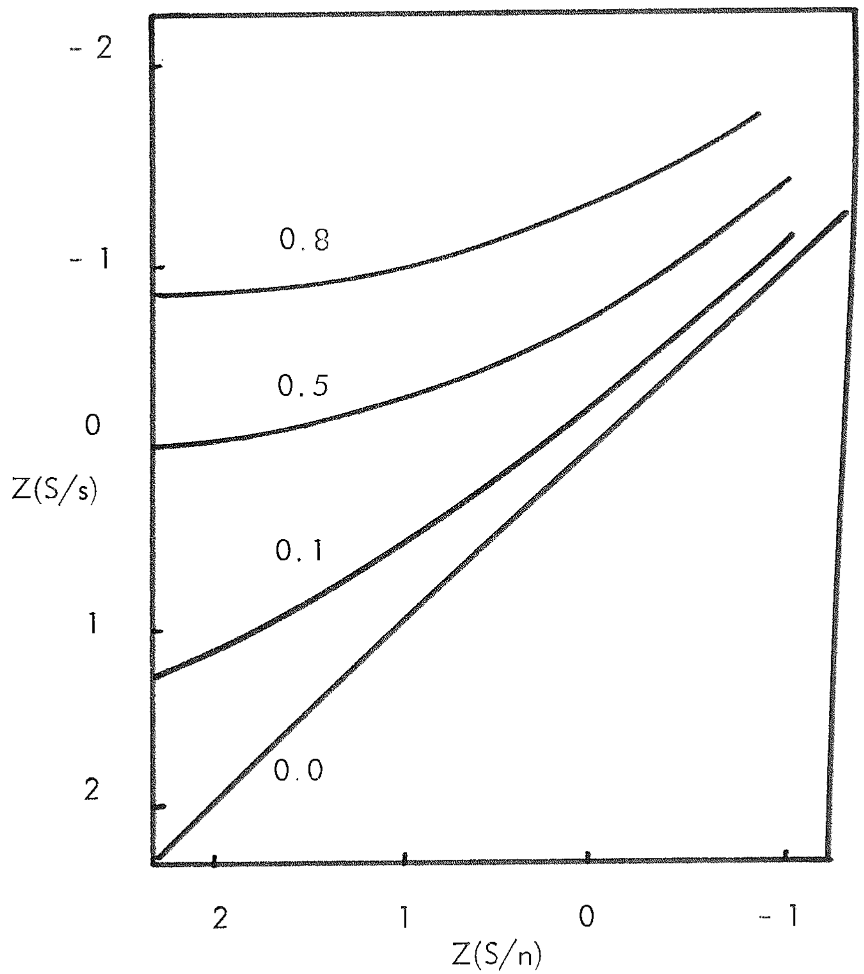


FIGURE 4 8 High Threshold Theory OCs plotted on normal deviate axes.

High threshold theory can be confidently rejected as a model for detection and monitoring behaviour for a number of reasons. Firstly, signal detection studies have shown that the index p does not remain invariant over conditions in which only response bias is varied, as would be expected of a 'pure' measure of sensitivity. Secondly, most of the available data is inconsistent with the straight line OC predicted by 2-HT (see, for example, Swets, Tanner and Birdsall, 1961). In monitoring tasks, the false alarm rate is usually low, and it is a typical finding that changes in false alarm rate are accompanied by much larger changes in hit rate (Broadbent and Gregory, 1963a). This is inconsistent with the 2-HT model, which predicts a smaller change in hit rate. Finally, when second choices are required in forced choice studies, these are not made at random, as predicted by 2-HT, but in accordance with TSD. (McNicol, 1972; Swets, et al. 1961).

For these reasons, we shall not consider high threshold seriously in our analysis of monitoring behaviour. Nevertheless, the theory has commanded considerable attention over the years as a reference point for the evaluation of other theories. Being derived from classical threshold theory, it has dominated much of the research thinking in psychophysics up to the late 1950's. The theory has been implicitly accepted by many researchers, especially those working in the field of monitoring behaviour. An examination of this literature reveals that false alarm data is seldom reported, or when it is, is given a minor treatment or given the 'guessing' tag, so that a high threshold model is implicitly assumed. The experiments of Broadbent and Gregory (1963a) and others have showed, however, the crucial importance of the false alarm rate and the invalidity of the 'guessing' treatment of false alarms.

The other major finite state model is the two-state low threshold (2-LT) theory of Luce (1963a). Low threshold theory differs from high threshold theory by assuming that the threshold can be exceeded by the presentation of noise alone, but only on a small proportion q of the noise trials (hence the term 'low-threshold'). Thus a two-way guessing mechanism is assumed such that the observed false alarm rate is either greater or smaller than the 'true' rate according as whether the observer says 'Yes' to a proportion of sub-threshold events or 'No' to a proportion of supra-threshold events. Using the same rationale as that used in deriving the 2-HT OC, it can be shown that the equations linking the observed hit and false alarm rates and the threshold proportions p and q are

$$P(S/s) = (p/q)P(S/n) \qquad 0 \leqslant p(S/n) \leqslant q$$

$$P(S/s) = p + \frac{(1-p)P(S/n)-q)}{(1-q)} \qquad q \leqslant p(S/n) \leqslant 1$$

These equations can be used to obtain the theoretical OC; it consists of two straight line segments intersecting at the point (q,p) in the unit square, and shown in Figure 4.9. The OC in z space is shown in Figure 4.10.

The OCs predicted by 2-LT are thus different to those of TSD, but in practice it may be difficult to distinguish clearly between the two models. In vigilance, the results of Broadbent and Gregory (1963a) and Egan, Greenberg and Schulman (1961) are consistent with the TSD model. However, because of the difficulty in determining on which of the two segments of the OC the data lie, and of locating the point q,p (due to the low false alarm rate), the 2-LT model cannot be entirely ruled out. A further difficulty arises in the assumptions inherent in the use of the rating method. In the formal treatment of OCs in 4.1, we saw that OC is constrained to describe a smooth monotonic function. In this context, Larkin (1965) has argued that the rating method is unsuitable for testing the predictions of 2-LT. When a large number of

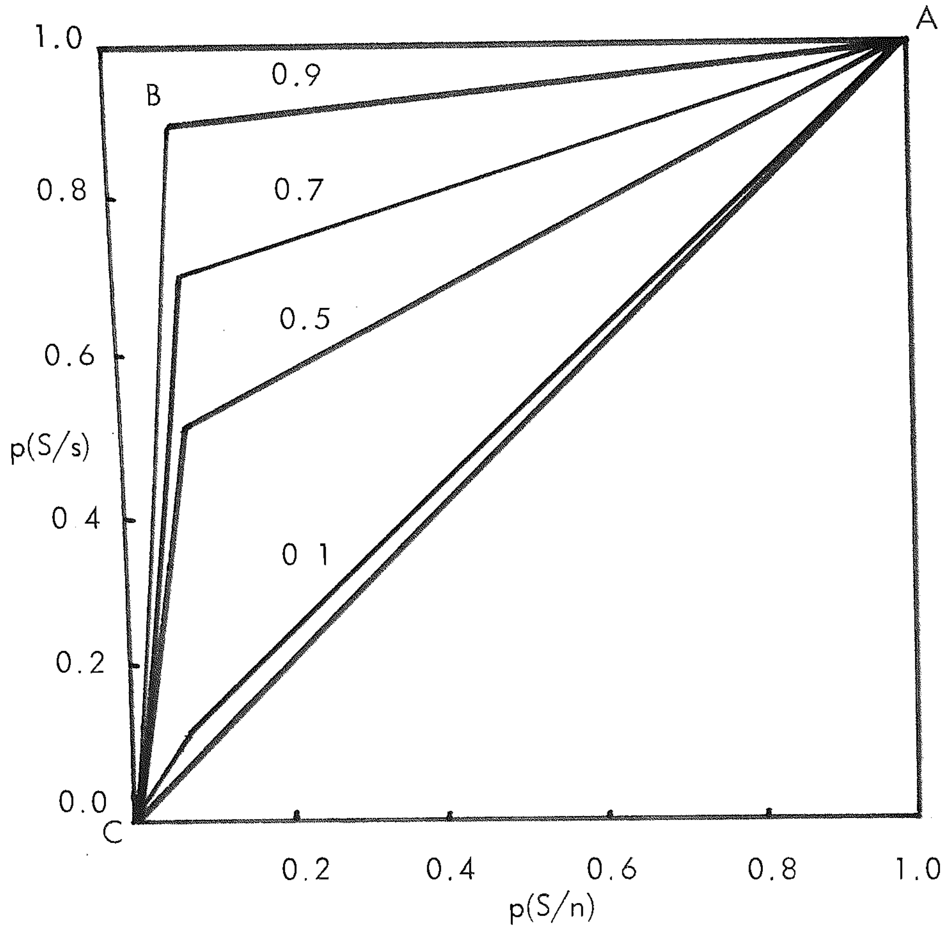


FIGURE 4.9 Theoretical operating characteristics predicted by Low Threshold Theory for q set to 0.08 (see text). The parameter numbers indicate values of p.

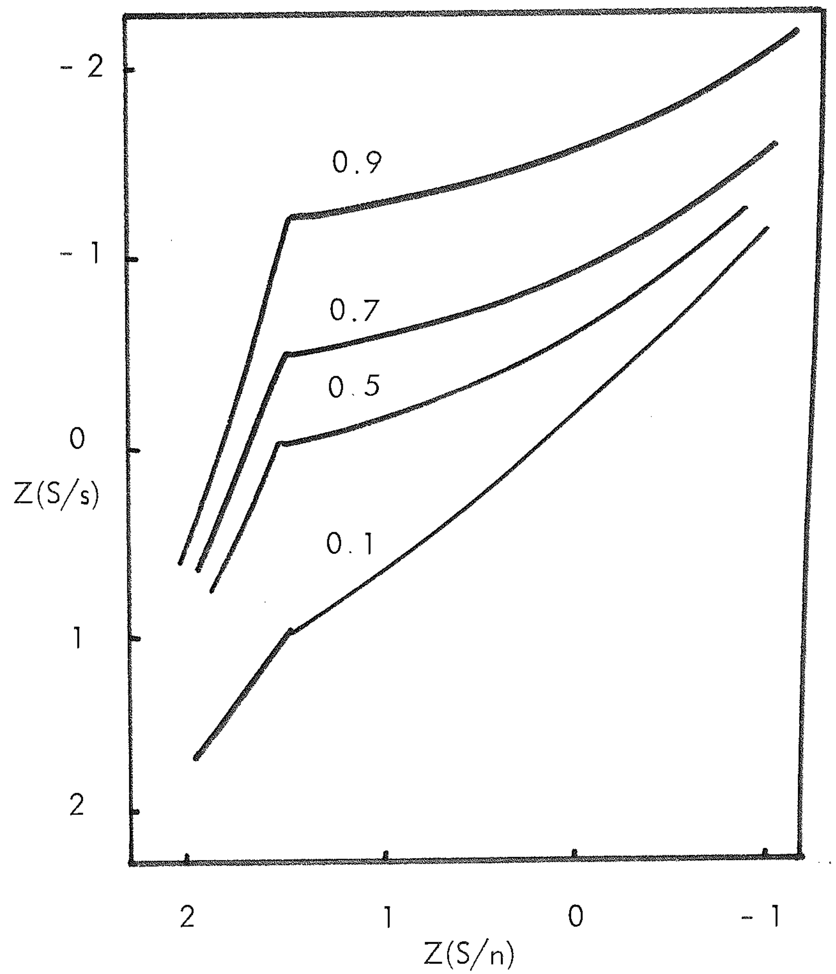


FIGURE 4.10 Low Threshold Theory OCs plotted on normal deviate axes.

rating categories are used, as for example by Watson et al. (1964), the smooth curved form of the OC may merely reflect the statistical dependence of the OC, rather than give an indication of the form of the representation of the internal states. The rejection of 2-LT thus requires the assumption that there is a corresponding mapping of the response (confidence) states and the detect states which form the input to the decision stage. However, there is no reason that such a strategy is used by the observer, and it is possible for him to mix strategies (from the 'pure' strategies in the theoretical OC, at A, B and C) such that the empirical OC describes a curved function in the unit square. Such a mapping was assumed by Nachmias and Steinman (1963) and Green and Moses (1966) (but see Wickelgren, 1968; Krantz, 1969). Luce and Green (1974) have recently shown how such strategies, or the influence of 'non-sensory' internal states on the response categories could produce a curved OC even if the sensory detect states are assumed to be dichotomised.

Therefore, while high threshold theory can be confidently rejected, low threshold can only be weakly rejected in favour of TSD. However, two further criticisms of the 2-LT model can be made, which tilt the balance in favour of TSD. The first criticism concerns the fact that in many experimental situations, including monitoring tasks, fairly high hit rates can be achieved with a negligible false alarm rate, which implies that $q \approx 0$. In this case 2-LT reduces to 2-HT, which has been shown to be untenable. The second criticism concerns the fact that q, which is the probability that noise alone exceeds the threshold, should depend on the nature of the noise only; yet Nachmias and Steinman (1963) obtained estimates of q which decreased as <u>signal</u> strength was increased. Krantz (1969) has also outlined similar criticisms of two state threshold theories and proposed instead a 3-state low-high threshold model (3-LHT) which is a combination of 2-HT and 2-LT. At this stage in the development of detection models, however, distinguishing between the predictions of models on the basis of empirical data becomes increasingly difficult, and more so as models with a higher number of states or a greater number of free parameters are considered (all multi-state quantal models of Norman, 1964, and the variable sensitivity model of Atkinson, 1963). This applies with greater force to the evaluation of monitoring performance, where, as we have seen, accurate OCs cannot be obtained at the data may be of 'rough' quality.

### 4.5.3 Continuous state models

Continuous state models include TSD-type models with non-Gaussian distributions, as well as the Choice Theory model of Luce (1959, 1963b). The exponential TSD model was first proposed by Green and Swets (1966) as an alternative to the unequal-variance Gaussian TSD model, since the latter model suffers from the drawback that likelihood ratio is non-monotonic with the decision axis. On the other hand, the exponential model has the advantages that likelihood ratio is monotonically related to the decision axis[5] and is a one-parameter[6] distribution. The signal and noise distributions in this model are

$$f(x/s) = ke^{-kx} \quad , \quad f(x/n) = e^{-x} \qquad k < 1,$$

where the parameter k completely specifies the distributions. Likelihood ratio is $f(x/s)/f(x/n)$ or $-ke^{-(k-1)x}$, which is a monotonic function of x. The hit and false alarm probabilities are

$$p(S/s) = \int_c^{\infty} ke^{-kx}dx = e^{-kc} \quad , \quad p(S/n) = \int_c^{\infty} e^{-x}dx = e^{-c} \quad ,$$

from which it follows that

$$p(S/s) = p(S/n)^k \quad ,$$

$$\text{or } \log p(S/s) = kp\ (S/n).$$

Hence the OC describes a power function in the unit square, and a straight line of slope k when plotted on logarithmic axes. The power relation in the exponential model was assumed by Egan et al. (1961) in their analysis of 'free-response' data.

The exponential model thus provides a simple, useful alternative to the Gaussian continuous state model, and because of its relation to counting processes, it may be an appropriate model for detection in situations where these processes are evident (see Luce and Green 1972), in the so-called free-response situations. Another useful advantage of this model in these situations is that it permits an OC analysis even though the absolute operating probabilities cannot be computed, since the logarithmic relationship enables an analysis in terms of the number of hits and false alarms (see also Broadbent, 1971, p.85).

The other probability distribution which has been considered (and which is sometimes assumed in the estimation of detection parameters, see 4.4.1) is the logistic distribution. This distribution is often taken as an approximation to the normal distribution since it can be specified in closed form. It is also naturally related to the other major continuous state model, Luce's Choice Theory, which makes certain predictions about the OC which are very similar to those predicted by the normal TSD model. In view of the similarity of the normal and logistic distributions, this is not too surprising. The inter-relationships between Choice Theory and the normal and logistic TSD models have been outlined recently by McNicol (1972) and so the treatment here is brief.

Choice theory assumes that alternative responses are associated with 'strengths' depending on whether signal or noise is presented. If the relative response strength associated with responding 'signal' is v when noise is presented and av when signal is presented, it can be shown that

$$P(S/s) = a/(a + v) \qquad P(S/n) = 1/(1 + v)$$

from which it follows that

$$\frac{P(S/s)}{1 - P(S/s)} = \frac{a \ P(S/n)}{1-P(S/n)}$$

This equation is analogous to that obtained if logistic distributions in the TSD model are assumed, and both are approximations to the function predicted by the normal TSD model (which of course, cannot be specified in closed form). The predictions of Choice Theory and TSD are thus remarkably close, although the former was formulated from entirely different starting points. Choice Theory has found its main area of application in the analysis of performance of so-called open ended tasks, where the set of possible responses is not specified a priori as in the Yes/No rating and forced choice tasks (see Broadbent, 1967; Ingleby, 1968).

Before leaving this section on continuous models, a brief mention may be made of Smith's (1968) Cost Theory of discrimination. Under this theory the concept of the 'cost' of psychophysical decision is introduced and related to both discriminability and bias. Smith incorporated the cost function into two existing detection models, TSD and Choice Theory. For the model extended from TSD, Smith proposed that the average cost is a linear function of the sum of the signal and noise probability densities; the cost function for TSD thus describes a trough-shaped curve with its maxima at the means of the distributions. Hence for criteria placed within the area of the trough (relatively lax criteria), a change in criterion implies an increase in cost, while cautious criteria (outside the trough area) can be increased with a concomitant loss in cost. Smith proposed this type of mechanism to account for the results of Broadbent and Gregory (1965) who, as we shall see, found that in a monitoring task cautious criteria increase with time at work while lax criteria do not (Smith's assumption being that observers try to reduce cost over time in a monitoring situation). We shall return to Broadbent and Gregory's findings in the next chapter, and review the evidence bearing on this interpretation.

### 4.5.4 Stochastic models

In this section we will very briefly consider Luce's (1966) stochastic model of detection, although some other models under this heading have also been proposed, such as the so-called 'neural-counter' models (McGill, 1967). In its original formulation, Luce's (1966) model was considered in relation to performance in free-response tasks, but it has recently been developed to have somewhat wider applications (Luce and Green, 1972).

In Luce's (1966) model, the occurrence of a signal is assumed to 'generate' a series of pulses (which can, in principle, be related to neural pulses) in the decision centre of the observer at a rate dependent on signal intensity. The pulses are assumed to describe a Poisson process. The pulse rate, which is the inverse of the mean inter-arrival pulse time, completely describes a Poisson process. The pulse rate parameters associated with signal and noise thus describe performance in a manner analogous to d' and $\beta$ in TSD. The derivation of these parameters is dependent on assumptions regarding the decision rule used, the simplest one being that each pulse activates a response. This was shown to be untenable, and more complex rules (such as one based on certain sums of the inter-arrival pulse times) have been proposed in later developments of the model (Green and Luce, 1971; Luce and Green, 1972).

Luce (1966) considered this model to be of particular value in the analysis of free-response tasks, which include some vigilance tasks in which the task is not divided up into signal and noise events (continuous rate tasks, see 2.2). Luce therefore thought it unfortunate that Poisson signal schedules have not been used in any studies of vigilance, since the use of rectangular signal schedules may introduce additional response biases (signal probability in time is constant only under a Poisson schedule).

Luce's model is difficult to evaluate from the point of view of monitoring behaviour because very few studies would meet the assumptions needed to use the model. The experimental evidence provided by Luce and Green themselves is by no means clear; in a number of studies aimed at directly testing this model, a number of inconsistent and complex results have been obtained (see Luce and Green, 1972), which suggest that either a reworking of the theory is required, or that further extension of the basic model is required by introducing additional assumptions into the original set.

## 4.6 A Decision Theory Model for Response Latencies

Thus far in our discussion of decision theory and detection behaviour, we have been concerned only with 'discrete' measures, that is, with measures of performance accuracy. However, 'continuous' measures, that is, measures of response latency, may also be recorded in detection and monitoring tasks. Response latencies in monitoring have not been interpreted within a decision theory framework. In this section we shall show that such an interpretation is possible, and shall outline a working model whereby response latency data in monitoring tasks may be analysed.

In section 4.5., we reviewed a number of different models of detection and discrimination. For the most part, these models have been concerned almost exclusively with the analysis of discrete or categorical performance measures, such as choice proportions or response probabilities. Only a few of the models make a provision for an analysis of the time taken to initiate a response, and generally do not consider the implications of response latency data for the detection model. More recently, this tendency for detection and latency analyses to be conducted separately has been countered in some theoretical studies by Thomas (1973; Thomas and Myers, 1972) in which the import of latency data for detection models has been considered.

Some of the principal latency models in choice reaction and detection tasks have recently been reviewed by, among others, Audley (1973), Laming (1973) and Welford (1971). Audley reviewed a number of models under some general headings: sequential decision models, communication theory models, preparation models, fast guess models, fixed-sample models, stimulus repetition models[7]. He concluded that none of the models could adequately account for all the data, and that some models made very similar predictions for the forms of the latency distributions; this is a familiar theme which we encountered in our discussion of detection models in 4.5. Audley emphasised the need to clarify the range of applicability of the various models; one basis of classification which he suggested was between tasks involving easy discriminations and those involving difficult discriminations or weak signals. He was of the opinion that "tasks involving difficult discriminations exhibit considerable differences from the usual choice reaction time task employing easily distinguishable stimuli: the main one being that, for difficult tasks, errors are made with longer average latencies than correct responses (for example, see Audley, 1970; Audley and Mercer, 1968; Pike, 1968), whereas easy tasks elicit quicker errors (for example, Laming, 1968; Egeth and Smith, 1967), and in some cases both slow and quick errors occur (Wilding, 1970). Whether the information processing strategies used for the two kinds of task are the same is the important question, and I do not believe that we yet have the data to resolve this issue" (Audley, 1973, p 511).

This is a point that has been considered previously in this thesis in relation to monitoring tasks, for which a distinction can be made between tasks with transient or weak signals and those with long duration or 'hold' signals (see Chapter 2). Although the points of difference between the information processing strategies used in the two cases are not fully understood, it appears that a decision theory analysis is more applicable to the former case; and hence there is a case for supposing that the same type of analysis can be applied to latency data in difficult-discrimination tasks. Such a distinction between easy and difficult tasks relating to the form of the speed-accuracy tradeoff and the relative latencies of correct and erroneous responses has also been recently proposed by Norman and Bobrow (1975), who distinguish between so-called <u>data-limited</u> and <u>resource-limited</u> processes. Although their paper is mainly concerned with the evaluation of performance tradeoff in time-shared situations, they make similar points to those of Audley (1973) about differences in the latency distributions and the speed accuracy tradeoff between tasks in which performance is limited by the quality of sensory information, and those in which performance can be improved by a re-allocation of the processing resources.

Monitoring tasks, are, for the most part, data-limited, or tasks where the quality of the evidence received by the observer is relatively imperfect. Thus an analysis which emphasises the influence of the quality of the evidence on both choice proportion and response time measures may be appropriate to such tasks. It is thus pertinent to ask whether signal detection theory, which has been applied to the analysis of choice probabilities, can be extended to the analysis of response times as well. There are several ways in which this can be achieved, and two broad classes of the latency models discussed by Audley (1973) can be considered to be consistent with decision theory: sequential decision models, and fixed sample models. A latency model falling in the latter category is considered here: there are some reasons why fixed sample models might be more appropriate for the analysis of latency data in monitoring tasks. Firstly, there is some evidence in the literature that fixed sample models are particularly relevant to task situations involving difficult discriminations with weak, transient signals (Gesheider, Wright and Evans, 1968) while sequential decision models appear to handle well data from tasks having easier discriminations and where performance can be improved by successive sampling of the input (Laming, 1968). Secondly, slower error responses are predicted by fixed sample models, whereas sequential decision models generally predict that errors have faster or the same latencies as correct responses (although with some modifications, slower errors can also be explained in some cases; see Laming, 1968, 1973). Some recent studies have shown that errors, especially false detections, have consistently longer latencies than correct responses in visual monitoring tasks (Colquhoun and Goldman, 1972; Davies and Tune, 1970, p.17; Parasuraman, 1975b; Parasuraman and Davies, 1975). Finally, since monitoring behaviour has been analysed in terms of TSD, it would be expedient if the same model can be used to analyse both detection and latency aspects of performance.

When latency measures have been considered in relation to monitoring behaviour, usually only detection latency has been examined (Buck, 1966). As Buck's review demonstrates, there is a tendency for detections and latencies to be inversely related, and Buck suggests that these two performance measures are governed by similar underlying mechanisms. Hence it would be clearly desirable for any theory of vigilance which attempts to explain vigilance phenomena in terms of decision processes to account for both detectability and response latency measures. Ideally, latencies associated with all response categories would be considered, but it has been usual to take measures of detection latency only. The reason for this is that in the usual monitoring situation, unlike the Yes/No detection situation, only one response, 'signal present' or 'Yes' has been required of subjects, so that latencies associated with the other response categories (correct rejections and omissions) have not been recorded. Latencies associated with false alarms may be recorded in the single response case, but have rarely been reported in vigilance (but see Colquhoun and Goldman, 1972). Latencies corresponding to responses made with different confidence levels may also be recorded. Although trends over time for detection latency have been reported, there are few data on latency changes with time on task for the other response categories (see Davies and Tune, 1970, pp. 17-18).

In the TSD analysis of detection performance in monitoring tasks, it is assumed that the observer takes a sample (an 'observation') of the 'evidence' presented on noise alone or signal plus noise trials (Green and Swets, 1966; McNicol, 1972). A decision is then made between the response 'Yes" and 'No' on the basis of a comparison between likelihood ratio at the observation point (or some monotonic function of likelihood ratio) and the response criterion. In other words, the closer the likelihood ratio is to the criterion, the less 'evidence' the observer has to respond positively to signals. This implies that the time the observer takes to respond may be related to the relative strength of the evidence, or the relative distance of the observation point from the criterion; thus the shorter the distance, the weaker, relatively speaking, the evidence, and hence, the longer the decision time.

The extension of TSD to include latency data can thus be made by assuming response latency to be inversely related to the distance, in decision space, from the response criterion. It should be emphasised that this interpretation of latency rests on the further assumpton that variations in the time taken to gather evidence (in the TSD sense) of differing strengths are small compared to the overall decision time. Hence this model is a fixed-sample model.

The assumption that input sampling time[9] does not form a major part of the observed response time appears reasonable as a first approximation, and in view of the reasons which we put forward regarding the applicability of fixed-sample models to monitoring tasks. Hence, with the further assumption relating latency to criterion distance, a response latency function of the form f(x - c) may be generated. It is possible to consider certain forms of such a function, just as we considered forms for the representation of the distribution of signal and noise in the TSD detection model. It is also possible to consider other features of the function, such as its symmetry about the criterion, and its response to changes in the form of the distributions in the TSD detection model. We will not offer any such theoretical speculations however, but merely state the model in its simplest form; any modifications will be considered in the light of the evidence we obtain in subsequent experiments. This is consistent with our treatment of TSD within the general framework of decision theory.

The general concept of a relationship between response latency and the degree of responsiveness or confidence in perceptual judgment was recognized as early as the beginning of this century, by, among others, Henmon (1911) and Lemmon (1927). More recent statements of this concept are to be found in Cartarette, Friedman and Cosmides, (1965), Emmerich, Gray, Watson and Tanis(1972), Gesheider, Wright and Evans (1968) and Thomas and Myers (1972).

For the decision theory model, the relation between response latency and criterion position may be easily worked out: for a change in criterion position, for example, to one of increased caution, the latencies of Yes responses or of responses made at confidence levels pointing to the detection of a signal, will, on average, increase, since the observer now demands that the evidence be stronger before he responds to signals. On the other hand, the latencies of No responses will, on average, decrease, since No responses will be distributed, on average, further away from the criterion than before. The increase in Yes response latencies and the decrease in No response latencies as the criterion shifts to the right along the decision axis is illustrated in the upper half of Figure 4.11. These predicted trends hold only if the relative distance between the means of the signal and noise distributions in the TSD model remains constant, that is, if detectability is constant. This latency model provides testable predictions concerning trends in latencies associated with different response categories, resulting from a change in criterion placement, given constant detectability, as for example has been demonstrated both as a result of time on task and with variations in signal probability.

On the other hand, if a change in signal detectability does occur, the model predicts that the trends in the latencies will be somewhat different. With a loss in detectability, for example, the distance between the means of the probability density functions is reduced, as shown in the lower half of Figure 4.11. A decrease in sensitivity will thus be

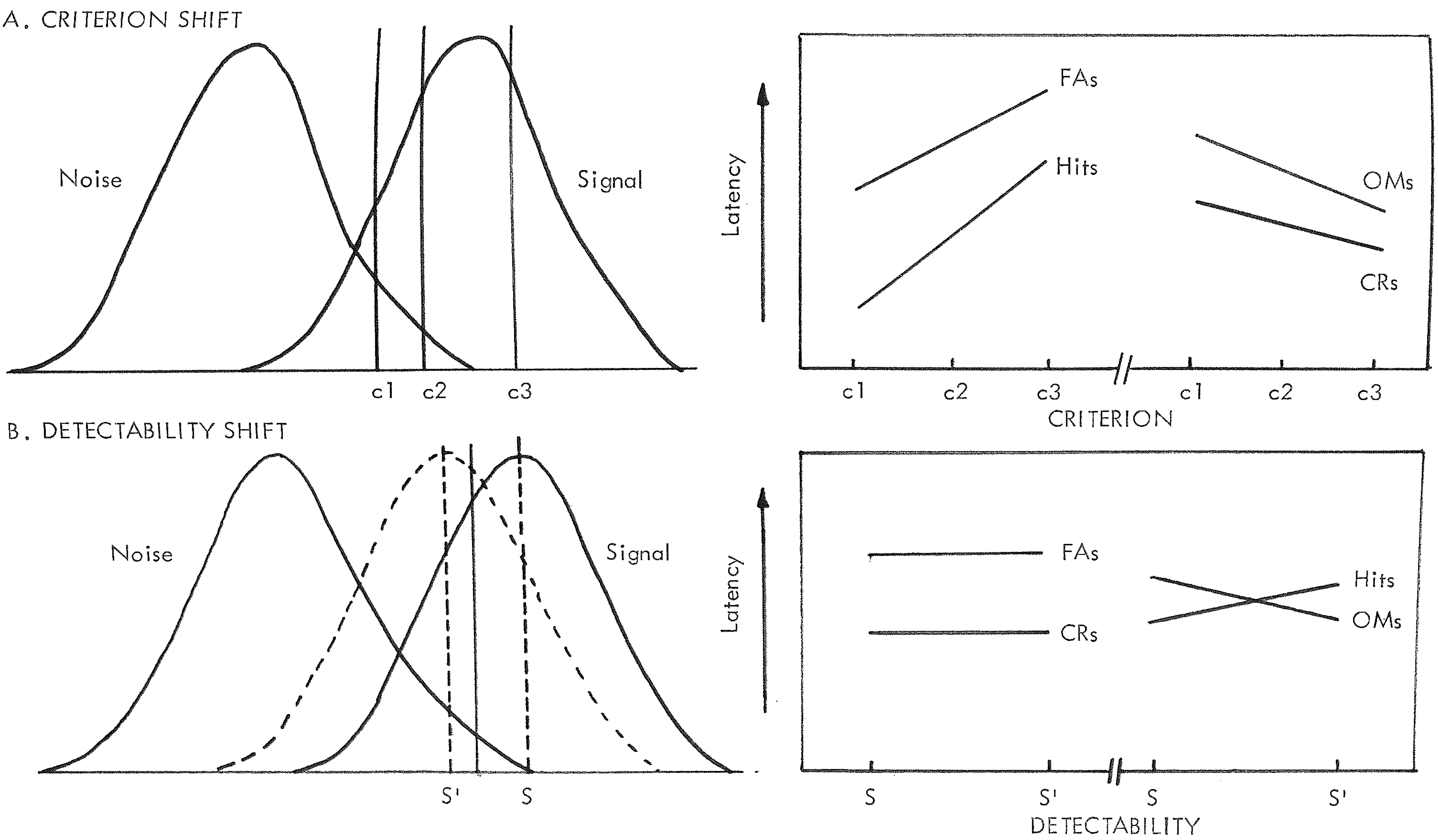


FIGURE 4.11 The predicted trends in the latencies of correct detections (Hits), false alarms (FAs), correct rejections (CRs) and omissions (OMs) as a function of independent shifts in criterion (A) and detectability (B). In (A), it is assumed that detectability remains constant, while a constant criterion cut-off (solid vertical line) is assumed in (B).

associated with a change in the mean latencies of responses to the signal distribution; for Yes responses the mean latencies will increase since, on average, Hit responses will be distributed nearer the criterion as the signal distribution shifts left, while the opposite will occur for No responses (omissions). Assuming criterion position remains invariant with respect to the noise distribution, latencies associated with responses to nonsignals (false alarms and correct rejections) will not change significantly. However, these predicted trends hold strictly only given the rather strong assumption that the criterion c is held constant with a change in detectability; we saw in Chapter 4 that the evidence on this point was equivocal, and it remained an empirical question whether the criterion or the likelihood ratio was held constant with a change in detectability. The decision theory analysis of latency data thus provides a potential means of clarifying this problem.

The relative ranking of the latencies we have shown in Figure 4.11 at each criterion position is arbitrary; the latency model does not specify differential latencies for Yes and No responses, unless asymmetric normal distributions are assumed. For instance, if signal variance is greater than noise variance, Yes response latencies may be shorter than No response latencies, at least for certain criterion positions. The model does, however, predict different mean latencies for correct and incorrect responses. If criterion placement is such that probability correct is greater than probability incorrect, error responses will be distributed nearer the criterion than correct responses and will have longer latencies. Hence within Yes responses, correct detections will have mean latencies shorter than those associated with false alarms, while for No responses, correct rejections will have shorter latencies than those associated with omissions. The prediction of different latencies or correct and incorrect responses distinguishes this model from random walk models which predict equal latencies for correct and incorrect responses (Audley and Pike, 1965).

We have thus provided a working model whereby response latency data in monitoring tasks may be interpreted. Although it is likely that it is probably too simple in its present form, the model does indicate that both detection and latency data may, in principle, be interpreted within the same decision theory framework.

## 4.7 Physiological Concomitants of Decision Processes

We have now examined a number of different aspects of the decision theory approach to detection behaviour. Such an approach proceeds, as we have seen, from a form of TSD based in statistical decision theory. Moreover this form of TSD may achieve greater efficacy in its potential ability to interpret both detection and latency data, as we have seen in the last section.

Given this approach to the analysis of detection and monitoring tasks, it is pertinent to ask whether there is any physiological evidence of the decision making activity in such tasks. It would be expedient to have available, for example, independent physiological indices of the sensitivity and bias aspects of performance. We could then, in principle, examine the decision making model by investigating the effects of certain variables on both these indices. Of course, the utility of this approach rests on the assumption that the physiological indices do in fact reliably reflect the psychological processes they are intended to measure, and that any variations in the indices do not proceed independently of the processes under study. As we shall see, it is difficult to identify any indices which meet these requirements completely; and even in using the best available index we may be faced with problems of ruling out alternative sources of variation in the index, and in justifying the psychological basis of such variations. Nevertheless, a physiological approach, when appropriately applied, may shed further light on the nature of the decision making processes in monitoring tasks.

### 4.7.1 Physiological correlates of vigilance

Physiological studies of vigilance abound in the literature. Several indices have been employed, including skin conductance and other electrodermal measures, various measures of the electroencephalograph (EEG) and the evoked potential, electromyographic measures, heart rate and the concentrations of certain substances in the blood, such as adrenalin and noradrenalin (see Davies and Tune, 1970, and Mackworth, 1969, for references). The majority of these studies have been interpreted within an arousal or activation framework. The rationale behind such an arousal-based approach is as follows: during a long monitoring session, the observer becomes less aroused as the task proceeds, and this might result in him missing some signals. If a physiological index of arousal also shows a trend over the run indicative of a drop in arousal, then there is some evidence for the influence of arousal on the vigilance decrement. This type of interpretation forms the basis for a number of reports in the literature. There are a number of difficulties with arousal theory and the design of these experiments that make an unequivocal interpretation of these studies somewhat difficult. Firstly, there is no way of exluding the possibility that the observed trend in the physiological index proceeds independently of the overt performance trends.[10] Hence, although the physiological index may indicate a loss in arousal with time at work, this does not establish a causal link with the decrement in detection rate or latency. Moreover, when some task variable has been varied to produce a change in performance, there may be little change in the recorded physiological index. For instance, both signal frequency and noise have been shown to affect detection rate in vigilance tasks, but there is no evidence from the physiological studies that the change in performance is accompanied by changes in arousal (Davies and Treacher, 1968; Eason , Beardshall and Jaffee, 1965; Stern, 1966). A further problem arises from the finding that performance may be efficient at both high and low levels of arousal, and that there are individual differences in the optimum level of arousal within the so-called U-shaped continuum (Davies and Hockey, 1966). Some further problems with arousal theory are discussed by Broadbent (1963) and Jerison (1967b).

An alternative approach to the type of experiment where parallel trends in performance and an arousal index are reported is the one where subjects are classified according to individual differences in a physiological index <u>and</u> performance. In three such studies, reported by Coles and Gale (1971), Siddle (1972) and Verschoor and Van Wierengen (1970), it was found that subjects with a high physiological reactivity on tonic and phasic measures of electrodermal activity had higher performance levels and lower decrements than low-reactive subjects. A problem with these vigilance studies, however, one common to many vigilance studies, is that false alarm data were not reported; hence the reported superiority of reactive subjects might not be due to a genuine superiority in sensitivity. This was the exact result obtained by Parasuraman (1975a) in a related context; in a short-duration tone discrimination task, it was found that although reactive subjects (as indicated by several electrodermal measures) detected a significantly greater number of tones at all confidence levels, their discrimination sensitivity (d') was almost exactly equal to that of low-reactive subjects. The apparent superiority of high-reactive subjects was entirely due to their relative riskiness in responding, and not due to a genuinely superior discriminative ability.

A further alternative in the design of psychophysiological experiments in vigilance is to investigate the relationship between short term or time-locked changes in physiological indices and performance indices. Some examples of this type of study are reported by Daniel (1967), Groll, (1966), Haider, Spong and Lindsley (1964) and Wilkinson, Morlock and Williams (1966). These studies used EEG measures, and found that while correct detections were associated with periods of high arousal, physiological activity just preceding an omission was indicative of lowered arousal. On the other hand, Opton (1964) found the reverse, his study indicating that omission errors were associated with EEG activation in certain paced inspection tasks, especially in older subjects. We shall, however, consider this type of study in further detail in the following sub-section.

### 4.7.2 Evoked potentials and decision-making activity

The studies reported by Haider et al. (1964) and Wilkinson et al. (1966) made use of averaged EEG measures, that is, of evoked responses, or evoked potentials. In both studies it was found that components of the evoked potential (EP), obtained by averaging the EEG over a number of trials, showed a trend over the work period which was indicative of a decrease in arousal; again, however, a reliable link between the physiological index and performance was not demonstrated. Moreover, while in both studies it was noted that the form of the EP was different according to the type of response trials in which the EP was averaged, the implications of this finding were not fully explored.

Since these early 'pioneering' studies, however, a number of investigators have emphasised the need to obtain separate EPs by pooling trials arising from different behavioural categories (Donchin and Lindsley, 1969; Paul and Sutton, 1973; Squires, Squires and Hillyard, 1975a;). Thus the recording of EPs contigent upon different types of response may provide a better means of examining the physiological concomitants of the 'underlying' information processing and decision making strategies than the gross averaging of EPs regardless of the type of response. This type of analysis of EPs has been applied to a number of substantive areas in psychology (for a bibliography, see Price and Smith, 1974); however little or no research has been reported in the area of monitoring behaviour (but see Ritter and Vaughan, 1969; Wilkinson and Haines, 1970).

The basic premise in the use of EPs in psychological research is, that by averaging the EEG in a 'time-locked' fashion to either the stimulus or response events, some information may be gained of the processing and decision making mechanisms linking these events. This itself depends on two assumptions: firstly, that the method of averaging the EEG provides a reliable index of the underlying cortical processes, and secondly that a link between the EP and a psychological process may be established. Although both these assumptions have been challenged, for example, in the former case by Sayers, Beagley and Henshall, (1974), and in the latter by Clark, Butler and Rosner (1970), it is beyond the scope of this review to consider these studies in detail [11]. Although only the second class of objection to the EP has been adequately answered (Donchin and Sutton, 1970; Paul and Sutton, 1973), we shall, as the majority of investigators in this area, take both assumptions to be valid.

Early investigators of the EP tended to think of these averaged brain potentials as purely reflecting cortical activity triggered by external stimuli. Since the influential report of Sutton, Braren, Zubin and

John (1965), however, much evidence has gathered that the EP may also be associated with data processing activity within the cortex, that is with 'psychological processes' rather than with 'sensory processes' alone. Certain 'late' components of the EP, in particular a positive-going complex peaking approximately 300 m.sec. after the stimulus (variously termed 'P3', 'P300' or 'association cortex potential'), has been reported to be extremely sensitive to the influence of a large number of 'psychological variables', while being relatively unaffected by the physical properties of the stimulus. These psychological variables include the concepts of 'information delivery','task relevance', 'stimulus uncertainty', 'selective attention', and others (see Karlin, 1970). It is clear that P300 reflects 'endogenous' cortical processes, but there is little consensus concerning their nature.

The P300 component of the association cortex potential has been distinguished from earlier components, which may be identified with sensory evoked potentials, on the basis of both latency range and intercranial origin (Vaughan and Ritter, 1970). Vaughan and Ritter also state that "the association cortex potentials differ importantly from the (sensory) evoked potential in respect to the factors determining their latency and amplitude: (sensory) evoked potential parameters being primarily defined by stimulus variables, and the association cortex potentials by task variables or stimulus significance" (Vaughan and Ritter, 1973, p. 141). The overall EP can thus be viewed as reflecting a dual process; firstly, a sensory stimulus processing stage, followed by a more labile, diffuse potential which, when present, represents either the code or the sign (Uttal, 1967) of some data processing activity associated with certain categories of stimulus and response. The major problem is in evaluating the reliability of the many psychological variables which have been examined. The literature is full of a confusing welter of results, and at least one major investigator in this field has been forced to remark "God only knows what P300 is!" (quoted in Naatanen, 1975). The same investigator has, however, proposed a similar two-process view of EPs (Hillyard and Picton, 1974) as that suggested by Vaughan and Ritter (1973).

In such a climate of uncertainty regarding the nature of P300, it might seem foolhardy to propose another 'variable' as the major one associated with P300. However, there is growing evidence from very recent studies that variations in <u>decision-making</u> activity may play an important role in determining P300 amplitude and latency (Davies and Parasuraman, 1976; Squires, Squires and Hillyard, 1975a, b). Decision making may be treated either in the specific sense of TSD (as suggested by Squires et al. 1975a,b), or in a more general sense, in which case, P300 may be related to the activity of a general purpose 'decision' or 'cognitive' processor (Donchin, Kubovy, Kutas, Johnson and Herning, 1973). In the former treatment, an analogy may be made between the two-process view of EPs and the two-process theory of signal detectability; in the latter, Donchin et al. (1973) draw the analogy of a floating point processor (attached to a general purpose computer) which might be invoked if the many incoming programs possess a single, broad criterion in common.

The two-process analogy between the EP and TSD is almost certainly too simplistic, but this gross relationship might serve to elucidate, to a first degree, some of the physiological processes underlying detection and discrimination behaviour. Certain points of similarity are present: for example, in the TSD model, sensitivity is usually constant for a given observer and for fixed sensory stimulation. In these conditions, the sensory evoked potential is also remarkably stable. On the other hand, variations in decision behaviour in the TSD model are due to variations in the response criterion, which may be affected by a large number of 'psychological' variables. The same psychological variables have been shown to influence the form of P300 and other late component potentials.

A number of recent studies have attempted to obtain an empirical basis for this analogy by recording EPs in signal detection tasks. Paul and Sutton (1972) for instance, manipulated the response criterion by varying both the a priori signal probability and the payoff matrix, and found that the amplitude of P3 was systematically related to the strictness of the response criterion. Similar results have been reported in experiments on auditory detection within an EP-TSD framework by Hillyard, Squires and associates (Hillyard, Squires, Bauer and Lindsay, 1971; Squires et al., 1975a,b). Perhaps the most impressive result demonstrated by this group is the one illustrated in Figure 4.12. Here the EPs to correctly detected signals are plotted in an order corresponding to the appropriate criterion cut-off (c) on the associated group of trials, and without regard to the method by which the criterion was derived (ratings or variations in signal probability). This result strongly supports the contention that for such tasks the amplitude of the P3 component is closely related to the subject's psychophysical decision that a signal is present.

While the majority of the EP studies we have reviewed have used measures of the amplitude of EP components, latency measures have received very little attention. Since P300 and the other components of the association cortex potentials have been related to aspects of decision making, the peaking time of such components might be expected to reveal some information about the temporal aspects of the decision activity; using Donchin et al.'s (1973) analogy, the computation time in the floating point processor might give us some further information regarding the nature of

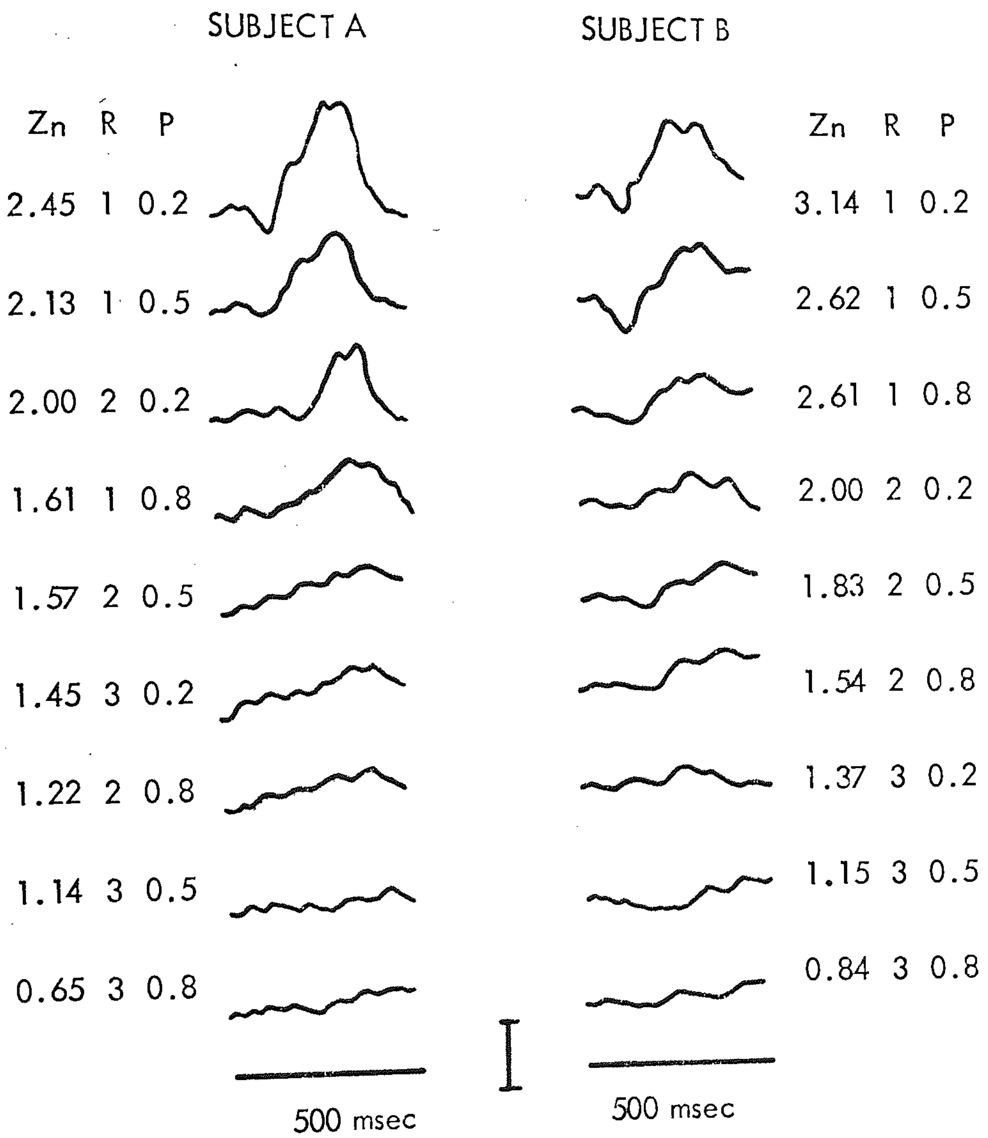


FIGURE 4.12 Evoked potentials averaged on Hit trials ranked by magnitude and objective criterion cut-off values (Zn) obtained from rating categories (R) or *a priori* signal probability (P). Calibration 10 microvolts for Subject A and 5 microvolts for Subject B, positivity upwards. Adapted from Squires, Squires and Hillyard, 1975a, Figure 4.

the program, and hence of the decision making process.

A few experiments have reported that late component latency measures are sensitive to both between and within subjects variations in decision time (Donchin and Lidsley, 1966; Squires et al., 1975a,b). Ritter, Simson and Vaughan (1972) also briefly reported a relationship between latency in a discrimination task and late component EP latencies. At this point, however, we have reached the borders of knowledge on this subject, and must seek fresh evidence in support for our speculations. We shall accordingly report an experiment later in this thesis where decision latencies and EP component latency measures are examined in relation to each other within a decision theory framework. Although the studies reviewed in this section are only suggestive, it may be possible to analyse EP measures in a monitoring situation within the type of model linking EPs to TSD which we have discussed.

In summary, we may conclude that EPs may provide a more 'central' index of cognitive activity than other physiological indices, and may be related to decision making activity, although the mechanisms underlying the relationship are not fully understood. Examination of the latencies of EP components, as well as the amplitudes, might serve to elucidate the nature of the relationship.

## 4.8 Assumptions in the Application of Decision Theory

Having discussed the several aspects of theories of detection and discrimination, we now return to a consideration of TSD and decision theory from the point of view of the implicit and explicit assumptions they make. The theory has not undergone a major revision since the original prescription of the model for the analysis of detection performance with weak auditory signals. Recently, efforts have been made to emphasize the generality of the theory (Swets, 1973), but the original model of Tanner and Swets (1954) may still be considered to hold. The problem thus arises of the suitability of the theory to describe psychological phenomena not envisaged in its original conception, and consequently, objections to the use of TSD have often been raised. The counter to this is that the original model was specified in too narrow a context, and that a generalized statement of TSD can be made in terms of the broader tenets of statistical decision theory.

Nevertheless, the problem of the validity of the assumptions in TSD and decision theory must be dealt with. In this context, it might be expected that the use of TSD in the analysis of monitoring behaviour, which is not unrelated to detection behaviour, would be less open to objection than the use of the theory in less related fields, such as recognition memory. However, if anything, the reverse is true, a number of objections having been raised to the use of TSD in monitoring behaviour. (Jerison, 1967; Wiener, 1973). Some of these have been useful, insofar as it is important to determine the limits of applicability of TSD (cf. the concepts of 'strong' and 'weak' applications of TSD in Broadbent, 1971). Others, however, have been mainly directed at the use of the TSD parameters d' and $\beta$ (Jerison, Pickett and Stenson, 1965), but as we have seen, a TSD type of analysis need not be restricted to the use of these parametric indices.

In considering the assumptions in the use of TSD, we may consider three sources: assumptions in the use of decision theory, those in the application of TSD to psychophysics, and those in the application of TSD to monitoring behaviour. A hierarchy of assumptions may thus be built up.

A. Decision theory assumptions

B. TSD assumptions in psychophysics

C. TSD assumptions in monitoring

## A. Decision theory assumptions

The use of statistical decision theory to describe human behaviour involves some very general assumptions about human functioning which may be disputed, but which are not particularly amenable to empirical test. Other assumptions involve the rationality and cooperation of the observer in an experimental situation, and the way in which he uses his decision rules to achieve his decision goals. We will list five assumptions regarding decision rules and goals:

1) The subject partitions the elements of the set of 'evidence' into response classes on the basis of a decision rule.

2) The decision rule is used to satisfy some goal, such as the minimization of error.

3) If the subject has at his disposal two decision rules P and Q such that P better satisfies the goal than Q, the subject consistently chooses P unless otherwise instructed.

4) The decision rules and goals inferred from performance data[12] bear a relation to those actually used by the subject.

5) Decision goals do not fluctuate from trial to trial.

## B. TSD assumptions in psychophysics

The assumptions under this heading include those under A as well as some additional ones:

1) The amount of sensory information received by the subject is limited (cf. 'data-limited' tasks, Norman and Bobrow, 1975).

2) The 'evidence' received by the subject can be uni-dimensionally represented.

3) The probability density distributions of signal and noise are Gaussian (although this need not be assumed a priori, see 4.2 and 4.5).

4) The criterion used by the subject remains constant in a given block of trials.

5) The various experimental paradigms of TSD (Yes/No, rating, forced-choice) are all formally equivalent (see 4.3).

## C. TSD assumptions in monitoring

The additional assumptions in the application of TSD to the analysis of monitoring behaviour are:

1) The subject has a clearly marked decision interval in which signal or noise may occur.

2) The evidence received by the subject conforms to that of the TSD model and is not distorted by other (possibly attentional) factors (again this need not be assumed *a priori*).

Having rather baldly stated these assumptions, let us consider their implications for a decision theory analysis of monitoring behaviour. It is evident some of the assumptions are of no consequence, some are met, while others may be questioned. As far as the assumptions in decision theory are concerned, probably the only ones which may be seriously questioned are assumptions 4 and 5, regarding the equivalence of the inferred decision rules to the actual decision rules, and the invariance of decision goals. It will be recalled that there is some evidence that subjects may not respond with the aim of maximising expected value, but appear to match the frequency of their responses to the frequency of signals (Thomas and Legge, 1970). Whether this holds for monitoring tasks is a matter for empirical investigation.

Assumption B1 is also worthy of closer consideration. The statistical limitation of the sensory information is a crucial assumption for a fixed sample theory of decision making, as in TSD. If the available sensory information is not limited, then performance need not lie within the convex set in Figure 4.1. This assumption is almost certainly not valid for the so-called unlimited-hold tasks, where the signal remains present until detected (see 2.2). The assumption of limited sensory information is also important as far as a decision theory analysis of response latency in monitoring is concerned, and we shall return to it at various other points in this thesis.

Apart from being a fixed-sample model, TSD is also a fixed-criterion model. Variations in the criterion are only considered across conditions where there are grounds to assume a change in response bias, such as across different signal probability levels. Recently, however, a few investigators have considered the effects of trial-to-trial criterion variation within a TSD context (Ingleby, 1968; Thomas, 1973; Thomas and Myers, 1972). Ingleby (1968) has shown that the effect of criterion variance is to increase the signal to noise variability in the TSD Gaussian model, so that apparent decreases in sensitivity might be obtained. Broadbent and Gregory (1967) put forward a similar interpretation to explain their finding that subjects have a decreased sensitivity to emotionally toned words. Thomas and Myers (1972) have also discussed some of the effects of criterion variance on the OC. On the whole, criterion variance has received very little attention in the literature. In the TSD

Gaussian model, estimating the effects of criterion variance is difficult since at least one further parameter has to be estimated. In a related context, the increase in the criterion over a block of trials is a worthy candidate for further investigation by detection theorists, especially since a decision theory analysis of monitoring reveals, as we shall see in the next chapter, that performance is associated with an increase in the criterion with increased time at work.

Of the other assumptions in A, B and C, some have been discussed previously, for example the assumption regarding the equivalence of TSD procedures. The assumptions in C will be given further consideration in the following chapter. It is clear that all the assumptions we have listed are not completely met, but we have circumscribed the types of task in which these assumptions may hold, and on the whole, we may assert that a non-rigorous form of TSD, based in decision theory, is appropriate for the analysis of monitoring behaviour. Some of the other assumptions we have listed need not be taken a priori, and can be discarded if the experimental data so dictates. What is important, however, is that these assumptions be considered and rejected, if necessary, in any particular application. In subsequent chapters we shall make frequent reference to them and attempt to justify our decision as to whether they influence or have no effect on the type of behaviour under study.

This section concludes the discussion of aspects of detection behaviour and decision theory. In the next chapter, we shall relate these considerations to the analysis of monitoring behaviour, and review a number of studies in the literature from a decision theory point of view.

## 4.9 Notes to Chapter 4

1 Theories may also be identified in which decision uncertainty arises out of the choice of particular stimuli from a larger set (Garner, 1974), or in a choice of responses as a 'function' of different stimuli (Anderson, 1974).

2 The OC represents the best possible performance only if stimulus information is limited (see Laming, 1973, p.77).

3 Whether the decision rule in TSD is based on likelihood ratio or on the evidence variable is a matter for continuing debate. The two are equivalent only in the equal-variance TSD model.

4 The terms sensitivity, detectability and discriminability will be used interchangeably, although strictly we speak of the sensitivity of the subject and the detectability or discriminability of the signal.

5 A monotonic relationship between likelihood ratio and the decision variable holds, not only for the normal and exponential distributions, but also for the chi-squared and gamma distributions, as well as for simple ramp and gate distributions (see Egan, 1975).

6 The parsimony gained by reducing the number of free parameters in a detection model to one may be cancelled by the fact that such models fit fewer data than two-parameter models.

7 Two models not reviewed by Audley (1973) which may be classified generally as 'sequential sampling or counting' models, are the ones proposed by Luce and Green (1972) and Vickers, Nettleback and Wilson (1972).

8 For recent elaborations of the sequential sampling model for latency, see Link and Heath (1975).

9 'Samples' or 'observations' may either refer to stimulus sampling in the 'observing response' sense, or in the sense of samples drawn from an iconic store of the stimulus (Sperling, 1960).

10 The problem also arises of deciding between physiological indices exhibiting differing trends over the monitoring period. Thus a number of studies have found that while one physiological index shows a trend indicative of a loss in arousal, for example a decrease in skin conductance, other physiological indices, simultaneously recorded, may show no change or an increase in arousal (Coules and Avery, 1966; Opton, 1964; Stern, 1966). This is the problem of <u>directional fractionation of response</u> (Lacey, 1967).

11 Methodological and psychological issues in EP research are treated in greater depth in Naatanen (1975) and Sutton and Tueting (1975).

12 The decision rule inferred from data gathered over a number of trials is not necessarily the one used by the subject for a single trial; the optimum decision rule for a single trial is the one which is correct for that trial, and this may not be the same as that inferred from block trial data.

# CHAPTER FIVE

## DECISION PROCESSES IN MONITORING BEHAVIOUR

## 5.1 Decision Theory Analysis of Monitoring Performance

We have seen that the theory of signal detectability (TSD) of Tanner and Swets (1954) may, in general, be treated in two different ways. We may choose to interpret the theory in a specific way, with reference to Gaussian probability distributions, or in a broader sense within the framework of statistical decision theory, in which case the forms of the probabilistic representation of stimuli need not be specified a priori. The former treatment may lead to the formulation of sensitivity and bias parameters such as d' and B, while the latter implies a more general analysis in terms of the operating characteristic (OC). In the application of TSD to the analysis of monitoring performance, most investigators have employed the specific form of TSD (but see Broadbent, 1971).

In most of these applications, empirical validations of the assumptions concerning the Gaussian form of the probabilistic representations have not been carried out. In consequence, some of the results obtained using the parametric measures d' and B may be unreliable; the question therefore arises as to whether they are preferable to the more traditional measures such as the detection rate and reaction time. Some investigators do not think they are (Jerison, 1967; Wiener, 1973), and object to the analysis of monitoring performance using TSD. However, their objections are mainly directed to the use of d' and B, and, as we have stated many times, a TSD analysis is not restricted to the use of these measures.

It can readily be acknowledge that d' and B are not the exact equivalents of the parameters of psychophysical detection, but then in monitoring situations we are not usually concerned with obtaining precise fits to psychophysical functions, but with evaluating the trends over time in both signal sensitivity and response bias. The main contribution of TSD has been to point out the need for examining both these aspects of performance, and to re-establish the importance of the false detection rate. TSD, or more correctly, decision theory, provides a sound basis for interpreting both correct and false detection data, and a means of 'encapsulating' both types of data in the OC. As we have also noted previously in 4.4.1, merely plotting a single pair of hit and false alarm probabilities provides more information than does a consideration of the 'traditional' hit rate alone.

An analysis of monitoring performance in OC terms need not be dependent on the restrictive assumptions underlying the use of d' and B. Such an analysis might also benefit from some newer techniques of isolating independent measures of sensitivity which are less affected by departures from the equal variance model of TSD, or from methods of obtaining non-parametric measures (see 4.4.1). We are primarily interested in knowing whether a given performance difference between two conditions represents a change in bias or a change in sensitivity, or possibly both, or neither. Such differences are better evaluated on the basis of OCs than by a consideration of hit rate alone. Figure 5.1 shows both sensitivity and criterion changes for OCs plotted in the unit-square and on double probability axes. As we shall see, a re-interpretation of the vigilance 'decrement' shows that performance changes within a monitoring session in most monitoring tasks are associated with the latter type of change, that is, a change in the response criterion at invariant sensitivity.

In signal detection tasks, the OC may be obtained by a number of different methods; some of these were reviewed in 4.3, in which it was concluded that the rating method was the most efficient procedure for the derivation of OCs, although its equivalence to the standard Yes/No paradigm has not been reliably demonstrated. For monitoring tasks, two studies have found only minor differences between d' values obtained by single response and rating procedures (Guralnick and Harvey, 1970; Loeb

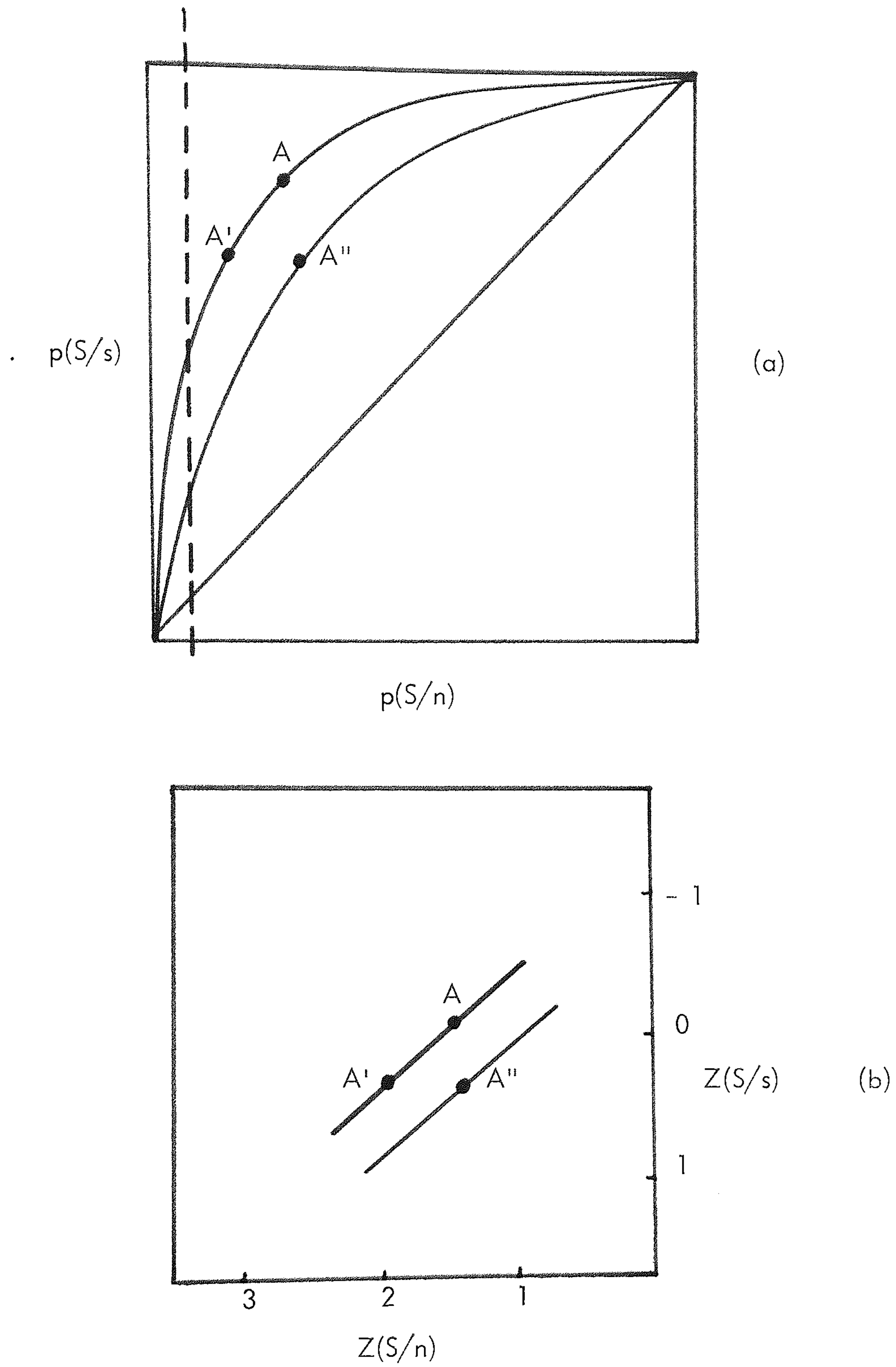


FIGURE 5.1 (a) An illustration of how, in principle, an operating characteristic (OC) analysis can be used to determine whether a given change in performance is due to a change in criterion or sensitivity. Movement along the OC from A to A' represents a change in criterion so that fewer events are responded to positively, while a movement from A to A" represents a fall in sensitivity, primarily. In practice, these changes may not be easy to distinguish if the false alarm rate restricts the OC to the left of the dotted line, although they may be easier to distinguish in z space, as in (b).

and Binford, 1964), although corresponding OCs were not compared (see also Appendix D).

The number of studies which have analysed monitoring behaviour in OC terms is small, but in most of these it has been found that the empirical OCs conform fairly well to the predicted TSD Gaussian form; it will be recalled from Chapter 4 that such OCs describe a continuous monotonic function in the unit square, and a straight line in z space. Such straight line OCs have been obtained in monitoring situations (Broadbent and Gregory, 1965, re-analysed by Mackworth, 1970; Loeb and Binford, 1964; Milosevic, 1974, and personal communication; see also Chapters 7 and 9).

Loeb and Binford (1964) derived both single-response and rating scale operating probabilities and plotted them on normal deviate co-ordinates. All the points were well fitted by the equation -

$$z(S/s) = 1.05\ z(S/n) + 2.94$$

which is in accordance with the equal-variance model of TSD. This assumption has not often been empirically checked in studies in the monitoring literature which have used the parametric TSD measures d' and B. While these results suggest that the departures from the assumption of equal variances are not great, further experimental work is required, possibly within a task classification framework, to enable an evaluation of the possible types of tasks for which the assumption does not hold. There is already evidence from signal detection tasks that visual tasks invariably have skewed OCs, but this needs to be confirmed and extended to monitoring tasks. At the same time, it should be noted that there is some evidence that the equal-variance TSD model holds up remarkably well even in the performance on certain operational monitoring tasks, such as industrial inspection (Sheehan and Drury, 1971; Wallack and Adams, 1969)[1].

One of the first studies to analyse monitoring performance within a TSD framework was reported by Egan, Greenberg and Schulman (1961a). They used highly trained subjects to examine the detection of tones in background random noise in 2-min. listening periods, separated by brief rest pauses. The subjects were required to respond with varying degrees of confidence in different listening periods. Egan et al. found that sensitivity, as measured by d', was invariant over conditions, and that the OC described a downward concave curve consistent with the Gaussian TSD model. The authors concluded that the temporal uncertainty of signals (which is also a major feature of more typical monitoring tasks) exerts a major influence on detection strategies in the so-called 'free-response' tasks; in another study this view was confirmed, as it was observed that there was a progressive fall in d' as the interval of time uncertainty during which signals might occur was increased (Egan, Schulman and Greenberg, 1961b). Mackworth (1970) has emphasized the relationship between the interval of time uncertainty and the rate of presentation of events in 'discrete' tasks. Hence, this study also provided some evidence for the importance of the <u>event rate</u> in determining sensitivity trends in monitoring; we shall consider these in a later section. However, Egan et al. found no change in d' between confidence levels, in accordance with TSD. Thus while the tasks used by Egan et al. were not monitoring tasks in the strict sense (because of their short duration and since rest pauses were provided), their studies are important in providing the first empirical evaluation of a TSD analysis in free response tasks, and for indicating the possibilities for similar analyses of prolonged monitoring tasks.

Broadbent and Gregory (1963a) and Mackworth and Taylor (1963) reported the first experiments in which TSD was applied to the analysis of performance in more typical monitoring tasks. Broadbent and Gregory used two monitoring tasks; in their visual task the detection of a brighter flash appearing on one of three flashing lights was required, while the auditory task involved the detection of a pure tone presented at unpredictable times. A TSD analysis of performance with both tasks showed clearly that the decrement in detections with time on task (the 'traditional vigilance decrement') was not due to a loss in sensitivity over time, but rather, was associated with an increase in the response criterion over the work period. This important result, which was also earlier suggested by Egan et al. (1961a), has since been confirmed in further experiments (Baddeley and Colquhoun, 1969; Broadbent and Gregory, 1965; Colquhoun, 1969; Levine, 1966; Milosevic, 1974; Williges, 1969).

On the other hand, Mackworth and Taylor (1963) reported that for subjects working with the Continuous Clock, a display in which brief cessations in the continuous movement of a clock hand have to be detected, there was a significant decline in sensitivity (d') over the monitoring period. Again, some studies have since also reported that there is a decrement in sensitivity within sessions, although this aspect of monitoring performance has not received as much attention in the literature (Deaton, Tobias and Wilkinson, 1971; Hatfield and Soderquist, 1969; Leob and Binford, 1968; Mackworth, 1965a, b), and remains to be verified.

There are thus two interpretations of the 'vigilance decrement' which receive support in the literature; the first, which is by far the more widely held view, is that sensitivity is invariant over a monitoring period, and there is an increase in the criterion, while the second proposes that there is a genuine decrement in sensitivity. These need not necessarily be conflicting views, if it can be demonstrated that they are relevant to different categories of task. It is therefore clearly important to establish the relative influence of criterion and sensitivity shifts, and the types of task in which these effects might appear. This forms the concern of the following sections and of some of the experimental work reported in this thesis. In the next two sections we shall consider various aspects of criterion and sensitivity shifts in monitoring.

## 5.2 Criterion Shifts and their Interpretation

We have seen that a TSD analysis of monitoring performance leads to a separation of sensitivity and the response criterion. The criterion can be taken to more or less equivalent to the criterion in signal detection, depending on whether one assumes the narrow or broad interpretation of TSD. In Chapter 4 we mentioned some independent variables which exerted important influences on the response criterion in signal detection tasks, such as the *a priori* signal probability, the payoff matrix, instructions, and certain types of feedback. These and other variables have also been shown to similarly affect the criterion in monitoring tasks, and thus we shall briefly examine them.

### 5.2.1 Variables affecting criterion placement

One 'variable' which has been shown to influence criterion placement in monitoring has already been mentioned: this is time on task, which leads to an increase in the response criterion. The increase has been shown to be maximal for strict criteria, that is for those responses associated with a high level of confidence (Broadbent and Gregory, 1963a, 1965; Levine, 1966; but see Milosevic, 1974, 1975). We shall consider the time on task variable in more detail when we come to interpret criterion shifts over time. However, a number of other variables, common to detection and monitoring tasks have been investigated:

#### Signal probability

The change in the criterion with a change in signal probability leads to the derivation of the OC, as we saw in Chapter 4. In monitoring tasks, although the complete range of the OC has not been spanned, it has generally been shown that an increase in signal probability leads to a riskier criterion, as predicted by TSD. We shall reserve a fuller consideration of the influence of signal probability until section 5.4, where the available evidence will be reviewed in greater detail.

Instructions

Another way of generating the empirical OC is the method where observers are asked to adopt different degrees of responsiveness in different sessions. Colquhoun (1967) used this method in a monitoring situation, by asking his subjects to report signals (sonar reports) which they were completely sure of, or, in a separate session, of anything, however doubtful. As predicted by TSD, the criterion was significantly lower in the 'doubtful' session, a finding which is inconsistent with a simple threshold model of monitoring behaviour (see 4.5). Williges (1973) has also reported that if observers are asked to be more or less cautious in a monitoring situation, B shifts in the appropriate direction.

Cost and values

Central to TSD and decision theory is the view that decisions are influenced by the relative costs and values attaching to alternative responses. Thus in making a statistical decision, we may be concerned with the costs attached to the making of Type 1 and Type II errors, and utilize these to maximize or minimize some goal value. In monitoring, there is some evidence that response probabilities are influenced by the values attached to the various response alternatives. Levine (1966) varied the costs attached to making a false detection or an omission, and found that B increased as both costs were increased, although there was no interaction with the rate of increment in B over the monitoring period. The finding that B increases even when the cost of missing a signal is raised is rather surprising since subjects should (ideally) adopt riskier criteria (that is, increase their positive response rate); this result is thus directly the opposite of that predicted by decision theory. Williges (1971) has also reported that the effects of costs and values of signals are not as marked, and sometimes not in the same direction as that predicted by TSD. On the other hand, Davenport (1969), using a vibtrotactile monitoring task, found that by varying the costs of false alarms and omissions and the value of a hit, B changed in the appropriate direction (although not as much as on an expected value model), while d' remained constant.

Other variables

Some other variables have been shown to influence the response criterion in monitoring, including knowledge of results (Wilkinson, 1964, re-analysed by Mackworth, 1970; Williges and North, 1972), noise (Broadbent and Gregory, 1963a, 1965), induced hyperthermia (Colquhoun and Goldman, 1972), sleep deprivation (Deaton et al. 1971) and other stresses (Poulton and Edwards, 1974).

Of these, the results of Broadbent and Gregory (1963a, 1965) are probably the most reliable, primarily because they employed a method enabling an analysis in terms of the OC. Their results are complex, but can be summarized by saying that the effect of noise is to move risky and strict criteria closer together (that is, to effect a contraction of the decision axis). We shall consider the implications of this result at various points in the following chapters.

### 5.2.2 Interpretation of criterion shifts

A number of variables influencing criterion placement in monitoring tasks have now been identified. Most of the results of the studies reviewed have been in agreement with the general predictions of TSD; the best established result is probably the increase in the response criterion with time at work and with diminished signal probability. However, while the latter effect and those due to some of the other variables reviewed (for example, instructions) can be readily interpreted on the basis of TSD alone, explanations of the effect of time at work must go beyond TSD and invoke certain theoretical constructs.

One construct which may be examined in relation to criterial shifts in monitoring is expectancy, which has also been proposed as a general mechanism underlying monitoring behaviour by some investigators (Baker, 1959; Deese, 1955; see also Broadbent, 1971). Support for expectancy theory, as taken within a decision theory context, has come from a number of important experiments demonstrating the critical importance of the expectancies established during training sessions for monitoring tasks (Colquhoun and Baddeley, 1964, 1967; Williges, 1969). These experiments showed that, for both visual and auditory tasks, if a training session using a high signal rate is given, then a greater within-session decrement in hits and false alarms is observed during the monitoring period than if a training session using a low signal rate appropriate to that used in the monitoring session is given. The frequency of signals expected by subjects, which is established during training, exerts an important influence on within-session trends in performance in the monitoring period. If therefore, subjects are trained with an inappropriately high signal rate, they find that signals occur much less frequently than they expected, and consequently revise their criteria towards greater strictness; this might therefore result in a sharp decline in both hits and false alarms and a rise in B, as has been reported by a number of investigators.

We might therefore suppose that if subjects were given appropriate training, and allowed to have practice sessions to stabilize their criterion, then there should be no increase in the criterion with time on task. In fact, however, an increase in B is still found, even though subjects have been 'expectancy matched' (Baddeley and Colquhoun, 1969; Milosevic, 1975; Parasuraman, 1975b; Williges, 1973). Do these results therefore argue against the expectancy approach?

No, since we can propose another form of expectancy, a variant of the mechanism originally proposed by Baker (1959) in a related context. Since signal probability is usually low in monitoring situations, it may be assumed that a subject always under-responds to signals (that is, for signal probability $p < 0.5$, $p(\text{Yes}) < p$ is implied by the expected value TSD model). If, in addition, we assume that the subject monitors his own responses in a self-feedback loop, and takes his response rate as reflecting the approximate signal rate, he will revise his criterion since he is under-responding. Revision of his criterion to one of greater strictness will result in a lower hit and false alarm rate, leading to further stricter revision of his criterion, and so on in a 'vicious circle' (Broadbent, 1971). This explanation can also be applied to the finding that there is a greater increase in the strict criterion than for lax criteria, if we assume that in a rating task the subject would be more likely to base his estimate of signal probability only on his confident (strict criterion) reports.[2]

Such an expectancy interpretation assumes that the observer follows the behaviour of the ideal TSD observer. Williges (1969, 1973) has proposed that this relationship holds, at least for weak signals. As we have seen in Chapter 4, there are, however, suggestions that subjects may not follow the TSD model of maximizing expected value; the often noted finding of the conservativeness of subjects (namely that they usually do not adopt as high or as low criteria as that predicted by TSD) may be an indication of probability matching strategies rather than those assumed by TSD (Thomas and Legge, 1970). It is also conceivable that a subject begins a monitoring session by over-responding, that is, sets riskier criteria than that used by the ideal TSD observer ($B < (1-p)/p$ for a symmetrical payoff matrix); then, by the 'vicious circle' argument, he (further) over-responds by feedback and revision of his criteria towards riskiness, thus exhibiting a decrease in B over time. However this would mean that the rate of positive (Yes) responses would be greater than the signal rate, and would become more so during the session, but there is no evidence of this in most of the experiments which have reported both hit and false alarm rates.

One implication of the 'vicious circle' expectancy approach is that an opposite trend in B over time should be observed if *a priori* signal probability is greater than 0.5 (assuming subjects begin by over responding initially). Broadbent (1971) considered this a crucial test of the expectancy approach, and, on the basis of an unpublished study by Simpson (1967), who obtained a within-session B increment even for $p > 0.5$, concluded that the expectancy explanation was not acceptable. However, Williges (1969, 1971, 1973) has shown in a series of experiments that there is no significant rise in the criterion with time at work if signal probability is increased beyond 0.5; these studies therefore support the vicious circle expectancy interpretation, and we shall consider them in some detail.

Williges (1969) used a task in which subjects were required to detect a periodic transient increase in the duration of intermittent visual stimuli. Critical signals were either presented with a low probability (.16), or a high probability (.84). These conditions were combined factorially with two conditions of pre-training in which subjects were either provided with accurate or inaccurate expectancies regarding mean signal frequency. In the 'accurate' condition, an appropriate signal rate was used to provide a set for the subjects (either .16 or .84), while in the 'inaccurate' condition subjects received signals with a probability of .5. The results for B conformed to that predicted by the expectancy model, and are illustrated in Figure 5.2. When subjects monitored under accurate instructions, B increased in the low signal probability condition and decreased (non-significantly) in the high

Aston University

Content has been removed for copyright reasons

FIGURE 5.2 Trends in likelihood ratio (B) as a function of pre-session expectancy (Accurate or Inaccurate) and signal probability (low or high). From Williges (1969, p. 65).

Aston University

Content has been removed for copyright reasons

FIGURE 5.3 Trends in B as a function of low (1/10) and chance level (1/2) signal probability conditions, and a condition in which signal probability changes in the session (1/10, 1/2). From Williges (1973, p. 183).

signal probability condition. Furthermore B was higher in the former condition. Both these results are as predicted by TSD and expectancy theory. When subjects receive an inaccurate set, B did not vary over time (the trend apparent in Figure 5.2 for the highest B values (High-Inaccurate) was not significant), and B was higher in the high signal probability condition. These results are thus consistent with the assumption that subjects use self-feedback to adjust their criteria during a watch. Williges (1969, 1976) has also interpreted his results as suggesting that trends in B represent attempts to reach optimal behaviour since the obtained B values approached the optimum B value for a symmetric payoff during the end of a session (see Figure 5.2).

There is thus somewhat more evidence for the expectancy theory approach than Broadbent (1971) put forward, and in a further experiment Williges (1973) confirmed his findings for both low and high signal probability conditions and for a condition in which signal probability changed halfway through a session (see Figure 5.3). The unexpected result by Simpson (1967) from which Broadbent (1971) concluded that the expectancy interpretation is untenable, cannot be easily explained, except by suggesting that individual differences may be a neglected factor in this area. Some subjects might exhibit a decrease in B over the run, while a fewer number of subjects show an increase; since there may be a 'cellar' effect for decreases in B, as Figures 5.2 and 5.3 appear to show, the larger increases in B might outweigh the decreases, and a group increment in B might be observed.

What evidence is there for alternative interpretations of the criterion increment in monitoring tasks? One alternative interpretation is one based on arousal theory; this has been considered in detail and rejected by Broadbent (1971), and so we will not consider it here, since there has been little fresh evidence since Broadbent's review. The general problem of arousal theory, which has been pointed out elsewhere (Broadbent, 1963; Jerison, 1967b), is the need to resort to the concept of over-arousal. It is almost a truism that virtually any result can be explained by arousal theory on the basis of the inverted-U function. This will be especially evident when we come to discuss sensitivity shifts in monitoring performance in the following sections. However, it should be emphasized that we are only rejecting arousal as a major theory for the interpretation of performance trends over the run; it remains of value in the explanation of monitoring performance changes between conditions where there are good reasons to suppose that arousal changes.

An explanation of the criterion increment in terms of the costs and values of responding has been suggested by Broadbent (1971); he proposed that there is a loss in the subjective gain associated with a correct detection as the work period progresses, resulting in an increment in B (on the expected value TSD model, the optimum value of B is inversely proportional to the gain of a hit). The data of Davenport (1969) and Levine (1966) suggest that variations in the cost of responding also influence criterion value between conditions. In a related context, cost is central to Smith's (1968) theory of detection, which, as we have seen in 4.5.3, may be used to explain the criterion increment. However, there is no direct evidence that the payoff matrix affects B trends during the monitoring period; and we have seen that the results of Levine (1966) and Williges (1971) are not entirely compatible with the predictions of the TSD expected value model. Williges suggested that signal probability has a greater influence on criterion bias than the payoff matrix, and the bulk of the evidence supports this view.

The results of Colquhoun and Baddeley (1964, 1967), therefore, and the complementary findings of Williges (1969, 1971, 1973) provide strong support for the expectancy approach to monitoring behaviour. To what extent such an interpretation suffices for other aspects of monitoring performance, and for other monitoring tasks, is not known, and forms a topic for investigation in subsequent chapters. The expectancy approach appears however, to provide the most parsimonious explanation for task situations where performance trends are associated principally with a criterion increment over the work period.

## 5.3 Variables Influencing Detection Sensitivity

It is clear from the results of a number of studies which we have reviewed so far, that in many monitoring tasks, the decrement in detections is associated with an increment in the response criterion rather than any loss in sensitivity. However, a number of studies have also reported sensitivity shifts in monitoring tasks. These studies are therefore important to evaluate on two counts: firstly, from a theoretical viewpoint, we wish to know whether the decision theory approach to monitoring behaviour is compatible with such tasks, and secondly, from a practical viewpoint, we need to examine the distinguishing characteristics of such tasks. We shall consider these two points in this and the following section, and return to them at various points in this thesis.

The first study reporting a sensitivity decrement within a monitoring session was by Mackworth and Taylor (1963). The task used in this study and others by Mackworth (1965a, b, 1968) was the Continuous Clock (Baker, 1963a), which is similar to the Mackworth Clock, but which has a continuously moving hand. The critical signal for detection in this task is a brief stoppage in the moving hand. It appears that the nature of the continuous observation required in this type of task is the relevant factor in the finding of a sensitivity decrement. This view was confirmed in a study by Mackworth (1968) in which the task was compared with the Mackworth Clock at two levels of signal rate. It was found that a decrement in sensitivity was obtained only for the Continuous Clock task. This result thus seems to indicate that if a task requires continuous observation of the display, there is a drop in effective sensitivity with time, or an effect of 'filtering', to use Broadbent's (1958) term. To what extent it is true of other sensory modalities, or of tasks with different types of signals, is not entirely clear; this is a major point for further investigation which is examined in some of the experimental work reported in this thesis. We shall see that a task classification analysis of different monitoring displays provides a solution to this problem.

At the same time, however, it seems clear that tasks requiring continuous visual fixation, such as the Continuous Clock, will exhibit a reliable decrement in sensitivity over the work period. This is also apparent, albeit indirectly, in a study reported by Stern (1966), in which subjects were required to continuously monitor a steady point of light and detect occasional movements of the light. Stern did not compute detectability indices, but it is clear from his data that there was an effective drop in sensitivity for the group of subjects as a whole, since while the number of hits declined over the session, the number of false alarms increased.

Another prominent feature of the tasks used by Mackworth and Stern, apart from the continuous nature of the tasks, is the high degree of <u>time uncertainty</u>; we mentioned previously that this feature may influence the effective sensitivity in a detection or monitoring situation (see 5.1). Mackworth and Taylor (1963) have also shown that sensitivity is lowered if the degree of uncertainty in the timing of signals is increased. However, in a related context, Laming (1973) has suggested that the reported loss in sensitivity in continuous monitoring tasks may be an artifact of the method used to measure sensitivity. He proposed that time uncertainty may be 'resolved', to a degree, by using continuously sampled data and that in continuous monitoring tasks, subjects will therefore employ a sequential sampling strategy rather than the fixed sample strategy implicitly assumed by TSD. This may certainly be true of continuous monitoring tasks, although it remains to be empirically demonstrated. The analysis of performance with such tasks might thus benefit from a 'sequential-decision' analysis, since these tasks have to be arbitrarily partitioned into 'observation intervals' in order to undertake a conventional TSD analysis.[3] However, the interpretation of the sensitivity decrement in terms of an artifact resulting from an inappropriate analysis can only provide a partial explanation, since sensitivity decrements have also been observed for discrete tasks with well defined observation intervals and weak signals. In such tasks, there is little evidence on which to choose the more complicated sequential sampling analysis in preference to the simpler fixed sample TSD analysis. Nevertheless, if the type of analysis carried out by Laming (1973) for reaction times proves of use in the analysis of detection performance in continuous vigilance tasks, this would provide an important contribution towards a task classification-based evaluation of the decision processes involved in diverse monitoring tasks.[4]

We have seen that a sensitivity decrement is obtained if a task requires continuous observation. However, decrements in sensitivity have also been reported for 'discrete' tasks, which only demand, in principle, intermittent observation (Bennedetti and Loeb, 1972; Binford and Loeb, 1966; Hatfield and Soderquist, 1969; Loeb and Binford, 1968; and others, see Chapter 12). Some authors have suggested that 'coupling' may be an important factor in determining sensitivity decrement (Loeb and Binford, 1968), while others have emphasized the importance of the stimulus event rate (Mackworth, 1970). We shall see that the 'coupling' interpretation does not tell the whole story, since sensitivity decrements may be observed for auditory as well as visual tasks; this explanation also suffers from the drawback that 'coupling' may be interpreted in opposing ways so that decrements in either visual or auditory tasks may be explained (cf. Deaton et al. 1971; Loeb and Binford, 1968).

On the other hand, there is evidence suggesting that the rate of stimulus presentation may play an important role in determining sensitivity decrements in monitoring (Loeb and Binford, 1968; Mackworth, 1965a). However, its critical influence on sensitivity shifts remains to be reliably demonstrated. In particular, there is no data pertaining to tasks in which confidence ratings have been used; this would provide badly needed information about the mechanisms underlying possible decrements in sensitivity over the work period. It is also clear that an organizational study of the tasks used in the literature, in the manner of the study reported by Levine, Romashko and Fleishman (1971), but from a decision theory point of view, would provide further clues as to the influence of event rate and other task variables on sensitivity decrements in monitoring tasks. We will postpone such a study until Chapter 12, where we will have an opportunity to evaluate some of the evidence bearing on this issue from some of the experiments reported in this thesis.

Before leaving this section on detectability shifts in monitoring tasks, we should briefly mention some other variables which may affect signal detectability in monitoring behaviour. Perhaps the most obvious of these is signal intensity or strength; empirical evidence for the effect of signal strength has been provided, incidentally, by Broadbent and Gregory (1965) and Williges (1973). Mackworth and Taylor (1963) have also shown that the sensitivity decrement in the Continuous Clock task is dependent on the initial detectability of the signal. This result appears to be related to Teichner's (1974) finding that the rate of decrement (in detections, not sensitivity) in simple visual monitoring tasks is related to the initial detectability of the signal, or the initial 'subjective signal strength'.

Signal detectability in a monitoring situation has been shown to be adversely affected if another task has to be performed at the same time (Taylor, 1965; Williges, 1969). Time sharing between different sources of the same task would also appear to lower detectability, but there is very little data on this point. Milosevic (1974) has, however, recently compared one-source monitoring with five-source monitoring for a visual task in which the occasional brighter flash in a series of flashes had to be detected. Sensitivity was significantly reduced in the multi-source task, but there was no change in the trend in d' over the run (d' was invariant over time in both conditions).

Small but reliable practice effects on d' have also been shown by Binford and Loeb (1966) and Colquhoun and Edwards (1970). This might be interpreted as an effect of learning, and such effects might be especially evident in more complex tasks. Learning effects are also important to evaluate from the point of view of tasks exhibiting sensitivity decrement; that is, if the sensitivity decrement is due to inadequate learning of the signal characteristics, the decrement might not appear in later sessions. In point of fact, the data of Binford and Loeb (1966) show just the opposite; nevertheless the effects of repeated sessions on monitoring behaviour are important to evaluate both from a theoretical and a practical viewpoint.

A few studies have indicated that monitoring efficiency is also reduced under stress or adverse environmental conditions. Reductions in d' with sleep deprivation have been reported by Wilkinson (1968) and Deaton et al. (1971). Finally, although they did not report data for d', Lewis, Baddeley, Bonham and Lovett (1970) reported that the percentage of correctly detected signals was significantly reduced while the percentage of false detections was roughly constant when subjects performed an auditory vigilance task (Wilkinson, 1968) in a traffic-polluted environment. This study implies that the effect of breathing polluting air was to reduce monitoring efficiency.

## 5.4 The Influence of Signal Probability and Event Rate on Decision Processes in Monitoring Behaviour

We have now considered the evidence relating to both criterion and sensitivity shifts in monitoring performance. The experiments of Baddeley and Colquhoun (1969) and Broadbent and Gregory (1965) have shown that signal probability is the major variable of interest with regard to criterion shifts; there is some less reliable evidence pointing to the importance of the event rate for determining sensitivity shifts. Thus far, we have only briefly considered the studies bearing on the effects of these two variables. Since the two variables have been used to support two opposing theoretical positions (that is, signal probability has been related to a decision theory model, while event rate forms a part of the 'observing response' model of monitoring behaviour), we shall consider them in more detail in this section.

### 5.4.1 The role of 'unwanted' signals

Monitoring tasks require the detection of signals which occur at quasi-random times over a prolonged period. The rate at which the signals occur influences the probability of their detection, higher signal rates generally resulting in improved detection performance (Davies and Tune, 1970). We have seen in 5.2 that signal rate influences performance primarily through a change in the response criterion; within-session changes in the criterion are also significantly affected by expectations, established in training and practice periods, of the relative frequency of signals, as shown by Colquhoun and Baddeley (1964, 1967). These authors also suggested that much of the early evidence on the effects of signal rate is unsatisfactory for two reasons (Baddeley and Colquhoun, 1969): firstly, that in many of the experiments, the rate of 'unwanted' non-signal events was often varied at the same time as the signal rate. Secondly, pre-test expectancy was not controlled, since, as was common in vigilance studies at the time, subjects were trained with an inappropriately high signal rate.

Colquhoun (1961) therefore proposed that it was important to distinguish between signals and 'unwanted' signals, and that the important factor in monitoring performance was not the absolute signal rate (or the absolute signal probability in time), but the conditional probability of a signal given an event (which is related to both the signal and 'unwanted' signal rates). Mackworth (1957) had also earlier pointed out that the effects of 'unwanted' signals warranted further investigation; he suggested that further work was required on three aspects of such events, their rate, regularity and similarity. Little or no work has been reported on the latter two aspects, but the effects of the rate of presentation of events have received increasing attention in recent years (Colquhoun, 1961, 1969; Jerison and Pickett, 1964; Mackworth, 1965a; Metzger, Warm and Senter, 1974; Moore and Gross, 1973; Taub and Osborne, 1968).

The effects of the rate of presentation of stimulus events on monitoring behaviour have been treated in two ways in the literature; either as a direct effect of the event rate itself (Jerison and Pickett, 1964; Loeb and Binford, 1968; Metzger et al., 1974, and others), or as an effect due to the resultant change in the conditional probability of a signal given an event[5] (Colquhoun, 1961, 1969). Since signal probability, signal rate and the event rate are related to each other in the ratio signal probability = signal rate/event rate, the three variables cannot be varied independently of each other; however two variables may be independently manipulated while the third is held constant, since the same change in signal probability can be effected by opposing changes in either signal rate alone or event rate alone. An interpretation of the effects of 'unwanted' non-signal events in terms of signal probability alone would thus, ideally, predict similar effects on performance due to a change in signal probability, whether the change is brought about by changing signal rate or event rate. On the other hand, an interpretation based on event rate alone, or one based on both signal and event rate, would, of course, consider changes in either of these independently.

The effects of signal probability and event rate on performance may be conveniently represented in graphical form by plotting a measure of performance against signal probability for different event rate levels. Such graphical plots can be used to assess the relative effects of signal probability and event rate on several measures of performance. Figure 5.4 shows effects on performance due to (a) an independent effect of signal probability, (b) an independent effect of event rate, and (c) both signal probability and event rate effects. It would be preferable to assess these effects in terms of a decision theory analysis of performance, but, as we have noted previously, very few of the relevant studies report false alarm or confidence rating data which would permit such an analysis.

### 5.4.2 Relative effects of signal probability and event rate

Colquhoun (1961) reported one of the first experiments in which, in accordance with Mackworth's (1957) suggestion, the effects of the 'unwanted' signal events were examined in a monitoring situation. Colquhoun varied the number of non-signal events in a visual inspection

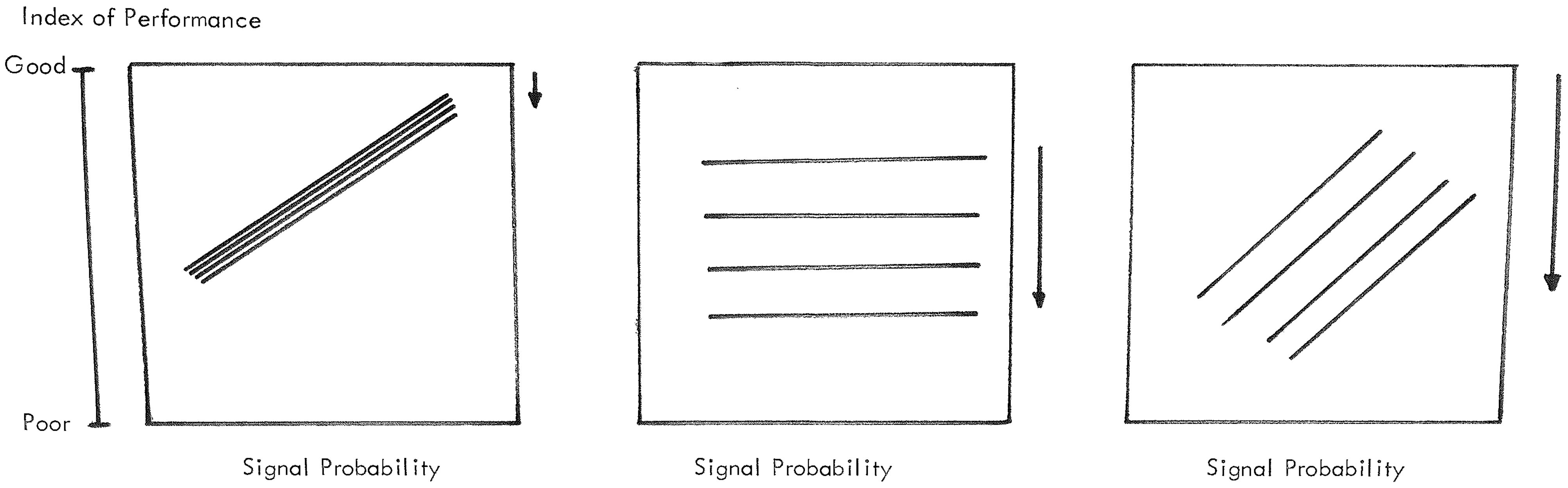


FIGURE 5.4 Representations of the effects of signal probability and event rate on performance. In each panel idealized functions of the effects of these two variables are displayed. In (a) an independent effect of signal probability is present, while in (b) an independent effect of event rate is present. Both variables are shown to influence performance in (c). The arrows indicate the direction of increasing event rate.

task in such a way that the effects of signal probability and event rate could be distinguished. He used a task in which the event was a strip displaying six green discs, the critical signal for detection being a display which contained a paler disc. Two event rates were combined with two signal rates in three conditions such that signal probability was either 0.5 or 0.08. The results indicated that the percentage of correctly detected signals was related to signal probability and unrelated to event rate; when event rate was held constant and signal probability decreased there was a significant drop in the percentage of signals detected, while when signal probability was held constant and event rate increased there was no change in performance. Colquhoun's results thus confirmed the importance of the effects on the non-signal events for monitoring performance, but suggested that the effects were seen through resulting changes in signal probability rather than event rate itself.

False alarm data were not reported in this experiment, but later experiments by Colquhoun using the same task not only confirmed these results, but also established their basis in response criterion changes (Baddeley and Colquhoun, 1969).

The effects of a change in the event rate on monitoring performance were first investigated by Jerison and Pickett (1964). They used a task in which the signal was an increased second jump of a bar of light on a CRT display. For a constant signal rate, probability of detection was reduced from 0.9 to 0.3 when the event rate was increased from 5/min. to 30/min., and there was also an increase in the decrement at the high event rate. However, since signal rate was held constant in this experiment, the results may be interpreted as an effect of signal probability, although the size of the effect is much larger than that usually reported for the same change in signal rate (see Baddeley and

Colquhoun, 1969). An interpretation in terms of signal probability, furthermore, is consistent with an analysis of performance with this task by TSD, which revealed that event rate served to increase B while leaving d' constant (Jerison, Pickett and Stenson, 1965).

Colquhoun (1969), using a task in which a brighter flash had to be detected in a series of intermittently flashing illuminations of a circular display, also found that there was a large drop in detections when the event rate was increased, for a constant signal probability, from 6 to 30/min. Like Jerison et al., Colquhoun (1969) found that event rate had no significant effect on d', but influenced B instead.

Mackworth (1965a) tested housewives on a visual monitoring task similar to that of Colquhoun (1969), except that there were two circular flashing sources, an event rate of 40/min. and an extremely high rate of 200/min. Unfortunately, Mackworth did not report a statistical comparison of performance levels between groups; an examination of the data shows that the major difference between groups was in the probability of detected signals, fewer signals being detected at the high rate. There appeared to be no significant difference in d' between low and high event rate groups. It was reported that a significant decrement in d' was obtained in the high event group while sensitivity was invariant over time in the low event rate group. Data using the B measure were not reported. However, since there was a reduction in false alarm rate, it appears that B would have increased, as reported in the experiments of Colquhoun (1969) and Jerison et al. (1965).

Similar experiments in which the event rate is varied for a constant signal rate have been reported by Guralnick (1973), Metzger et al. (1974) and Moore and Gross (1973). Probability of detection was found to be inversely related to the event rate in all three studies, although the effect was not significant in the Moore and Gross (1973) experiment. Sensitivity and criterion measures were not utilized in these studies.

The results of this group of experiments, in which the effects of event rate for a constant signal rate were examined, are therefore completely consistent with Colquhoun's (1961) original and novel view that it is the conditional signal probability rather than the signal or event rates that is the important factor in monitoring performance. It is unfortunate that the investigators in these studies, especially the more recent ones, did not attempt to assess the effects of event rate independently of signal probability, in view of the importance of the findings of Colquhoun (1961) and Baddeley and Colquhoun (1969). The results of these studies do not therefore, establish the critical, independent influence of the event rate, and do not clarify the relative effects of event rate and signal probability on decision processes in monitoring.

However, three further studies have, in fact, examined the effects of signal probability and event rate in an independent manner, as was first done by Colquhoun (1961): these are the studies of Jerison (1965; also summarized in Jerison, 1967b), Loeb and Binford (1968) and Taub and Osborne (1968).

Jerison (1967) repeated his original experiment (Jerison and Pickett, 1964) using conditions where event rate was varied for a fixed value of signal probability. His results are graphically indicated in Figure 5.5 (b). This shows that the original effect of event rate was again observed, independent of signal probability; an effect of signal probability alone only appears at high event rates, as indicated by the relative slopes of the lines FG and AH. Colquhoun's (1961) data have

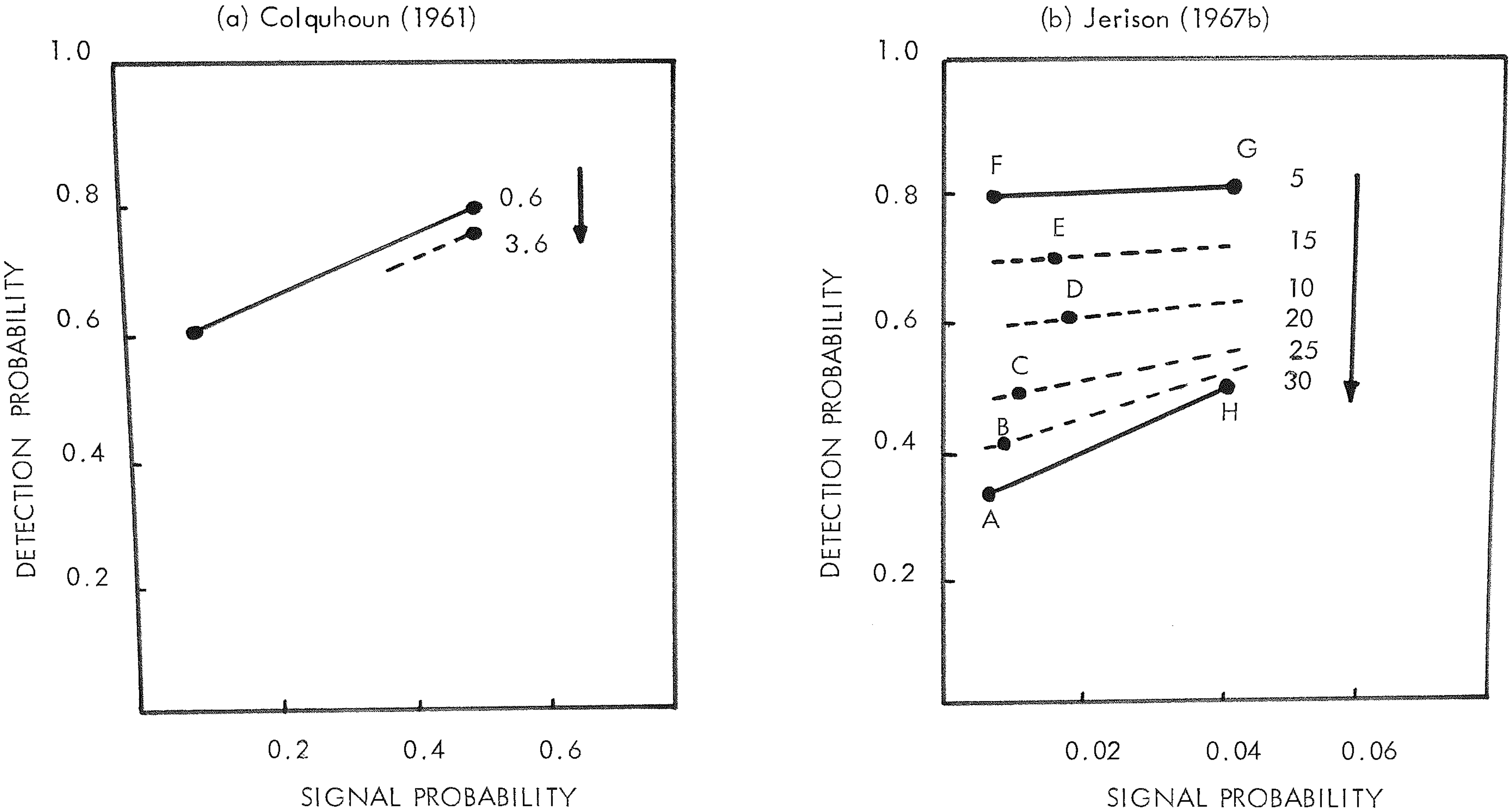


FIGURE 5.5 The results of Colquhoun (1961) and Jerison (1967b) represented in the manner of Figure 5.4 to indicate the independent effects of signal probability and event rate. The parameter numbers indicate values of event rate in events per minute.

also been plotted in Figure 5.5 (a) for comparison purposes. In making such a comparison, three possibilities emerge: 1) that Jerison's findings are completely inconsistent with Colquhoun's, 2) that the respective results are not so much inconsistent as indicative of differences in certain task factors, and 3) different mechanisms are operative in the two studies, and these are obscured by the examination of detection scores only. We shall consider all these possibilities in the following sub-section.

Let us return however, to the other two studies which have independently manipulated signal probability and event rate. In the first of these (Taub and Osborne, 1968), two experiments using a modified version of the Mackworth Clock were reported. In both experiments, it was observed that for a fixed signal probability, there was a marked decrease in detection probability with an increase in signal probability. Their results have been represented in Figure 5.6 (a); this shows virtually no effect of signal probability as such.

Unfortunately, neither Jerison nor Taub and Osborne reported data using sensitivity and bias measures, but this was done by Loeb and Binford (1968). Two tasks were used in this study; in the visual task an analogue of the Mackworth Clock (using lamps arranged in a circle) was used, while in the auditory task detection of a transient increase in intensity of an intermittent white noise pulse was required. The results of this important study are illustrated in Figures 5.6 (b, c) and 5.7 (a, b), and can be summarized as follows: both signal probability and event rate independently affected detection probability, although the effects for event rate were larger; however, while there was only a change in B with a change in signal probability, there was a reduction in d' with an increase in event rate, especially for the visual task.

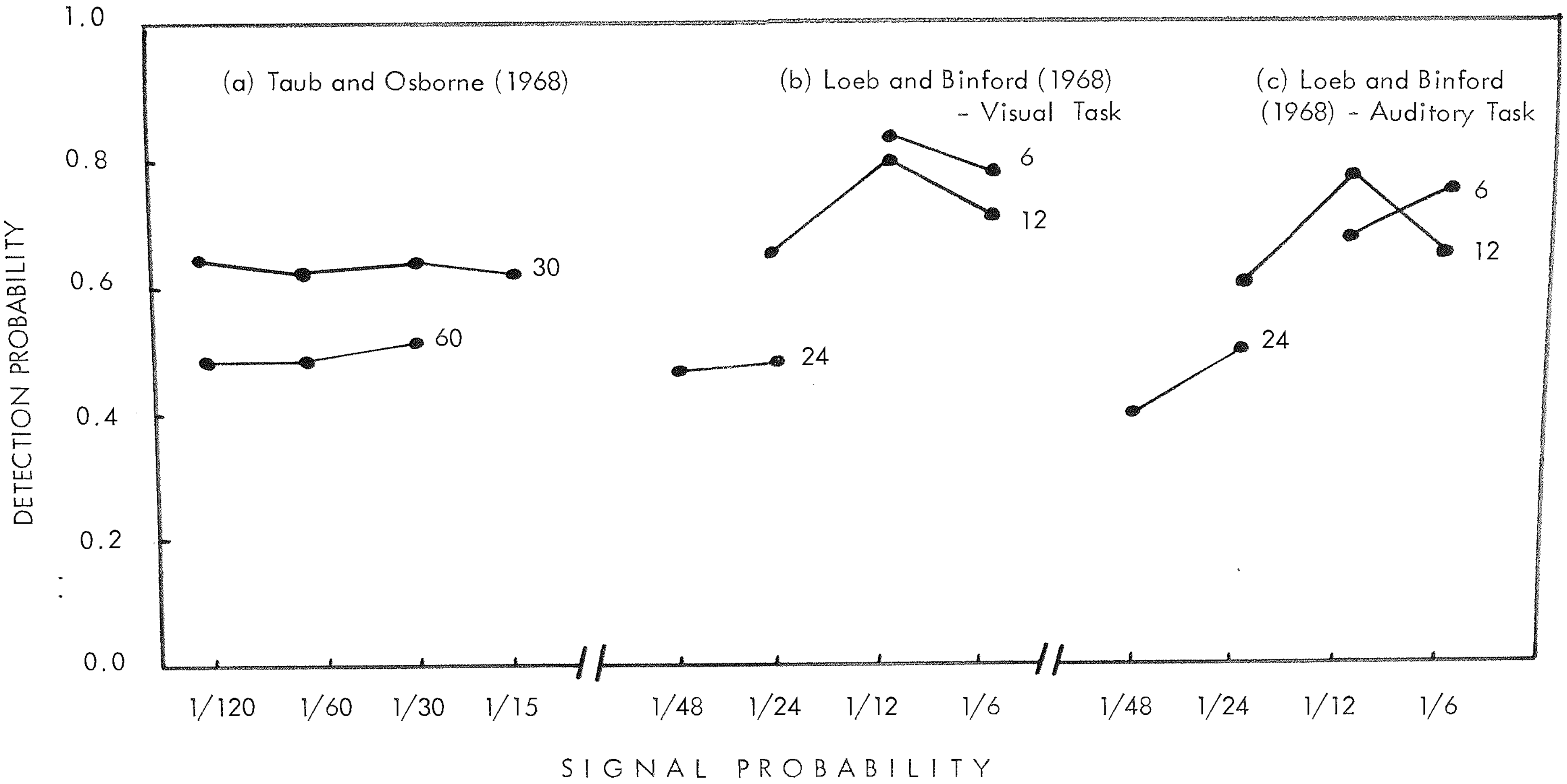


FIGURE 5.6 The results of Taub and Osborne (1968) and Loeb and Binford (1968) represented in the manner of Figure 5 4 to illustrate the independent effects of signal probability and event rate . The parameter numbers indicate event rates in events per minute.

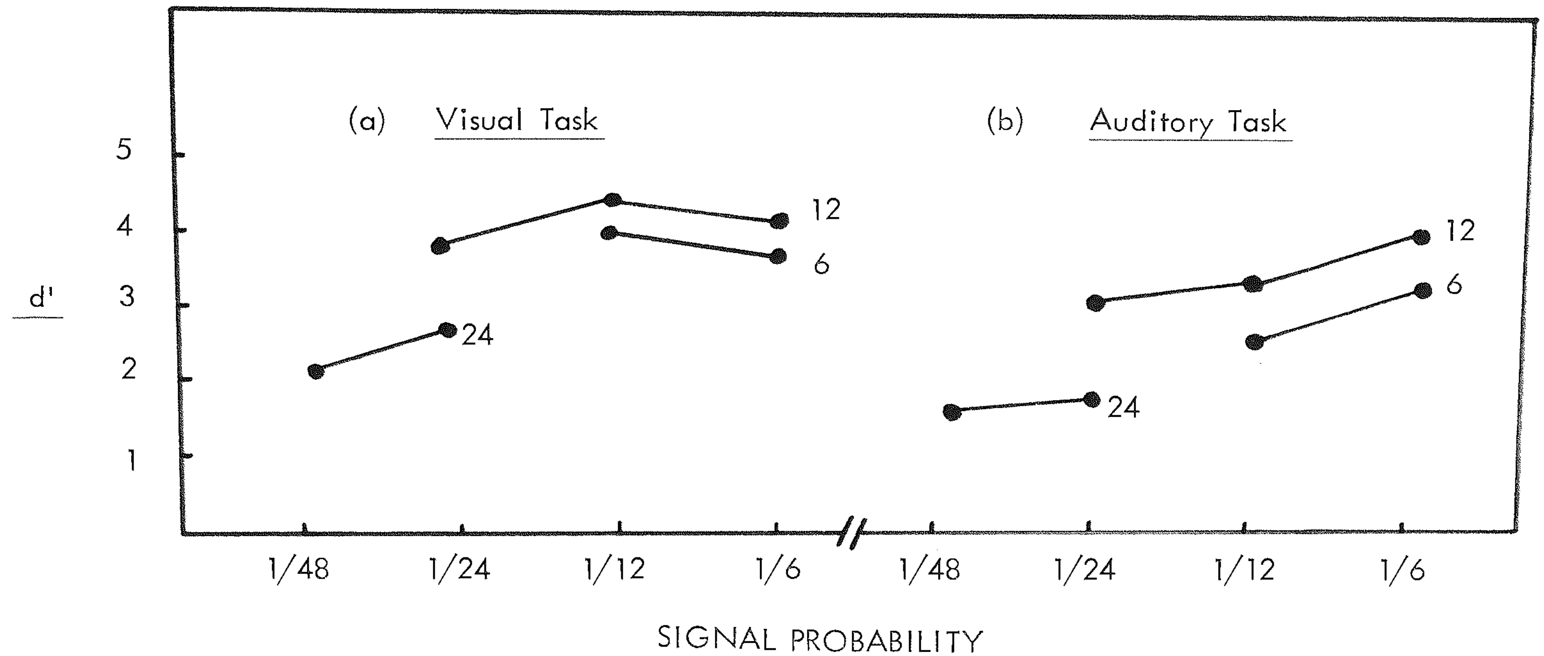


FIGURE 5.7 The results of Loeb and Binford (1968) represented to indicate the independent effects of signal probability and event rate. The parameter numbers indicate values of event rate in events per minute.

### 5.4.3 Interpretation of the results: an approach with task classification and decision theory

We have now reviewed a number of studies concerned with the effects of signal probability and event rate on monitoring behaviour. These studies fall into two general classes: in the first, event rate is varied for a fixed signal rate, so that there is a simultaneous change in signal probability. These studies reported an inverse relationship between event rate and probability of detection; however this result cannot be claimed to support theoretical positions based on event rate, such as the 'observing response' model of Jerison (1967b), since, in view of the change in signal probability, the results may be explained in terms of the expectancy hypothesis of Colquhoun (1961) and Baddeley and Colquhoun (1969). In the second class of study, signal probability and event rate were independently varied; and in these experiments more reliable evidence for the effects of event rate emerged.

However, there are a number of conflicting results, some of which may only be resolved once experiments using decision theory measures are carried out. In particular, the discrepancy between the results of Colquhoun (1961) and Jerison (1967b) needs to be explained. Jerison (1966) suggested that the visual search requirement of Colquhoun's task might be a possible intervening factor, but Colquhoun (1966) repeated his 1961 experiment using the same task with only two sources instead of six and found essentially the same result as he had before. We may however, examine two other aspects of Colquhoun's (1961) study. Firstly, very low event rates were used, in the range 0.6 - 3.6/min., while Jerison (1967) and Loeb and Binford (1968) only observed an effect of event rate for rates greater than about 20/min. Secondly Colquhoun's task (both in 1961 and the 1966 versions) involved the detection of a stimulus quality (hue or increased size) in two or more stimulus sources presented simultaneously, that is, a detection within the same event. As we have seen in

Chapter 2, this type of task can be thought of as a task requiring 'flexibility of closure', rather than 'perceptual speed'; this latter task dimension appears to be a component of the other studies reporting event rate effects in the literature.

Hence it is apparent that while the effects of signal probability are fairly clear and straightforward, the nature of the effects of event rate and its independence of signal probability have not been reliably demonstrated. Since these two variables appear to influence the two major and opposing mechanisms in the decision theory model of monitoring, (that is, criterion shifts and sensitivity shifts, it is particularly important to establish the mechanisms underlying their effects. It is clear that data from confidence ratings will have to be provided to clarify further the nature of the decision mechanisms affected; none of the studies we have reviewed have done so.

There also appear to be certain differences attributable to differences in the type of monitoring task used, and a study of the effects of signal probability and event rate within a task classification approach would therefore appear essential. There is also the problem of sensitivity decrements over the monitoring period, and we have noted that event rate may be implicated here too. We have emphasized the need for a task classification approach to the specification of tasks yielding a sensitivity decrement.

The study of the effects of signal probability and event rate thus raises a number of substantive questions about decision processes in monitoring behaviour, to which we must seek more reliable and comprehensive answers. It is indeed unfortunate that theory appears to have run way ahead of the data in this area; there is a very real need for the provision of more 'basic' data, that is, data relating to the influence of these variables on sensitivity and bias trends over the monitoring period, and how these are affected by factors such as sense modality, source complexity, type of signal discrimination and so on. The report by Loeb and Binford (1968) goes some way towards this, but did not attempt to check the validity of the equal-variance TSD Gaussian model on which the reported results depended.

This review has therefore thrown up several lines for further experimentation which we shall follow up in the work to be reported in the following chapters. In particular, the review has indicated that further work is required within the framework of decision theory and task classification, and this is one reason why we introduced these topics in the early chapters (the other reasons being related to other aspects of monitoring behaviour, such as the consistency of individual differences in monitoring performance).

This section concludes the introductory part of this thesis. In the next Chapter we shall outline some of the research techniques which we shall use in the experimental part of this thesis, which spans Chapters 7 to 12. In Chapter 7 we shall examine some of the questions which we have raised in this and preceding sections, in particular the issues raised in the review of the effects of signal probability and event rate on decision processes in monitoring behaviour.

## 5.5 Notes to Chapter 5

1 Baker and Schuck (1975) have examined some other applications of TSD to performance in inspection and other industrial tasks.

2 Smith (1968) has proposed an alternative interpretation of criterion

shifts in vigilance based on the concept of 'cost' (see 4.5.3).

3 Mackworth (1965b) has shown that the choice of a particular value for the decision or observation interval does not significantly alter the trend in d' over time, although the absolute values of d' may change.

4 The sequential sampling model of Luce and Green (1972), which has some points of similarity to Laming's (1973) model, has had only limited success in interpreting latency data in vigilance tasks (see 4.5.4).

5 We will refer to the conditional probability of a signal given an event as simply 'signal probability'; but this should not be confused with 'signal probability' referring to the absolute probability, in time, of the occurrence of signals.

# C H A P T E R  S I X

## RESEARCH METHODS

## 6.1 Research Outline

In the preceding chapters, various aspects of monitoring behaviour have been considered from two main standpoints, task classification and decision theory. Some of the theoretical and practical consequences of these two approaches to the assessment of performance were outlined in Chapters 1 to 5, and several possible areas for further research emerged in these considerations. In this section, the investigations designed to follow up these lines of enquiry are very briefly described.

The main interest in this thesis is the influence of and interaction between task categorization and decision processes in monitoring performance. It has been argued that such a combined approach may enable more reliable evaluations to be made of different aspects of performance in a wide range of monitoring tasks. The experiments reported in this thesis were accordingly concerned, firstly, with the analysis of decision processes in individual monitoring situations, or more specifically, with the changes in monitoring efficiency and decision criteria due to prolonged work and other critical variables, and the interpretation of these changes; and secondly, with the specification of the various types of task in which these changes in decision-oriented activity occur. Another major factor behind the studies was the requirement, expressed in Chapter 2, for an empirical validation of taxonomic systems for monitoring tasks. Thus the experiments were also carried out in order to validate the task categories comprising the task classification system described in 2.3.

The present research effort commenced in October 1973 with an investigation of performance in visual 'speed' and 'closure' tasks. The research strategy at the time was to follow the lines of experimentation and theoretical implications suggested in the results of this preliminary study. A series of interlocking studies were conceived, each relating to and, hopefully, extending the two central standpoints of the thesis.

Eight experiments and an organizational study were carried out. Experiments 1, 2, 4 and 7, were concerned, in broad terms, with sensitivity and criterion trends in different monitoring tasks, and the influence of task categorization. In these experiments different aspects of the decision theory approach outlined in Chapters 4 and 5 were also examined. In particular, in Experiments 1, 4 and 6, the latencies of different response categories were examined to investigate the reliability and generality of the decision theory latency model described in Chapter 4. Furthermore, in Experiment 6 the relation between temporal measures of decision making activity and latency measures derived from evoked potentials averaged to different behavioural categories was also examined.

In Experiments 3 and 5, the influence of task classification on the consistency of individual differences in performance in different monitoring tasks was examined. Experiment 8, which is reported in Appendix B, was concerned with the consistency in performance within tasks across repeated sessions.

Finally, a task classification analysis of the studies reporting sensitivity and criterion data is reported in Chapter 12. This study was primarily carried out to examine whether the generalizations derived from the previous experimental studies were consistent with data reported in the monitoring literature.

A total of 9 monitoring tasks combining various categories in the task classification system were investigated. Task specifications and other methodological details are given in the remaining sections of this chapter.

## 6.2 Instrumentation and Programming

The experiments reported in this thesis were carried out 'on-line' using a Digital Equipment Corporation (DEC) PDP-8/E minicomputer. A computer was used mainly because of the resultant ease in data handling and control of monitoring displays. Although a number of different displays using both visual and auditory stimuli were required, each display was controlled in the same general way in a 'closed-loop' system incorporating the computer and peripherals, the display, and the response unit (subject). The functioning of this system is described prior to a description of the more specific types of apparatus used.

### 6.2.1 The general experimental system

Figure 6.1 displays the elements and illustrates the functioning of the general experimental system. The PDP-8/E computer was used to perform three main functions: 1) control of all stimulus events (time sequencing), 2) acquisition of response data on each trial and over a prolonged monitoring period, and 3) initial analysis of data, generally restricted to the computation of summary statistics and individual operating characteristics. Functions 1 and 2 are on-line activities, while function 3 was carried out 'off-line' at the end of each monitoring session.

These functions were effected with the help of a Control Unit interfacing the subject to the computer, as shown in Figure 6.1. The control of the monitoring displays was restricted to the time sequencing of stimulus events. The Control Unit also received response signals from the subject, and these were relayed to the computer, where they were coded and stored. A single master program, CONTROL, controlled the execution of the experiment and the acquisition of data. A program SCORE carried out function 3. Typically, about 100,000 units of raw data output were obtained in each experiment.

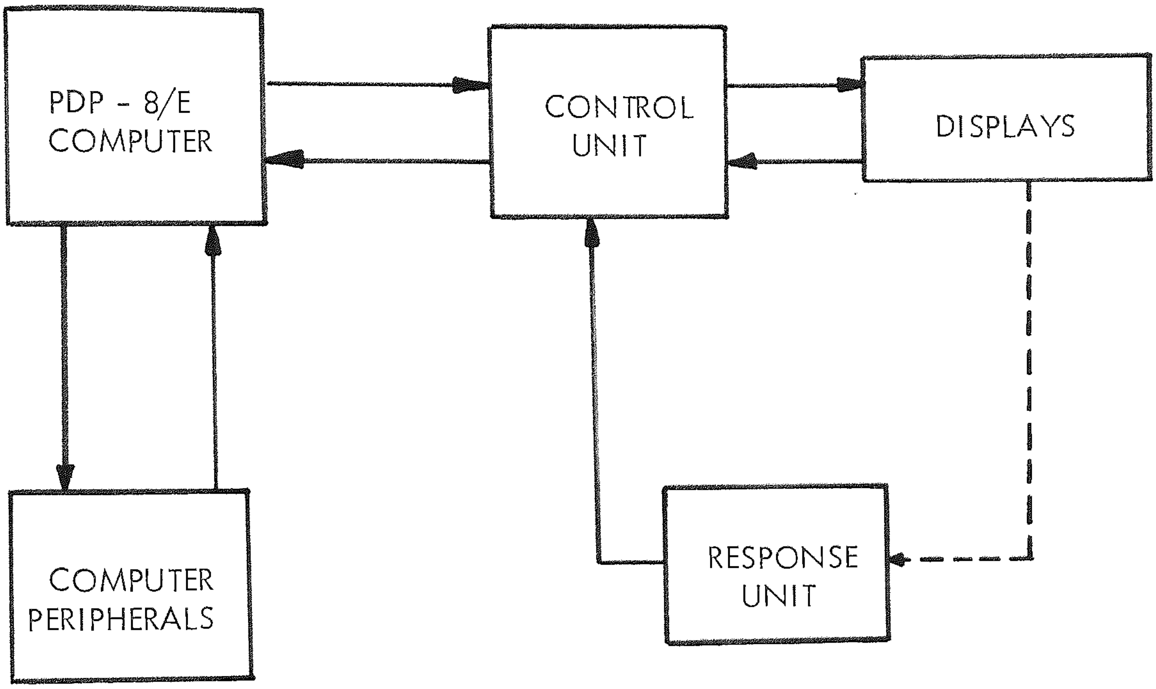


FIGURE 6.1 Block diagram of the major elements of the experimental system.

### 6.2.2 Apparatus

A more detailed diagram of the experimental system is shown in Figure 6.2. The PDP-8/E and Control Unit directed all timing pulse sequences to and from the display apparatus; these pulses were the same for all displays. The Control Unit incorporated the following devices; 1) a self-starting or triggerable astable multibrator, 2) frequency selector (0.01 to 1.0Hz), 3) pulse counter, 4) 100 msec. and 500 msec monostable multivibrators, 5) closing contact reed-relays, and other supporting logic circuitry. The timing pulse sequences could be initiated either manually or under program control. The timing diagram of the major events, from the starting pulse to the response, is given in Figure 6.3.

The Response Unit consisted of the response buttons, which closed micro-switches when depressed. These activated relays transmitting the response information back to the PDP-8/E via the Control Unit and a relay buffer. The relays enabled the 12-bit interface interrupt facility of the PDP-8/E, which was programmed to record the type and latency of each response (the latter being measured correct to 1 millisecond by a DK8-EP programmable real-time clock). Communicative control with the PDP-8/E was maintained with an ASR-33 Telytype. Programs were read in using a PC8-EB High Speed Paper Tape Reader and Punch.

The visual displays were derived from a Kodak Carousel projector fitted with a Lafayette variable duration shutter. Both the projector and the shutter could be activated either electronically or manually. The shutter incorporated a pair of contacts which closed when the shutter was exposed (the total delay time between trigger pulse onset and the contacts closing was estimated not to exceed 15 μsec). This was used to signal to the computer the time of stimulus onset, and to start the real-time clock. Timing pulses recorded on one channel of 4-channel tape achieved

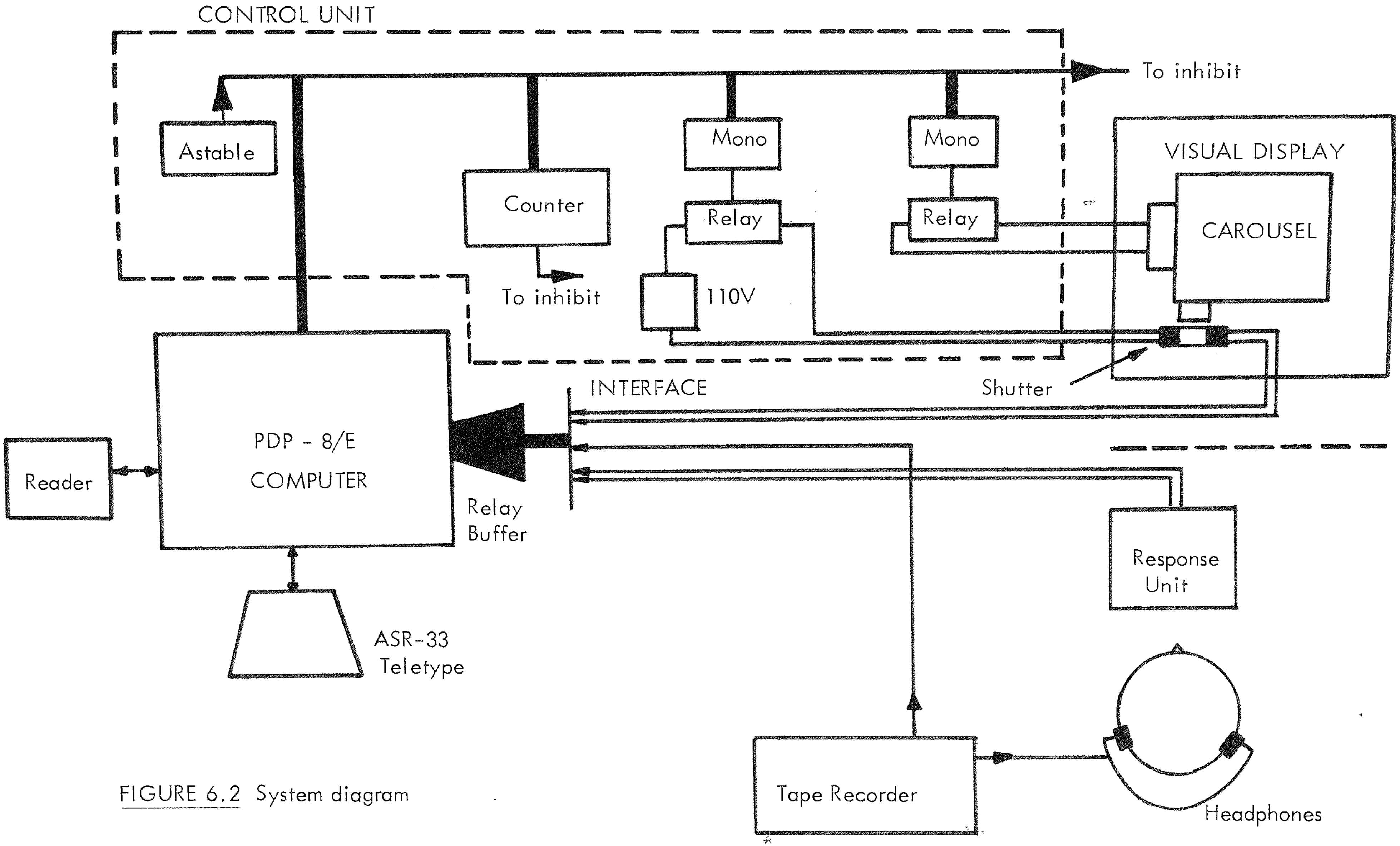


FIGURE 6.2 System diagram

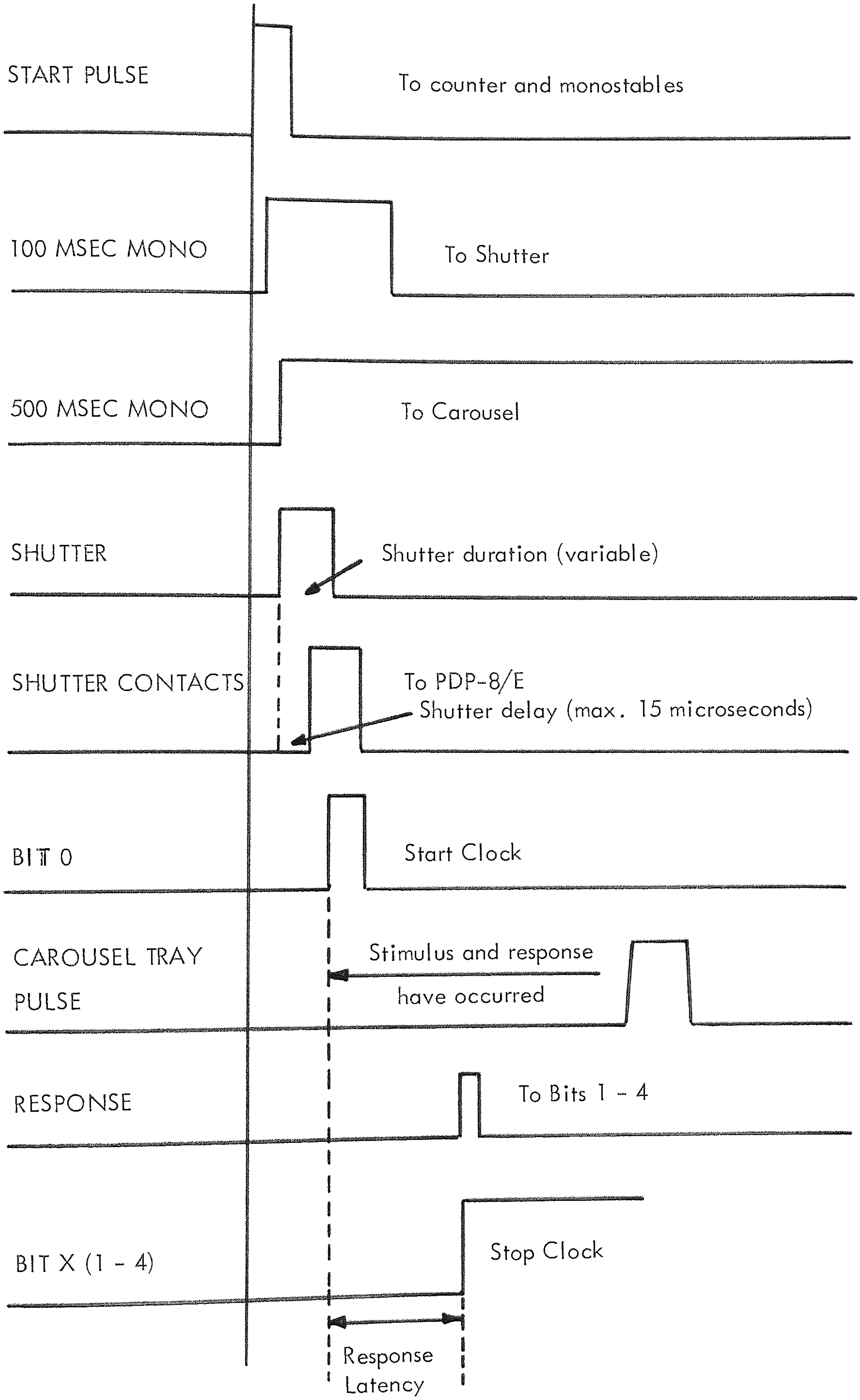


FIGURE 6.3 Timing diagram (not to scale).

the same function for the auditory displays. Auditory stimuli were presented binaurally over Sharp HA10 headphones. The stimuli were pre-recorded onto tape on a 4-channel tape recorder and played back at each monitoring session. For both auditory and visual displays, the exact stimulus schedule was punched on to 8-channel paper tape and read into the PDP-8/E prior to any training or monitoring session.

The microswitches in the Response Unit shorted a response line associated with each response to a common line, with all lines being led to the computer. The number of lines differed according to the response mode used. As has been noted previously, three types of response mode, single, binary and rating, may be used. For binary response tasks, two response buttons labelled 'YES' and 'NO' were used. A more 'traditional' single response button, labelled 'DETECT', was also employed in the single response mode. For the rating mode, four categories of response were used; the four response buttons provided were labelled, from left to right, 'CERTAIN YES', 'DOUBTFUL YES', 'DOUBTFUL NO' and 'CERTAIN NO' (see Figure 6.4).

The START button on each response box initiated the monitoring period. Subjects were instructed not to rest their fingers on the response buttons in the inter-stimulus intervals, but at a point below the buttons. For the rating mode, subjects were instructed of the meaning attaching to the labels above each response button, and they were encouraged to use each response category, although they were not provided with any information regarding the expected frequency of different response categories, apart from the information, common to other response modes, that signals would be less frequent that non-signals.

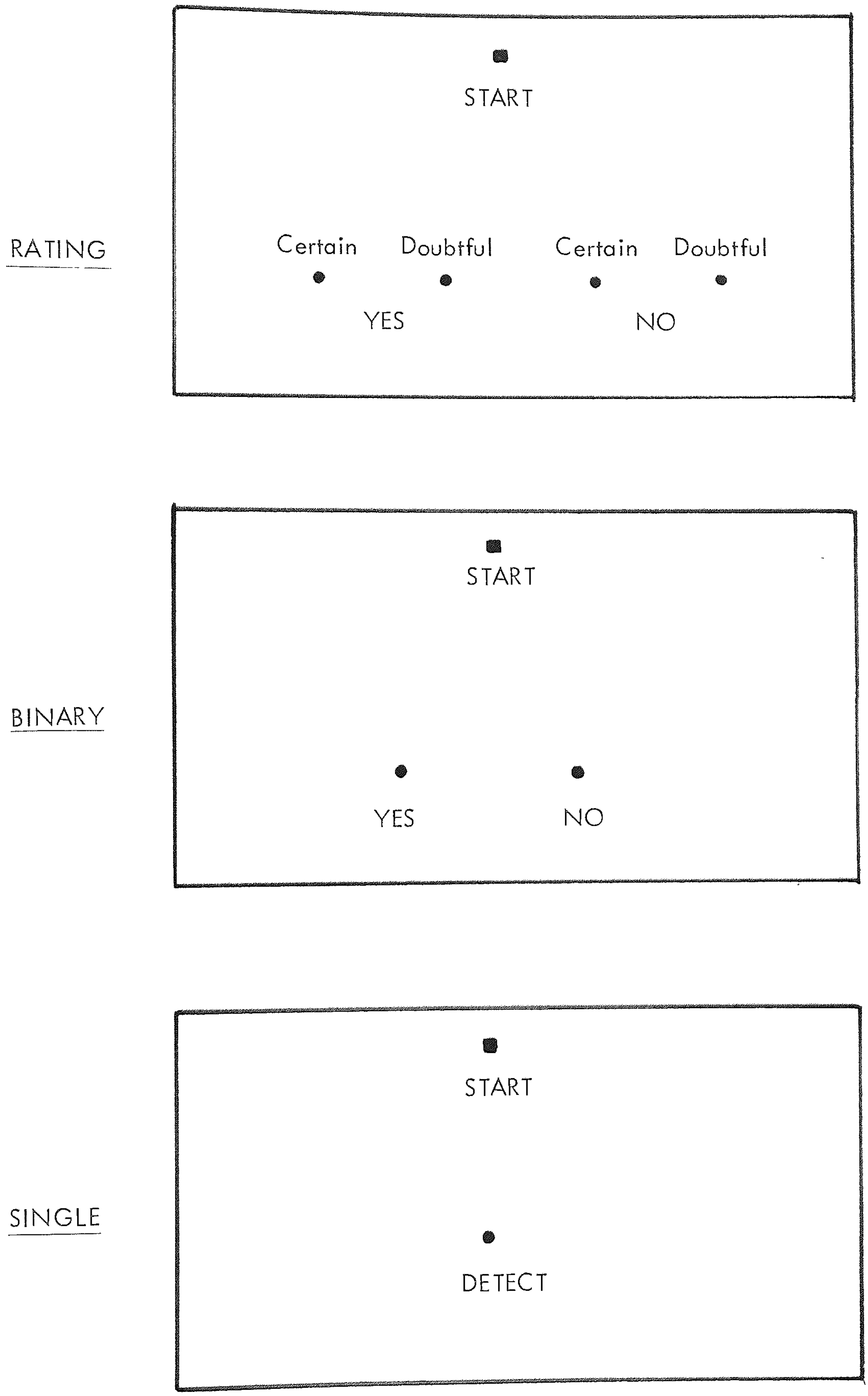


FIGURE 6.4 Response panels for Single, Binary and Rating response modes

| Program | Type | Functions | Language |
|---|---|---|---|
| CONTROL | On-line | Control of timing and scheduling; acquisition of responses; printing of raw data (choice data and associated latencies) | Assembly Language |
| SCORE | Off-line | Computation of data averages in successive time blocks and (optional) of operating characteristics | FOCAL-8 |
| REACT 1* | On-line | Acquisition and sampling of EEG | Assembly Language |
| SUMAVE | Off-line | Selective averaging of EEG | Assembly Language |
| PLOTER | Off-line | Outputting of evoked potentials | FORTRAN IV |

TABLE 6.1 Computer programs used in the experimental studies (*this program was written by Ian Webb, formerly of the Neuropsychology Unit, Department of Applied Psychology, University of Aston).

### 6.2.3 Programs and output format

Table 6.1 lists the programs used in the different experiments. As mentioned previously, CONTROL and SCORE were used for on-line control and off-line, initial data analysis, respectively. These programs were used in every study. In addition, the programs REACT 1, SUMAVE and PLOTER were used in the recording and analysis of evoked potentials in Experiment 6. At the end of a monitoring session, the PDP-8/E outputted the response data accumulated during the monitoring period. CONTROL coded each response, and stored it along with the associated latency in addressable locations. This data was output in pairs, as shown in Figure 6.5. In this figure, the asterisk refers to a 'null response'. It was arranged that on those occasions when the subject failed to respond to a stimulus within 1.95 seconds a 'null response' was recorded (this does not apply to the single response case). In practice, such occasions

Single

2 0564
2 0548
3 0669
2 0554
2 0601
. . . . . . .

Code

2 Hit
3 False alarm

Binary (Yes/No)

5 0596
5 0613
5 0587
5 0592
2 0600
5 0634
5 0619
5 0698
5 0677
2 0610
5 0613
5 0651
4 0716
5 0612
5 0508
5 0564
3 0726
5 0504
* 1950
5 0702
5 0675
. . . . . . .

Code

2 Hit
3 False alarm
4 Omission
5 Correct rejection

Rating

9 0575
9 0619
9 0724
8 0756
2 0699
9 0753
9 0724
8 0796
7 0911
8 0753
3 0712
9 0652
9 0721
9 0652
9 0612
8 0632
4 0734
9 0673

Code

| | Rating 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Signal | 2 | 3 | 4 | 5 |
| Nonsignal | 6 | 7 | 8 | 9 |

FIGURE 6.5 Sample output formats for data using the Single, Binary and Rating response modes (the first digit in each pair codes the response, while the second number gives the associated response latency in milliseconds).

were rare; when they occurred, this response was either reassigned as an omission or a correct rejection, according to the type of stimulus, on the grounds that a failure to respond represents a negative 'response'.

### 6.2.4 Task layout and environment

The physical arrangement of the apparatus is displayed in Figure 6.6. Each subject sat in a comfortable chair with arm rests in a quiet, draught-free room measuring about 3m by 5m. A large screen separated the subject from the computer and other apparatus. The subject was seated about 2m from this screen and about 1m from the visual display. The only apparatus in the immediate vicinity of the subject and which he could see was the display, the Response Unit, and the headphones through to which broad-band (50 - 5000 Hz) noise was relayed. The spectrum level of the noise was approximately 60 dB (re 0.0002 ubar). This sound pressure level of noise is within the range of acceptable ambient levels in which there is no appreciable effect on monitoring performance, 50-80 dB (Blackwell and Belt, 1971). Effects of noise on monitoring performance are generally only observed for noise intensities greater than about 70-80 dB (see Broadbent, 1971; Davies, 1968). As far as possible, other ambient conditions were kept stable during a monitoring session and between sessions, although a strict record of the relevant environmental variables was not kept. Room temperature was, however, measured on each occasion. All experiments were carried out at temperatures in the range 64 to 71$^{\circ}$F.

### 6.2.5 Monitoring tasks

A total of nine monitoring displays were used in the experiments reported in this thesis. The tasks were designed to minimize differences along task dimensions not considered in the task classification system of 2.3. Thus, none of the tasks had a visual search requirement, and short-duration ('transient') stimuli were employed. The displays incorporated

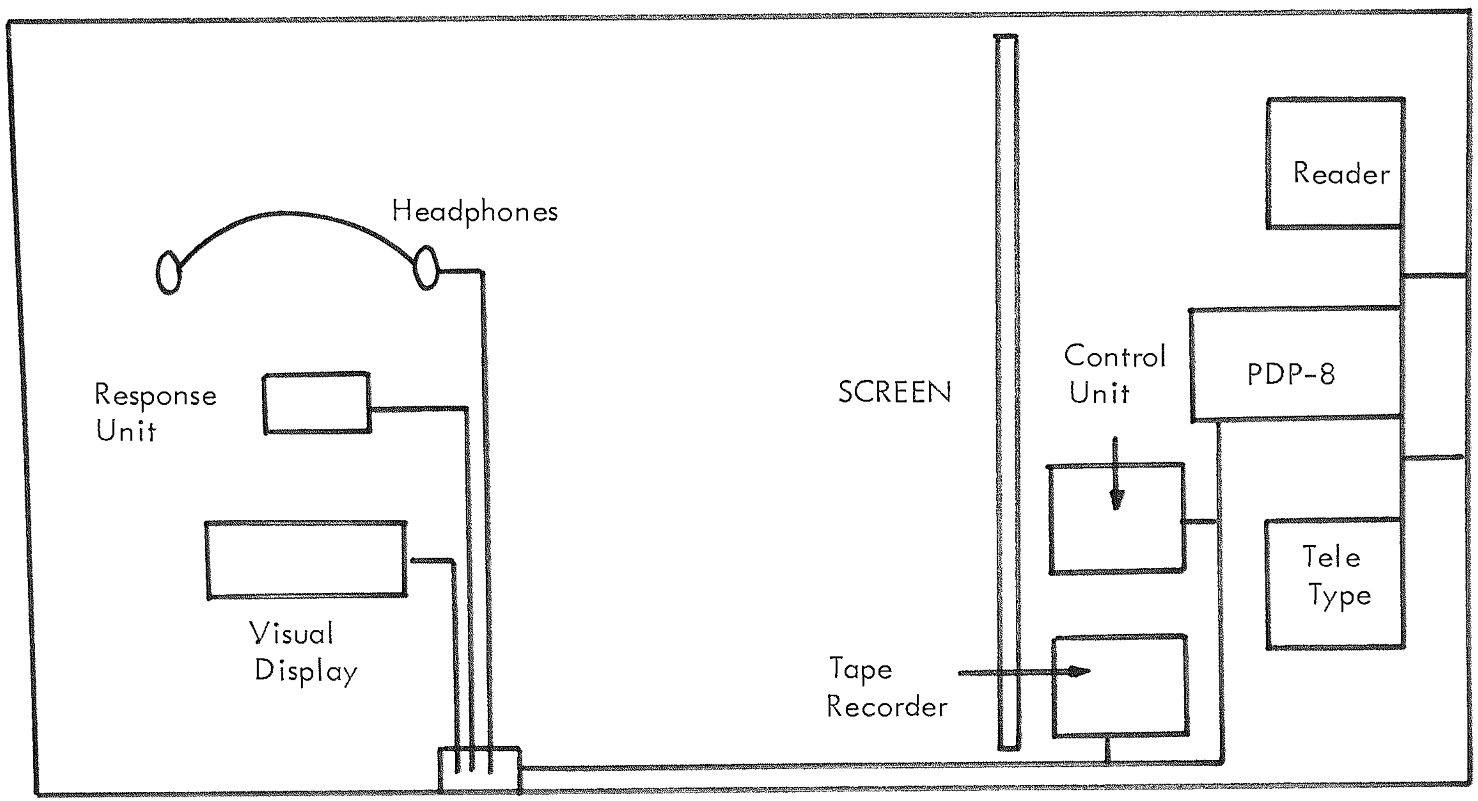


FIGURE 6.6 Layout of apparatus (not to scale).

different combinations of the four task dimensions comprising the task classification system, modality (visual or auditory), type of signal discrimination (speed or closure), source complexity (single or multi-source) and the time course of events (low or high rate). Task descriptions are given below.

1. Visual 'speed' task 1 (VS1)

In this task, subjects were required to detect a periodic decrease in the intensity of an intermittently flashing light source. The source was derived from the Carousel by interposing a 3cm circular perspex disc between the Carousel light source and the lens. The luminance of this source was attenuated by using Kodak-Wratten neutral density filters. These were fixed in slides and placed in the Carousel tray. The decrease in intensity that defined a signal was .05 log ft.L (luminance values were measured using an empirically determined equation, see Appendix G). The light flashes were presented against a matt black background. This, and the low illumination level ensured that contrast effects were minimized. The light flashes (100 msec. duration), were presented to the right eye in approximately Maxwellian view, at $2^{o}$ of visual angle.

2. Visual 'speed' task 2 (VS2)

In this task subjects monitored a screen intermittently displaying two vertical lines each 10 cm long. The lines were originally drawn in black ink on cards and then photographed to yield several slides. The signal was a decrease in the horizontal separation of the two lines by approximately 1 cm. Stimulus duration was 0.2 sec. Subjects sat approximately 1m. from the display, the line of sight being symmetrical and perpendicular to the displayed lines. Signal and non-signal slides were photographed from only two original ink drawings, but some inherent 'noisiness' in the display was nevertheless present due to small perturbations in the mechanical system of the Carousel projector.

3. Visual 'closure' task 1 (VC1)

This task used the same general display as in VS1, but the signal for detection was the occasional appearance of a 1cm pink circle at the centre of the light flashes. Hence, a discrimination process spanning successive events was not required, rather a pure detection process within each event (trial). In other respects, the task was similar to VS1.

4. Visual 'closure' task 2 (VC2)

The slide and screen display of VS2 was used in this task. However, the critical signal for detection was a central 0.5 cm gap in both vertical lines. Again, this task required the pure 'closure' detection of signals presented in a single trial, rather than of a change in a stimulus relative to preceding stimuli. In other respects, this task was similar to VS2.

5. Auditory 'speed' task (AS): 6. Auditory 'closure' task (AC)

The stimuli used in these tasks were 'pure' sinusoidal tones of 500 milliseconds duration. The tones were gated with electronic switches and recorded on 4 channel tape. The rise and decay times of the tones were not accurately measured, but were estimated not to exceed 15 msec. Tone frequency was 1000 Hz. The tones were recorded in background white noise. The signal to noise ratio was taken as $E/N_o$, where E is the signal energy, and $N_o$ is the noise power density (per cycle per second). $E/N_o$ was measured with a full-wave rectifier registering 'true rms' values. An $E/N_o$ value of approximately 30, or 14.8 dB was used for the standard tones. In the AS task, the subjects were required to detect a 1 dB increase in the intensity of the 1000 Hz tones, which were presented intermittently over headphones at a rate determined by the experimental condition (usually, every 2 or 4 seconds). In the AC task, the 1000 Hz pure tone had to be detected within 60 dB noise bursts of 0.5 second duration which were pulsed every 2 or 4 seconds. Signal strength was associated with an $E/N_o$ value of 19 (12.8 dB). The listening conditions were the same as in AS. The difference between the tasks is that in AS the 'signal' was specified as a change in the standard tone, whereas a pure detection of the tone within the noise burst was required in AC.

7. Multi-source 'speed' task (MS)

This task was a modification of the single source visual speed task VS1. It was designed so as to retain the distinguishing feature of speed tasks (comparison of temporally successive events) for a multi-source task by introducing source redundancy. The task involved the monitoring of two circular sources separated by about 35 cm in the horizontal plane, for occasional decreases in intensity relative to the preceding non-signal event. The sources were lit intermittently, and signals appeared on both sources. This was a high-rate task, with subjects having to monitor 30 events/min. The signal schedule, the light intensity levels, the stimulus sizes and other features of the source and background were identical to that in VS1. The rationale behind the use of this task is described in Chapter 12.

8. Multi-source 'speed-closure' task (MSC)

This task was identical to MSC except that signals appeared on only one source at a time, the discrimination process thus presumably being restricted to a single event, since signal and non-signal information was presented on the two sources. Signals appeared at either source in a pseudo-random fashion, but an equal number of signals were presented at each source.

9. Multi-source 'closure' task (MC)

This task was the dual-source version of the single source closure task VC1. As in MS and MSC, two circular light sources were lit simultaneously, and intermittently, but the critical signal for detection, as in VC1, was a circular pink spot appearing at the centre of one of the sources. As in MSC, an equal number of signals appeared at either source in a pseudo-random fashion. All other features of the task were identical to that of VC1.

## 6.3 Experimental Procedures

A standard testing procedure was used for all the subjects who were tested. The standard procedure included instructions and training sessions, as well as 'expectancy matching' sessions, prior to work on the main monitoring session. The block diagram in Figure 6.7 shows the sequence of activities during the typical experimental session, which lasted, on average, about $1\frac{1}{4}$ hours. Monitoring sessions were usually 45 min. long.

Two experimental sessions were held each day, in the morning and in the afternoon. Morning sessions were held between 10.00 a.m. and 11.30 a.m. and afternoon sessions between 3.00 p.m. and 4.00 p.m. In any particular experimental group, subjects were divided equally between morning and afternoon sessions. Although time of day effects have been reported to affect several aspects of human performance (Baddeley, Hatter, Scott and Snashall, 1970; Blake, 1967; Hockey and Colquhoun, 1972), no systematic effects were observed for the monitoring tasks used in this thesis. Due to the relatively low sample sizes and the presence of other variables which were controlled for (for example, sex, and testing sequence), time of day was not included as a variable in the statistical analysis of the data, but there was rarely more than about 7% difference in mean performance levels between morning and afternoon sessions.

Experimental sessions were held during the period October 1973 to December 1975 in two laboratories. Experiments 1, 2, and 6 were carried out in the main Departmental building (College House), while Experiments 3, 4, 5, 7 and 8 were carried out in a laboratory in Kyrle Hall. The room referred to in 6.2.4 describes the second laboratory, but the layout and general environment in the first laboratory were very similar to that described in 6.2.4 and Figure 6.6.

| | Time into session (minutes) |
|---|---|
| FAMILIARIZATION | 0 |
| INITIAL INSTRUCTIONS 1 | 2 |
| TRAINING PERIOD 1 | 5 |
| TRAINING PERIOD 2 | 10 |
| PRACTICE SESSION | 15 |
| FINAL INSTRUCTIONS 2 | 25 |
| MONITORING SESSION | 30 |
| REST/COFFEE/INTERVIEW | 75 |

FIGURE 6.7 Test procedures.

### 6.3.1 Instructions and training

When a subject arrived for the experimental session, he or she was allowed to become familiar with the experimental room and the monitoring display. The experimenter (the author) engaged the subject in general conversation while completing the initialization of the computer, and then informed the subject of the general nature of the task. These initial instructions emphasized that the task would be 'quite long', and would involve the 'detection of relatively infrequent signals which had to be discriminated from other more frequent non-signal events'. At this point, the subject was given examples of signals and non-signals and encouraged to ask any questions. These initial instructions were the same for all subjects and experiments.

Subjects were then instructed in the proper use of the response unit. The complexity of the response requirement varied between experiments, as pointed out in 6.2.2. Generally, the greatest instruction time was spent in explaining the use of the rating scale. After this subjects were given two training sessions.

The monitoring literature is not in agreement as to the best method of training for monitoring tasks (see Davies and Tune, 1970; Wiener, 1975b), although there is some evidence that optimal response criteria may be reached by providing appropriate knowledge of results, while detection sensitivity may be improved by training procedures incorporating signal cueing (Annett and Paterson, 1967). In the experiments reported in this thesis, both cueing and knowledge of results were provided in two separate training periods.

In the first training period, the subject was given a sequence of 50 trials with 20 signal trials randomly inserted among these. The subject was verbally cued before a signal occurred. At the end of this training period, most subjects had a clear idea of the distinction between signals and non-signals.

In the second training period, the same number of trials were given, but without cueing. The subject was asked to respond verbally when he detected a signal. Immediate knowledge of results were provided. A criterion performance level for this training session was set, with Hit probability (min) = .7, and false alarm probability (max) = .1. If the subject failed to meet this criterion, the training session was repeated. Such cases were very rare, only 5 subjects having to be retrained.

Following the two training periods, the subject was given an 'expectancy matching' task, which accurately sampled the features of the main monitoring task, except for length, which was 10 min. The subject worked alone, and was told that the task was intended as practice for the main session. Colquhoun and Baddeley (1964, 1967) have emphasized the importance of pre-test expectancy in monitoring tasks.

After this, a final set of instructions were read to the subject. These were standard, except that they varied according to the response requirement. In the rating scale case, it was especially ensured that the subject used the rating categories consistently and understood their meaning. The final instructions emphasized that the monitoring session would be the same as the practice session, but would be longer. Subjects were also instructed to respond as quickly as possible, and to give equal weight to the need to detecting signals and avoiding false detections.

### 6.3.2 Monitoring sessions

Following the training and practice periods, the subject rested while the experimenter initialized the computer for the monitoring session, and checked the apparatus. Subjects were asked to give up their wrist watches. They were not informed about the length of the session, except that it would be 'quite long'; however, since they were previously told that the complete experimental session would last about $1\frac{1}{2}$ hours, they had some indication of the length of the monitoring period. Subjects were told to begin the session when they were ready by depressing the 'START' button.

During the monitoring session (usually 45 min. long), the subject worked alone, and no knowledge of performance was provided. No specific instruction about smoking was given, but only 3 subjects lit cigarettes during the monitoring session. These subjects' performance data were not included in the analysis of group data.

At the end of the session the subject was provided with a hot beverage, and an informal, unstructured interview was then conducted. The general questions asked related to performance on the task and to the general state of the subject, e.g., whether adequate sleep had been had for the previous nights, smoking and drinking habits, etc. Before leaving, subjects were paid for their services.

## 6.4 Subjects

Subjects were recruited with the help of notices displayed on various notice boards situated in Aston and Birmingham Universities. Most of the subjects recruited this way were undergraduate or postgraduate students, but not all subjects used were students. The remaining subjects were generally people in regular employment, but who could spare a morning or afternoon to participate in an experiment. Table 6.2 indicates that, on average, about 30% of the subjects were not students.

| Experiment | Subjects | Males | Females | Students | Psychology Students | Median Age | Age Range |
|---|---|---|---|---|---|---|---|
| 1 | 100 | 100 | 0 | 55 | 30 | 20 | 17-26 |
| 2 | 8 | 8 | 0 | 6 | 6 | 22 | 19-24 |
| 3 | 30 | 30 | 0 | 24 | 16 | 23 | 17-26 |
| 4 | 60 | 30 | 30 | 50 | 28 | 21 | 17-27 |
| 5 | 30 | 30 | 0 | 15 | 10 | 23 | 19-28 |
| 6 | 8 | 8 | 0 | 8 | 8 | 21 | 18-24 |
| 7 | 36 | 18 | 18 | 30 | 10 | 22 | 18-27 |
| 8 | 20 | 20 | 0 | 14 | 6 | 23 | 17-26 |
| Total | 292 | 244 | 48 | 202 | 114 | | |
| Percentage | | 83% | 17% | 70% | 39% | | |

TABLE 6.2 Subject numbers and other subject characteristics in the different experiments.

This table also lists some other characteristics of the subject samples in the different experiments.

In each study, either male subjects only, or both male and female subjects were employed. In the latter case, equal numbers of males and females served in each experimental condition. The age range of both male and female subjects was about 17-28, with some variation between studies. Sex differences in monitoring performance were generally very small (see also Appendix E).

A rigorous screening procedure was not used, but only subjects with good vision (as assessed by subjective report) were employed. Subjects were also asked if they suffered from hearing loss. Partial selectivity was ensured since the recruiting notice emphasized that "if you have good eye-sight and hearing and would like to earn about 75p for $1\frac{1}{2}$ hours of work, contact ....". Most of the subjects who answered this notice had not previously served in an experiment on monitoring behaviour (although when asked what was meant by 'monitoring' or 'vigilance' behaviour, most gave satisfactory answers). Of all the subjects used, however, about 30% had previously served as subjects in other experiments. The time lag between participation in one of the experiments in this thesis and other experiments was not generally less than a month.

## 6.5 Methods of Analysis

Several dependent measures were utilized in the analysis of the performance data. Davies and Tune (1970) have specified a number of different measures, such as correct detections, commission errors, etc., which might be used in the analysis of monitoring performance. In section 4.5, a number of measures derived from, or consistent with, statistical decision theory were also described. These include the d' and B measures of the theory of signal detectability (TSD), as well as continuous measures of performance, such as response time. In the experiments in this thesis, both discrete (choice) and continuous (response time) measures were considered.

### 6.5.1 Performance measures

In Chapter 4 we saw that the use of measures derived from decision theory leads to a consideration of operating characteristics (OCs). An analysis of monitoring performance in OC terms need not make any explicit , *a priori* assumptions about the 'underlying' discrimination process. In the experiments reported in this thesis, performance measures consistent with this type of analysis were used. These included both 'raw' (or 'traditional') and derived measures.

Two raw detection performance measures used were the probability of correct detections (Hits), and the probability of false alarms (FAs). The probability of an FA was taken as the ratio of the number of false alarms to the total number of stimulus events in a given time period (i.e., an 'observation period' covered one stimulus event or trial). The TSD measures d' and B were computed off-line using these Hit and FA probabilities. When these were either 1 or 0, a technique described in Appendix F was used to derive estimates of the probabilities.

For experiments in which the rating method was employed, a four category rating scale was used, as described previously. The four categories 1, 2, 3 and 4 corresponded to the responses 'Certain Yes', 'Doubtful Yes', 'Doubtful No', and 'Certain No', respectively. It was assumed that the ratings mark cut-offs of 'criteria' that are used consistently by subjects. For a four category scale, three categories of criteria can be distinguished. These are denoted, respectively the strict (1), medium (2) and lax (3) criteria. Each criterion defines the responses made at confidence levels <u>above</u> and including that criterion level. The strict criterion includes only the 'Certain Yes' responses, the medium criterion defines both 'Certain Yes' and 'Doubtful Yes' responses, while the lax criterion includes the responses 'Certain Yes', 'Doubtful Yes' and 'Doubtful No'. The Hit and FA measures were categorized according to these criteria; hence three pairs of Hit and FA probabilities were obtained for a given time block. These were used to plot the empirical OC.

The various performance measures used are listed in Table 6.3. It should be noted again that the symbol 'B' is used, for convenience, to refer to

| MEASURE | RESPONSE MODE Single | Binary | Rating | |
|---|---|---|---|---|
| Probability of a Hit | p(H) | p(H) | p(H) above c1, c2, c3 | |
| Probability of a FA | p(FA) | p(FA) | p(FA) above c1, c2, c3 | |
| Criterion cut-off | c | c | c1, c2, c3 | |
| Sensitivity (Detectability) | d' | d' | $d_a$ | |
| Log likelihood ratio | log B | log B | log B1, log B2, log B3 | |
| Detection latency | DL | DL | DL | |
| False alarm latency | FAL | FAL | FAL | (At different |
| Omission latency | - | ML | ML | confidence levels) |
| Correct rejection latency | - | CRL | CRL | |

TABLE 6.3 The performance measures used in the experimental studies.

likelihood ratio, which is more commonly represented by '$\beta$' (Beta). Summary values of the response measures were printed by SCORE at the end of a monitoring session. Mean values were computed over 15-min. time blocks. SCORE also gave standard deviation values for the latency measures, although these were only considered to provide a general indication of the degree of variability in latency within subjects.

### 6.5.2 Notes on the statistical treatment of data

Randomized factorial designs were used in most of the experiments reported in this thesis. In two experiments, where the performance consistency of subjects serving in more than one experimental condition was of interest, designs incorporating correlated observations or 'repeated measures' were used. At any event, in all experiments, performance data were analysed for successive time blocks, so that each analysis of variance (ANOVA) included a 'repeated measures' time blocks factor. For example, in Experiment 1, a three factor ANOVA was used to evaluate the effects of the factors signal probability, task type and time blocks. This factorial design, which assumes linear additive effects, has been described in detail elsewhere (Kirk, 1968; Winer, 1962). in ANOVAs for the 'raw' hit and false alarm data, the scores were adjusted appropriately to compensate for differences between conditions in the signal and event rates.

Treatment effects were tested for significance with the F statistic, as usual. One of the assumptions inherent in its use is that the variances of the experimental error components within the different treatment populations are the same. Fmax tests (see Winer, 1962, p. 92) were used to test for the homegeneity of the error variances associated with non-repeated factors. The departure from homogeneity was generally not great enough to invalidate the F statistic. In the evaluation of the effects associated with repeated factors, the conservative F test of Greenhouse and Geisser (1959) was used (this test avoids the need to carry out the lengthy computation of the variance-covariance matrices of error variance).

When homogeneity assumptions were violated to a greater degree, appropriate transformations of the scale of data were carried out. These are considered separately where they were used. The FA and B measures were log transformed as the distributions of these measures between subjects were generally skewed. Finally, strength of association indices ($w^2$) are occasionally reported for the effects of certain independent variables on performance. The index suggested by Keppel (1973), which is independent of the number of factors in the ANOVA, was used.

These statistical considerations complete this chapter on research methodology. The next six chapters describe the experiments and task classification analyses. Task categorization and decision processes were investigated in relation to visual monitoring in the two experiments reported in the next chapter.

# CHAPTER SEVEN

# TASK CLASSIFICATION AND DECISION PROCESSES IN VISUAL MONITORING

## 7.1 Introduction

Task classification has been identified as a useful tool for the efficient description and evaluation of performance in different tasks. On the basis of a review of various taxonomic approaches, and their effectiveness in providing adequate descriptors of performance differentiation, a task classification system for monitoring tasks comprising four task 'dimensions' was outlined. The components of this classification scheme are the speed-closure or signal type dimension, sense modality, source complexity and the time course of events. The speed-closure dimension was viewed as the primary component of the classification system. As yet, an empirical basis for such a classification scheme has not been provided, nor has its efficiency as a taxonomy of monitoring tasks been examined. These two requirements are a major concern of the experiments reported in the present and succeeding chapters.

Two experiments are reported in this chapter. In Experiment 1, a preliminary evaluation was carried out into the feasibility of the speed-closure dimension of the task classification system in the description and analysis of performance as a function of time at work and of selected independent variables. Following from an important result obtained in this experiment a further investigation of the influence of the signal type dimension on performance sensitivity was carried out in Experiment 2.

Experiment 1 follows directly from the considerations of decision processes in monitoring behaviour in Chapter 5. In that chapter it was concluded that the application of the theory of signal detectability (TSD) within a statistical decision theory framework may provide a fuller understanding of the mechanisms underlying performance trends in monitoring tasks. We saw however, that there were several points of uncertainty regarding decision processes in monitoring behaviour; in particular, the independence of the sensitivity and response criterion aspects of performance has not been established, and little is known of the relative effects of criterion and sensitivity shifts in relation to different types of monitoring task. We shall address ourselves to these issues in the first experiment reported. More specifically, since criterion and sensitivity shifts have been associated with changes in signal probability and event rate, Experiment 1 investigated the influence of these variables on decision processes in visual monitoring within a task classification framework (event rate has also been related to one component of the task classification system, but for the present it will be treated as an independent variable).

A feature of the approach to the analysis of decision processes in this thesis concerns the treatment of response latency data. We have previously noted that a consideration of response latencies might provide a further understanding of the relative effects of criterion and sensitivity shifts, and their interaction, if any, with task demands. A conceptual model for interpreting these shifts was outlined in Chapter 4 (section 4.6). Experiment 1 therefore also investigated the adequacy of this model in terms of its ability to interpret time-related changes in the decision times associated with different types of response in a monitoring task.

An additional impetus for Experiment 1 was provided by the results of a preliminary pilot study. This pilot study was originally concerned with equating task difficulty on two visual tasks, one of which was classified as a speed task and the other a closure task. The two tasks were such that they were matched, as far as possible, on all task dimensions except the speed-closure dimension. In the visual speed task, a decrease in the horizontal separation of two vertical lines had to be detected, while in the visual closure task the critical signal for detection was a small pink circle appearing at the centre of an intermittently flashing circular light source. In both tasks the rate of stimulus presentation was 15/min. Task difficulty was equated by varying signal duration to give approximately equal d' values for both tasks, for both 'alerted' and prolonged performance, and for a group of three subjects. Subsequently, another three subjects were tested on the same tasks, but in which the event rate was raised to 30/min. It was observed that for this group of subjects, a significant decrement in sensitivity (d') occurred, but only for the speed task. For the closure task, there was no effect of event rate and no sensitivity decrement. Accordingly, this result, which was unexpected, and obtained somewhat by chance, was further investigated in the first experiment reported. In the main experiment, however, another two speed and closure tasks were used, to see whether the result was merely related to the features of the apparatus used in the pilot tasks, or was a general result relating to the hypothesized differences in signal discrimination type between speed and closure tasks.

In Experiment 1 therefore, the effects of the speed-closure dimension and of signal probability and event rate on decision processes in monitoring performance were investigated. More specifically, the experiment was concerned with the following issues:

1) Is there a sensitivity decrement over time at high event rates?
2) If so, for what kinds of task does the effect appear?
3) How are criterion changes related to the differing demands of speed and closure tasks?
4) What are the relative effects of signal probability and event rate?
5) What are the performance trends over time in the decision latencies in a monitoring task, and how may response latency data be interpreted?

## 7.2 Experiment 1

Summary

One hundred subjects were assigned to one of ten groups, each of which were tested with either a visual speed or visual closure task, under one of five conditions combining different levels of signal probability (SP) and event rate (ER). The results generally endorsed the feasibility of the task classification system, in that different performance functions were obtained for different task categories. The principal finding was that there was a decrement in sensitivity over the monitoring period, and an overall mean reduction in sensitivity, for the speed task under a high ER. This effect did not appear for the closure task. The demonstration of the sensitivity shift was not dependent on the use of the TSD statistic d', but was also apparent using a quasi-operating characteristic analysis. Criterion shifts were not differentially related to task demands. Response latency data were consistent with a decision theory model, and with the results obtained using discrete measures of performance accuracy. Overall, the results were interpreted as providing preliminary endorsement of the task classification and decision theory approaches to the analysis of monitoring performance.

### 7.2.1 Method

Ten experimental groups were used in Experiment 1. Each group worked with either a speed or a closure task, and under one of five conditions combining different levels of SP and ER (see Table 7.1). The experimental design permitted the effects of SP and ER to be assessed independently over three levels of each variable. Several response measures were utilized in the analysis of performance, including response latency measures. The binary response mode was used to enable the recording of latencies associated with all four categories of response as in the Yes/No task.

Subjects

The subjects were 100 adult males, ranging in age from 17 to 26 years. All of them had good vision and hearing with no subject requiring to wear corrective spectacles or contact lenses. The subjects were randomly assigned to ten equal groups. Further characteristics of the subject sample for this experiment are listed in Table 6.2.

Apparatus

The apparatus and general experimental system were as described previously in 6.2, with all experimental control and data handling being carried out by the PDP-8/E computer. Two visual monitoring tasks were used, a visual speed task (VS1) and a visual closure task (VC2). In VS1 subjects were required to detect a decrease in the intensity of an intermittently flashing circular light source. In VC2, the subjects had to detect a central gap of 0.5 cm. in two vertical lines each 10 cm. long when projected. More detailed descriptions of the tasks are given in section 6.2.5 of Chapter 6. It should be noted however, that in VC2, a discrimination process within a trial is only required, while in VS1, a detection of a change in a stimulus, and hence, a comparison over trials, is required. Furthermore, the superficial physical similarity between these tasks and the tasks used in the pilot is inverted in relation to the task categories: that is, in the pilot study, the speed and closure tasks used slide and flash presentation respectively, while in the main experiment this relation was reversed. This was to provide a means of checking the generality of the result obtained in the pilot study.

In both tasks, the binary response mode was used, with subjects being required to respond either 'Yes' or 'No' to each event. This is a response procedure that is somewhat different to that commonly used in monitoring tasks, but as noted in Chapter 4, there is some evidence suggesting that the use of two response keys does not lead to performance functions significantly different to those obtained using the single response method (see also Appendix D).

| Condition | Signal Rate | Event Rate | Signal Probability |
|---|---|---|---|
| I | $0.53\dot{3}$/min | 8/min | $0.06\dot{6}$ |
| II | 1.0/min | 15/min | $0.06\dot{6}$ |
| III | 2.0/min | 30/min | $0.06\dot{6}$ |
| IV | $0.53\dot{3}$/min | 15/min | $0.03\dot{5}$ |
| V | $0.26\dot{6}$/min | 15/min | $0.01\dot{7}$ |

Variable Event Rate

Variable Signal Probability

TABLE 7.1 Signal probability (SP) and event rate (ER) levels defining Conditions I to V. The lines link the two sets of conditions in which ER is varied for a fixed SP (Conditions I, II and III), and vice versa (Conditions II, IV and V).

Finally, in both speed and closure tasks, signals were presented irregularly, with inter-signal intervals being randomized within successive 15-min. blocks. The same signal schedule was preserved for each time block. Each monitoring session was 45 min. long. Signal rates were 0.26, 0.53, 1.0 or 2.0/min, and event rates were 8, 15, or 30/min., depending on the experimental condition.

## Procedure

Subjects were randomly assigned to one of ten groups; half of these worked with a speed task and the other half with a closure task. The five conditions of the experiment are defined in Table 7.1. In conditions I, II and III, ER was varied, while SP was held constant at

0.066, while in conditions II, IV and V, ER was fixed at 15/min., and SP varied. The five conditions thus represent independent variation of SP and ER over three levels, with one common condition (II).

The experimental procedures were as described previously in section 6.3 of Chapter 6. Each subject received extensive training and expectancy matching practice as outlined in 6.3.

### 7.2.2 Results

#### Data analysis

The mean probabilities of correct detections (Hits) and false alarms (FAs) were derived for each subject in successive 15-min. time blocks of each 45-min. monitoring period. The TSD parameters d' and log B were computed from these operating probabilities, but an interpretation of these indices in terms of sensitivity and criterion changes was always made with reference to the corresponding quasi-operating characteristic. The mean response latencies associated with all four categories of response (correct detections, false alarms, correct rejections and omissions) were also recorded.

In the statistical analysis of the data, each performance measure was analysed in two analyses of variance (ANOVAs), one for independent ER effects (comparing conditions I, II and III), and the other for independent SP effects (comparing conditions II, IV and V). These two ANOVAs are denoted the ER ANOVA and SP ANOVA, respectively. The factors in each ANOVA (a 2 x 3 x 3 split-plot design) were Tasks (speed or closure task), ER or SP (three levels each), and Time Blocks (three levels), with 'repeated measures' on the Time Blocks factor. The notation used for these ANOVAs is shown in Table 7.2. Note that the symbol B refers to either SP or ER, depending on which ANOVA is considered, while both ANOVAs share the same factors A and C. Furthermore, the levels of factor

B from B1 to B3 represent an increase in ER in the ER ANOVA, and a decrease in SP in the SP ANOVA. Furthermore it should be noted that the symbol B is also used to represent the likelihood ratio Beta.

Each performance measure was analysed in each of the two ANOVAs, and the results are treated separately for the different measures. Since a large number of ANOVAs were obtained (16), only those associated with the main results of interest are listed in the main body of this chapter; however, the other ANOVAs are listed in full in Appendix C1.

Correct detections and false alarms

The mean probabilities of Hits, p(H), in each 15-min. time block and for each task and condition are displayed in Figure 7.1(a,b). The raw detection data were analysed in the two ANOVAs, as described previously, to

| ER ANOVA (Conditions I, II, III) | | | SP ANOVA (Conditions II, IV, V) | | |
|---|---|---|---|---|---|
| A | Tasks | | A | Tasks | |
| A1 | Speed task | | A1 | Speed task | |
| A2 | Closure task | | A2 | Closure task | |
| B | ER | | B | SP | |
| B1 | Condition I: | 8/min. | B1 | Condition II: | 0.066 |
| B2 | Condition II: | 15/min. | B2 | Condition IV: | 0.035 |
| B3 | Condition III: | 30/min. | B3 | Condition V: | 0.017 |
| C | Time Blocks | | C | Time Blocks | |
| C1 | 1st 15-min. time block | | C1 | 1st 15-min. time block | |
| C2 | 2nd 15-min. time block | | C2 | 2nd 15-min. time block | |
| C3 | 3rd 15-min. time block | | C3 | 3rd 15-min. time block | |

TABLE 7.2 Notation for SP and ER ANOVAs (see text).

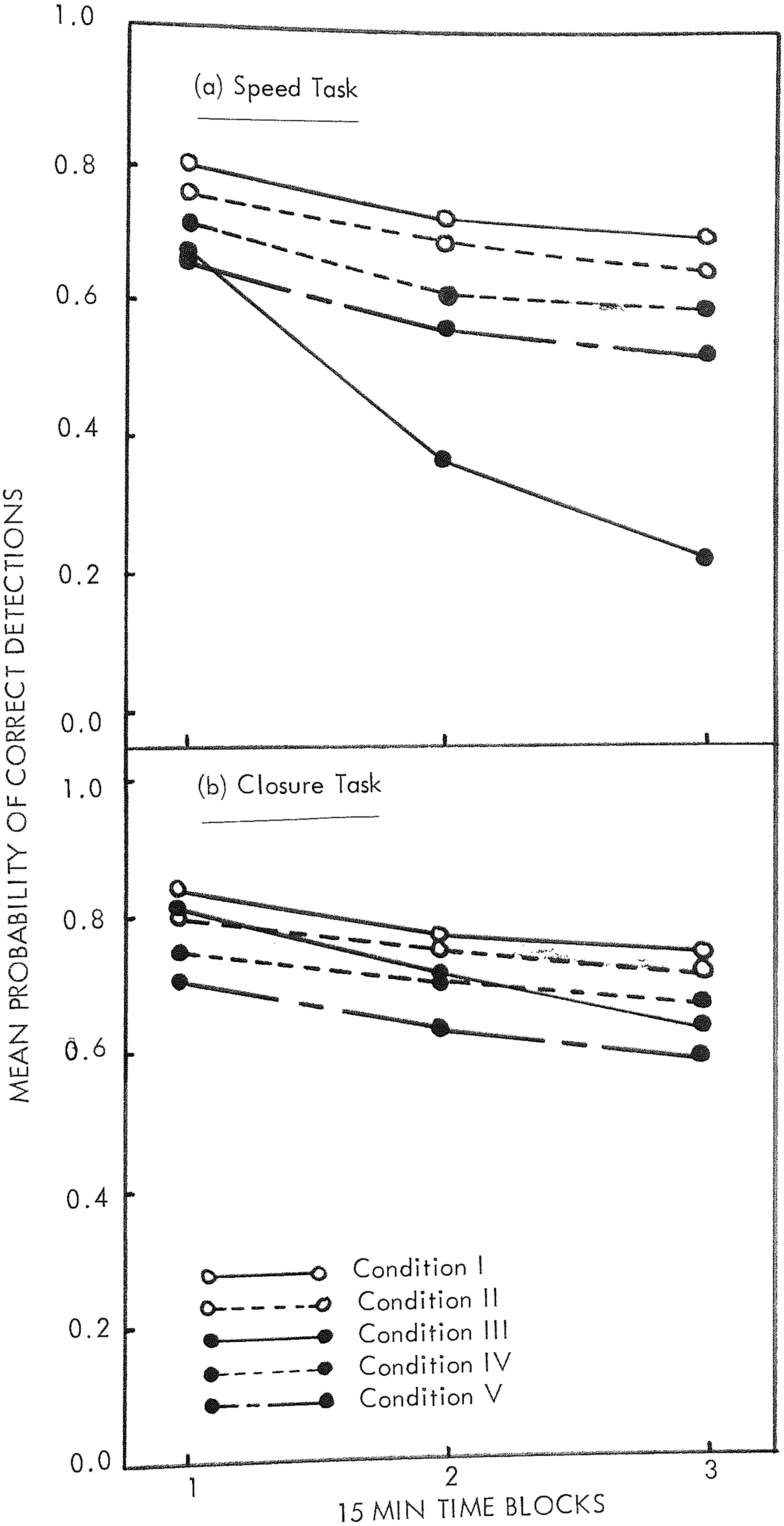


FIGURE 7.1 The mean probability of correct detections as a function of time blocks and conditions, and for both speed and closure tasks.

examine the independent effects of SP and ER. The summary table for the ER ANOVA for Hits is given in Table 7.3. Three main effects were obtained, for Tasks ($p < .001$), ER ($p < .001$) and Time Blocks ($p < .001$). In addition, the ER x Time Blocks and the Tasks x ER x Time Blocks interactions were also significant (both, $p < .01$).

These results confirm the general impression gained from an inspection of Figure 7.1, the principal result being that in the high ER condition (III), there is a greater decrement and an overall reduction in Hits, but for the speed task only. This is also apparent in the tests for the simple effects of ER and the ER x Time Blocks interaction (see Table 7.3).

| SOURCE | SS | DF | MS | F | p | $w^2$ |
|---|---|---|---|---|---|---|
| A (Tasks) | 406.5200 | 1 | 406.5200 | 14.794 | .001 | |
| B (Event Rate) | 1129.8090 | 2 | 564.904 | 20.559 | .001 | .395 |
| AB | 570.5191 | 2 | 286.759 | 10.436 | .001 | .239 |
| SWG | 1483.7474 | 54 | 27.477 | | | |
| C (Time Blocks) | 789.115 | 2 | 394.557 | 40.185 | .001 | |
| AC | 56.951 | 2 | 28.475 | 2.900 | ns | |
| BC | 300.145 | 4 | 75.036 | 7.642 | .01 | |
| ABC | 299.185 | 4 | 74.796 | 7.618 | .01 | |
| C x SWG | 1060.396 | 108 | 9.818 | | | |
| Simple effects | | | | | | |
| B at A1 (Speed Task) | 1647.705 | 2 | 823.852 | 29.893 | .001 | .491 |
| B at A2 (Closure Task) | 52.622 | 2 | 26.311 | 0.958 | ns | |
| A at B1 | 0.947 | 1 | 0.947 | 0.034 | ns | |
| A at B2 | 8.067 | 1 | 8.067 | 0.294 | ns | |
| A at B3 | 968.025 | 1 | 968.025 | 35.231 | .001 | |
| SWG | 1483.747 | 54 | 27.477 | | | |
| BC at A1 | 367.885 | 4 | 91.971 | 9.368 | .01 | |
| BC at A2 | 31.445 | 4 | 7.861 | 0.800 | ns | |
| C x SWG | 1060.396 | 108 | 9.818 | | | |

Comparisons between means (Tukey's HSD Test)

| Comparison | q | p |
|---|---|---|
| B1 - B2 at A1 | 1.405 | ns |
| B1 - B3 at A1 | 10.565 | .01 |
| B2 - B3 at A1 | 9.160 | .01 |

$q_{.05}(2,30) = 2.89$
$q_{.01}(2,30) = 3.89$

TABLE 7.3 ER ANOVA and comparisons between means for Hits in Conditions I, II and III.

Three levels of ER were used in Conditions I, II and III: 8/min., 15/min. and 30/min. However, the effect of ER was apparent only between Conditions II and III, that is, for an increase in ER from 15 to 30/min. For both tasks, the mean number of hits over an entire monitoring session was not significantly different between Conditions I and II, as indicated by Tukey's Honestly Significantly Difference (HSD) test (see Kirk, 1968).

The significant effect due to the Tasks factor indicates that the number of Hits in the closure task was greater than that in the speed task. However, a test of simple effects of this factor at each ER level indicated that 99.1% of its variance was accounted for by the simple effect at the highest ER, the simple effects at the lower ER levels being non-significant. Hence there were no significant differences in the overall level and decrement in the probability of a Hit between speed and closure tasks, except in condition III, the high ER condition.

For the SP ANOVA (see Table 7.4), there were significant effects due to SP ($p < .05$) and Time Blocks ($p < .01$) only. The effect of a reduction of SP was to depress the mean probability of Hits fairly uniformly in all three time blocks and for both tasks. While there was a small but significant decrement in Hits in all three SP conditions, the decrement was not significantly different between tasks and over SP levels.

The false alarm data is given in Figure 7.2 (a, b), which displays the <u>median</u> probabilities of FAs, p(FA), as a function of time blocks for both tasks and all conditions. It should be noted that the probability values were based on a differing number of nonsignal trials in the various conditions; the number in any time block were 112, 210, 420, 217 and 221 for conditions I, II, III, IV and V respectively.

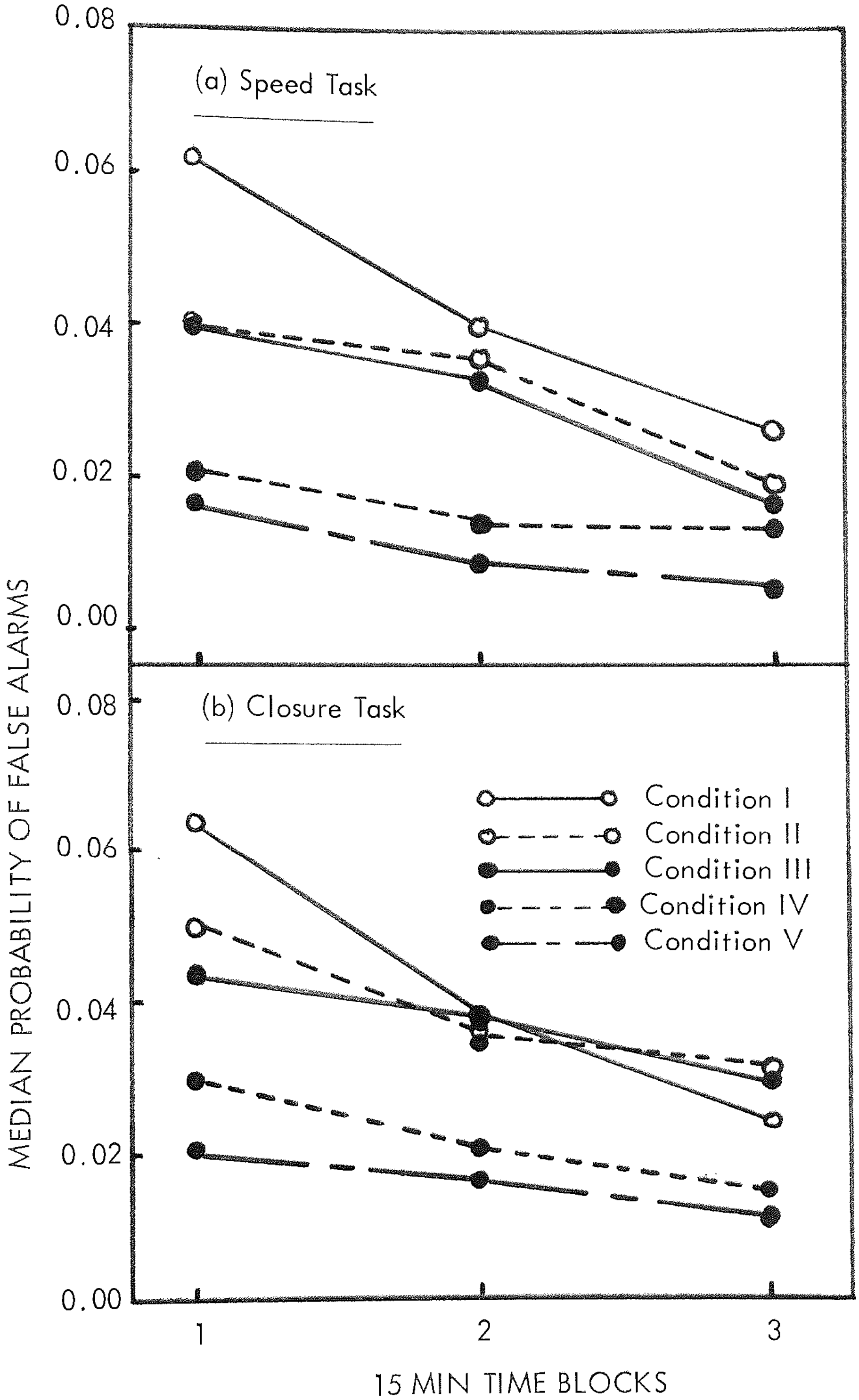


FIGURE 7.2 The median probability of false alarms as a function of time blocks and conditions, and for both speed and closure tasks.

| SOURCE | SS | DF | MS | F | p | $w^2$ |
|---|---|---|---|---|---|---|
| A (Tasks) | 2.903 | 1 | 2.903 | 0.7992 | ns | |
| B (Signal Probability) | 23.693 | 2 | 11.847 | 3.2614 | .05 | .070 |
| AB | 0.088 | 2 | 0.044 | 0.0121 | ns | |
| SWG | 196.157 | 54 | 3.633 | | | |
| C (Time Blocks) | 22.880 | 2 | 11.440 | 0.677 | .01 | |
| AC | 0.374 | 2 | 0.187 | 0.158 | ns | |
| BC | 0.101 | 4 | 0.025 | 0.021 | ns | |
| ABC | 0.585 | 4 | 0.146 | 0.123 | ns | |
| C x SWG | 127.680 | 108 | 1.182 | | | |

Comparisons between means (Tukey's HSD Test)

| | q | p | |
|---|---|---|---|
| B1 - B2 | 2.186 | ns | |
| B2 - B3 | 2.242 | ns | $q_{.05}(2,60) = 2.83$ |
| B1 - B3 | 4.430 | .01 | $q_{.01}(2,60) = 3.76$ |

TABLE 7.4 SP ANOVA and comparisons between means for Hits in Conditions II, IV and V.

The distribution of FAs between subjects was skewed in most conditions, but after transformation by log (X + 1), the raw scores did not violate the homogeneity assumptions of the ANOVA (as indicated by Fmax tests; see also section 6.5.2 in Chapter 6).

The ER ANOVA gave significant effects for the Time Blocks factor only ($p < .001$; see summary Table C1.1 in Appendix C1). Figure 7.2 shows an apparent decrease in p(FA), but this was not significant. It is notable that ER had no significant effect on the false alarm probability.

The SP ANOVA for FAs indicated that in both tasks there was a significant decline in the probability of FAs with time on task ($p < .001$), and with a reduction in SP from condition II to V ($p < .01$; see Table C1.2 in Appendix C1).

There were no differences in FAs between tasks in any of the five conditions, and there were no significant interaction effects. These results therefore indicate that while there was a consistent decline in FAs with time on task, the only effect of conditions was that associated with the decrease in p(FA) with a decrease in SP.

## Sensitivity and the response criterion

Sensitivity, or signal detectability, was estimated by computing values of d', the TSD statistic whose derivation has been outlined previously. Since the binary response mode was used in this experiment, it was not possible to evaluate the goodness of fit of the TSD Gaussian model to the data. However, in the following subsection we shall outline an analysis of the data using a quasi-operating characteristic which indicates that the results obtained using d' are not significantly dependent on an acceptance of the TSD model. The same is true, although to a lesser extent, of the results for log B, which was used to estimate the criterion.

The sensitivity and criterion data is shown in Table 7.5 (Condition V was excluded from TSD analysis because of the low number of signal trials in this condition). The raw data were analysed, as before in the ER and SP ANOVAs.

The ER ANOVA for d' gave significant effects for Tasks, ER and the interaction between ER and Tasks ($p < .001$ in each case; see Table 7.6). The interaction between ER and Time Blocks was also significant ($p < .05$), but the interaction between Tasks and Time Blocks and the three-way interaction were not significant.

These results show fairly clearly that a decrement in sensitivity over the work period was obtained in only one condition: that is, in the high

| Condition | Sensitivity (d') Speed Task | Sensitivity (d') Closure Task | Criterion (log B) Speed Task | Criterion (log B) Closure Task | Log Bo |
|---|---|---|---|---|---|
| I | 2.57 2.70 2.75 (2.67) | 2.69 2.48 2.94 (2.70) | .293 .595 .767 (.552) | .407 .535 .620 (.521) | 1.146 |
| II | 2.51 2.42 2.52 (2.48) | 2.52 2.48 2.53 (2.51) | .497 .693 .871 (.687) | .470 .672 .844 (.662) | 1.146 |
| III | 2.40 1.68 1.52* (1.87) | 2.68 2.50 2.50 (2.56) | .649 .871 .979 (.833) | .451 .803 .884 (.713) | 1.146 |
| IV | 2.64 2.49 2.62 (2.58) | 2.57 2.56 2.67 (2.59) | .752 .948 1.03 (.908) | .635 .805 1.06 (.833) | 1.434 |

TABLE 7.5 Mean values of sensitivity (d') and criterion (log B) in successive time blocks for speed and closure tasks in Conditions I, II, III and IV (values in brackets indicate mean performance levels averaged over time blocks; log Bo is the optimum value of log likelihood ratio, see text).

ER condition III, and for the speed task only. This condition is indicated by an asterisk in Table 7.5. Although the three-way interaction was not significant, the tests for simple effects shown in Table 7.6 indicate that there was a sensitivity decrement for this case only, and an overall reduction in d'. However, between conditions I and II (ER levels of 8/min. and 15/min.), there was no change in d' in either task. These results therefore demonstrate that a decrement in sensitivity over a monitoring period is obtained only at the higher event rates, and for speed tasks only.

The SP ANOVA for d' gave no significant effects for any factor (see Table C1.3, Appendix C1); d' was approximately invariant between tasks, with time on task and between conditions II and IV.

The data for log B are also shown in Table 7.5. The ER ANOVA for this measure gave significant effects due to Time Blocks ($p < .001$) and ER ($p < .05$), there being no other significant effects (see Table C1.4 in Appendix C1). Log B declined significantly over time and increased with ER.

The SP ANOVA for log B showed that there was a significant increase in the response criterion with time blocks ($p < .001$) and with the reduction in SP from condition II to condition IV ($p < .01$). Both these results are in general agreement with the predictions of decision theory, although there was no effect on the <u>trend</u> in log B of a change in SP. The log B data were consistent with Williges' (1973) ideal observer hypothesis, since the obtained log B values approached the optimum values (log $B_0$) in the last 15-min. time block, as shown in Table 7.5 (log $B_0$ = log $((1 - p)/p)$, where p = signal probability). However, due to the lack of a means of testing the validity of the TSD equal-variance model, it is rather risky to interpret the <u>absolute</u> values of B in relation to ideal values; as for the case of d', we shall restrict ourselves to interpretations of the <u>trends</u> in either of these measures.

| SOURCE | SS | DF | MS | F | p | $w^2$ |
|---|---|---|---|---|---|---|
| A (Tasks) | 2.8476 | 1 | 2.8476 | 12.5139 | .001 | |
| B (Event Rate) | 6.8280 | 2 | 3.414 | 15.0029 | .001 | .318 |
| AB | 4.4378 | 2 | 2.2189 | 9.7150 | .001 | .225 |
| SWG | 12.2880 | 54 | 0.2276 | | | |
| C (Time Blocks) | 1.0325 | 2 | 0.5163 | 2.1130 | ns | |
| AC | 0.5312 | 2 | 0.2656 | 1.0870 | ns | |
| BC | 3.1423 | 4 | 0.7856 | 3.2161 | .05 | |
| ABC | 1.2883 | 4 | 0.3208 | 1.3129 | ns | |
| C x SWG | 26.3822 | 108 | 0.2443 | | | |

Simple effects

| | | | | | |
|---|---|---|---|---|---|
| B at A1 (Speed Task) | 10.6766 | 2 | 5.3383 | 23.4593 | .001 |
| B at A2 (Closure Task) | 0.5892 | 2 | 0.2946 | 1.2946 | ns |
| A at B1 | 0.0138 | 1 | 0.0138 | 0.0610 | ns |
| A at B2 | 0.0123 | 1 | 0.0123 | 0.0544 | ns |
| A at B3 | 7.2593 | 1 | 7.2593 | 31.9011 | .001 |
| SWG | 12.2880 | 54 | 0.2276 | | |
| C at B1 | 0.7593 | 2 | 0.3786 | 1.55 | ns |
| C at B2 | 0.0911 | 2 | 0.0455 | 0.1863 | ns |
| C at B3 | 3.3244 | 2 | 1.6622 | 6.8040 | .01 |
| C x SWG | 26.3822 | 108 | 0.2443 | | |

Comparisons between means (Tukey's HSD Test)

| Comparison | q | p |
|---|---|---|
| B1 - B2 at A1 | 2.174 | ns |
| B1 - B3 at A1 | 9.454 | .01 |
| B2 - B3 at A1 | 7.088 | .01 |
| B1 - B2 at A2 | 2.193 | ns |
| B1 - B3 at A2 | 1.623 | ns |
| B2 - B3 at A2 | 1.570 | ns |

$q_{.05}(2,30) = 2.89$
$q_{.01}(2,30) = 3.89$

TABLE 7.6 ER ANOVA and comparisons between means for d' in Conditions I, II and III.

## Operating characteristics

The results using the TSD indices d' and log B have indicated that in all but one of the conditions used in this experiment, performance trends over the monitoring period are associated with an invariant sensitivity and an increasing response criterion over time. Only in the high ER condition III, and for the speed task only, was a significant decrement in sensitivity observed.

The use of the d' and log B indices rests, as has been noted, on the validity of the equal-variance TSD model. A check on this model is usually made by deriving an empirical operating characteristic (OC). The experimental procedure used in this experiment does not permit the OC to be derived in the usual way. However, by pooling together hit and false alarm data from different time blocks and conditions, and plotting the Hit and FA probabilities in z space (as in a normal probability OC), some further insight into the nature of the criterion and sensitivity shifts may be gained. These z plots, or quasi-OCs are shown in Figure 7.3 (a, b) for the speed and closure tasks. Each co-ordinate (z(H), z(FA)) represents the mean of the z scores of p(H) and p(FA) averaged over subjects, rather than the z score corresponding to the mean p(H) and p(FA) probabilities. As in the construction of group OCs, the method of averaging the z scores reflects the central tendency of a group of subjects better than the method of averaging the probabilities.

The data points for conditions I, II and IV for the speed task (Figure 7.3a) and for all conditions for the closure task (Figure 7.3b), are approximately fitted by straight lines of near unit slope (about 0.8), thus indicating a rough constancy in sensitivity over time blocks and between conditions; the data points in these conditions for the speed task lie within two d' lines differing by only about 0.2 (for the closure task the variation is somewhat greater). The operating probabilities are thus related to each other in a manner consistent with a change in criterion and an invariance in sensitivity with time blocks and with a change in SP.

For condition III for the speed task, however, a decrement and an overall reduction in sensitivity is apparent, since the points in Figure 7.3a (joined by the heavy dashed line) cross several d' lines. Hence, although we cannot take as correct the absolute values of d' as a measure of

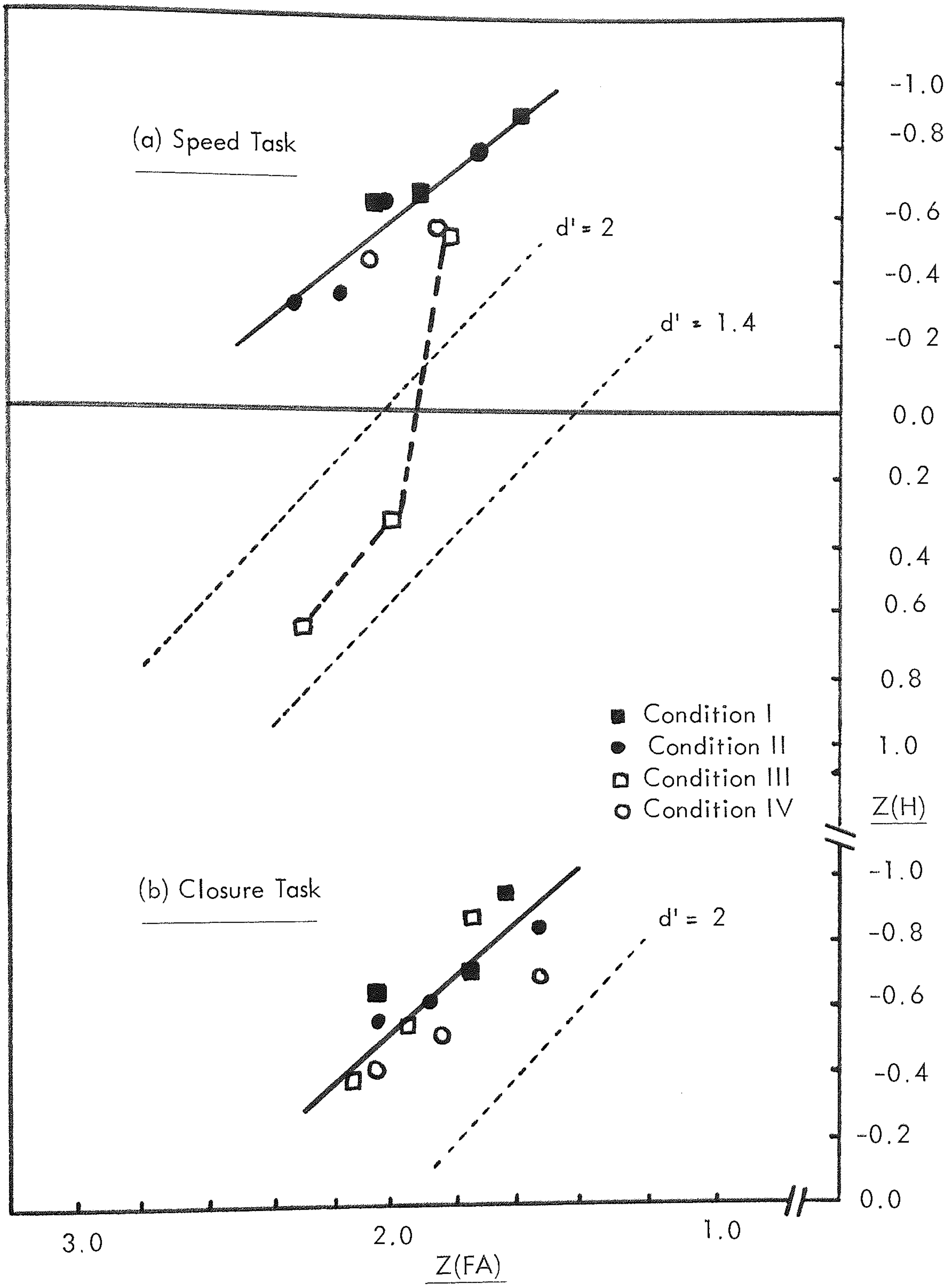


FIGURE 7.3 Group 'z-plots' of the mean normal deviate hit (Z(H)) and false alarm (Z(FA)) values in corresponding time blocks and conditions, and for the speed (a) and closure (b) tasks. In (a) the solid line has been fitted by eye to the points for Conditions I, II and IV only.

sensitivity, the trend in d' does reflect a reliable trend in sensitivity.

A further illustration of the reduction in sensitivity in condition III is shown in Figure 7.4, in which mean performance levels are compared across conditions II, III and IV using the analysis developed by Pollack and Norman (1964). This analysis, which was discussed in Chapter 4 (section 4.4.1), delimits the regions in the unit square in which the symmetric OC passing through a single operating point is constrained to lie, so that an approximate measure of sensitivity may be obtained. We will not derive such a measure, but this analysis for, admittedly 'rough' data (because of the low false alarm rate), nevertheless shows that an OC passing through the condition III data point must be associated with a lower sensitivity than that at condition II or IV. On the other hand, the plot shows, as has been noted previously in the results for d', that conditions II and IV cannot be distinguished on the basis of detectability, but that performance in the latter condition is associated with a higher response criterion.

As far as such changes in the response criterion are concerned, it is probably true that although a meaning cannot be attached to a particular log B value, trends in this measure do reflect the trends in the response criterion. We have previously noted that it may be instructive to note the trends in the criterion cut-off c, before interpreting log B values in terms of criterion bias. For the SP conditions, the trends in c are consistent with those in log B, an increase in c being accompanied by an increase in log B, with time on task and with a reduction in SP. However, for the ER conditions, a significant increase in log B was obtained for an increase in ER, while there was little change in the criterion cut-off c. This is shown in Table 7.7. It should also be noted that there was no significant effect of ER on FA probability. These results therefore suggest that with a reduction in sensitivity, the

| Dependent Measure | Speed Task Event Rate 8 | 15 | 30 | Closure Task Event Rate 8 | 15 | 30 |
|---|---|---|---|---|---|---|
| d' | 2.67 | 2.48 | 1.87 | 2.70 | 2.51 | 2.56 |
| log B | 0.55 | 0.69 | 0.83 | 0.52 | 0.66 | 0.71 |
| c | 1.89 | 1.91 | 1.98 | 1.82 | 1.81 | 1.92 |

TABLE 7.7 Mean values of sensitivity (d'), criterial likelihood ratio (log B) and criterion cut-off (c) as a function of event rate for speed and closure tasks.

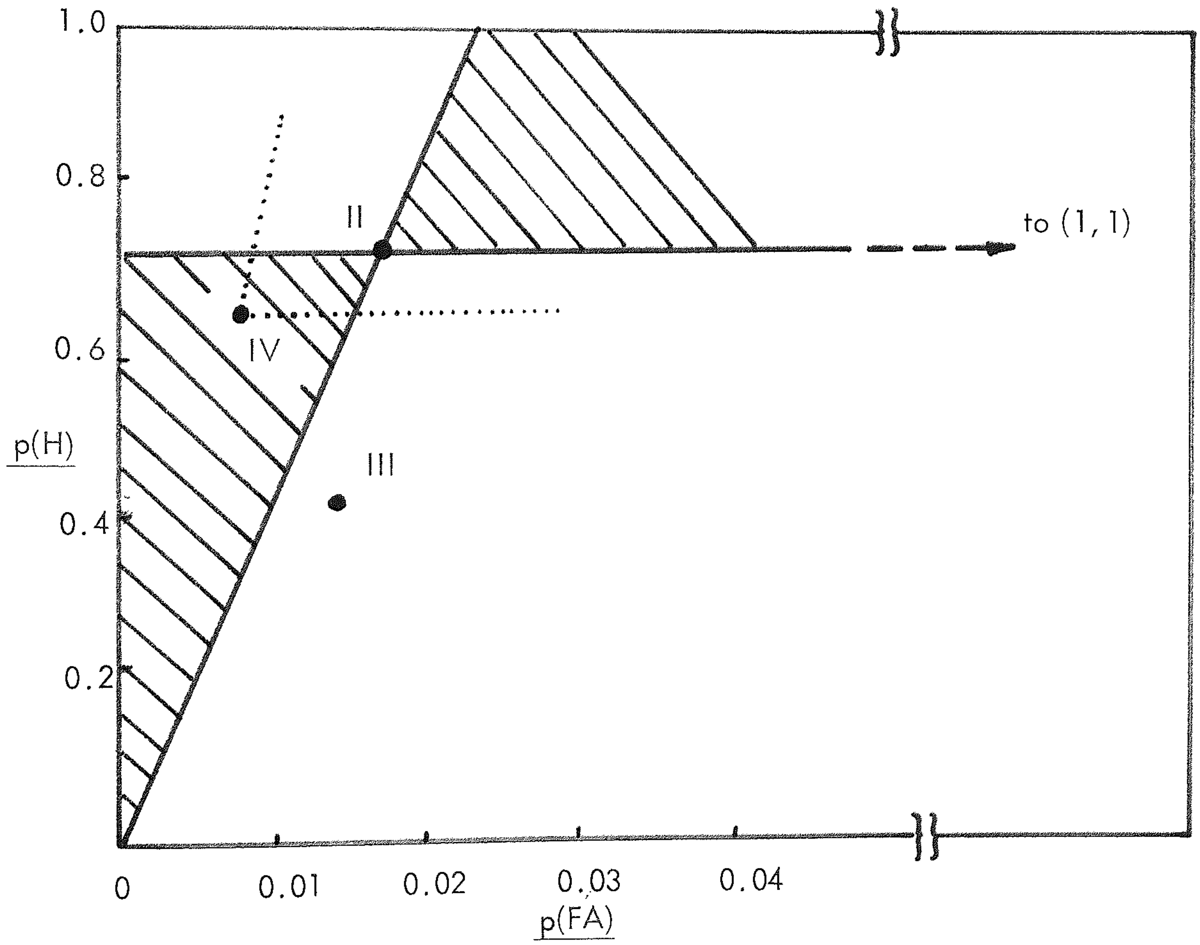


FIGURE 7.4 Operating probabilities in Conditions II, III and IV and the regions in the unit square to which symmetric operating characteristics passing through II (and, for the dotted lines, for IV) are constrained (hatched area).

criterion cut-off is held constant rather than B; so that the decrease in d' is seen in a change in B. This finding will be discussed further in a later part of this chapter, and with reference to the evidence obtained from latency data, which is now considered.

Response latencies

Four measures of response latency were available: Hit or detection latency (DL), false alarm latency (FAL), omission latency (ML) and correct rejection latency (CRL). For each subject, a latency measure for each response category was taken as the average of the corresponding individual response latencies over a 15-min. period. The number of responses going into the means varied between subjects and response categories, but each mean value was based on at least two responses within a 15-min. time block. Since there were a number of instances in conditions I and V where fewer than two incorrect responses (false alarms and omissions) occurred, these conditions were excluded from the analysis.

Mean response latency values, averaged over subjects in each condition, are displayed in Figure 7.5 and 7.6. It is immediately apparent that while the latencies associated with (correct or incorrect) positive (Yes) responses increased with time on task, the latencies associated with negative (No) responses decrease or remain stable over time blocks. To test for these trends, as well for the effects of SP and ER, each latency measure was analysed, as before, in the SP and ER ANOVAs (the only difference being that for these analyses the SP and ER factors had only two levels each due to the exclusion of conditions I and V).

The eight summary tables of the SP and ER ANOVAs for the four latency measures are listed in full in Appendix C1 in Tables C1.5 to C1.12. For convenience, the significant results are listed in Table 7.8. This table shows the direction and significance levels associated with the effects

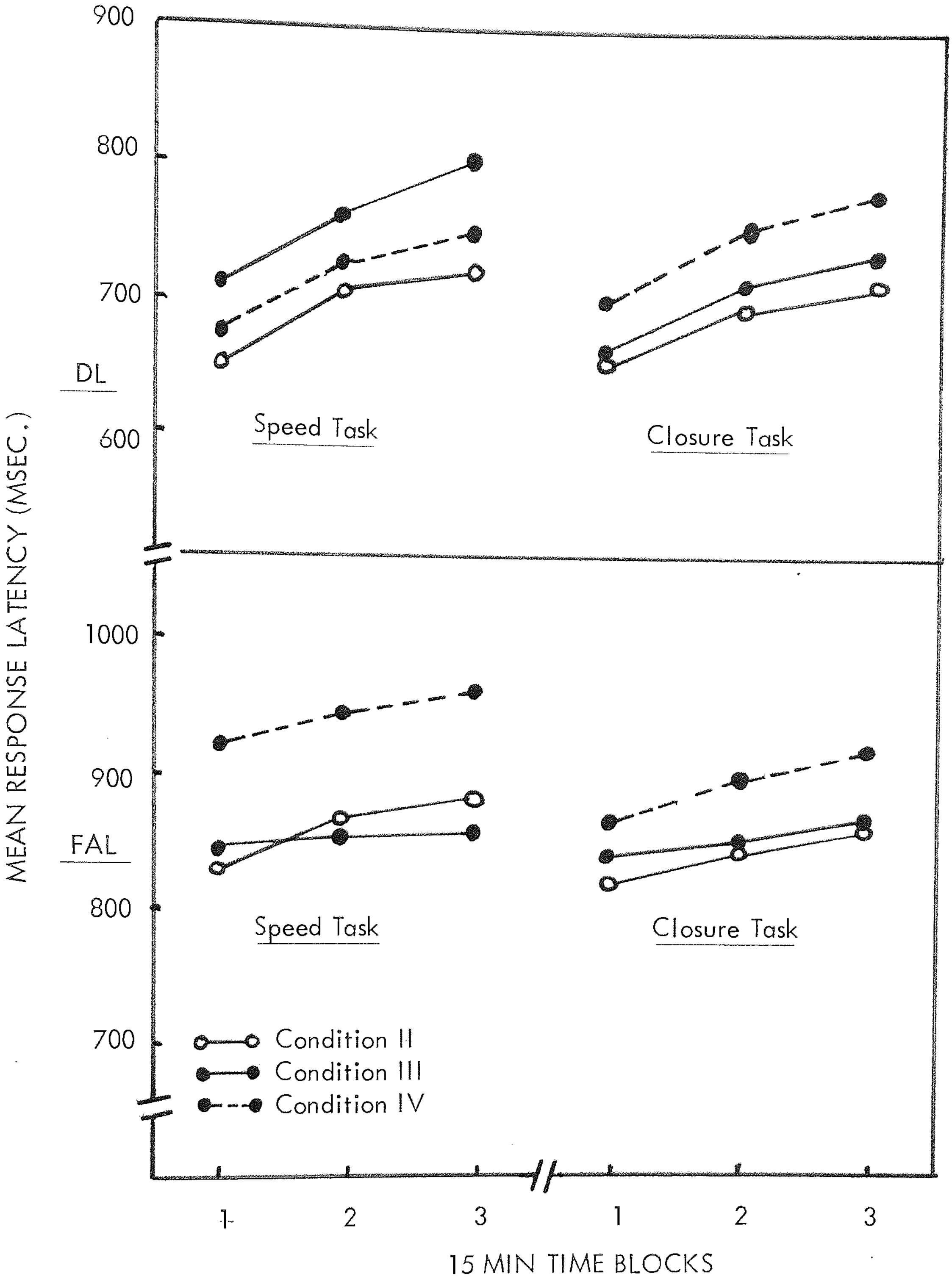


FIGURE 7.5 The mean latencies of correct and incorrect Yes responses (DL and FAL respectively) as a function of time blocks and conditions, and for both tasks.

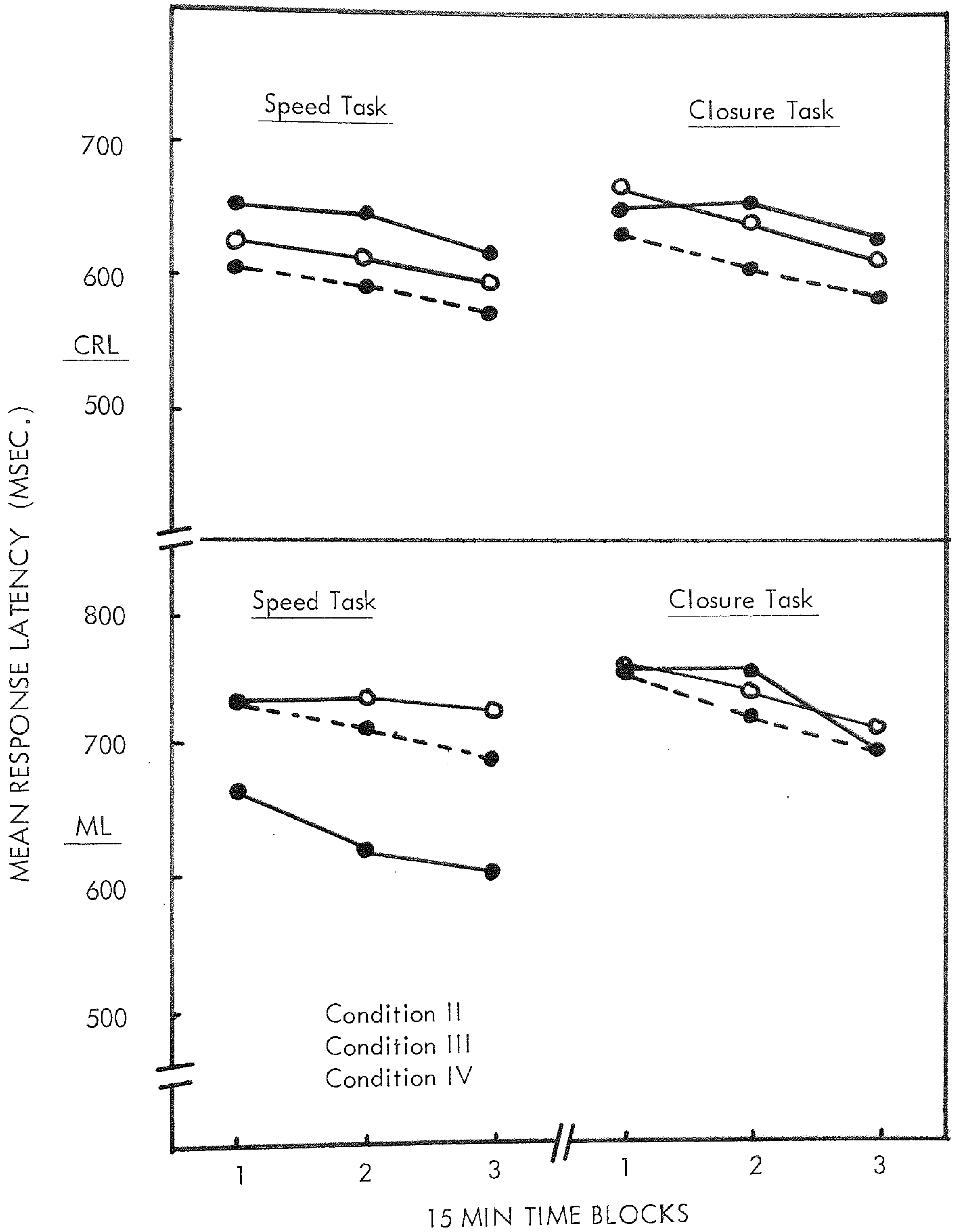


FIGURE 7.6 The mean latencies of correct and incorrect No responses (CRL and ML, respectively) as a function of time on task and conditions, and for both tasks.

of the three independent variables SP, ER and Time Blocks and their interactions on each latency measure. The various results may be summarized as follows:

1) For <u>both</u> the speed and the closure task, and in <u>each</u> of the conditions II, III and IV, there was a significant increase in the mean latencies associated with Yes responses (DL and FAL), and a significant decrease in the mean latencies associated with No responses (CRL and ML). These trends over time on task are consistent with an increase in the criterion, and are in agreement with the predictions of the decision theory latency model outlined in Chapter 4 (section 4.6).

2) With an increase in SP (from condition IV to II), there was a significant decrease in the mean latencies of Yes responses (DL and FAL). At the same time, there was an increase in the No latencies CRL and ML, as

| | ANOVA FACTORS | | |
|---|---|---|---|
| Response Latency | SP | ER | Time Blocks |
| DL | Decrease<br>$p < .05$ | Increase[1]<br>$p < .05$ | Increase<br>$p < .001$ |
| FAL | Decrease<br>$p < .01$ | No change<br>ns ($F = 0.00$) | Increase<br>$p < .05$ |
| CRL | Increase<br>ns ($F = 1.27$) | No change<br>ns ($F = 0.49$) | Decrease<br>$p < .001$ |
| ML | Increase<br>ns ($F = 0.47$) | Decrease[2]<br>$p < .05$ | Decrease<br>$p < .001$ |

<u>TABLE 7.8</u> Summary of effects of SP, ER and Time Blocks on each of the four latency measures (see text; df = 1,36 for each factor; 1 for speed task only, simple effect, $F = 18.04$, $p < .001$; 2 for speed task only, simple effect, $F = 9.99$, $p < .01$).

shown in Figure 7.6, but the increase was not significant in either case.

3) With an increase in ER (from condition II to III), the only significant effects obtained were for the mean latencies of responses to signals (DL and ML). For each latency measure, there were significant interactions between ER and Tasks, indicating that ER had opposing effects on DL and ML, but for the speed task only. There were no significant effects of ER on the latencies of responses to non-signals (CRL and FAL).

4) There were no other reliable interaction effects.

5) The only effect of Tasks was for DL and ML in conditions II and III; this was due to the fact that with the increase in ER between these conditions, both DL and ML significantly changed in opposing directions (see 3) above), but for the speed task only.

These results therefore demonstrate that the effect of an increase in the response criterion, whether due to prolonged work at the task or due to a reduction in signal probability, was to significantly increase the mean Yes response latencies and, in some cases, to significantly decrease the mean No response latencies. On the other hand, with a decrease in sensitivity due to increased ER, there was a significant increase in the mean detection latency and a significant decrease in the mean omission latency, while the latencies of responses to non-signals remained invariant. We shall discuss these results in further detail in the next section.

### 7.2.3 Discussion

Task classification emerges as an important tool in the description and evaluation of monitoring performance in different task situations. In both a preliminary pilot study and in Experiment 1, performance functions for a number of dependent measures were found to be reliably different, in a consistent manner, between different task categories. The principal result was that for speed tasks with a high rate of stimulation, as compared to closure tasks and 'low-rate' speed tasks, there is a reliable decrement and an overall reduction in perceptual sensitivity over the monitoring period. The fact that this result was observed for different speed and closure tasks (as used in the pilot study and in Experiment 1), and for both discrete (accuracy) and continuous (latency) measures, testifies to its generality. The results can therefore be interpreted as providing a preliminary endorsement of the utility of the task classification approach to evaluating monitoring performance as a function of different task categories and selected independent variables, as originally suggested by Fleishman (1972) and Levine, Romashko and Fleishman (1971).

However, the results of the present experiment are not in agreement with the more specific results of Levine et al. (1971) with regard to the effects of signal rate on performance. Levine et al. reported that an interaction between the effects of signal rate, task category and time on task emerged when 53 studies in the pre-1969 monitoring literature were classified according to whether they used speed or closure tasks. They reported that for low signal rates less than 1/min., closure tasks exhibited a sharper decrement in hits and had a lower mean performance level than speed tasks; for moderate signal rates (1 to 2/min.), speed tasks demonstrated a decrement while no decrement was obtained for closure tasks; finally, for signal rates greater than 2/min., Levine et al. found no differences between speed and closure tasks.

In the present experiment, the range of signal rates used (0.26 to 2.0/min.) covers the three signal rate categories of Levine et al., but no differential trends between speed and closure tasks were noted. However, an examination of the tasks included in the report of Levine et al. indicates that the discrepancy between the studies may be due to certain methodological inadequacies in the experiments in the literature which have investigated signal rate effects. As noted in 5.4, Baddeley and

Colquhoun (1969) have pointed out that much of the pre-1969 research on signal rate is unreliable because subjects were not 'expectancy matched', and because the number of unwanted signals was often varied at the same time as the number of signals. Some of these studies (e.g. Jenkins, 1958; Kappauf and Powe, 1959) were included in the Levine et al. analysis. Baddeley and Colquhoun (1969; Colquhoun and Baddeley, 1964, 1967) have, however, demonstrated that pre-test expectancy and signal probability underlie the signal rate effect. The present results also demonstrate the effects of signal probability, and show further that they are relatively independent of task demands as defined on the speed-closure dimension. Levine et al's. results appear to be contaminated by uncontrolled variables not identified in their classification procedure. These results therefore indicate that there may be certain pitfalls in the use of task classification to integrate research findings unless methodological deficiencies and important uncontrolled variables are identified in the classification process.

Fleishman (1972) has, however, suggested that task classification may provide a profitable framework for carrying out research in any human performance area, in its potential capability of identifying gaps in the literature, and of revealing new functional relationships. The present findings that within-session decrements in sensitivity are restricted to speed tasks with high stimulation rages are in accordance with these suggestions. We shall reserve a more detailed discussion of the types of task in which a sensitivity decrement is obtained, and of the theoretical mechanisms underlying the decrement, until Chapters 12 and 13. In these chapters a specification is provided of the types of display yielding sensitivity decrements, after a consideration of the evidence from the present experiment, from Experiments 4 and 7 (in which sensitivity decrements are examined in relation to auditory and multi-source visual monitoring, respectively), and from a task classification analysis of all the studies reporting sensitivity data. At this stage, however, we shall merely note that high event rates lead to a decrement in sensitivity, but for speed tasks only. In these tasks, the signal is specified only as a change in a repetitive stimulus, so that the discrimination process may be assumed to require temporal integration of information to a greater extent than in closure tasks, where all the information needed to discriminate the signal is presented in a single event. The results of Experiment 1 alone do not allow a firm interpretation of sensitivity decrement to be made on the basis of this difference between speed and closure tasks, but in subsequent chapters we shall see that it is a crucial factor in the determination of sensitivity shifts in monitoring performance.

As far as criterion shifts are concerned, the results of Experiment 1 are in agreement with previous findings that there is an increase in the response criterion both as a result of continued work at the task, and with a reduction in SP (Baddeley and Colquhoun, 1969; Broadbent and Gregory, 1965; Williges, 1973). The present results, however, also show that an analysis of response latencies also leads to the conclusion that the criterion shifts over the monitoring period and with a change in SP. Furthermore, the results show fairly clearly that while the criterion significantly decreases with an increase in SP, it is relatively unaffected by ER, which, as we have seen, has a significant effect on sensitivity. Thus, independent changes in SP and ER lead to different effects on performance; SP primarily changes the criterion, while ER mainly affects sensitivity. These results therefore serve to clarify some of the confusing results obtained in previous investigations of the effects of these variables, where, as noted in Chapter 5, SP and ER have not been independently varied, and their effects on both criterion and sensitivity have often not been reported (but see Baddeley and Colquhoun, 1969: Loeb and Binford, 1968).

Hence, we cannot, in a sense, speak of the relative effects of SP and ER (which were considered in Chapter 5), since they influence theoretically independent aspects of performance. As far as the probability of detection is concerned however, both variables significantly affect performance, but ER appears to have a greater effect on detections than SP (for the speed task only). The substantial decrease in detection probability with an increase in ER is consistent with previous studies (Jerison and Pickett, 1964; Loeb and Binford, 1968; Metzger, Warm and Senter, 1974). A corresponding increase in SP does not lead to as great a change in detection probability. Furthermore, in a study in which the effects of SP over a number of levels were examined, Baddeley and Colquhoun (1969) found that although SP had a consistent effect on detection probability and criterion placement, an 18-fold increase in SP from .02 to .36 improved the mean detection probability by only about .17. However, it is clear from the results of Colquhoun and Baddeley (1964, 1967) and Williges (.969, 1973) that expectancy-linked variations in SP have important effects on within-session trends in performance, particularly in task situations where sensitivity is stable over time on task.

Analysis of the response latency data gave results generally consistent with those obtained using the 'discrete' measures of performance accuracy. We have also established some empirical findings not previously reported; that is, with time on task in a monitoring situation, there is an increase in the latencies of both correct and incorrect positive (Yes) responses, and a decrease in the latencies of negative (No) responses. The same trends in the latencies were observed for a reduction in SP. These results may thus be interpreted in terms of the decision theory latency model put forward in Chapter 4 as providing additional support for the view that with increased time on task or with a reduction in SP, there is an increase in the strictness of the criterion adopted by subjects, rather than any change in their ability to discriminate signals.

Under conditions in which a change in discriminative ability is obtained, however, the latency model predicts somewhat different effects on response latency; again the result that only the latencies of responses to signals were significantly affected by a change in ER is consistent with the predictions of the model. The only effect of ER on latencies was for performance on the speed task, and this again harmonizes with the discrete performance data.

The finding that there is no change in the latencies of responses to non-signals with a change in sensitivity suggests that the criterion cut-off position remains constant, when measured with respect to the mean of the underlying non-signal distribution. Analysis of the discrete false alarm probabilities showed that this was so; and hence it appears that in the situation examined in Experiment 1, it is the criterion cut-off c rather than the criterial likelihood ratio B which is held constant when discriminability varies. This result is in agreement with the suggestions of Hardy and Legge (1968) and McNicol (1972), but inconsistent with the results of Broadbent and Gregory (1963b) and Moray and O'Brien (1967), who found that B was invariant with a change in d'. It seems likely, therefore, that the specification of a bias change cannot be made by using only one parameter, and that the effects of a change in discriminability may well be different depending on the nature of the variable effecting the discriminability change. Thus in some cases, likelihood ratio may be held constant when discriminability changes, while the criterion cut-off is held constant in others. On the basis of a probability matching model, Thomas and Legge (1970) suggested that p(Yes) may be invariant with discriminability, but this prediction is clearly falsified by the results of Experiment 1.

Another problem in interpreting the effects of ER on response bias should be noted. This concerns the finding that although false alarm probability remained constant with a change in ER (from which it was inferred that c remains constant), the absolute number of false alarms increased. In this epxeriment it so happened that the increase in the number of false alarms was approximately the same as the increase in the number of non-signals, so that there was little change in false alarm probability. Loeb and Binford (1968) and Mackworth (1968, 1970) have, however, reported that false alarm probability decreases with an increase in ER (and hence that B increases). These results therefore appear to somewhat limit the generality of the finding that criterion cut-off is unchanged with a change in ER.

While the effects of ER on the response latencies were in agreement with the predictions of the latency model, there are some difficulties in interpreting within-session trends in latencies in the high ER condition. In this condition, a decrement in sensitivity and an increase in the criterion over time on task was observed. If latency is a function of the relative distance from the criterion, then we would expect that there would be a greater increase in detection latency and a greater decrease in omission latency in this condition than in the condition where only the criterion increases (the low ER condition II). Although such differential trends were discernible (see Figures 7.5 and 7.6), the interaction between ER and time blocks was not significant for either of these latency measures. However, while this result does not support the latency model, it clearly does not falsify it and point to the acceptance of some other model, such as the 'perceptual vigilance' model of Buck (1966); on the basis of this model, we would expect an increase in the latency of omission responses, since these would be associated with declining 'perceptual vigilance'. The present results, which show that the latencies of negative responses decrease with time on task, are inconsistent with this model.

Finally, let us consider a result obtained by Colquhoun (1969), which is not in agreement with the results we have obtained for detection latency. Colquhoun found that for a visual monitoring task, an increase in ER was associated with a decrease in detection latency, which is the opposite result to the one we have obtained. There are some procedural differences between his experiment and the one we have reported, such as the fact that the same subjects served under different ER levels in Colquhoun's experiment (whereas independent groups were used in our experiment). However, it would appear that the crucial factors accounting for the discrepancy relate to the event rate levels used and the instructions given to subjects. Subjects in the present experiment were instructed to respond as quickly as possible, even in the lowest ER condition I (for which latency data were not reported). It is not known whether such instructions were given in Colquhoun's experiment. It may be that subjects deliberately take longer to respond when the event rate is very low, and the inter-event interval is rather long, as in the low ER condition used by Colquhoun and in condition I in the present experiment. Examination of the latency data in this condition revealed that some subjects did indeed take longer to respond than those in condition II, but overall the mean latency was about the same as in condition II. Since the same subjects served in all conditions in Colquhoun's experiment, it is possible that the decrease in latency with increased ER reflects a change in the response set adopted by subjects.

Overall, therefore, the results of Experiment 1 provide general support for the task classification and decision theory approaches to the analysis of monitoring behaviour, and, more specifically, suggest that sensitivity decrements are obtained for speed tasks but not for closure tasks. This latter result is examined in greater depth in Experiment 2.

## 7.3 Experiment 2

Summary

Eight male subjects were tested on either a visual speed task or a visual closure task under conditions of a high stimulation rate, as in Experiment 1. Unlike Experiment 1, however, confidence ratings were used, and subjects were tested on two successive days. Results indicated that there was a reliable sensitivity decrement for the speed task only, on both days, and for responses at any confidence level. An operating characteristic analysis revealed no systematic trends in the underlying signal-non-signal distribution variance ratio with time on task. The results were interpreted as firmly establishing the principal result of Experiment 1, and discounting the effects of learning factors on sensitivity shifts.

### 7.3.1 Introduction

The principal result of the previous experiment was the finding of a sensitivity decrement for speed tasks under high stimulation rates. The TSD measure d' was used to demonstrate this decrement, although it was also apparent using other approximate operating characteristic analyses. However, in order to enable a reliable operating characteristic (OC) analysis of sensitivity shifts to be made, an experiment repeating a part of Experiment 1 with the addition of confidence ratings was planned.

An OC analysis permits, in principle, a closer examination of the sensitivity decrement in terms of the underlying signal and non-signal distributions, and to see whether there is any systematic change in the signal to non-signal variance ratio with time on task. Variations in this ratio affect d', and thus the results of Experiment 1 do not make clear whether the sensitivity decrement reflects a genuine decline in perceptual efficiency at all criterion levels, or a change in the variance ratio. In certain cases an artifactual decrement in sensitivity might be observed if d' is measured at only one criterion level (as in Experiment 1) and the variance ratio increases. This is illustrated in Figure 7.7.

Another feature of Experiment 2 is that subjects were tested on two occasions on the same task. This was carried out to examine whether the differential sensitivity trends between speed and closure tasks were related to factors such as practice and learning. It might be the case that the sensitivity decrement observed for the speed task in Experiment 1 was due to the subjects having inadequate practice at the task, although extensive training periods and expectancy matching practice sessions were given to subjects working with both the speed and the closure tasks.

### 7.3.2 Method

Subjects

Eight male subjects ranging in age from 19 to 24 years participated in the experiment. They were assigned to two equal groups. Other characteristics of these subjects are listed in Table 6.2.

Apparatus

The same speed and closure tasks which were used in Experiment 1 were utilized (see 6.2.5 for task descriptions). Unlike Experiment 1, however, a four-category rating scale was used in place of the binary response panel; the rating response panel had four response buttons labelled, from left to right "Certain Yes", "Doubtful Yes", "Doubtful No" and "Certain No" (see 6.2.2 and Figure 6.4).

For both tasks, stimuli were presented at a rate of 30/min., and signals were presented irregularly at a mean rate of 2/min. The monitoring period lasted 45/min.

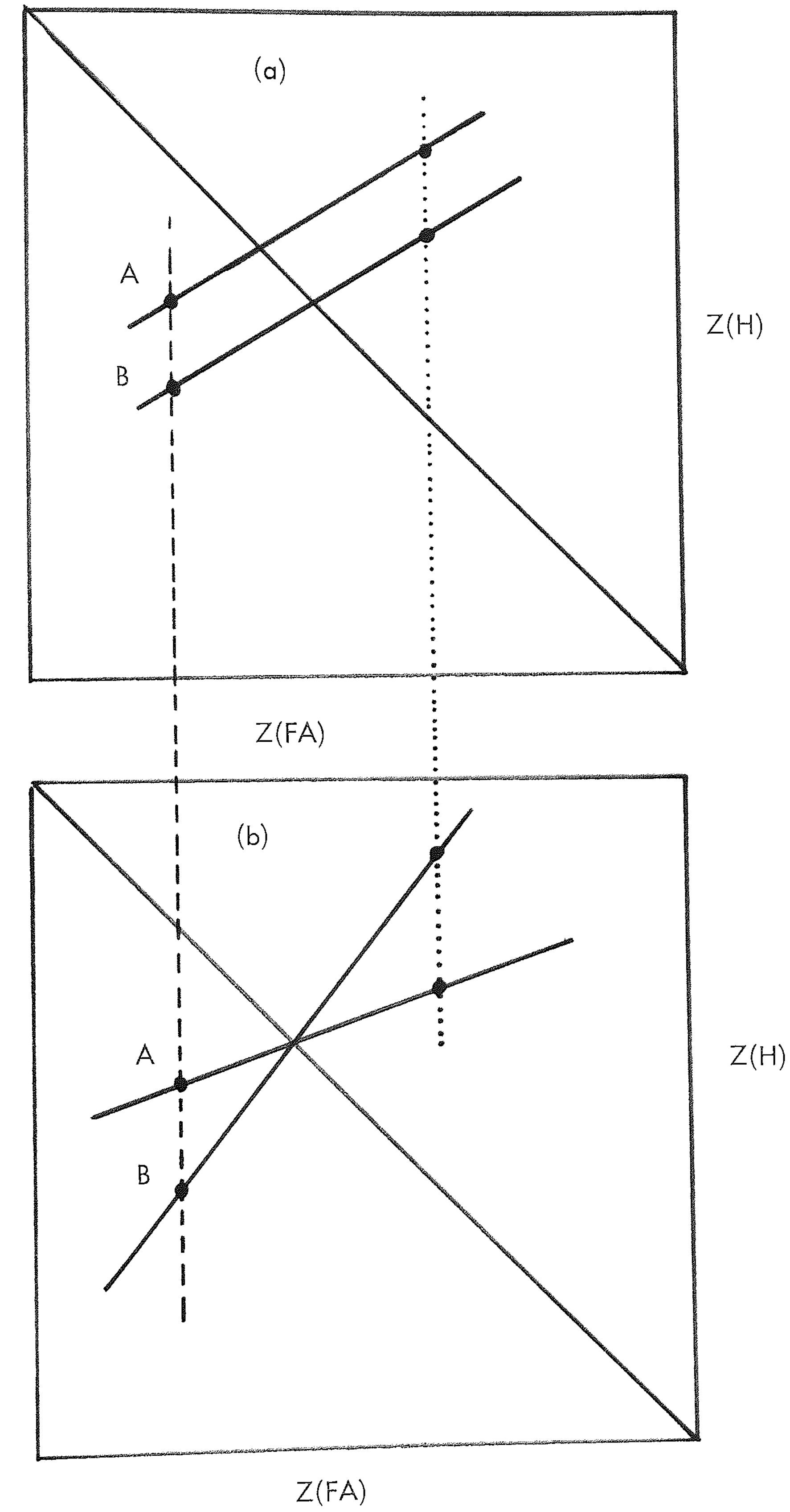


FIGURE 7.7 Illustration of how a decrement in sensitivity between two conditions A and B may be either due to a genuine reduction in detectability (a), or due to a change in signal to noise variance ratio (b). If sensitivity is measured at only criterion point (represented by dashed line) d' decreases in (a), but increases if sensitivity is measured at another point (dotted line). Note that in (b) the index d is constant.

Procedure

Half the subjects worked with the speed task, while the other half worked with the closure task. Each subject was tested at the same time on two successive days. Before each session, the same training and expectancy matching practice as described previously was given to each subject (see also 6.3).

### 7.3.3 Results

The use of a four-category rating scale in Experiment 2 enabled the derivation of three-point OCs for each subject. These were computed separately for each 15-min. time block and for both sessions, so that six OCs per subject and a total of 48 OCs were available. Twenty four of these OCs are displayed in Figures 7.8 and 7.9, for four subjects, two of whom worked with the speed task, while the other two worked with the closure task. These figures show the OCs plotted on double-probability or z co-ordinates; they were derived by assuming that the four-category rating scale creates three response criteria $c_1$, $c_2$, $c_3$ with which the cumulative Hit and FA probabilities may be partitioned to yield three pairs of operating probabilities.

The individual OCs display a fairly high degree of skew, but there appears to be no systematic trend in skewness over time blocks or between tasks. On the other hand, there is a marked difference between the forms of the speed and closure task OCs; the speed task OCs exhibit a declining sensitivity over time blocks (this is why separate lines have been fitted by eye to the OC points in successive time blocks), while the individual OCs for the closure task indicate that sensitivity is fairly stable over time blocks. For the closure task OCs the operating probabilities in successive time blocks lie closely distributed along a single line (or on separate lines of differing slope but of approximately the same detectability).

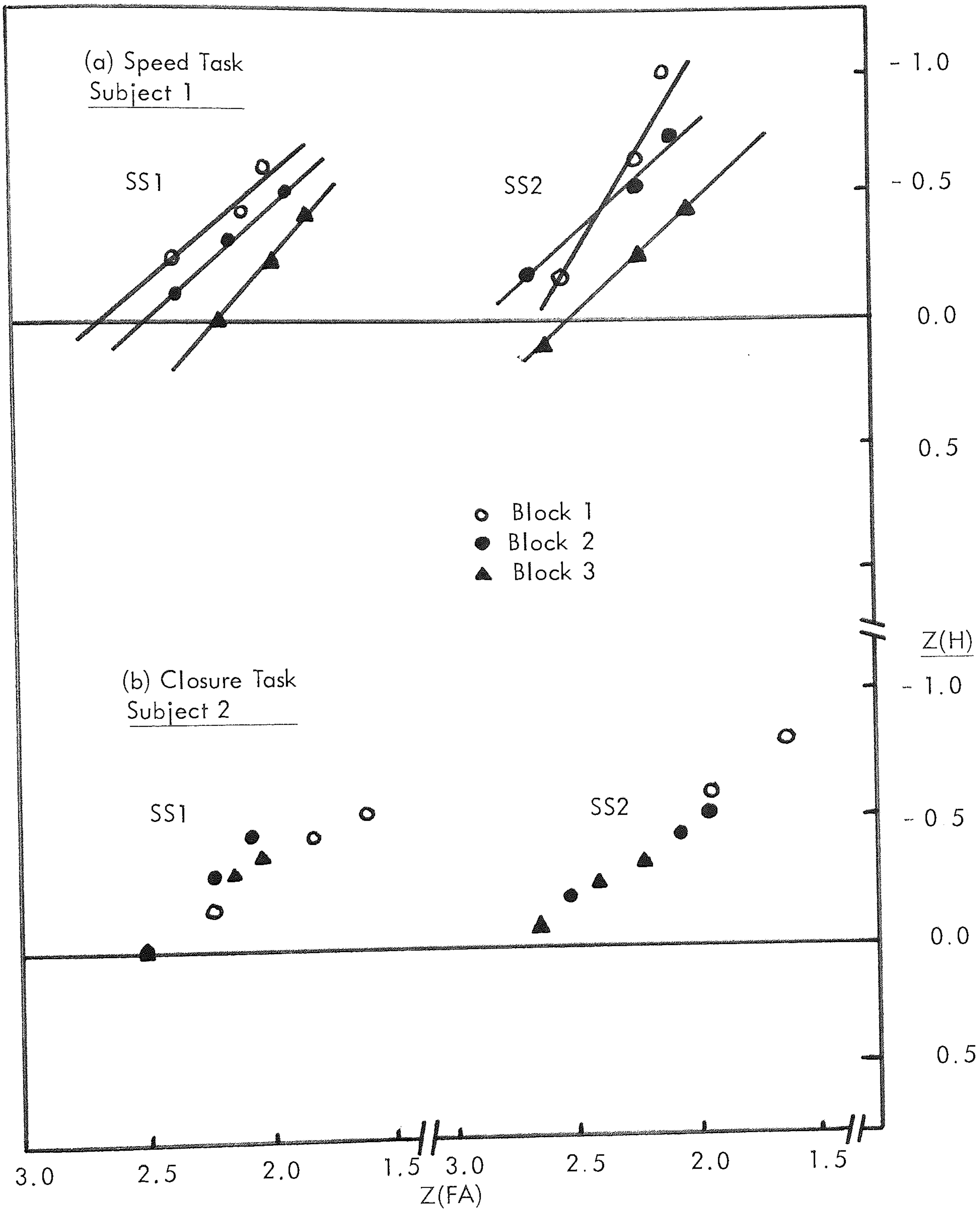


FIGURE 7.8 Individual operating characteristics for two subjects in successive time blocks and sessions (SS).

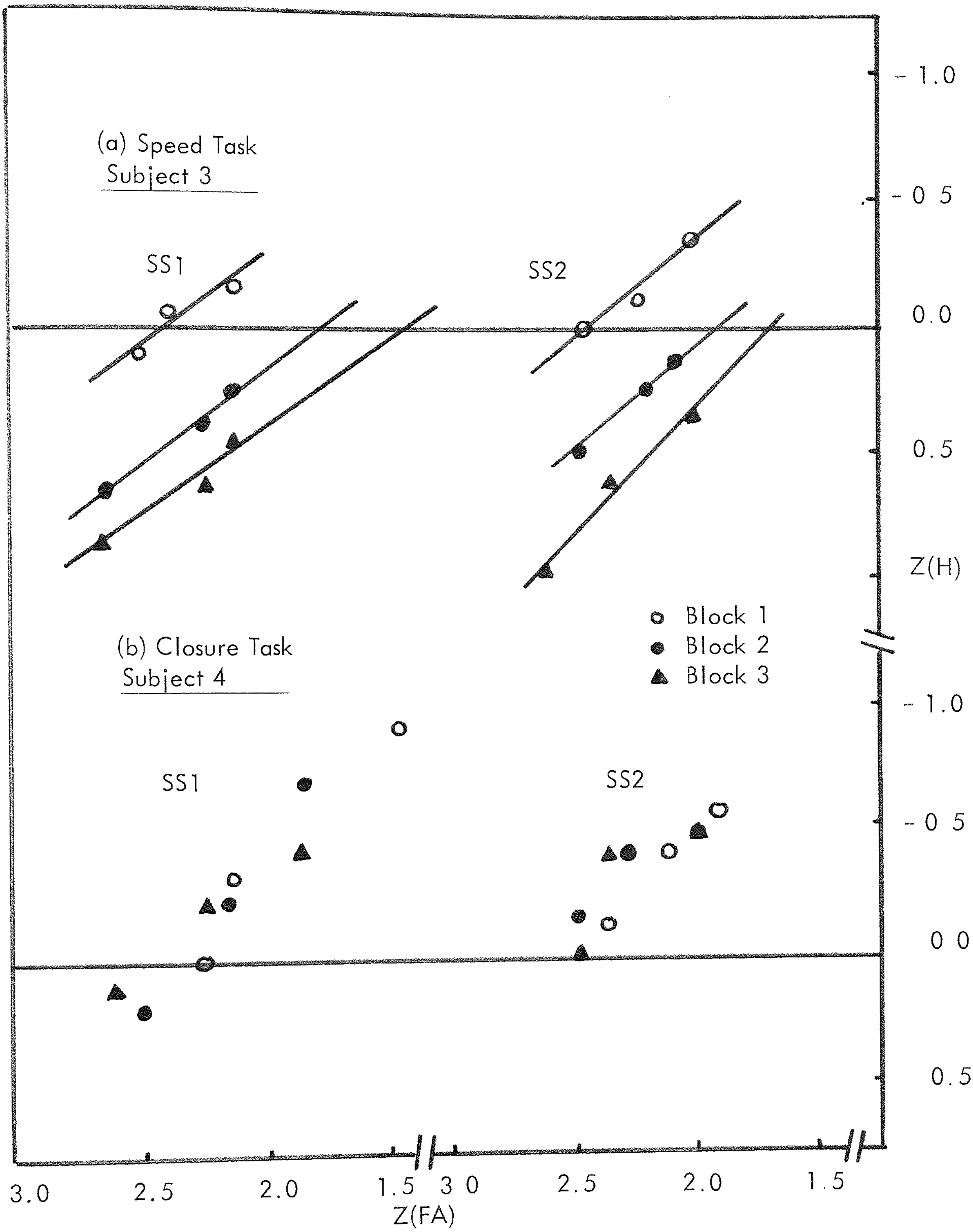


FIGURE 7.9 Individual operating characteristics as in Figure 7.8 for a further two subjects.

| Dependent Measure | Speed Task Session 1 | | | Speed Task Session 2 | | | Closure Task Session 1 | | | Closure Task Session 2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $d_a$ | 2.34 | 2.09 | 1.76 | 2.50 | 2.34 | 1.98 | 2.37 | 2.43 | 2.39 | 2.50 | 2.41 | 2.49 |
| log B1 | 1.12 | 1.17 | 1.15 | 1.14 | 1.22 | 1.29 | 1.03 | 1.13 | 1.31 | 1.21 | 1.30 | 1.38 |
| log B2 | 0.89 | 0.90 | 0.94 | 0.89 | 0.98 | 1.12 | 0.70 | 0.85 | 0.98 | 0.85 | 0.92 | 1.12 |
| log B3 | 0.70 | 0.74 | 0.77 | 0.64 | 0.74 | 0.83 | 0.38 | 0.59 | 0.67 | 0.59 | 0.67 | 0.76 |

TABLE 7.9 Mean values of sensitivity ($d_a$) and log likelihood ratio at each criterion level (log B) in successive time blocks, and in both sessions and tasks.

The form of these OCs therefore clearly indicates that a sensitivity decrement is obtained only for the speed task. The decrement is maintained in a subsequent session. The only effects of repeated testing on the OCs seem to be a slight increase in the 'strict' criteria and, possibly, an increase in detectability. As far as the sensitivity decrement over time blocks is concerned, however, repeated testing did not influence the form of the decrement for any of the subjects tested.

Sensitivity was quantified by taking values of $d_a$ in each time block and for each subject. The rationale behind the use of this measure was explained in Chapter 4. The mean $d_a$ values in each time block and for both sessions and tasks are tabulated in Table 7.9. This table also gives the log B values associated with each criterion $c_1$, $c_2$ and $c_3$ (corresponding to the strict, medium and lax criterion, respectively). Each of these measures were analysed in a 2 x 2 x 3 analysis of variance (Tasks x Sessions x Time Blocks), with 'repeated measures' on the latter two factors.

The summary table for the ANOVA for $d_a$ is given in Table 7.10. Two reliable effects were obtained, Time Blocks, and Tasks x Time Blocks. Tests for the simple effects of the Time Blocks factor (see bottom half of table) showed that there was a reliable decrement in $d_a$ with time blocks on the speed task only ($p < .01$). The effects due to Sessions were marginally reliable ($p < .10$), indicating that there was a small increase in mean sensitivity over sessions in both tasks.

The summary tables for the ANOVAs for the log B measures are listed in Appendix C2, in Tables C2.1 to C2.3. There was a significant increase in all the criterion values with time on task. In addition, there was a significant increase in log B2 with sessions ($p < .05$), but similar increases in the other two criterion measures were only marginally significant ($p < .10$). There were no other significant effects.

Finally, in order to examine whether there were any consistent trends in OC skew, the ratio of the signal to non-signal standard deviations (taken as the inverse of the OC slope) was analysed in the same way as the other measures. The summary table for the ANOVA is given in Table C2.4 in Appendix C2. No significant effects due to any factor emerged.

### 7.3.4 Discussion

The results of Experiment 2 confirm the previous findings that under a high stimulation rate, a decrement in sensitivity over the monitoring period is observed in speed tasks alone, and not in closure tasks. The use of an OC analysis clearly demonstrates that subjects exhibit a loss in perceptual efficiency over time in a speed task, irrespective of the criterion level at which they work. Thus, although the obtained OCs were skewed in the direction of greater signal variance, as suggested by Green and Swets (1966) and Mackworth (1970), there was no systematic trend in the degree of skew. Hence the type of artifact resulting from measuring d' at a single criterion level, which was shown in Figure 7.7, cannot be taken to explain

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.83741 | 1 | 0.83741 | 3.7668 | ns |
| SWG | 1.33389 | 6 | 0.22232 | | |
| B (Sessions) | 0.24368 | 1 | 0.24368 | 5.0855 | ns(.10) |
| AB | 0.05880 | 1 | 0.05880 | 1.2272 | ns |
| B x SWG | 0.28749 | 6 | 0.04792 | | |
| C (Time Blocks) | 0.60140 | 2 | 0.30070 | 7.0177 | .05 |
| AC | 0.65382 | 2 | 0.32691 | 7.6294 | .05 |
| C x SWG | 0.51418 | 12 | 0.04285 | | |
| BC | 0.00420 | 2 | 0.00210 | 0.0798 | ns |
| ABC | 0.02795 | 2 | 0.01398 | 0.5307 | ns |
| BC x SWG | 0.31598 | 12 | 0.02633 | | |
| Simple effects | | | | | |
| C at A1 (Speed Task) | 1.25520 | 2 | 0.62760 | 14.6469 | .01 |
| C at A2 (Closure Task) | 0.05501 | 2 | 0.02750 | 0.6418 | ns |
| C x SWG | 0.52418 | 12 | 0.04285 | | |

TABLE 7.10 Analysis of variance for $d_a$.

the sensitivity decrement observed in Experiment 1. The emergence of the same differential within-session sensitivity shift between speed and closure tasks in a pilot study, in Experiment 1 and in the present study serves to establish the reliability of this finding. These results also suggest a possible empirical basis for the hypothesized differentiation between the speed and closure categories, which was originally proposed by Levine, Romashko and Fleishman (1971).

The results of Experiment 2 show, furthermore, that repeated testing did not eliminate the within-session decrement found for the speed task, although there was a marginal improvement in sensitivity over sessions. This result is consistent with our observation that the extensive training and expectancy matching procedures used were sufficient to minimize the effects of practice at the task (see Appendix B). The sensitivity decrement thus does not appear to be significantly influenced by learning factors, although it may be argued that the use of only two sessions in the present experiment is not sufficient for the effects of such factors to be observed. However, the results of Binford and Loeb (1966) can be taken to support the present finding. These authors used an auditory 'high-rate' speed task analogous to the visual speed task used here, and found that although sensitivity improved over sessions, a with-session sensitivity decrement was observed in the later sessions of an eight-session testing cycle.

Binford and Loeb (1966) also found that there was an increase in the criterion over sessions; a marginal increment in log B was also observed in the present experiment. Moreover, an increment over time on task was observed at all three criterion levels. This result is inconsistent with the result of Broadbent and Gregory (1963a, 1965), who found that lax criterion values did not change significantly over the monitoring period. On the other hand, Milosevic (1974, 1975) has recently reported time-related increases in criteria associated with risky responses. It would appear that criterion movement is dependent on the particular values of criteria chosen by subjects and on whether a sensitivity decrement over the monitoring period is observed. There is some evidence that when criterion values are fairly high, as in the present experiment, increments in the criterion values over time on task may be obtained even for so-called risky or lax criteria (Broadbent, 1971).

## 7.4 Conclusions

The results of Experiments 1 and 2 may be interpreted as providing general support for the twin approaches to the analysis and evaluation of monitoring performance adopted in this thesis, that is, task classification and decision theory. We may, however, list some more specific conclusions in relation to the questions posed in the introduction to this chapter:

1) There is a reliable sensitivity decrement over the monitoring period

when the stimulation rate in the task is high. The decrement is apparent at all criterial levels, and is not significantly influenced by changes in the signal to non-signal variance ratio, which exhibits no systematic trend over the monitoring period.

2) The decrement in sensitivity at high event rates is obtained only for speed tasks, in which the critical signal demands a detection of a change in a repetitive stimulus. The decrement for this type of task is not eliminated if subjects serve in a second monitoring session.

3) There is an increase in the stringency of the response criteria with increased time at work and with a reduction in signal probability. However, these changes do not appear to be related to task demands as defined on the speed-closure dimension.

4) Independent variations in signal probability and event rate lead to independent variations in performance; the former variable mainly affects the response criterion, while the latter primarily affects sensitivity.

5) With time on task, there is a significant increase in the latencies of positive responses and a significant decrease in the latencies of negative responses. An interpretation of response latencies within a decision theory latency model is consistent with the data. Variations in response latency with a change in criterion, and, to a lesser extent, with a change in sensitivity, are in agreement with the predictions of a latency model which assumes an inverse relation between latency and the relative strength of the evidence pointing to or against the presence of a signal.

# CHAPTER EIGHT

## CONSISTENCY OF INDIVIDUAL DIFFERENCES IN VISUAL MONITORING

## 8.1 Introduction

The results of Experiments 1 and 2 led to the general conclusion that when monitoring tasks are classified according to the task dimensions of a task classification system, the impact of an independent variable on performance is a function of the task type or category. These results therefore suggest that group performance functions are consistent between tasks classified in similar categories of the task classification system.

A further, more direct test of the validity of the task classification dimensions would be to examine the consistency of individual performances on tasks classified either similarly or differently on one dimension of the task taxonomy. For example, if the hypothesised feature differences, or differences in 'ability requirements' between speed and closure tasks are to be considered important, a greater consistency (higher correlation) in individual differences in performance should be observed for pairs of similarly classified tasks than for differently classified tasks. Such an investigation is required if an empirical basis for the proposed task classification dimensions is to be established.

A further impetus for an investigation of the consistency of individual differences in different monitoring tasks within a task classification framework stems from the observation that in previous studies of inter-task correlations, several inconsistent results have been obtained, and that these inconsistencies may be related to task factors. In a review of these studies in Chapter 3 (see 3.2), we noted that the early view that individual differences in monitoring performance are highly task specific (Buckner, Harabedian and McGrath, 1960) could be challenged on the basis of more recent results (Hatfield and Soderquist, 1970; Tyler, Waag and Halcomb, 1972; see also 3.2). These more recent studies have controlled task factors more closely by equating tasks for the difficulty and type of signal discrimination. It thus seems likely that tasks which exert similar demands are more likely to share common performance variance than tasks which do not, although none of the existing studies provide a direct confirmation of this view. In this chapter we report an experiment which examines this suggestion with regard to the speed-closure dimension of the task classification system.

## 8.2 Experiment 3

### Summary

Three groups of subjects were tested on pairs of visual monitoring tasks classified according to the speed-closure dimension of the task classification system. Groups 1 and 2 worked with two speed and two closure tasks respectively, while Group 3 worked with a speed and a closure task. Results indicated that individual differences were highly consistent for performances on similarly classified tasks (Groups 1 and 2), despite dissimilarities in displays, but significantly less so for tasks classified across the speed-closure dimension (Group 3). The results point to the importance of the task classification approach for the evaluation of the consistency of individual differences in monitoring tasks, and may be interpreted as establishing an empirical basis for the speed-closure dimension.

### 8.2.1 Method

This experiment reports an investigation into the consistency of individual differences in the performance of visual monitoring tasks classified according to the speed-closure dimension. Two tasks in each category were used, the visual speed and closure tasks from Experiment 1, and a further two tasks used in the pilot study to Experiment 1. An experimental design was utilized in which the possible effects of testing sequence could be examined.

Subjects

Thirty male subjects, aged 17 to 26 years, were invited to participate in the experiment. They were randomly assigned to one of three equal groups. Table 6.2 gives some other characteristics of the subject sample in this experiment.

Apparatus

Details of the aparatus and the experimental situation are given in Chapter 6 (section 6.2). Four monitoring tasks were used. In each task the same 'single' response mode was used, and there was no visual search requirement.

Visual speed task 1 (VS1): In this task subjects were required to detect a decrease in the intensity of an intermittent circular light source. The decrease in intensity that defined a signal was .05 log ft.L. The light flashes were presented to the right eye at an approximate angle of $2^{\circ}$.

Visual speed task 2 (VS2): Slide presentation was used in this task, the critical signal for detection being a decrease in the horizontal separation of two vertical lines each 10cm. long when projected. The decrease in separation that defined a signal was 1cm. Subjects sat 1m. from the display.

Visual closure task 1 (VC1): This task used flash presentation as in VS1, but the signal for detection was a central pink spot appearing occasionally within a light flash.

Visual closure task 2 (VC2): In this task the same mode of presentation as in VS2 was used, but the critical signal was a central 0.5cm. gap in both lines. In other respects the task was similar to VS2.

More detailed descriptions of these tasks are given in 6.2.5. In each task, the event rate was 15/min, and signals were presented irregularly at a mean rate of 1/min. Each monitoring session lasted 45 min. In a pilot study and on the basis of previous work, task difficulty was adjusted to give approximately equal d' values for each task.

Procedure

Each group of subjects worked with a different pair of tasks, as follows:-

Group 1 : VS1 and VS2

Group 2 : VC1 and VC2

Group 3 : VS1 and VC1

Hence groups 1 and 2 worked with two speed and two closure tasks respectively, while group 3 worked with differently classified tasks. Within each group half the subjects performed the tasks in the order shown (sub-group A), while the other half performed them in the reverse order (sub-group B). Subjects worked at the tasks at roughly the same time of day of two successive weeks. Before each monitoring session subjects were trained thoroughly and given a 10-min. expectancy matching task which accurately sampled the features of the task used in the main session (see also 6.3).

Experimental design

The design utilized in this experiment was particularly chosen so that possible effects due to repeated testing could be evaluated. With this design, three correlation coefficients for each dependent measure may be computed, for the three groups. A more efficient design from the point of view of the correlations would be to test the same subjects on all four tasks, since this would give more information on the inter-task correlations. With such a design however, a larger subject sample would be required in order to test for all possible sequence effects (at least 2 x 4! = 48). An alternative would be to take a sample of the 24 possible sequences of testing, as in a counter-balanced design. However, even if 'digram counterbalancing' (Wagenaar, 1969) were used, sequence effects would only be partially accounted for. For both the latter two designs, extraction of unambiguous information regarding performance correlation would be difficult if interactions between tasks and sequence of testing were obtained. Such effects, if obtained, are easier to interpret in the two-task design we have adopted.

### 8.2.2 Results

Five performance measures were utilized in the data analysis. These were the mean probability of correct detections (Hits), the mean probability of false alarms (FAs), and the derived theory of signal detectability (TSD) indices d' and B. In addition, an average measure of detection latency (DL) was also used.

Performance data for the three groups are displayed in Table 8.1 as a function of time on task (15-min. blocks) and task category. An initial analysis of the data was carried out by analysing each dependent measure in a three factor analysis of variance (ANOVA), for each group separately. The ANOVA factors were Testing Sequence, Tasks and Time Blocks, with 'repeated measures' on the last 2 factors. The complete summary tables for the 15 ANOVAs are listed in Appendix C3 in Tables C3.1 to C3.15.

For Group 1, the only significant effect for each measure was for the Time Blocks factor, there being a significant decline with time on task in Hits ($p < .001$) and FAs ($p < .001$) and a significant increase in log B ($p < .01$) and DL ($p < .01$); there was no significant change in d' over time. There were no significant effects associated with the Sequence and Tasks factors.

Essentially the same results were obtained for Group 2, who also worked with two similarly classified tasks. The same performance trends as for Group 1 were obtained, with Hits ($p < .01$) and FAs ($p < .05$) declining, and log B ($p < .01$) and DL ($p < .05$) increasing with time on task; d' was

| | GROUP 1 | | GROUP 2 | | GROUP 3 | |
|---|---|---|---|---|---|---|
| | VS1 | VS2 | VC1 | VC2 | VS1 | VC1 |
| Dependent Measure | | | | | | |
| p(H) | .827 .747 .680<br>(.751) | .807 .767 .720<br>(.764) | .793 .700 .673<br>(.722) | .787 .733 .673<br>(.731) | .807 .760 .707<br>(.758) | .760 .720 .733<br>(.738) |
| p(FA) | .055 .037 .027<br>(.040) | .057 .042 .025<br>(.041) | .054 .041 .026<br>(.037) | .051 .037 .031<br>(.040) | .048 .036 .033<br>(.039) | .044 .039 .032<br>(.038) |
| d' | 2.59 2.50 2.50<br>(.253) | 2.52 2.47 2.63<br>(2.54) | 2.51 2.42 2.46<br>(2.47) | 2.49 2.49 2.41<br>(2.46) | 2.58 2.53 2.45<br>(2.52) | 2.49 2.40 2.55<br>(2.48) |
| log B | .350 .590 .821<br>(.587) | .368 .547 .784<br>(.566) | .405 .690 .765<br>(.620) | .430 .666 .761<br>(.619) | .404 .610 .664<br>(.559) | .487 .613 .665<br>(.588) |
| DL (msec) | 618 660 694<br>(657) | 627 658 696<br>(660) | 615 642 667<br>(641) | 644 674 712<br>(677) | 654 676 713<br>(681) | 632 647 692<br>(657) |

TABLE 8.1 The mean probabilities of Hits, p(H), and false alarms, p(FA), mean values of d' and log B, and mean detection latency (DL) as a function of successive time blocks, for each task within groups (see text; figures within brackets indicate mean performance levels averaged over time blocks).

invariant over time blocks. As for Group 1, no sequence or interaction effects were obtained. However, a significant effect of Tasks was obtained for the DL measure; subjects took longer, on average, to respond to signals in the VC2 tasks than in the VC1 task ($p < .05$).

Group 3 differed from Groups 1 and 2 in that they worked with differently classified tasks. However, performance trends over time were the same as before, there being a significant decrement in Hits ($p < .05$) and FAs ($p < .05$), a significant increase in log B ($p < .01$) and DL ($p < .001$), and no significant change in d' with time on task. There were no sequence, task or interaction effects present.

These results may thus be summarized by saying that no effects attributable to testing sequence were obtained in any of the Groups, and that the performance trends over time were as expected, there being no change in sensitivity, while the response criterion increased with time on task.

In the further analysis of the data, product moment correlations were computed between mean performance levels (averaged over the monitoring period) for task pairs within Groups. These are displayed in Table 8.2 for the five dependent measures and for the three Groups. Correlations were also computed between corresponding performances in each 15-min. subperiod of the monitoring sessions. Correlations for only the Hits, FAs and d' measures are given in Figure 8.1, since, as we shall see, only these measures reflected the effects of task factors.

Table 8.2 shows that uniformly high and significant correlations were obtained for all performance measures for Groups 1 and 2, who worked with similarly classified tasks. On the other hand, for Group 3, who worked on a speed and a closure task, the inter-task correlations were low and not significantly different from zero. However log B, and to a lesser extent, DL, were correlated across tasks in all three groups, thus indicating that these aspects of performance are apparently independent of the speed-closure classification. As far as the detectability of the signal is concerned, however, and for the probability of Hits and FAs, individual differences were consistent only for tasks equated on the speed-closure dimension.

Given the finding that the inter-task correlations are high and significantly different from zero for Groups 1 and 2, but not for Group 3, it is natural to examine whether the Group 3 correlations are significantly different from the Group 1 and 2 correlations. There is no exact way of doing this, but Fisher's r to Z transformation may be used as a rough test. This test can be used to examine whether the correlations obtained from independent samples N1 and N2 of two populations are significantly different from each other (see Hays, 1963). The results of such an analysis, comparing Group 3 correlations with those of Groups 1 and 2 are shown in Table 8.3 in the form $Z_p - Z_q$, corresponding to transformations of the coefficients $r_p$ and $r_q$. Comparisons were restricted to the three measures which reflect the influence of task factors, namely Hits, d' and FAs. Three of the six comparisons in Table 8.3 are significant at the .05 level, one at the .1 level and two are not significant ($p < .2$). The results of this analysis thus do not entirely satisfy the hypothesis that the Group 1 and 2 correlations are significantly higher than those in Group 3. Fisher's test, however, rests on the assumption that the bivariate distribution of the pairs of scores comprising a correlation coefficient approximates a Gaussian distribution. The test also assumes the normality of the Z sampling distribution for low sample sizes. Neither of these assumptions were checked rigorously, but an inspection of the bivariate distribution of scores showed that the normality assumption was unlikely to be met. Furthermore, the formula given by Fisher

| Dependent Measure | Group 1 (VS1 & VS2) | Group 2 (VC1 & VC2) | Group 3 (VS1 & VC1) |
|---|---|---|---|
| Hits | .85*** | .61* | .29 |
| FAs | .85*** | .73** | .25 |
| d' | .70* | .85*** | .37 |
| log B | .89*** | .66* | .70* |
| DL | .67* | .62* | .53 |

TABLE 8.2 Inter-task correlations for five dependent measures and for each group (see text; * $p < .05$; ** $p < .01$; *** $p < .001$).

| Dependent Measure | Z1 - Z3 (Group 1) | Z2 - Z3 (Group 2) |
|---|---|---|
| Hits | .96* | .41 |
| d' | .48 | .87* |
| FAs | 1.00* | .69+ |

(Standard error of difference = .5345)

TABLE 8.3 Differences in Fisher's Z values, comparing Group 3 (Z3) correlations with either Group 1 (Z1) or Group 2 (Z2) for three dependent measures (* $p < .05$; + $< .1$).

for the standard error of the Z difference scores is an approximation for large sample sizes, and the relatively small sample size used here (N = 10) probably gives an overestimate of the standard error.

The difference between the Group 3 and the Groups 1 and 2 correlations can also be illustrated by examining the individual scores of subjects in each group. Figure 8.2 displays the individual rank profiles for the d' measure for Groups 1, 2 and 3. It is immediately apparent that individual performances are less consistent in Group 3 than in Groups 1 and 2.

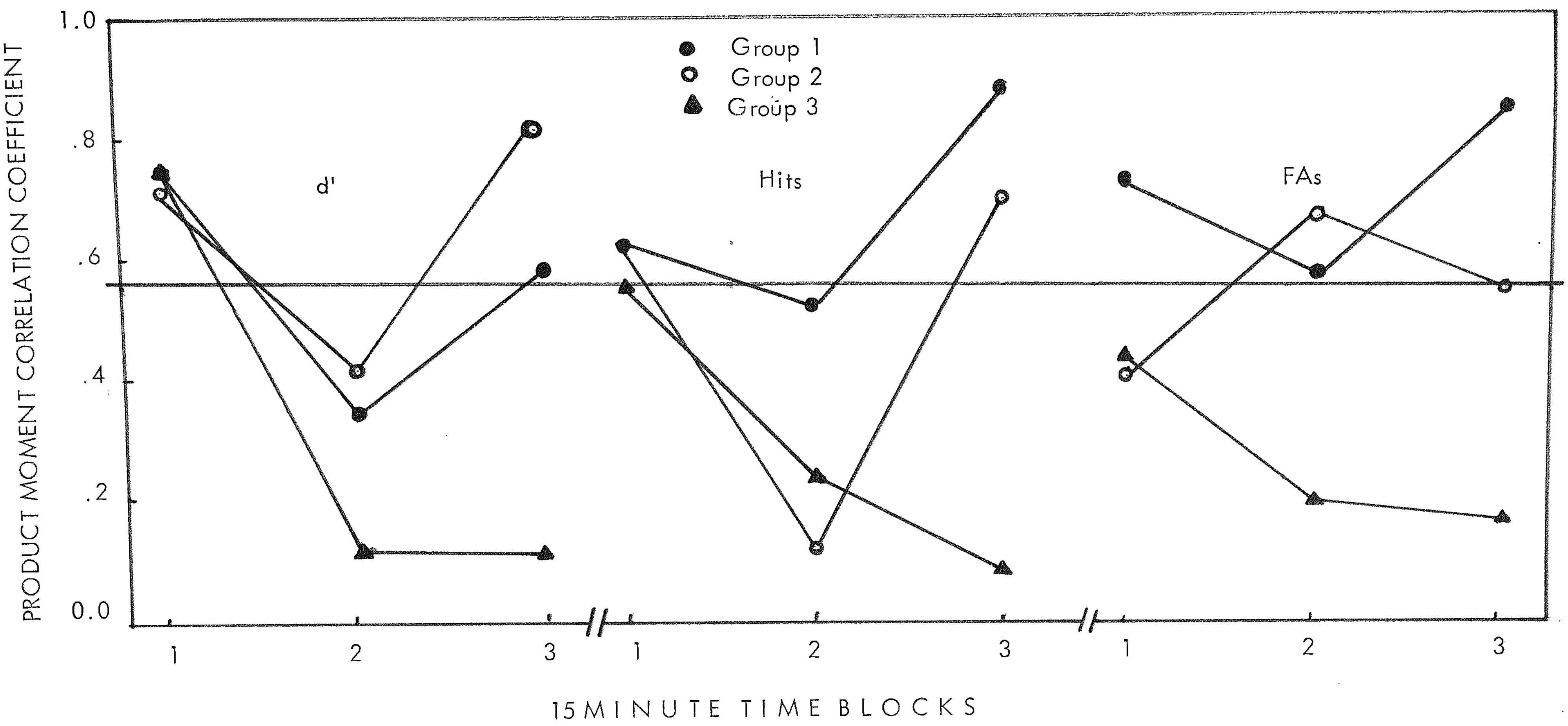


FIGURE 8.1 Inter-task correlations of three performance measures as a function of corresponding time blocks and groups (see text; correlations above horizontal midline are significantly greater than 0, $p < .05$).

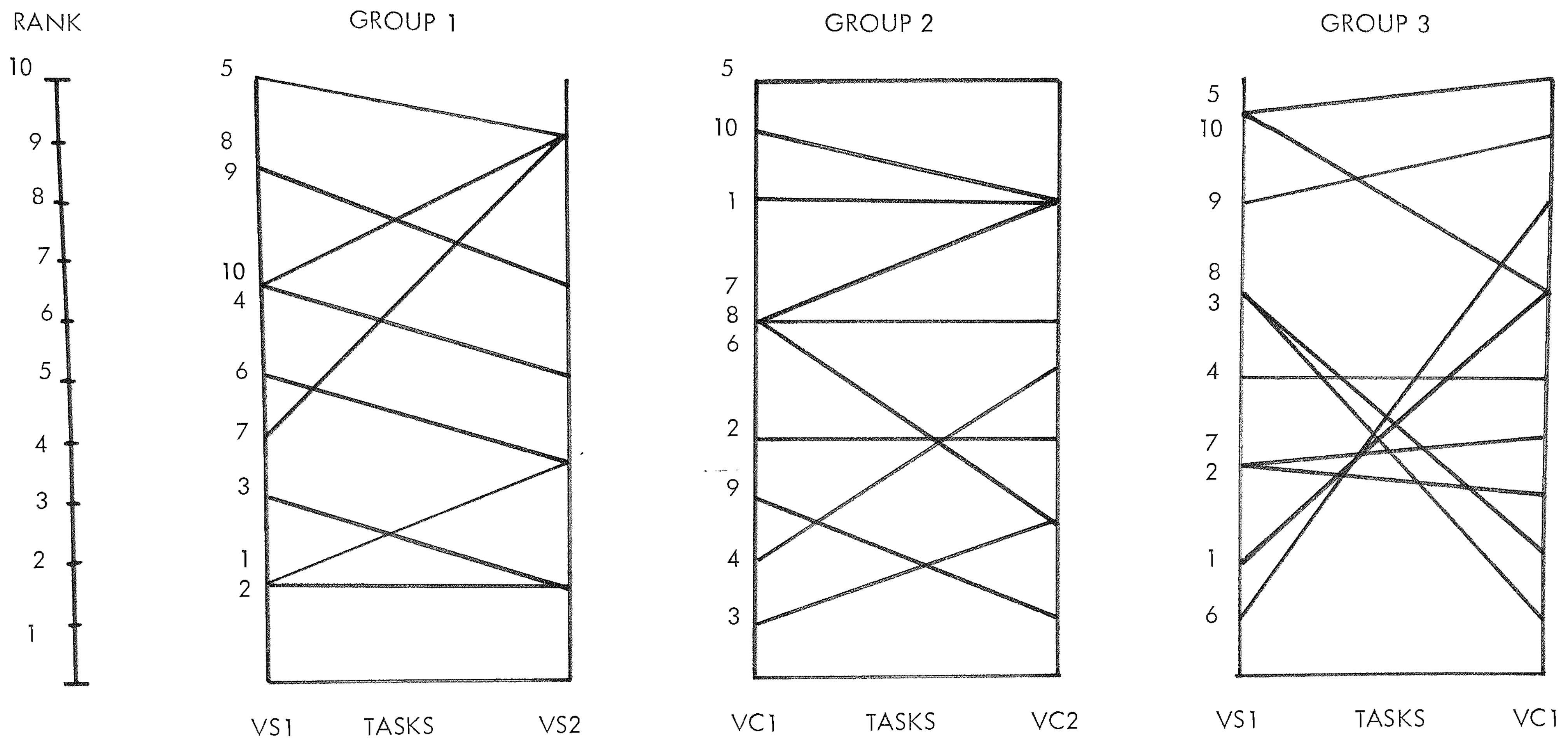


FIGURE 8.2 Rank profiles of mean d' values for individual subjects within groups (high rank signifies high d'; numbers beside profiles refer to individual subjects within groups, ten subjects per group; subjects 1 to 5 comprise subgroup A, while subjects 6 to 10 comprise subgroup B, see text).

Nevertheless, the profiles for Group 3 indicate that, as for Groups 1 and 2, the best two or three performers on one task are also among the best three on the second task. This suggests that these subjects are able to meet the 'ability requirements' of both speed and closure tasks. On the whole, however, individual performances are not highly correlated across tasks for Group 3.

Finally, Figure 8.1 displays correlation coefficients for the Hits, FAs and d' measures as a function of time blocks and groups. The general indication is that the performance correlations for Groups 1 and 2 were highest in the last 15 minutes of monitoring, and lowest for Group 3. In the first 15-min. time block, the correlations were moderately high for each group. Apparently, the divergence in the last 15-min. time block indicates that the dependence of the inter-task correlations on task classification categories is greatest during the later stages of a monitoring session.

### 8.2.3 Discussion

The results of Experiment 3 show that for visual monitoring tasks equated for detectability, uniformly high and significant intramodal performance correlations are obtained when tasks are equated on either category of the speed-closure dimension, but not otherwise. It is therefore apparent that the consistency of individual differences in performance on different visual tasks is strongly dependent on task factors as defined on the speed-closure dimension. Fleishman (1972) and Levine, Romashko and Fleishman (1971) have suggested that consistency in group performance functions between tasks might only be obtained if tasks are matched on certain categories of a task classification system. The results of Experiment 3 confirm and extend this proposal to the consistency of individual differences in performance.

The finding of high correlations for tasks equated for d' is in agreement with previous results (e.g. Hatfield and Soderquist, 1970; Tyler et al., 1972), but the finding of low inter-task correlations for dissimilarly classified tasks has not been demonstrated before. It was previously suggested that low correlations have been obtained in some early studies because task factors had not been controlled and dissimilar tasks had been used (e.g. Baker, 1963). The present results therefore provide support for this suggestion. However, they do not point to the importance of task similarity _per se_, but to task similarity on specified task dimensions. A closer examination of the tasks used in the present study reveals that high correlations were obtained for task pairs which are physically dissimilar but similarly classified (VS1 and VS2 in Group 1 and VC1 and VC2 in Group 2). VS1 requires detection of signals in a series of intermittent light flashes, while VS2 requires detection of signals in a series of slide presentations, but in both tasks the critical signals are such that a _change_ in the stimulus event has to be detected. VC1 and VC2 are also physically dissimilar in that VC1 uses flash presentation while VC2 uses slide presentation, but both tasks are classified as closure tasks, in which signal and non-signal stimuli are presented within the same event. On the other hand, Group 3 worked with two physically similar but differently classified tasks VS1 and VC1, both of which used flash presentation. Yet it is for this group that low inter-task correlations were obtained, while for Groups 1 and 2, who worked with physically dissimilar but dimensionally equivalent tasks, high correlations were obtained. Hence it can be concluded that task type factors are more important than task similarity _per se_, and that individual differences in monitoring performance are not so much task specific (Buckner and McGrath, 1963b), as _task type specific_.

The results therefore demonstrate that while about 50-70% of the variance in performance sensitivity is shared between tasks equated for task demands on the speed-closure dimension, there is only about 14% common variance for tasks which are not matched for their demand characteristics. These results also apply to the task correlations in sub-sections of the monitoring period. We have noted however, that the divergence between the correlations for matched tasks and those for unmatched tasks is greatest in the latter half of the monitoring period. This result is consistent with the finding of Experiment 1 that the difference between performance on a speed task and that of a closure task under high event rate conditions is greatest in the later stages of the monitoring session. We may tentatively conclude, therefore, that the influence of the speed-closure dimension on group and individual performance levels is particularly significant only for extended time periods, that is, for most monitoring situations.

The finding of no significant order effects in this study is probably related to the extensive training and expectancy matching procedures used, and to the fact that testing was repeated at the same time and same day of successive weeks. In several previous studies investigating inter-task correlations in monitoring performance, training periods have been short, and subjects have not generally received expectancy matching pre-task practice, which has been shown to markedly influence performance in the main monitoring session (Colquhoun and Baddeley, 1964, 1967). In at least one study (Tyler et al., 1972), subjects were required to work in two monitoring sessions separated by only a short rest period (3 min.), resulting in particularly severe order and transfer effects. Furthermore, Binford and Loeb (1966) have found that Hit and FA scores stabilize only after about four monitoring sessions. These results suggest that the amount of training is an important factor in determining subsequent performance, and therefore, indirectly, in determining the inter-task performance correlations. In the experiment we have reported however, with the benefit of extensive pre-task practice and weekly testing, there was no evidence of contamination due to practice variables (see also Experiment 8 in Appendix B, which examines the effects of practice on some of the tasks used in Experiment 3, and which also gives reliability values which may be compared with the performance correlations obtained in Experiment 3).

Of the five performance measures utilized in Experiment 3, log B was the only one for which significant inter-task correlations were obtained for all three experimental groups. This result is in general agreement with the results of previous studies (e.g. Gunn and Loeb, 1967; Loeb and Binford, 1971; see also Type B studies in Chapter 3, section 3.2). It therefore appears that subjects' response criteria are consistently maintained in different monitoring tasks, irrespective of task demands as defined on the speed-closure dimension. This interpretation is consistent with the result from Experiment 1 that response criteria are not reliably different for speed and closure tasks, but must be tempered by the fact that the log B measure is very sensitive to deviations from the equal-variance TSD model, which we have used to compute values of B.

There are no difficulties, on the other hand, with the interpretation of the results for the other performance measures. Overall, they point to the importance of the speed-closure dimension for the determination of performance consistency in different visual monitoring tasks.

## 8.3 Conclusions

Individual differences in monitoring performance on different visual tasks are highly consistent when tasks are matched on either category of the speed-closure dimension of the task classification system, but not otherwise. These results establish an empirical basis for the speed-closure dimension, and endorse the task classification approach to describing both group and individual levels of performance on different monitoring tasks.

# C H A P T E R   N I N E

## TASK CLASSIFICATION AND DECISION PROCESSES IN AUDITORY MONITORING

## 9.1 Introduction

A dominant theme present in the experiments reported thus far has been the interactive influence on performance of task demands, as identified on the speed-closure task dimension. However, in all three experiments reported, the modality dimension of the task classification system was not considered, only visual monitoring tasks being used. The experiment reported in this chapter is a further investigation of the interaction between task categorization and decision processes in monitoring behaviour, but with reference to auditory monitoring.

Since modality also forms part of the task classification system, such an experiment is also of interest from the point of view of the relative demands of visual and auditory monitoring tasks. A specific result established in Experiments 1 and 2 is that visual detection sensitivity is a function of task demands defined by the speed and closure categories. Since visual monitoring may impose certain additional ('observing response' or 'coupling') demands on the observer's limited capacity system, there is a case for further investigating this result for auditory tasks, where, other things being equal, it has been proposed that such demands have a reduced influence (Elliott, 1960; Jerison, 1970).

The experiment reported in this chapter was also carried out with a view to further examining the relationship between response latency and the response criterion, which was observed in the results of Experiment 1. These provided preliminary evidence for the decision theory model for response latencies in monitoring. A more rigorous test of the reliability and generalizability of such a model would be to examine response latency data in relation to confidence rating criteria. Such data is also required to establish whether a fixed-sample model of latency is adequate to interpret latency data in monitoring tasks, or whether models incorporating input sequential sampling mechanisms such as the various random-walk models (Audley, 1973) need to be considered.

## 9.2 Experiment 4

Summary

Six groups of subjects were tested in 45-min. monitoring sessions on either a speed or a closure auditory task under one of three conditions independently combining signal probability (SP) and event rate (ER) levels. Monitors used a four-category rating scale in responding. Results showed that the effects of both SP and ER were dependent on task demands as defined on either category of the speed-closure dimension. Analysis of operating characteristics indicated that a within session loss in sensitivity was obtained for performance in the speed task under a high ER. Response latency was a function of confidence level, and the latency data was consistent with a fixed sample decision theory model. The results are discussed in relation to task classification procedures and the relative demands of visual and auditory monitoring.

### 9.2.1 Method

Experiment 4 can be viewed as an extension of Experiments 1 and 2 to auditory monitoring performance. As in Experiment 1, the effects of signal probability (SP) and event rate (ER) on performance were investigated in relation to the speed and closure task categories, but for auditory tasks. The same general design as in Experiment 1 was utilized, except that only two levels of SP and ER were employed.

#### Subjects

Thirty males aged 17 to 27 years and thirty females aged 20 to 25 years participated in the experiment. The subjects were randomly assigned to six experimental groups, with the restriction that each group contained an equal number of male and female subjects. Further characteristics of the subject sample are given in Table 6.2.

Apparatus

The apparatus and experimental environment were as described in section 6.2. Control of stimulus timing and all data recording were handled on-line by the PDP-8/E computer. Two auditory tasks were employed, an auditory speed task (AS), and an auditory closure task (AC). Detailed descriptions of each task are given in 6.2.5. In the AS task, subjects were required to detect an increase in intensity of an intermittent 1KHz tone which was presented every 2 or 4 seconds, depending on the event rate used. The increase in intensity that defined a signal tone was approximately 1 dB. In the AC task, the same 1KHz tone had to be detected within a noise burst which was pulsed at the same rate as in AS. Hence signal detection in AS required the detection of a change in a stimulus in comparison to previous stimulus presentations, which has been characterized as a distinctive feature of speed tasks; while in AC, the signal had to be detected within a single event, which is a feature of closure tasks.

In Experiment 1, observers were required to respond Yes or No on each trail during the monitoring period. In the present experiment, as in Experiment 2, the response requirement was extended to include two more response categories of intermediate confidence. The resulting four-category rating scale was represented on four response buttons labelled "Certain Yes", "Doubtful Yes", "Doubtful No" and "Certain No" (see 6.2.2 and Figure 6.4).

Signals were presented irregularly at a mean rate of 0.53, 1.0 or 2.0 (per min.), depending on the experimental condition, while event rate was

15 or 30 per min. Task duration was 45 min. in all conditions.

Procedure

Each of the six groups of subjects worked with either the speed or the closure task under one of three conditions combining different levels of SP and ER (see Table 9.1). In Conditions I and II, ER was held constant at 15/min. while SP was increased from 0.035 to 0.066, while in Conditions II and III, SP was fixed at 0.066, while ER was increased from 15 to 30/min. Apart from the fact that only two levels of SP and ER were utilized, these values are the same as in Experiment 1 (Conditions I, II and III form a subset of the five conditions in Experiment 1, and are equivalent to Conditions IV, II and III in Experiment 1, respectively).

The experimental procedure was as described previously in section 6.3. Subjects received the same training and expectancy matching practice in each condition.

### 9.2.2 Results

Data analysis

The analysis of the data was generally the same as in Experiment 2, in

| Condition | Signal Rate | Event Rate | Signal Probability |
|---|---|---|---|
| I | 0.533/min | 15/min | 0.035 |
| II | 1.0/min | 15/min | 0.066 |
| III | 2.0/min | 30/min | 0.066 |

Variable Event Rate

Variable Signal Probability

TABLE 9.1 Signal probability (SP) and event rate (ER) levels defining Conditions I, II and III. The lines link the two sets of conditions in which ER is varied for a fixed SP (Conditions II and III) and vice versa (Conditions I and II).

which confidence ratings were also employed. Hit and False Alarm (FA) probabilities above each of the three criteria created by the four-category rating scale were computed in each 15-min. time block; these are referred to as p(H1), p(H2), p(H3) and p(FA1), p(FA2), p(FA3) for Hits and False Alarms and the three criteria, respectively. These raw probabilities were normalized to yield double-probability (normalized) operating characteristics (OCs) for each subject.

Statistical analysis of the data was as in Experiment 1, except that the SP and ER factors had only two levels each. Performance measures were analysed separately in two analyses of variance (ANOVAs), one for independent SP effects (SP ANOVA), and the other for independent ER effects (ER ANOVA). Each ANOVA had three factors: Tasks (speed or closure), SP or ER, and Time Blocks.

Correct detections and false alarms

The mean Hit probabilities above each criterion are displayed in Figure 9.1 as a function of time blocks and conditions. A steady decline in p(H) is apparent in all conditions for the strict and medium criteria, but there appears to be less decrement in 'risky' or lax Hits. However, statistical analysis of the data showed that for the ER conditions (II and III), there was a significant decline with time on task in all three Hit measures (see Tables 9.2, 9.3 and 9.4). The summary tables also indicate that there was a significant decrease in the mean probability of a Hit at all criterion levels with an increase in ER ($p < .001$ in each case), but for the speed task only (see also Figure 9.1). This is the same result as in Experiment 1. However, unlike Experiment 1, there was no effect of ER on the decrement in Hits over the monitoring period.

The SP ANOVAs for Hits indicated that there was a significant increase in p(H) with SP for all three criteria (see Tables 9.5, 9.6 and 9.7). The

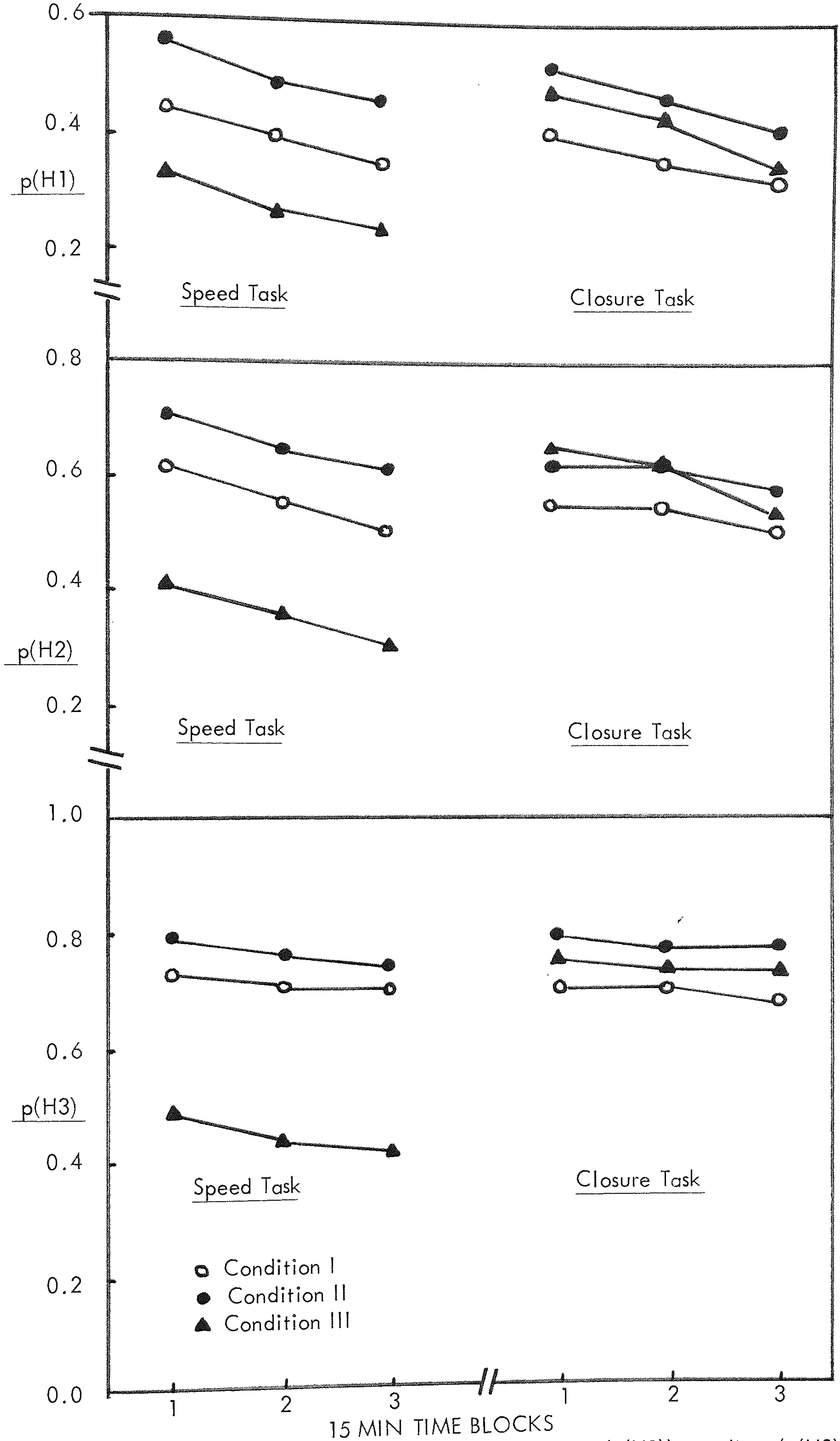


FIGURE 9.1 The mean probability of Hits above strict (p(H1)), medium (p(H2)) and lax (p(H3)) criteria as a function of time blocks and conditions, and for both tasks.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 54.675 | 1 | 54.675 | 6.4513 | .05 |
| B (Event Rate) | 484.008 | 1 | 484.008 | 57.1101 | .001 |
| AB | 216.008 | 1 | 216.008 | 25.4877 | .001 |
| SWG | 305.100 | 36 | 8.4750 | | |
| C (Time Blocks) | 241.516 | 2 | 120.758 | 96.1792 | .001 |
| AC | 7.350 | 2 | 3.675 | 2.9270 | ns |
| BC | 4.117 | 2 | 2.058 | 1.6394 | ns |
| ABC | 0.617 | 2 | 0.308 | 0.2456 | ns |
| C x SWG | 90.400 | 72 | 1.256 | | |

TABLE 9.2 ER ANOVA for Hits above the strict criterion (H1).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 276.033 | 1 | 276.033 | 23.7127 | .001 |
| B (Event Rate) | 662.700 | 1 | 662.700 | 56.9294 | .001 |
| AB | 520.833 | 1 | 520.833 | 44.7423 | .001 |
| SWG | 419.067 | 36 | 11.641 | | |
| C (Time Blocks) | 179.617 | 2 | 89.808 | 77.4086 | .001 |
| AC | 4.017 | 2 | 2.008 | 1.7310 | ns |
| BC | 3.350 | 2 | 1.675 | 1.4437 | ns |
| ABC | 0.817 | 2 | 0.408 | 0.3520 | ns |
| C x SWG | 83.533 | 72 | 1.160 | | |

TABLE 9.3 ER ANOVA for Hits above the medium criterion (H2).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 672.133 | 1 | 671.133 | 50.5504 | .001 |
| B (Event Rate) | 790.533 | 1 | 790.533 | 59.4552 | .001 |
| AB | 496.133 | 1 | 496.133 | 37.3136 | .001 |
| SWG | 478.667 | 36 | 13.296 | | |
| C (Time Blocks) | 32.916 | 2 | 16.458 | 10.2391 | .01 |
| AC | 7.217 | 2 | 3.608 | 2.2448 | ns |
| BC | 1.317 | 2 | 0.658 | 0.4096 | ns |
| ABC | 0.817 | 2 | 0.408 | 0.2540 | ns |
| C x SWG | 115.733 | 72 | 1.6074 | | |

TABLE 9.4 ER ANOVA for Hits above the lax criterion (H3).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 3.6401 | 1 | 3.6401 | 8.4571 | .01 |
| B (Signal Probability) | 21.4208 | 1 | 21.4208 | 49.7679 | .001 |
| AB | 0.0667 | 1 | 0.0667 | 0.1550 | ns |
| SWG | 15.4950 | 36 | 0.4304 | | |
| C (Time Blocks) | 12.2220 | 2 | 6.1110 | 19.538 | .001 |
| AC | 0.0347 | 2 | 0.0174 | 0.0556 | ns |
| BC | 0.1820 | 2 | 0.0910 | 0.4178 | ns |
| ABC | 0.3500 | 2 | 0.1750 | 0.8035 | ns |
| C x SWG | 22.5200 | 72 | 0.2178 | | |

TABLE 9.5 SP ANOVA for Hits above the strict criterion (H1).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 1.704 | 1 | 1.704 | 2.5632 | ns |
| B (Signal Probability) | 15.052 | 1 | 15.052 | 22.6371 | .001 |
| AB | 0.154 | 1 | 0.154 | 0.2318 | ns |
| SWG | 23.938 | 36 | 0.6649 | | |
| C (Time Blocks) | 9.570 | 2 | 4.785 | 24.0881 | .001 |
| AC | 0.679 | 2 | 0.3395 | 1.7090 | ns |
| BC | 0.061 | 2 | 0.0305 | 0.1540 | ns |
| ABC | 0.227 | 2 | 0.1135 | 0.5710 | ns |
| C x SWG | 14.303 | 72 | 0.1987 | | |

TABLE 9.6 SP ANOVA for Hits above the medium criterion (H2).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.01 | 1 | .010 | 0.0103 | ns |
| B (Signal Probability) | 8.427 | 1 | 8.427 | 8.6841 | .01 |
| AB | 0.7053 | 1 | 0.7053 | 0.7268 | ns |
| SWG | 34.9354 | 36 | 0.9704 | | |
| C (Time Blocks) | 1.1212 | 2 | 0.5606 | 1.5851 | ns |
| AC | 0.1415 | 2 | 0.0708 | 0.2001 | ns |
| BC | 0.1745 | 2 | 0.0873 | 0.2475 | ns |
| ABC | 0.2882 | 2 | 0.1441 | 0.4074 | ns |
| C x SWG | 25.4646 | 72 | | | |

TABLE 9.7 SP ANOVA for Hits above the lax criterion (H3).

effect of time at work in the SP Conditions (I and II) was to depress p(H) for strict and medium criteria, but not for the lax criterion. Additionally, there were significant differences between tasks for p(H1), detection performance being superior in the closure task (see Table 9.5 and Figure 9.1). However the Tasks factor was not significant for the other two Hit measures, and there were no significant interaction effects.

Hence the effects of SP and ER were to significantly improve and depress, respectively, the mean probability of detection, while leaving the decrement relatively unchanged. The effect of ER was significant only for the speed task.

The median probabilities of False Alarms are displayed in Figure 9.2 as a function of time blocks and conditions.

The ER ANOVA for FAs showed that the only effect of ER on p(FA) was for lax criterion FAs, which declined significantly with increased ER (see Tables 9.8, 9.9 and 9.10). There was a significant decrement with time on task in all three criterion FA measures, but for p(FA2), there was a significant interaction between Time Blocks and ER ($p < .05$; see Table 9.9), implying that p(FA2) declined in Condition II, but not in Condition III. In all other conditions and for each measure, there was a significant decline in p(FA) over time blocks (see SP ANOVAs in Tables 9.11, 9.12 and 9.13). There was an increase in p(FA) at each criterion level with an increase in SP; however, a significant interaction between Tasks and SP was obtained for p(FA3), indicating that SP had a significant effect on 'risky' FAs for the closure task only. This result apparently demonstrates an interaction between SP and task demands as defined by the speed and closure categories.

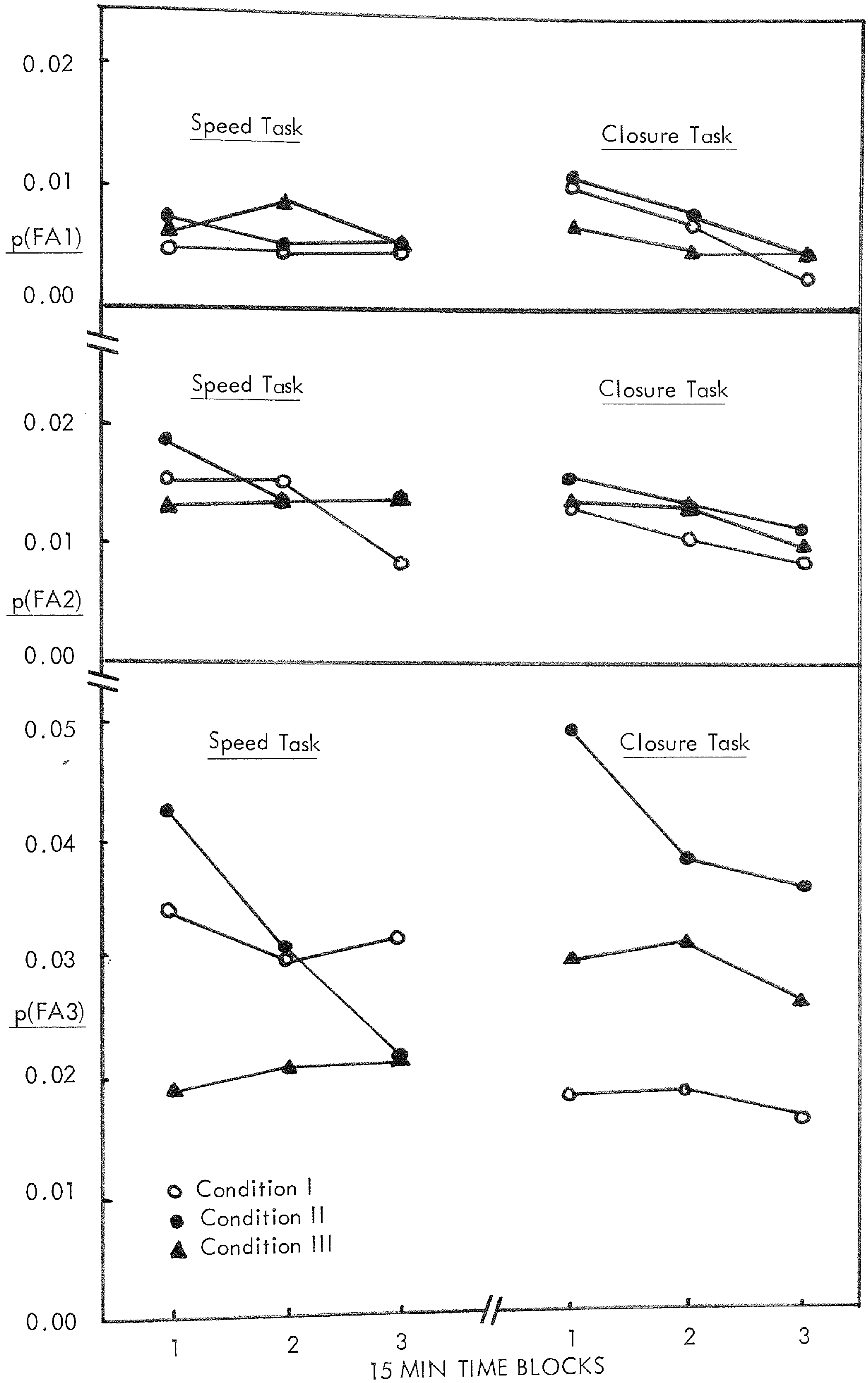


FIGURE 9.2 The median probability of false al arms above strict (p(FA1)), medium (p(FA2)) and lax (p(FA3)) criteria as a function of time blocks and conditions, and for both tasks.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.00000 | 1 | 0.00000 | 0.0000 | ns |
| B (Event Rate) | 0.00027 | 1 | 0.00027 | 0.0060 | ns |
| AB | 0.05633 | 1 | 0.05633 | 1.2557 | ns |
| SWG | 1.61503 | 36 | 0.04486 | | |
| C (Time Blocks) | 0.30105 | 2 | 0.15052 | 17.7601 | .001 |
| AC | 0.02546 | 2 | 0.01273 | 1.5020 | ns |
| BC | 0.04298 | 2 | 0.02149 | 2.5356 | ns |
| ABC | 0.03429 | 2 | 0.01714 | 2.0227 | ns |
| C x SWG | 0.61023 | 72 | 0.00848 | | |

TABLE 9.8 ER ANOVA for False Alarms (log transformed) above the strict criterion (FA1).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.00037 | 1 | 0.00037 | 0.0077 | ns |
| B (Event Rate) | 0.06394 | 1 | 0.06394 | 1.3337 | ns |
| AB | 0.00784 | 1 | 0.00784 | 0.1636 | ns |
| SWG | 1.72586 | 36 | 0.04794 | | |
| C (Time Blocks) | 0.12720 | 2 | 0.06360 | 9.6782 | .01 |
| AC | 0.04740 | 2 | 0.02370 | 3.6067 | ns (.10) |
| BC | 0.08866 | 2 | 0.04433 | 6.7457 | .05 |
| ABC | 0.01576 | 2 | 0.00788 | 1.1992 | ns |
| C x SWG | 0.47317 | 72 | 0.00657 | | |

TABLE 9.9 ER ANOVA for False Alarms (log transformed) above the medium criterion (FA2).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.37856 | 1 | 0.37856 | 6.3712 | .01 |
| B (Event Rate) | 0.50181 | 1 | 0.50181 | 8.4455 | .05 |
| AB | 0.00001 | 1 | 0.00001 | 0.0002 | ns |
| SWG | 2.13905 | 36 | 0.05942 | | |
| C (Time Blocks | 0.08714 | 2 | 0.04357 | 5.0621 | .05 |
| AC | 0.00522 | 2 | 0.00261 | 0.3034 | ns |
| BC | 0.05918 | 2 | 0.02959 | 3.4382 | ns (.10) |
| ABC | 0.06292 | 2 | 0.03146 | 3.6554 | ns (.10) |
| C x SWG | 0.61967 | 72 | 0.00861 | | |

TABLE 9.10 ER ANOVA for False Alarms (log transformed) above the lax criterion (FA3).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.05043 | 1 | 0.05043 | 1.1445 | ns |
| B (Signal Probability) | 0.28033 | 1 | 0.28033 | 6.2623 | .05 |
| AB | 0.00016 | 1 | 0.00016 | 0.0037 | ns |
| SWG | 1.58622 | 36 | 0.04406 | | |
| C (Time Blocks) | 0.62572 | 2 | 0.31286 | 41.4218 | .001 |
| AC | 0.03021 | 2 | 0.01511 | 2.0001 | ns |
| BC | 0.02348 | 2 | 0.01174 | 1.5544 | ns |
| ABC | 0.00316 | 2 | 0.00158 | 0.2093 | ns |
| C x SWG | 0.54382 | 72 | 0.00755 | | |

TABLE 9.11 SP ANOVA for False Alarms (log transformed) above the strict criterion (FA1).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.00140 | 1 | 0.00140 | 0.0248 | ns |
| B (Signal Probability) | 0.23497 | 1 | 0.23497 | 4.1644 | .05 |
| AB | 0.01141 | 1 | 0.01141 | 0.2022 | ns |
| SWG | 2.03122 | 36 | 0.05642 | | |
| C (Time Blocks) | 0.44787 | 2 | 0.22394 | 34.5753 | .001 |
| AC | 0.00671 | 2 | 0.00336 | 0.5181 | ns |
| BC | 0.00344 | 2 | 0.00172 | 0.2652 | ns |
| ABC | 0.02386 | 2 | 0.01193 | 1.8416 | ns |
| C x SWG | 0.46633 | 72 | 0.00648 | | |

TABLE 9.12 SP ANOVA for False Alarms (log transformed) above the medium criterion (FA2).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.00008 | 1 | 0.00008 | 0.0011 | ns |
| B (Signal Probability) | 0.62785 | 1 | 0.62785 | 8.3238 | .01 |
| AB | 0.39445 | 1 | 0.39445 | 5.2295 | .05 |
| SWG | 2.71545 | 36 | 0.07543 | | |
| C (Time Blocks) | 0.14943 | 2 | 0.07472 | 6.0745 | .05 |
| AC | 0.02472 | 2 | 0.01236 | 1.0050 | ns |
| BC | 0.02318 | 2 | 0.01159 | 0.9424 | ns |
| ABC | 0.01067 | 2 | 0.00534 | 0.4338 | ns |
| C x SWG | 0.88559 | 72 | 0.01230 | | |

TABLE 9.13 SP ANOVA for False Alarms (log transformed) above the lax criterion (FA3).

Operating characteristics

Individual operating characteristics (OCs) were obtained by plotting the normalized values of corresponding pairs of p(H) and p(FA) in each 15-min. time block. A representative sample of OCs from twelve subjects is shown in Figures 9.3 and 9.4. Group OCs are displayed in Figure 9.5; these OCs were plotted from the means of the individual normalized operating probabilities, that is,

$$\text{Group}\left\{Z(H),\ Z(FA)\right\} = \left\{\overline{Z(p(H))},\ \overline{Z(p(FA))}\right\} \neq \left\{Z(\overline{p(H)}),\ Z(\overline{p(FA)})\right\}$$

Figures 9.3 and 9.4 indicate that in Conditions I and II, and in Condition III for the closure task, the OC points from successive time blocks all lie approximately on the same straight line (or on different lines, but which have the same detectability). There is thus no indication of a within-session sensitivity loss in these conditions, but a change in the response criteria towards greater strictness. A movement in all three criteria is apparent, but there appears to be less movement for the lax criterion.

| | Conditions | | | |
|---|---|---|---|---|
| Criterion | I | II | III-Speed | III-Closure |
| Goodness of fit[1] | 70 | 85 | 80 | 80 |
| OC slope near unity[2] | 85 | 80 | 70 | 70 |
| OC slope unity | 45 | 55 | 50 | 30 |
| Sensitivity decrement[3] | 25 | 15 | 70 | 30 |

TABLE 9.14 Proportions of subjects (%) in each condition with operating characteristics (OCs) meeting goodness of fit, OC slope and sensitivity decrement criteria (1 mean square orthonormal error $< 0.1$; 2 $d/\Delta\sigma$ in the range 10 - $\infty$; 3 fall in $d_a$ of greater than 0.2 between 1st and 3rd time block).

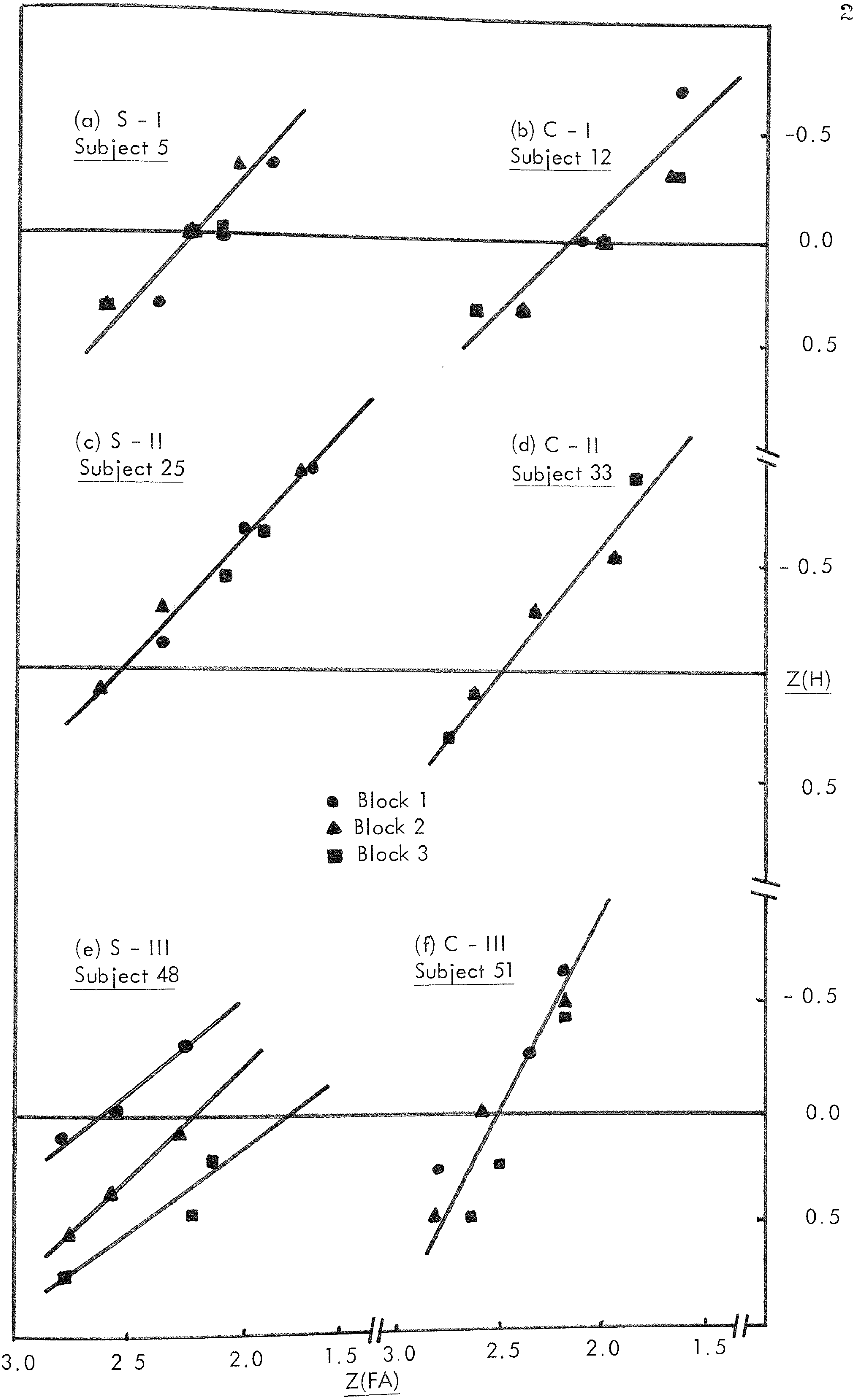


FIGURE 9.3 Individual operating characteristics for the speed (S) and closure (C) tasks in Conditions I, II and III.

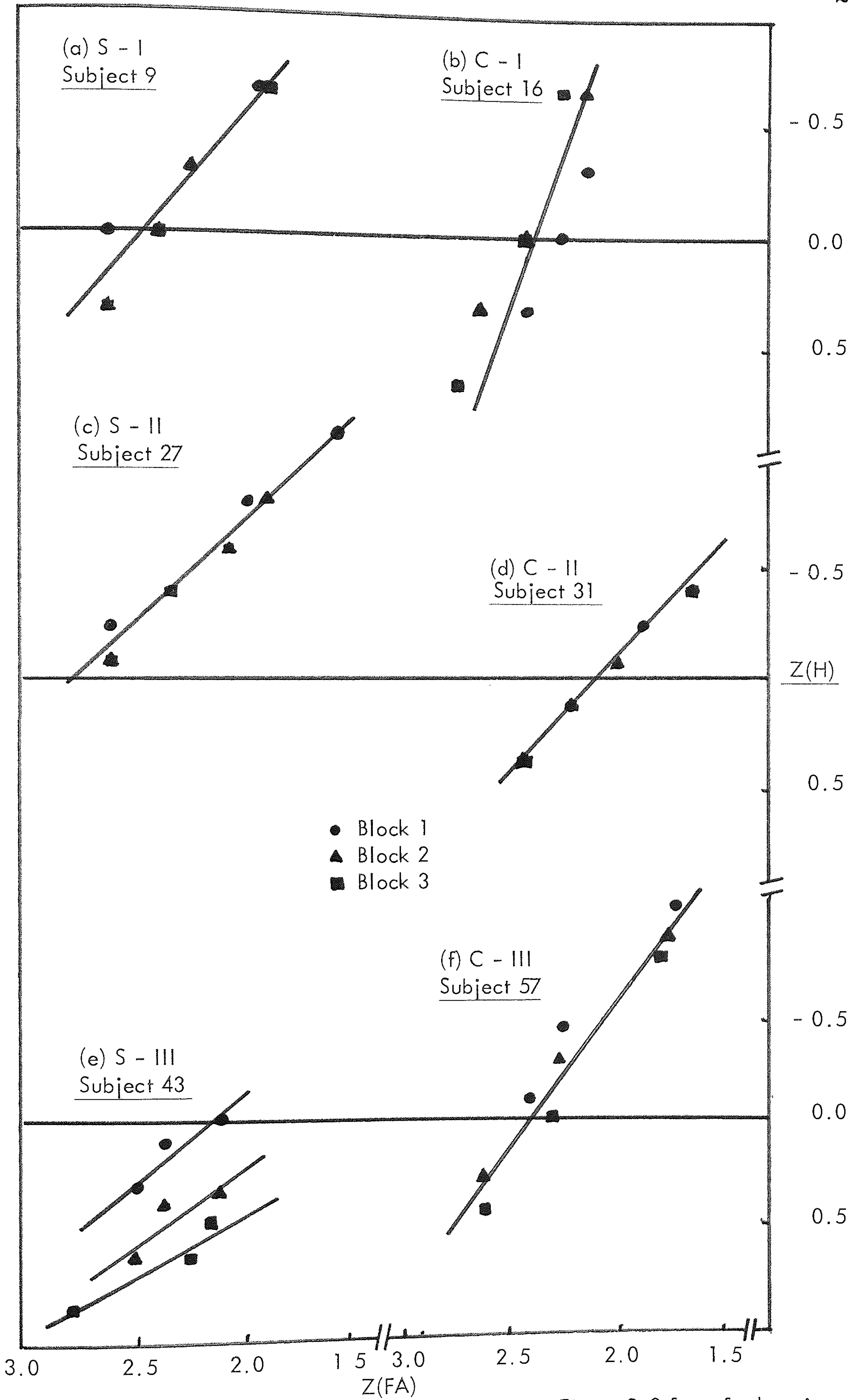


FIGURE 9.4 Individual operating characteristics as in Figure 9.3 for a further six subjects.

A sensitivity decrement over time blocks is apparent only for subjects working with the speed task under a high ER, as in Figure 9.3(e) and Figure 9.4(e). The OCs for these two subjects were representative of the group of subjects in this condition, and the group OC in Figure 9.5(e) also shows the same loss in sensitivity. This is also apparent in Table 9.14, in which, among other features, the proportion of subjects who exhibited sensitivity trends different to that typical of their group is listed for each condition; 70% of the subjects in the group working with the speed task at a high ER had a within-session sensitivity ($d_a$) loss of 0.2 or greater, while the majority of subjects in the other conditions exhibited no sensitivity decrement.

Table 9.15 lists the mean values of $d_a$ as a function of time blocks and tasks for the three conditions. These data show the same sensitivity decrement in one condition as indicated in the examination of the OCs. Statistical analysis of the data confirmed the finding of a sensitivity decrement for the speed task in the high ER condition only. The ER ANOVA for $d_a$, shown in Table 9.16, gave significant effects for each factor and interaction, reflecting the substantial effect due to the fall in sensitivity in the one condition only (indicated by an asterisk in Table 9.15). The SP ANOVA confirmed that there was no significant effects due to any factor in the SP Conditions (I and II). The summary table is given in Appendix C4 in Table C4.1.

The decrement in sensitivity in the speed task under a high ER was not related to changes in the signal to noise variance ratio over the monitoring period. This was confirmed by examining the individual signal to non-signal variance ratios in different time blocks and in each condition. There was no consistent trend in the variance ratios over time blocks for either task. Although the use of only a three-point OC

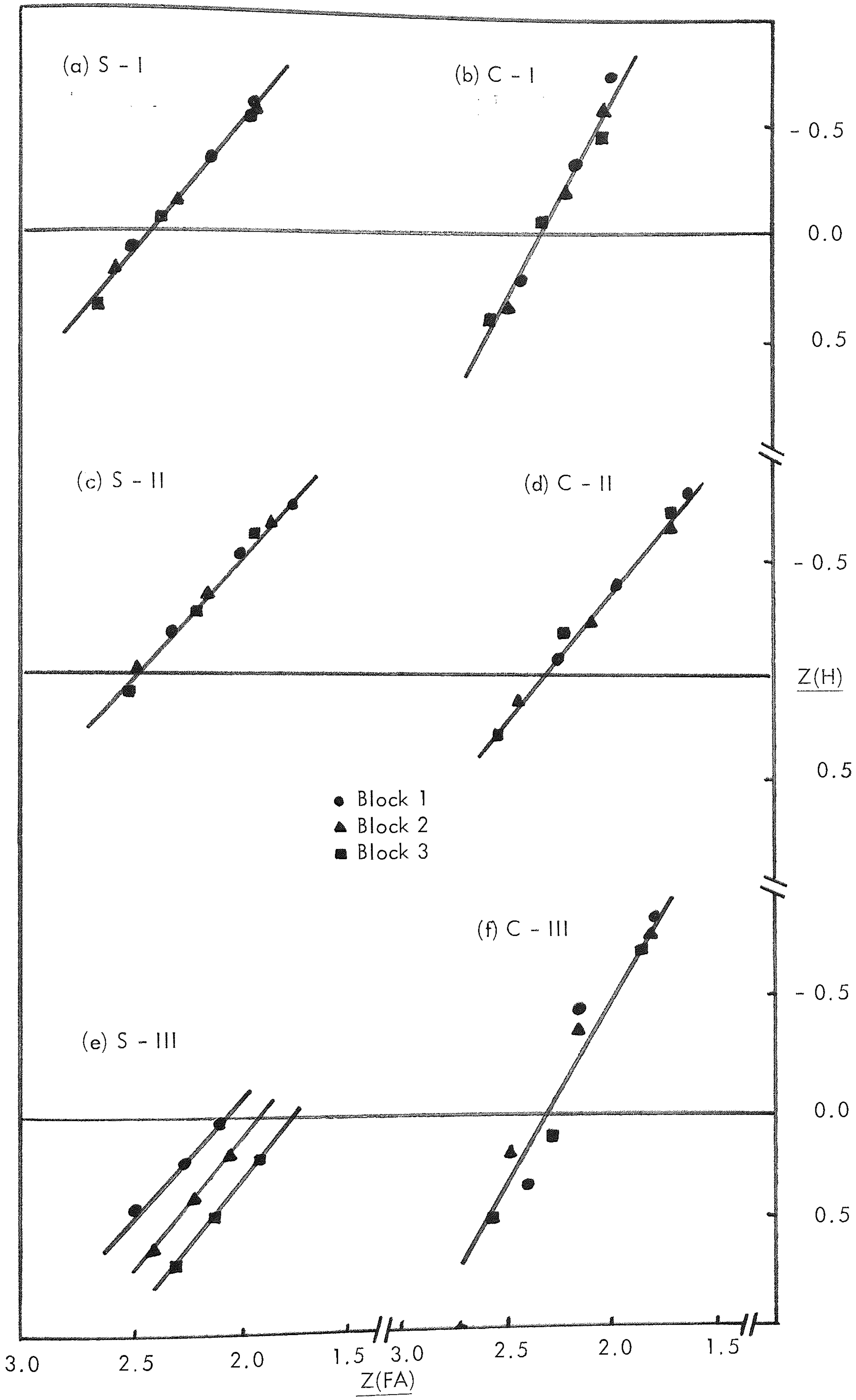


FIGURE 9.5 Group operating characteristics for the Speed (S) and Closure (C) tasks in Conditions I, II and III.

| Condition | Speed Task 1 | 2 | 3 | Closure Task 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| I | 2.48 | 2.47 | 2.43 | 2.42 | 2.41 | 2.42 |
| II | 2.56 | 2.57 | 2.53 | 2.44 | 2.46 | 2.43 |
| III | 2.10 | 1.83 | 1.79* | 2.47 | 2.54 | 2.46 |

TABLE 9.15 Mean values of $d_a$ in successive 15-min. time blocks (1, 2, 3) for the speed and closure tasks in Conditions I, II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 1.67560 | 1 | 1.67560 | 14.3633 | .01 |
| B (Event Rate) | 2.77856 | 1 | 2.77856 | 23.8179 | .001 |
| AB | 3.69603 | 1 | 3.69603 | 31.6825 | .001 |
| SWG | 4.19971 | 36 | 0.11666 | | |
| C (Time Blocks) | 0.11719 | 2 | 0.08559 | 9.3315 | .01 |
| AC | 0.18950 | 2 | 0.09475 | 10.3300 | .01 |
| BC | 0.08233 | 2 | 0.04117 | 4.4880 | .05 |
| ABC | 0.12184 | 2 | 0.06092 | 6.6414 | .05 |
| C x SWG | 0.66041 | 72 | 0.00917 | | |

TABLE 9.16 ER ANOVA for $d_a$.

limits the reliability of these values, they nevertheless indicated that the sensitivity decrement was obtained irrespective of criterion level, and did not reflect a change in the variability of the signal effect relative to the noise effect. Table 9.14 shows also, furthermore, that most of the individual OCs had slopes not greatly different from 1 (as defined by the criterion that $d/\Delta\sigma$ should be greater than 10 for near-unit slope OCs).

Figure 9.6 displays the mean values of log likelihood ratio (log B) at each criterion as a function of time blocks and conditions. These data were analysed, as for the other measures, in the SP and ER ANOVAs. A rather complex set of results were obtained in these (six) ANOVAs, and, for convenience, the significant effects are listed in Table 9.17. The

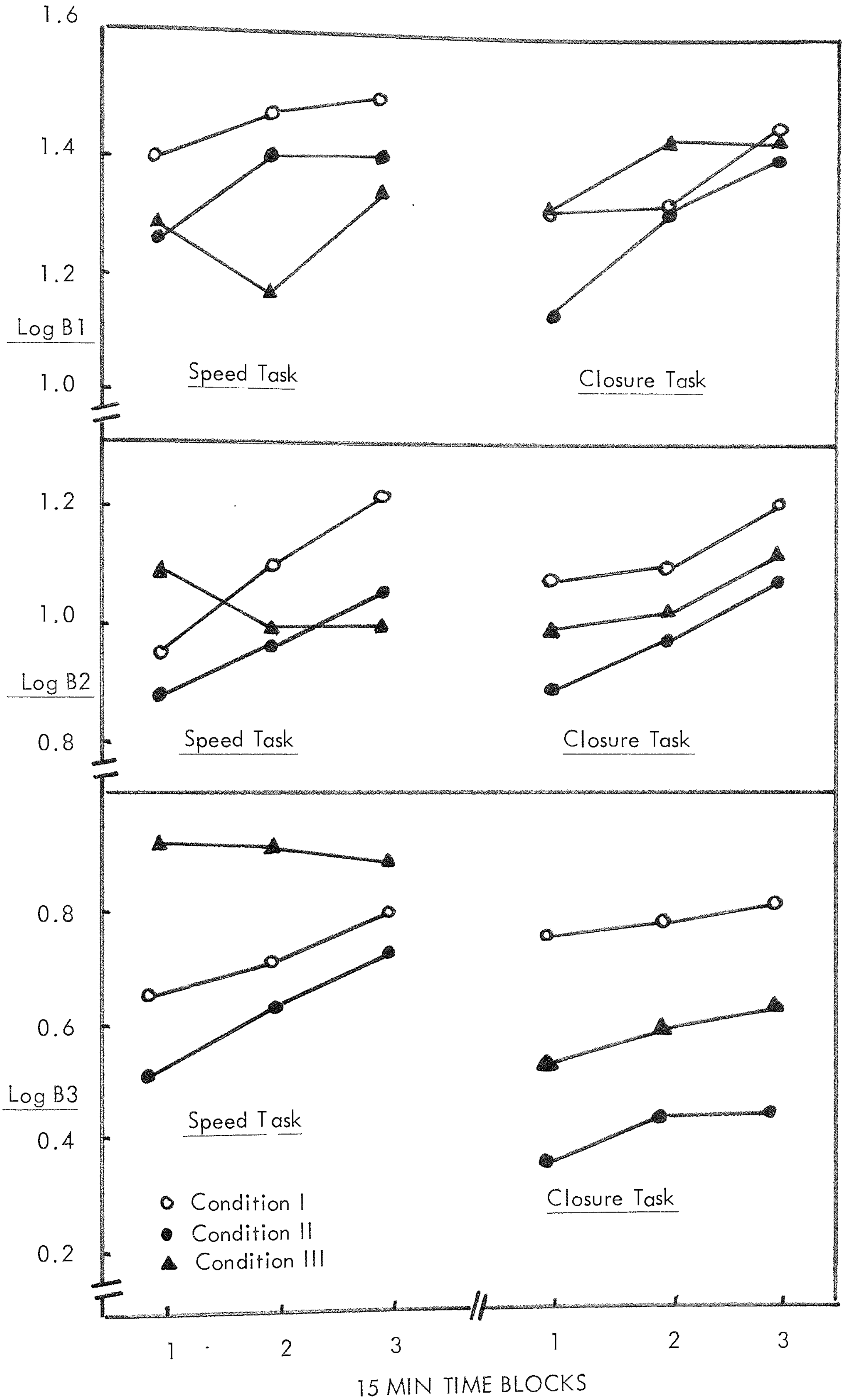


FIGURE 9.6 Mean values of log B for strict (log B1), medium (log B2) and and lax (log B3) criteria as a function of time blocks and conditions, and for both tasks.

| | SP ANOVA | ER ANOVA |
|---|---|---|
| log B1 | Tasks *<br>SP *<br>Time Blocks *** | Time Blocks **<br>ER x Time Blocks * |
| log B2 | SP *<br>Time Blocks *** | Time Blocks *<br>ER x Time Blocks * |
| log B3 | SP **<br>SP x Tasks * | Tasks **<br>ER *<br>Time Blocks *<br>Tasks x ER x Time Blocks * |

TABLE 9.17 Summary of significant effects obtained in the SP and ER ANOVAs for the three log B measures (* $p < .05$; ** $p < .01$; *** $p < .001$).

complete ANOVA tables are given in Tables C4.2 to C4.7 in Appendix C4.

An inspection of Table 9.17 and of Figure 9.6 reveals that in conditions I and II there was a significant increase in log B with time on task for the strict and medium criteria (log B1 and log B2 respectively) but not for the lax criterion (log B3). The effect of SP was to reduce log B for all three criteria, but for log B3 the effect was restricted to performance on the closure task, as the significant interaction between Tasks and SP indicates.

For the ER ANOVA, the results were consistent for the two higher criteria log B1 and log B2; there was a significant increase in both measures with time on task, but for the low ER condition only, as the significant interaction between ER and Time Blocks confirms. For log B3, however, further effects for Tasks, ER and the interaction between Tasks, ER and Time Blocks were obtained. This apparently indicates that while there was no increase in log B3 with time blocks in Condition III in the speed task, an increment was obtained in the closure task (see also Figure

9.6). The results for log B can be summarized as follows:-

1) With time on task, there was a significant increase in log B for the strict and medium criteria only, and for the low ER conditions I and II only.

2) An increase in SP significantly reduced log B for all three criteria.

3) An increase in ER had no significant effect on log B1 or log B2, but significantly increased log B3.

4) There was no increase in log B with time blocks in the high ER condition, except for log B3 for the closure task.

Response latencies

The response latency distributions were divided so that latencies associated with correct and incorrect positive (Yes) and negative (No) responses could be defined at one of two levels of confidence. Thus ratings 1 and 4 were used to derive the confident Yes and No response latencies, while the intermediate ratings 2 and 3 were used to derive the response latencies associated with a lower confidence level. Figure 9.7 shows that on average, response latency was higher at the intermediate confidence levels than at the high confidence levels. The same inverted-U relation as shown in Figure 9.7 was obtained in the other conditions and for both tasks.

For the purposes of the statistical analysis of the latency data, the responses were collapsed over the two confidence levels and a mean latency measure was used. This was because there were a number of occasions, especially in the low SP Condition I, where subjects had only a few responses in the intermediate response categories, thus precluding a strict comparison of latency across confidence levels, and across conditions for the low-confidence latencies. By collapsing responses across ratings 1 and 2, and across ratings 3 and 4, the mean

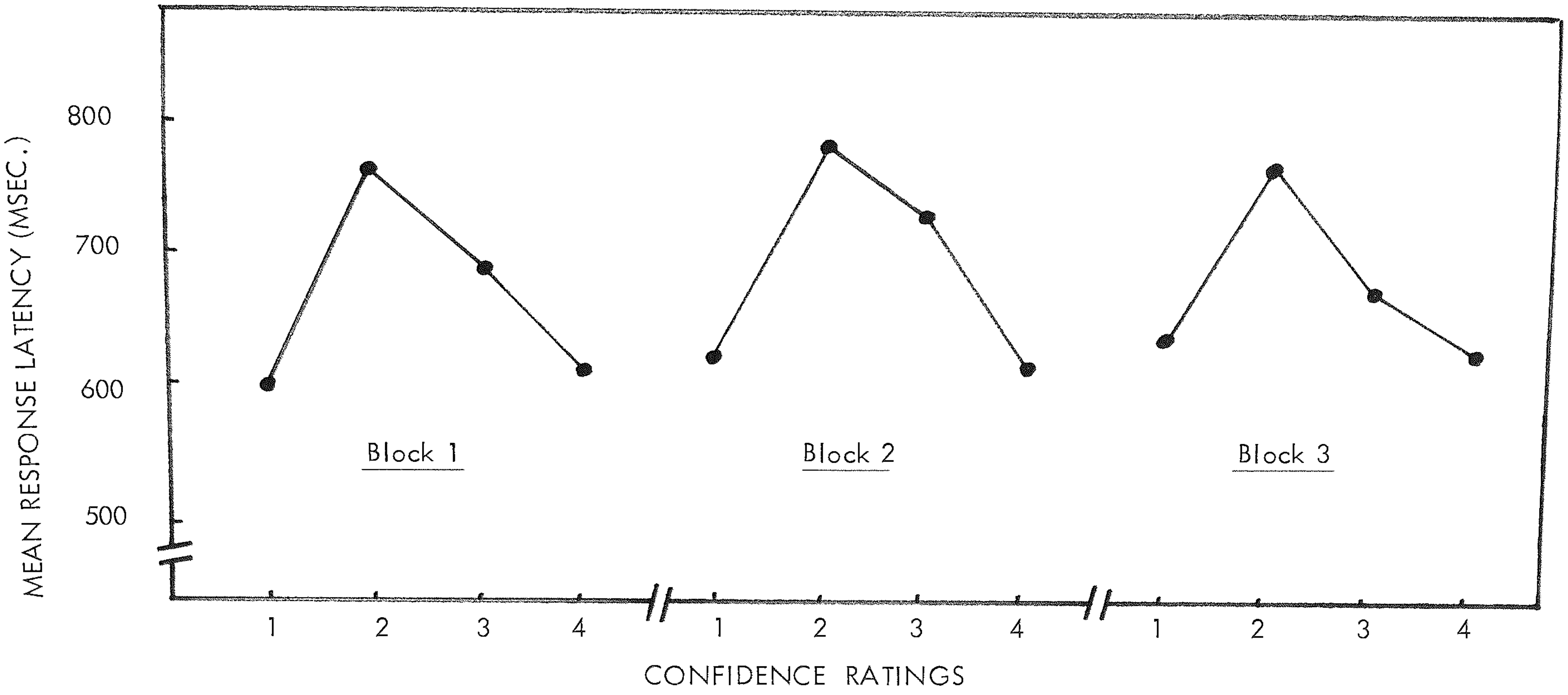


FIGURE 9.7 Mean response latency as a function of confidence rating and time blocks. Data from Condition II for the speed task.

| Response Latency | ANOVA FACTORS[1] SP | ER | Time Blocks |
|---|---|---|---|
| DL | Decrease<br>$p < .001$ | Increase[2]<br>$p < .01$ | Increase<br>$p < .001$ |
| FAL | Decrease<br>* ns ($F = 3.86$) | No change<br>ns ($F = .317$) | Increase<br>$p < .01$ |
| CRL | Increase<br>* ns ($F = .508$) | No change<br>ns ($F = .149$) | Decrease<br>* ns ($F = 1.86$)[3] |
| ML | Increase<br>$p < .001$ | Decrease<br>$p < .05$ | Decrease<br>$p < .05$ |

TABLE 9.18 Summary of effects of SP, ER and Time Blocks on each of the four latency measures (see text; 1 df = 1,36 for each factor; 2 speed task only, simple effect $F = 11.05$, $p < .01$; 3 $F(.05) = 4.12$).

latencies associated with the four categories of responses in the Yes/No paradigm were derived: detection latency (DL), false alarm latency (FAL), correct rejection latency (CRL) and omission latency (ML). For DL and CRL the contribution to the mean latency was greater for the more numerous confident responses (1 and 4), but for FAL, there were approximately equal contributions from ratings 1 and 2.

Figures 9.8 and 9.9 display the mean response latencies as a function of time blocks and conditions. Each latency measure was analysed as before in the ER and SP ANOVAs. The full summary tables for the ANOVAs are given in Appendix C4 in Tables C4.8 to C4.15. Table 9.18 identifies the direction and reliability of the effects of SP, ER, Time Blocks and their interactions with Tasks, if any, for each of the four latency measures.

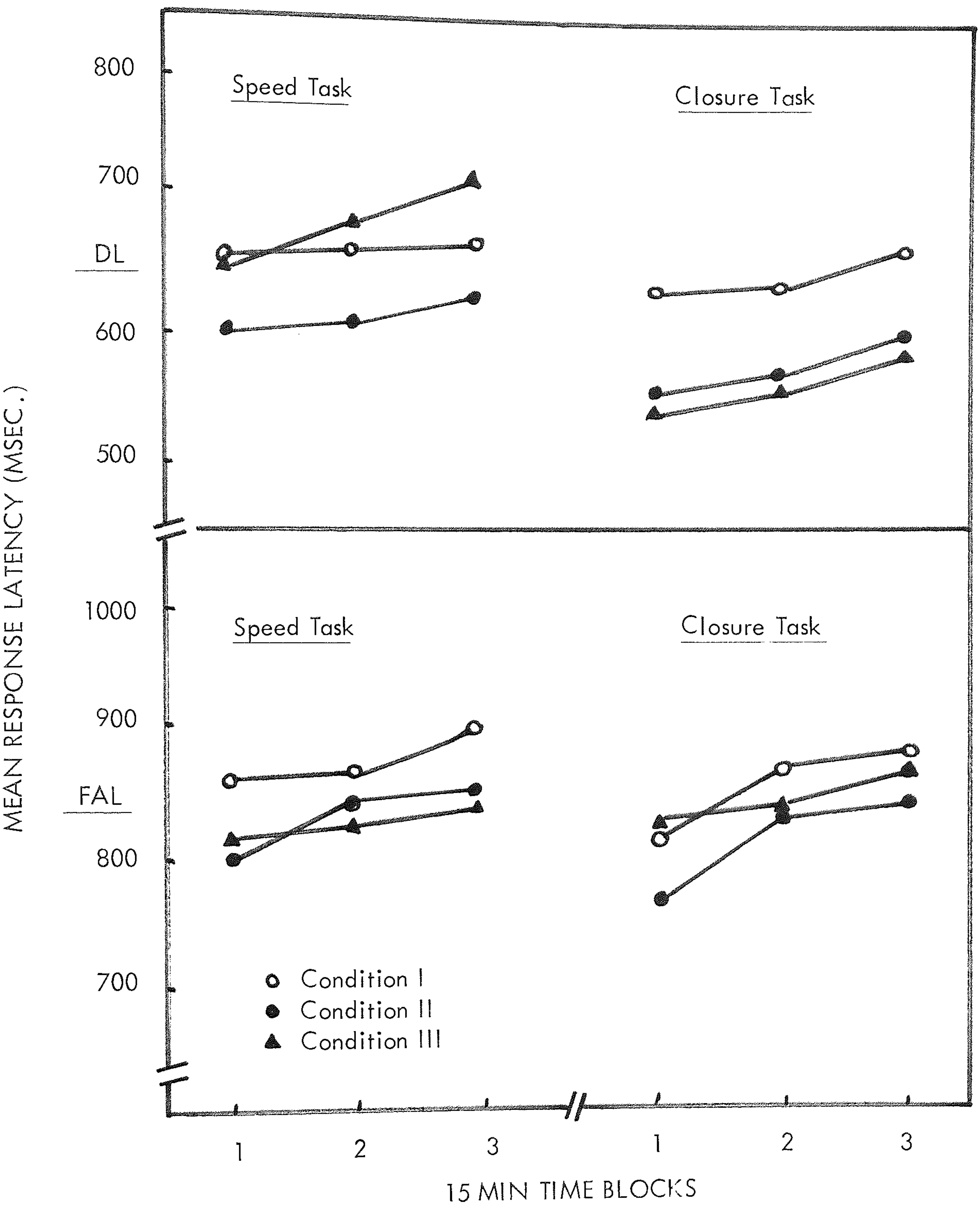


FIGURE 9.8 The mean latencies of correct and incorrect Yes responses (DL and FAL, respectively) as a function of time blocks and tasks.

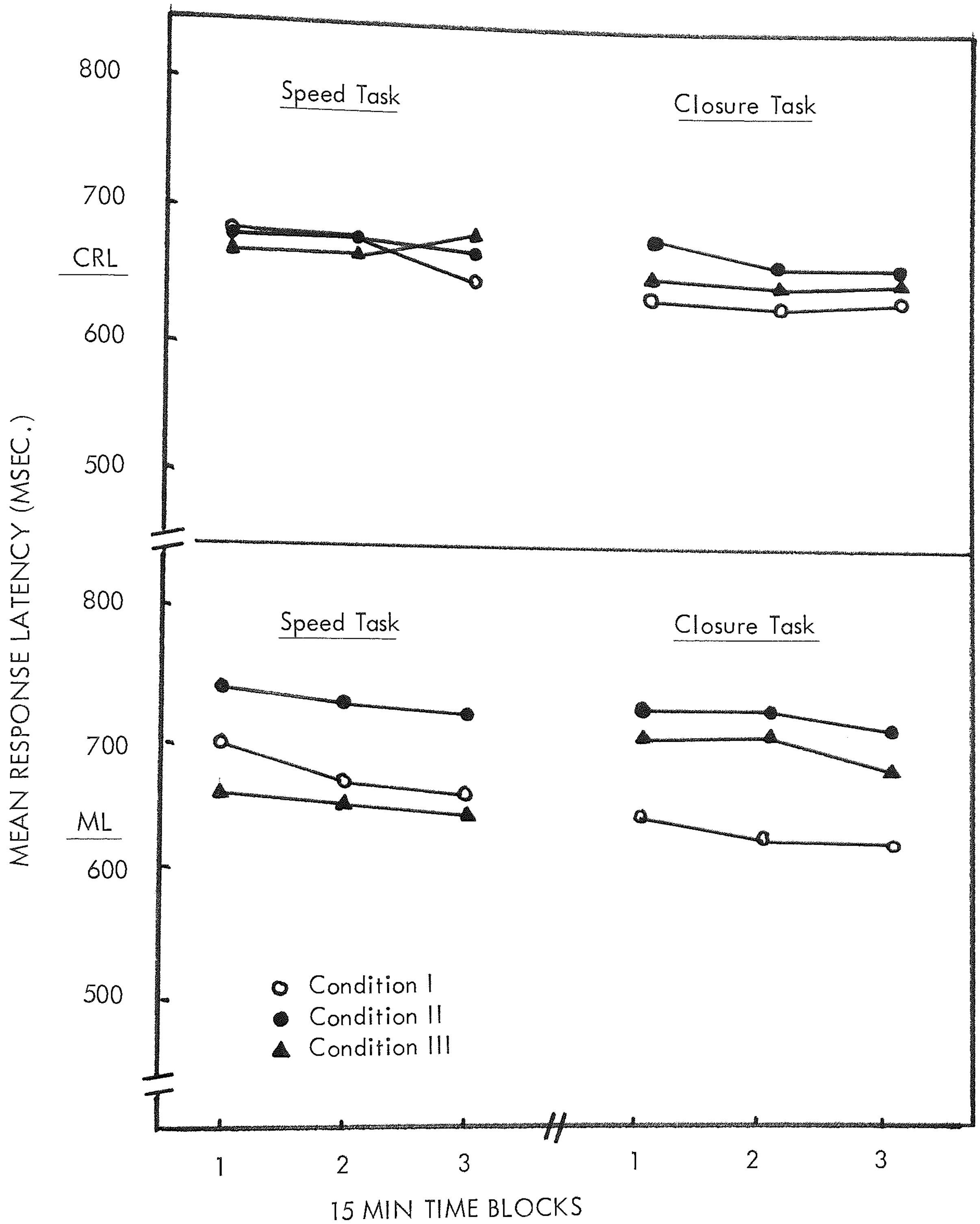


FIGURE 9.9 The mean latencies of correct and incorrect No responses (CRL and ML, respectively) as a function of time blocks and tasks.

Generally similar effects on latency were obtained as in Experiment 1; and the direction of the effects were also similar to those predicted by the latency model, although some of these were not reliable (as indicated by asterisks in Table 9.18). The results may be summarized as follows:

1) For both tasks, and in each condition, there was a significant increase with time on task in both Yes response latencies (DL and FAL), and a decrease in both No response latencies. However, for No latencies, only the decrease in ML was reliable.

2) With an increase in SP, there was a decrease in both Yes latencies, but the decrease in FAL just failed to reach significance at the .05 level. At the same time, there was an increase in both No latencies, but only the increase in ML was reliable.

3) With an increase in ER, there was no significant change in the mean non-signal response latencies (CRL and FAL), but there were reliable changes in the mean signal response latencies; DL significantly increased and ML significantly decreased with an increase in ER. For DL, there was also a significant interaction between Tasks and ER, indicating that the effect on DL was restricted to the speed task (see Table 9.18). However the same interaction was not significant for ML.

These results on latency are thus in general agreement with the predictions of the latency model, despite the unreliability of some effects, although they were in the predicted directions. The results may be interpreted as lending support to the view that performance changes over time and with a change in SP are characterized by a change in the decision criterion, and that a change in detectability is obtained in only one condition, and for the speed task only. We shall discuss these results in further detail in the following section.

### 9.2.3 Discussion

The results of Experiment 4 are in broad agreement with those of the previous experiments, and provide general support for the dual task classification-decision theory approach to the analysis of monitoring performance. In particular, the results further establish the reliability of the interaction between two dimensions of the task classification system, the type of signal discrimination and the time course of events, in the determination of sensitivity trends over the monitoring period. As in Experiments 1 and 2, Fleishman's (1972) suggestion that independent variables may interact with task demands in their influence on performance, was borne out.

The emergence of the same interaction effect for auditory monitoring tasks in the present experiment, as well as for visual tasks used in Experiments 1 and 2, clearly points to the generality and reliability of the effect, at least as far as single-source monitoring tasks are concerned. It is clear that the mechanisms behind the sensitivity decrement are not therefore purely related to the demands of visual monitoring. On the other hand, this does not imply that the mechanisms need be the same in the two cases; we have noted previously in Chapters 5 and 7 that in some monitoring situations, sensitivity decrement may be purely a function of the peculiar demands of certain visual tasks, such as visual search or continuous visual fixation, which cannot be related to sensitivity decrements in auditory tasks. For the many other types of monitoring situation, however, a common mechanism may be present. In Chapter 7 it was suggested that the mechanism was related to the temporal demands of the discrimination processes involved in the detection of signals, although this hypothesis was not elaborated. The present results greatly strengthen this explanation by showing that sensitivity decrement in auditory tasks may be interpreted in a similar manner.

As noted in Chapter 7, we shall reserve a more detailed discussion of the theoretical mechanisms underlying the sensitivity decrement, and of the types of tasks to which the decrement is restricted, until Chapters 12 and 13, after considering the evidence from four of the experiments in this thesis (Experiments 1, 2, 4 and 7) and from the experiments reported in the literature. These chapters also discuss the various theoretical constructs such as 'observing responses' (Jerison, 1970) and 'coupling' (Loeb and Binford, 1968) which have been proposed in relation to performance decrement in monitoring tasks. At this stage, it can be noted that while these constructs have some heuristic value, they cannot explain all the results obtained (nor, for that matter, a number of other results reported in the literature, see Chapter 12). On the basis of constructs of the 'observing response' type, performance on auditory tasks may be hypothesized to be less susceptible to sensitivity decrement that visual tasks. Yet the present results, along with some other studies in the literature in which speed-type auditory tasks have been employed (Deaton, Tobias and Wilkinson, 1971; Bennedetti and Loeb, 1972; see also Table 12.8) show that the differing demands of visual and auditory tasks (in which the observing response and coupling constructs are based) do not provide the sole basis for the explanation of performance decrement in monitoring tasks.

The results of the Experiment 4 with respect to criterion shifts and the relative effects of SP and ER are consistent with Experiments 1 and 2 and with previous studies (Baddeley and Colquhoun, 1969; Broadbent and Gregory, 1965; Loeb and Binford, 1968). Each of these independent variables influence theoretically independent aspects of performance, SP affecting the response criterion, and ER, in some tasks at least, affecting signal detectability. This interpretation is also strengthened by complementary evidence gained from an analysis of latency data, which we shall discuss further below.

Performance trends over time in the various conditions were generally consistent with the view that in monitoring tasks there is an increasing bias towards negative responding with time on task, or an increase in the stringency of the response criterion. This was true for responses made at strict and medium levels of confidence. In detection theory terms, there was a significant increase with time at work in log B for the strict and medium criteria, but not for the lax criterion. This result is similar to the one obtained by Broadbent (Broadbent and Gregory, 1963a, 1965), but the overall picture is somewhat complicated by the complex set of results obtained for log B in the ER conditions (see Table 9.17). In these conditions, there was a significant increase in log B in the low ER condition only, and for the strict and medium criteria only. It is difficult to interpret this result, as well as the obtained interaction between tasks and SP for the log B3 measure. This interaction indicated that the effect of SP on the risky criterion was significantly greater in the closure task than in the speed task. While this result again points to the influence of task demands, it has a low reliability, and it remains unclear why it was not observed for the other log B measures. The general conclusion for log B, as far as one can be made therefore, is that with time at work there is a significant increase in the strict and medium criteria, and, in some conditions, in the lax criterion. This results is in broad agreement with the results of Broadbent and Gregory (1963a, 1965), but also does not rule out the possibility that lax criterion movements over a monitoring session may also be present, as found in two experiments by Milosevic (1974, 1975). Criterion shifts in monitoring tasks are discussed in further detail in Chapter 13.

Analysis of the response latency data from a decision theory point of view gave results which were generally consistent with those obtained in Experiment 1. Furthermore, as a purely empirical result, the finding that Yes response latencies increase with time on task, while No response latencies decrease or remain constant is confirmed as a general feature of both visual and auditory monitoring. The effects on the response latencies of the two independent variables SP and ER were also consistent with the predictions of the working model outlined in 4.6, although the statistical reliability of all the effects was not established. The latencies of correct No responses appear to be particularly insensitive to the effects of task variables. The decision theory model predicts that CRL should decrease with an increase in the criterion (whether as a result of time on task or reduced SP), but although small decreases in mean latency were noted, overall no reliable effects on CRL were obtained. A similar tendency was noted in Experiment 1. This is probably due to the fact that in low signal probability tasks (as in all the monitoring tasks used in this thesis), criterion variation does not significantly affect the relative uncertainty in the decision to respond negatively to non-signals. In decision theory terms, a variation in criterion position will not radically affect the mean 'distance' of correct No 'observations' from the criterion when the false alarm rate is low, (as in most monitoring situations) so that there will be little resulting variation in CRL.

The assumption of a relationship between response latency and relative distance from the criterion was generally upheld in the results of the present experiment. Further evidence for this relation was provided by the finding that latency was a function of confidence level, with confident responses being much faster than less confident ones. It therefore seems clear that in the type of monitoring situation investigated, response latency is directly related to the relative strength of the received evidence on individual trials, relative to the critical level at which the signal is just reported. As noted in Chapter 4, such an interpretation of response latency represents a fixed sample model of latency, which proposes that the overall response latency is relatively unaffected by any variations in input sampling time (that is, in decision theory terms, in the time taken to accumulate the evidence). Neither the results of this experiment nor those of Experiment 1 provide any direct evidence for a substantial input sampling component of response latency, although, as Broadbent (1973) has suggested, a two-process model, with the input sampling stage feeding into a later decision stage, can be envisaged. We will not discuss such a model at this stage, but along with some recent observations of Broadbent (1975, 1976), shall consider some relevant latency models for monitoring tasks in Chapter 13. We may conclude this chapter by noting that response latencies in 'limited-hold' visual and auditory monitoring tasks are consistent with a fixed sample latency model based in, and extended from, the theory of signal detectability.

## 9.3 Conclusions

The results of Experiment 4 provide further support for the central theme of this thesis: firstly, that task classification is an important tool for the description and analysis of monitoring performance in different tasks and as a function of selected independent variables, and secondly, that decision theory provides a reliable analysis of both accuracy and latency aspects of performance. The more specific conclusions relating to sensitivity decrement, criterion shifts, and the interpretation of latency data in monitoring tasks are almost the same as those of Experiment 1 (see 7.4). Sensitivity decrement is not restricted to visual tasks, but may be observed for auditory tasks in which a speed-type of signal discrimination is combined with a high stimulus presentation rate.

Criterion shifts are observed for all confidence levels, but the movement in risky criteria is less than that for strict criteria.

# CHAPTER TEN

## INTERMODAL CONSISTENCY OF MONITORING PERFORMANCE

## 10.1 Introduction

The influence of task demands as defined on the speed-closure dimension, on decision processes in auditory monitoring performance was demonstrated in the results of the experiment reported in the previous chapter. The speed-closure task dimension describes reliably different categories of performance for both visual and auditory monitoring tasks; and thus the results obtained so far would seem to endorse the claim, made in Chapter 2, for the inclusion of this task dimension in a task classification system for monitoring tasks.

We have also seen that the results of Experiment 3 could be interpreted as providing an empirical basis for the speed-closure dimension; that is, as opposed to taxonomic approaches where classification categories are proposed on the basis of intuitive descriptions of tasks (as in some of the so-called 'utilitarian' taxonomies, see Chapter 2). However, since only visual tasks were investigated in Experiment 3, it is not clear whether the influence of task dimensions on the consistency in monitoring performance also obtains for performances across modalities. The experiment reported in this chapter accordingly directly follows up the results of Experiment 3, being concerned with the effects of the speed-closure classification on intermodal correlations in monitoring performance.

We have previously stated that "individual differences in monitoring performance are not so much task specific as task type specific" (see Chapter 8, section 8.2.3). This was originally intended to apply to the speed and closure signal types in visual tasks. To what extent this is true of cross-modal performance, however, remains to be determined. Furthermore, it might be proposed that modality specific factors are also important; indeed, 'from the armchair', it would appear that the difference between the demands of visual and auditory tasks are greater than between speed and closure tasks of the same modality, but this is a matter for empirical resolution.

The question of the effects of task demands is also important because it bears on some general theoretical issues about monitoring behaviour. Two general hypotheses about monitoring performance may be entertained: that performance is determined, on the one hand, by a general 'vigilance factor' (Tyler, Waag and Halcomb, 1972), and, on the other, by the unique characteristics of individual monitoring situations (Buckner, Harabedian and McGrath, 1960; Holland, 1963). The theoretical position adopted in this thesis lies between these two extremes. It proposes that, among other things, monitoring performance is a function of task demands, which may be identified on the basis of a few, non-overlapping task categories. Partial support for this view was indicated in the results of Experiments 1 and 3. The experiment reported in this chapter represents a further examination of this approach to the description of monitoring performance.

## 10.2 Experiment 5

### Summary

An investigation was carried out into the effects of the speed-closure and modality dimensions on the consistency of individual differences in different monitoring tasks. Two groups of men each worked with a visual and an auditory task matched on either category of the speed-closure dimension, while a third group worked with a pair of tasks classified across both dimensions. The results showed that intermodal correlations in performance are fairly high and significant for tasks matched on the speed-closure dimension, but almost zero for unmatched tasks. The results complement those of Experiment 3, and are discussed in relation to

hypotheses about the specificity of individual differences and the relative effects of signal type and modality.

### 10.2.1 Method

The experimental paradigm used in this study was similar to that employed in Experiment 3. As in that study, two speed and two closure tasks were considered, but within each signal type category a visual and an auditory task were used. Although sequence effects were not obtained in Experiment 3, transfer effects might be more likely when subjects switch from visual to auditory monitoring, and so the same design as in Experiment 3 was retained.

#### Subjects

Thirty male subjects, aged 19 to 28 years, served in the experiment. They were randomly assigned to three equal groups. Other characteristics of the subject sample are given in Table 6.2.

#### Apparatus

The apparatus and the experimental system were as described in section 6.2 in Chapter 6. Four monitoring tasks were used, the visual speed and closure tasks VS(1) and VC(1) from Experiment 3, and the auditory speed and closure tasks AS and AC, from Experiment 4. Reference may be made to these experiments for the task descriptions; the detailed descriptions are given in section 6.2.5.

The task variables signal rate (1/min.), event rate (15/min.) and duration (45 min.) were the same in each task. Task difficulty was adjusted, as before, to give roughly equal detectability for the four tasks. The 'single' response mode was used in each task.

Procedure

Each group of subjects worked with a different pair of tasks, one visual and one auditory, as follows:

| | | |
|---|---|---|
| Group 1 | : | VS and AS |
| Group 2 | : | VC and AC |
| Group 3 | : | VS and AC |

Thus, as in Experiment 3, tasks were matched on either category of the speed-closure dimension in Groups 1 and 2, but not in Group 3. Within each group, half the subjects performed the two tasks in the order shown, while the other half performed them in the reverse order. Testing was repeated at roughly the same time of day of the week following the first session. Subjects received the same standard training and expectancy matching practice as described previously (see 6.3).

### 10.2.2 Results

Four performance measures were utilized in the analysis of the data: the mean probability of Hits and false alarms (FAs), and the derived measures d' and log B. Each measure was computed within successive 15-min. time blocks of the monitoring sessions. The performance data are given in Table 10.1, as a function of time blocks and tasks.

A separate 2 x 2 x 3 (Testing Sequence x Tasks x Time Blocks) analysis of variance (ANOVA) was carried out for each group and for each measure. The summary tables are listed in Tables C5.1 to C5.12 in Appendix C5.

For Group 1, there was a significant decline in Hits and FAs ($p < .001$ in each case) and a significant increment in log B ($p < .001$) with time on task. Furthermore, there was a significant effect of Tasks for FAs, with a significantly greater proportion of FAs being made on AS than on VS ($p < .05$).

| | GROUP 1 | | GROUP 2 | | GROUP 3 | |
|---|---|---|---|---|---|---|
| | VS | AS | VC | AC | VS | AC |
| p(H) | .880 .847 .740<br>(.822) | .900 .880 .760<br>(.847) | .767..727 .700<br>(.733) | .833 .807 .733<br>(.791) | .787 .747 .693<br>(.742) | .807 .793 .747<br>(.782) |
| p(FA) | .088 .066 .029<br>(.061) | .116 .069 .036<br>(.074) | .046 .032 .024<br>(.034) | .063 .058 .029<br>(.050) | .044 .034 .026<br>(.035) | .056 .045 .031<br>(.044) |
| d' | 2.62 2.59 2.60<br>(2.60) | 2.55 2.71 2.58<br>(2.61) | 2.57 2.54 2.61<br>(2.57) | 2.63 2.59 2.56<br>(2.59) | 2.68 2.69 2.61<br>(2.66) | 2.51 2.59 2.5<br>(2.56) |
| log B | .098 .218 .720<br>(.345) | -.044 .070 .632<br>(.249) | .543 .727 .819<br>(.663) | .291 .360 .741<br>(.464) | .525 .673 .877<br>(.692) | .408 .496 .757<br>(.553) |

TABLE 10.1 The mean probability of Hits, p(H), and of false alarms, p(FA), and mean values of d' and log B as a function of successive time blocks, for each task within groups (see text; figures within brackets indicate mean performance levels averaged over time blocks).

For Group 2, the decline in FAs with time blocks was significant ($p < .05$), while the decline in Hits was marginally significant ($F = 4.55$; $df = 1,8$, $p < .10$ conservative test; $df = 2,16$, $p < .05$ risky test; see Table C5.5). There was a significant increase in log B with time on task ($p < .05$). Significantly more FAs were made on AC than VC within Group 2 ($p < .05$).

For Group 3, there was a significant decline in Hits ($p < .05$) and FAs ($p < .01$) and a significant increase in log B ($p < .01$) with time at work. There were no significant Task effects.

For all three groups, there was no significant change in d' over time blocks (see also Table 10.1). In addition, there were no reliable effects due to sequence of testing, although there were a few occasions when such effects approached significance (see Appendix C5).

The results of the various ANOVAs therefore indicate that, as far as performance trends over time are concerned, the expected results were obtained, and there were no contaminating effects due to sequence of testing.

| Dependent Measure | Group 1 (VS & AS) | Group 2 (VC & AC) | Group 3 (VS & AC) |
|---|---|---|---|
| Hits | .60* | .64* | .11 |
| FAs | .74** | .30 | .34 |
| d' | .63* | .67* | -.01 |
| log B | .62* | .33 | .33 |

TABLE 10.2 Intermodal correlations for four dependent measures and for each group (*$p < .05$; **$p < .01$).

Table 10.2 gives the product moment correlations of mean performance levels (averaged over time blocks) between modalities, for each group. These correlations show that individual differences in sensitivity and in the mean probability of Hits are fairly highly correlated across modes, (Groups 1 and 2), but almost completely uncorrelated for Group 3, who worked with unmatched tasks (on the speed-closure dimension). False alarms and log B were moderately correlated across tasks in all three groups.

Figure 10.1 shows the correlation coefficients when performances in corresponding subsections of the monitoring period are considered. No consistent pattern in the trend of the correlations over the monitoring period emerges. For d' in Group 3, individual differences between modes are uncorrelated throughout the monitoring period, which is a surprising result in view of the previous finding that performances are highly correlated for short detection periods even for unmatched tasks. However, for Hits, the same pattern of correlations as in Experiment 3 was obtained.

### 10.2.3 Discussion

Intermodal correlations in performance between visual and auditory monitoring tasks are fairly highly correlated for tasks matched on either the speed or closure category of the signal type dimension. For tasks classified across both signal type and modality dimensions, however, the results of Experiment 5 indicate that individual differences in monitoring performance are almost completely uncorrelated. These results therefore complement those of Experiment 3, and suggest that the speed-closure dimension of Theologus and Fleishman (1971) is an important factor in the determination of performance consistency, when either performances within or across modalities are considered.

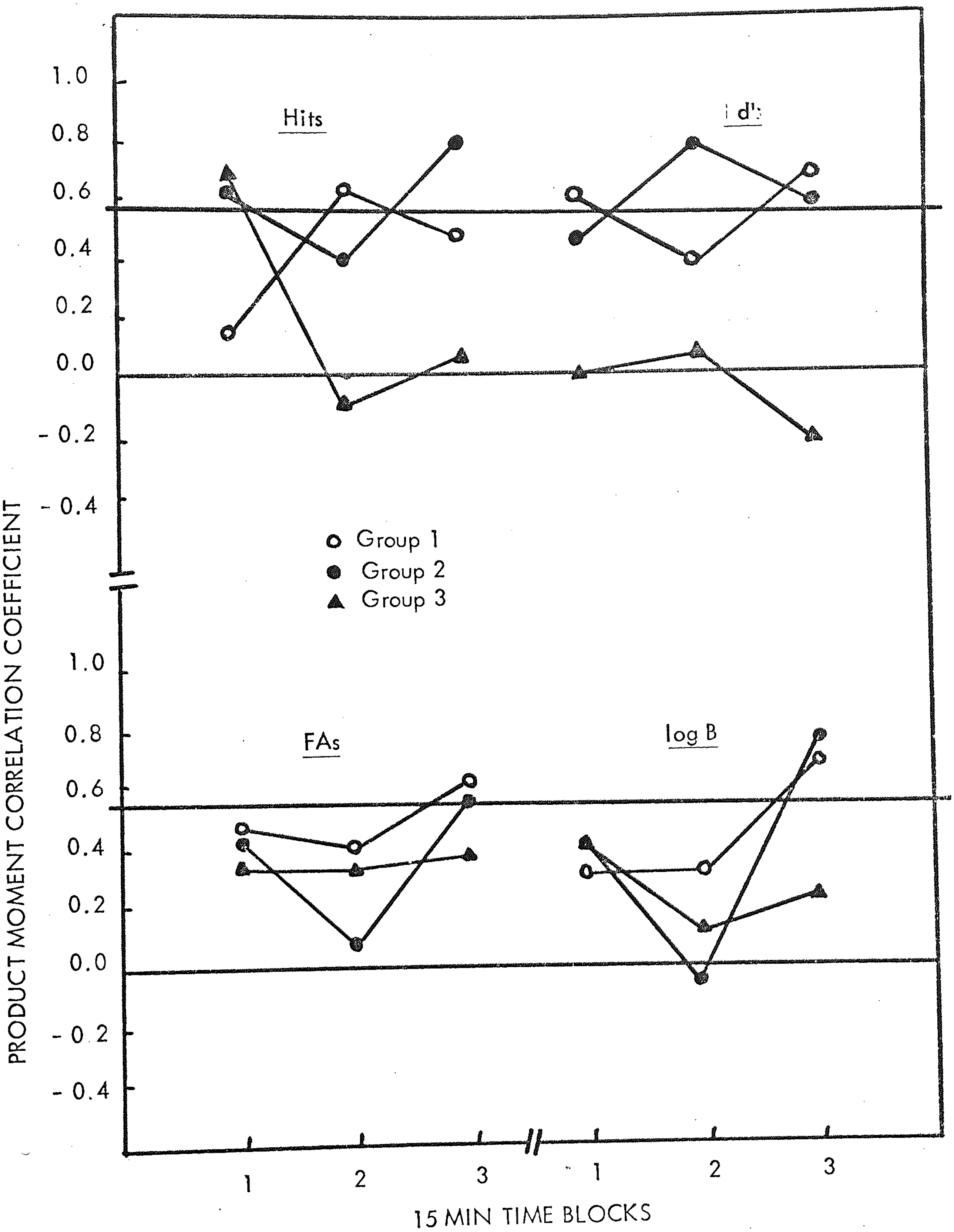


FIGURE 10.1 Inter-task correlations for each group for four dependent measures (see text; correlations above horizontal midlines are significantly greater than 0, $p < .05$).

In a review of recent studies of intermodal performance consistency in monitoring (section 3.2, Chapter 3), it was seen that a number of experiments have found that visual and auditory performances are significantly correlated if task difficulty is matched on an individual or group basis (Hatfield and Loeb, 1968; Loeb and Binford, 1971; Tyler et al., 1972; see Type B studies in section 3.2). In general, these studies were concerned with the view that performances in different monitoring tasks should be correlated, if the status of 'vigilance' as an independent factor is to be preserved. In this thesis, on the other hand, it has been proposed that the limits of performance consistency have to be established by empirical investigations, not by fiat; on such considerations, we can therefore conclude that the generality of the 'vigilance' factor is not compromised by the occurrence of low correlations for certain tasks, but merely reflects the influence of differing task demands. Furthermore, the finding of high correlations for the previously reviewed Type B studies reflects the fact that these studies employed very similar monitoring tasks. Task difficulty is only a crucial factor insofar as it relates to the relative difficulty of different task demands. Thus task difficulty was controlled in all three groups in the present experiment and in Experiment 3, yet for the groups working with tasks of different signal type, the inter-task correlations were low and not significant. Task difficulty is therefore a necessary but not sufficient condition for determining whether performances on two monitoring tasks correlate or not.

Finally, a general feature of the present results is that the obtained correlation coefficients were consistently lower than the correlations observed for visual tasks in Experiment 3. For Groups 1 and 2, for example, only about 36-40% of the variance in performance sensitivity (d') was shared between matched tasks of different modalities, as compared to corresponding values of about 65-70% for tasks in the same modality in Experiment 3. The correlations for Group 3 were also lower in Experiment 5 than in Experiment 3. Therefore, while modality does not appear to be a dominant factor, in itself, in determining consistency of performance in monitoring situations (at least when the influence of other factors, such as search requirements, are not considered), there is some evidence for mode specificity even in simple monitoring tasks. On balance, therefore, task modality may be retained as a task dimension in the task classification system outlined in this thesis.

## 10.3 Conclusions

Intermodal correlations for visual and auditory monitoring tasks are fairly high and significant for tasks matched on either category of the speed-closure dimension. For tasks classified across both the modality and signal type dimensions, performance consistency across tasks is almost zero. However, there is also some evidence pointing to the influence of modality specific factors. Task difficulty is a necessary but not sufficient condition for determining performance consistency between different tasks.

# C H A P T E R   E L E V E N

## RESPONSE AND EVOKED POTENTIAL LATENCY CORRELATES OF DECISION PROCESSES IN VISUAL MONITORING

## 11.1 Introduction

In the experimental investigations reported thus far, it has been argued that different aspects of monitoring behaviour may be profitably analysed within the framework of decision theory. In particular, the results of Experiments 1 and 4 have demonstrated that changes in response latency in monitoring tasks are, for the most part, consistent with the predictions of a decision theory model of latency based on and extended from the theory of signal detectability (TSD). In this chapter we shall consider response latency and the decision theory model in somewhat greater detail, and from a psychophysiological point of view.

In a previous discussion of the psychophysiological concomitants of decision processes in detection and monitoring tasks (section 4.7), some reasons as to why the evoked potential may be regarded as a potentially reliable index of decision-making activity, superior to other physiological measures, were proposed. As a purported measure of brain electrical activity, the evoked potential (EP) may provide a more 'central' index that other measures; furthermore, such activity may be precisely time-locked to externally observable events such as stimulus onset or response evocation within short intervals, so that a close link can be established, in principle, between the physiological activity and the decision-making activity. Against these advantages are the disadvantages that the EP cannot be recorded on a single trial, but is an average measure, and that it may be difficult, in practice, to record EPs without radically changing the experimental situation under study. However, a more compelling factor for the use of EPs stems from the results of some recent studies indicating a relationship between certain EP measures and behavioural measures of decision-making processes, (Hillyard, Squires, Bauer and Lindsay, 1971; Paul and Sutton, 1972; Squires, Squires and Hillyard, 1975a, b). These studies have demonstrated a reliable relationship between the amplitudes of 'late' EP components and decision criteria in signal detection tasks. In particular, it has been shown that the amplitude of a late component of the EP (generally referred to as P3 or P300) is directly related to the confidence of a decision. Paul and Sutton (1972) reported that the relationship between P3 amplitude and the decision criterion was the same whether criterion variation was effected by changes in a priori signal probability or in the payoff matrix; and Squires et al., (1975a, b) have shown that the same relationship holds if different criteria are generated with the use of a rating scale.

The EP studies reviewed in 4.7 have been concerned mainly with amplitude measures of the EP. Since we are interested in response latencies, there is somewhat greater intuitive appeal in considering EP latency measures. By commencing the averaging of the electroencephalograph (EEG) with the onset of a stimulus and terminating it selectively according to different responses, an estimate of the averaged electrical activity following the stimulus can be obtained, and thus, hopefully, of the decision-making activity preceding the response. If the time when the response is made varies, a change in the form of the preceding electrical activity might be expected to occur; given the assumption of a relationship between decision criteria and the EP, this might take the form of a concomitant variation in the time taken for the major EP components to develop or peak. Of course, whether these changes occur in a systematic fashion remains to be demonstrated. There is some evidence to suggest that the latencies of certain EP components are sensitive to both within and between-subject variations in response latency (Donchin and Lindsley, 1966; Ritter, Simson and Vaughan, 1972), but otherwise the relationships between EP latency and response latency remains to be fully explored. In the present experiment such an investigation into response and EP latencies was carried out for the different responses made during a monitoring task.

## 11.2 Experiment 6

Summary

Eight subjects performed a 40-min. visual speed monitoring task in which they were required to detect a decrease in the brightness of an intermittently flashing light source. Response latencies and the evoked potentials (EPs) were recorded separately for the different response categories in the binary response mode. EPs were also recorded selectively according to the responses made on a four-category confidence rating scale for a further two subjects. Results showed that there was a significant increase in both correct and incorrect Yes response latencies with time on task, and a decrease in incorrect No response latencies. The increment in Yes latencies was reliably related to increases in the latencies of late components (greater than about 200 msec.) of corresponding EPs. Furthermore, all subjects took longer, on average to make an incorrect Yes response (false alarm) than a correct one (Hit), and this increase in decision time was also reliably related to an increase in the late component latencies of EPs averaged to false alarms over those of EPs averaged to Hits. Finally, the increase in decision time for 'doubtful' responses as compared to 'confident' responses was reflected in similar differences in latency between EPs averaged separately to these behavioural categories. The results were interpreted as supporting the view that late EP component latencies may, under certain conditions, represent temporal correlates of decision-making activity in monitoring tasks.

### 11.2.1 Method

In this study, the relationships between decision latencies and evoked potential latencies were examined in relation to performance in a monitoring task. Evoked potentials were recorded separately for the four response categories of the binary response mode, as were the response latencies. EPs associated with responses of different confidence level were also recorded for a further two subjects who worked on the same task using a four-category rating scale.

Subjects

The subjects were eight young adult males aged 18 to 24 years who served in the main experiment. A further two male subjects, aged 21 and 22 years, served in the condition in which the rating scale was used. All subjects had normal electroencephalographs (EEGs) and no visual abnormalities. The subjects were students who were familiar with EEG recording procedures, having previously participated in EEG experiments.

Apparatus and procedure

The apparatus used was as described earlier, except that a 10 bit analog-digital (A-D) convertor was used for EEG digitization and subsequent averaging. The same visual monitoring task, VS1, as that used in Experiment 1, 2 and 3 was used, except that a flash duration of 33 msec. was employed. This task required the detection of a decrease in brightness of a intermittent circular light flash. The light flashes were presented to the right eye in approximately Maxwellian view at a visual angle of $2^{o}$. Subjects fixated a pink dot appearing at the centre of the flashes and subtending 19' arc at the right eye. Event rate was 15/min. and the mean signal rate 3/min. This relatively high signal rate was used to ensure that an adequate number of responses for EEG averaging would be obtained. Task duration was 40 min.

The EEG was recorded continuously from scalp electrodes at the right occiput (02 in the 10/20 International Electrode System) and at the right earlobe (A2) with a ground electrode at the left earlobe (A1). Silver-silver chloride stick-on electrodes were attached to cleansed areas of the scalp with collodion glue and filled with electrode gel. Contact resistance was adjusted to be always below 5K ohms, and on average about 4K ohms. After electrode resistances had been checked the electrode cups were sealed with more collodion glue to minimize the changes of the electrolyte drying up during the monitoring session.

Scalp signals were augmented by a pre-amplifier and the EEG was monitored with an SLE Galileo recorder. The bandpass of the system was 1 - 50 Hz. A low frequency cut-off at 1 Hz gives 3 dB attenuation of the signal at 1 Hz and more than 3 dB (6 dB per octave roll-off) at lower frequencies; thus with a interstimulus interval of 4 secs., this results in any inter-stimulus 'slow-wave' activity, such as the contingent negative variation (CNV), being attenuated.

The EEG was digitized at the rate of 1 sample every 5 msec. for 500 msec. following stimulus onset (allowing a factor of 5, this sampling rate gives unaliased reproduction of frequency components up to 40 Hz). The digitized sample values were stored on magnetic (DEC) tape for subsequent summation. A program SUMAVE summated the sample values separately according to the behavioural response, and PLOTER plotted the averaged EPs on a Calcomp X-Y Plotter (see also 6.2.3).

The EP was calibrated by feeding a 20 msec. 5 $\mu$V pulse through the recording system prior to flash onset. A continuous paper record of the EEG was also obtained. The time onset of artifacts due to muscle activity and eye blinks were marked on the paper record. On the few occasions in which these artifacts were present <u>within</u> the averaging epoch, the corresponding sampled EEG values were excluded from the averaging process.

Before and after each monitoring session, noise runs were conducted to obtain estimates of the EP noise level and the extent of artifact contamination. Subjects were instructed to view the circular stimulus for about 2 minutes while randomly responding to the soft clicks that preceded flash onset (these were not audible in the monitoring session since background noise was relayed over headphones). The flash light was turned off in these noise runs, so that the resulting EP represents the sum of the background EEG and artifacts, including possible motor potentials. However, the latter are maximal in vertex and rolandic sites and are unlikely to have contaminated the EPs since occipital derivation was used. Figure 11.2 shows the noise run EPs before and after a monitoring session for three subjects. It can be seen that the noise level is within $\pm$ 2.5 uV. With a mean EP amplitude to flashes of 10 - 15 uV, this represents a signal to noise of about 5:1, which is acceptable. Theoretically, noise levels can be infinitely improved given perfectly uncorrelated background noise; with this assumption, it can be shown that signal to noise ratio can be improved by a factor of $\sqrt{N}$ given N additional samples. In practice, however, since the EEG does not describe a perfect Gaussian process, a limit of about 7:1 signal to noise ratio is only possible.

Artifact rejection was limited to visual inspection of the EEG record. However, a high pre-amplifier gain was used so that the chance of a large artifact contributing to the EP was minimized. It should also be noted that pilot studies revealed that there were only minor differences in the EPs to bright and dim flashes when these were recorded in no-response conditions. Finally, control runs, in which the subjects were instructed to merely count critical signals or to delay response by 1 sec., indicated that overt responding did not contaminate the EP.

The sequence of events within each experimental session was as follows:

| Time into session | |
|---|---|
| 0 | Subject arrives. Electrodes attached. |
| 10 | Electrode resistances and base EEG activity checked. |
| 15 | Electrodes sealed. Task familiarization. |
| 17 | Initial instructions. |
| 20 | Training period 1. |
| 25 | Training period 2. |
| 30 | 10-min. practice (expectancy matching) session. |
| 40 | Noise run A. |
| 45 | Final instructions. Short rest period. |
| 50 | Monitoring session begins. EEG and EPs recorded. |
| .. | |
| .. | |
| 90 | End of monitoring session. Noise run B. |
| 95 | End of experimental session. |

### 11.2.2 Results

Detection measures

Performance scores were averaged for each subject for the two halves of each 40-min. monitoring session to yield the hit and false alarm probabilities, p(H) and p(FA), and the TSD indices d' and B. Performance trends over the two halves of the task were analysed using Wilcoxon's matched pairs signed-ranks test (Siegel, 1956). Table 11.1 gives the mean values of the detection measures for both halves of the task.

Statistical analysis revealed that the probability of hits declined marginally over the two halves of the monitoring session ($p < .10$), while the probability of false alarms declined significantly ($p < .05$). For the TSD measures, there was a significant increase in B over time ($p < .01$), while d' did not change significantly between the two halves of the task. Thus, apart from the marginal significance level for the probability of hits measure, the results for the detection measures were in agreement with expectations.

| Dependent Measure | 1st Half | 2nd Half |
|---|---|---|
| p(H) | 0.663 | 0.542 |
| p(FA) | 0.117 | 0.085 |
| d' | 1.66 | 1.50 |
| B | 1.78 | 2.48 |

TABLE 11.1 Mean values of hit and false alarm probability, and of d' and B in each half of the monitoring task.

| Latency Measure | 1st Half | 2nd Half |
|---|---|---|
| DL | 693 | 751 |
| FAL | 895 | 973 |
| CRL | 776 | 769 |
| ML | 819 | 804 |

TABLE 11.2 Mean response latencies (msec.) for the four categories of response, detection latency (DL), false alarm latency (FAL), correct rejection latency (CRL) and omission latency (ML) for both halves of the task.

## Response latencies

The mean response latencies for the four categories of response, detection latency (DL), false alarm latency (FAL), correct rejection latency (CRL) and omission latency (ML), are given in Table 11.2. As before, latency differences between the two halves of the task were tested for significance using Wilcoxon tests.

The results indicated that there was a significant increase in the mean latencies of Yes responses with time on task ($p < .01$ for DL, $p < .05$ for FAL). This increase in latency for correct positive responses was observed in all eight subjects, and six subjects had longer FA latencies in the second half of the task. At the same time, there was a drop in the mean latencies of both negative (No) responses, but only the drop in ML was significant ($p < .05$).

Comparisons were also made between the mean latencies of the different response categories. As is apparent in Table 11.2, FAL is substantially greater than DL. This increase in decision time for incorrect over correct Yes responses was significant in both halves of the task ($p < .01$ in each case). Furthermore, the mean latency of incorrect No responses (ML) was significantly greater than that of correct No responses (CRL), in each half of the task ($p < .01$ in each case). Finally, omission latency was significantly greater than detection latency, but only in the first half of the task ($p < .01$). The results of these pair-wise comparisons in latency may be summarized in the relation

$$\text{FAL} > \text{ML} > \text{DL} = \text{CRL}$$

where $>$ represents "significantly greater than' and = "not significantly greater than".

This relation applies for both halves of the task, except that, as noted before, DL was not significantly different from ML in the second half of the task. This is probably because criterion placement was such that $p(H) \approx 0.5$ in this half (mean $p(H) = 0.54$). This is further illustrated in Figure 11.1, in which the latencies of correct (Hits) and incorrect (omissions) responses to signals for individual subjects are plotted against response probability. It is evident that incorrect responses (omissions) have longer latencies when hit probability is greater than 0.5, as predicted by the decision theory latency model. For the region around $p(H) = 0.5$, hit and omission latencies overlap.

Evoked potentials

Evoked potentials were averaged separately according to response category,

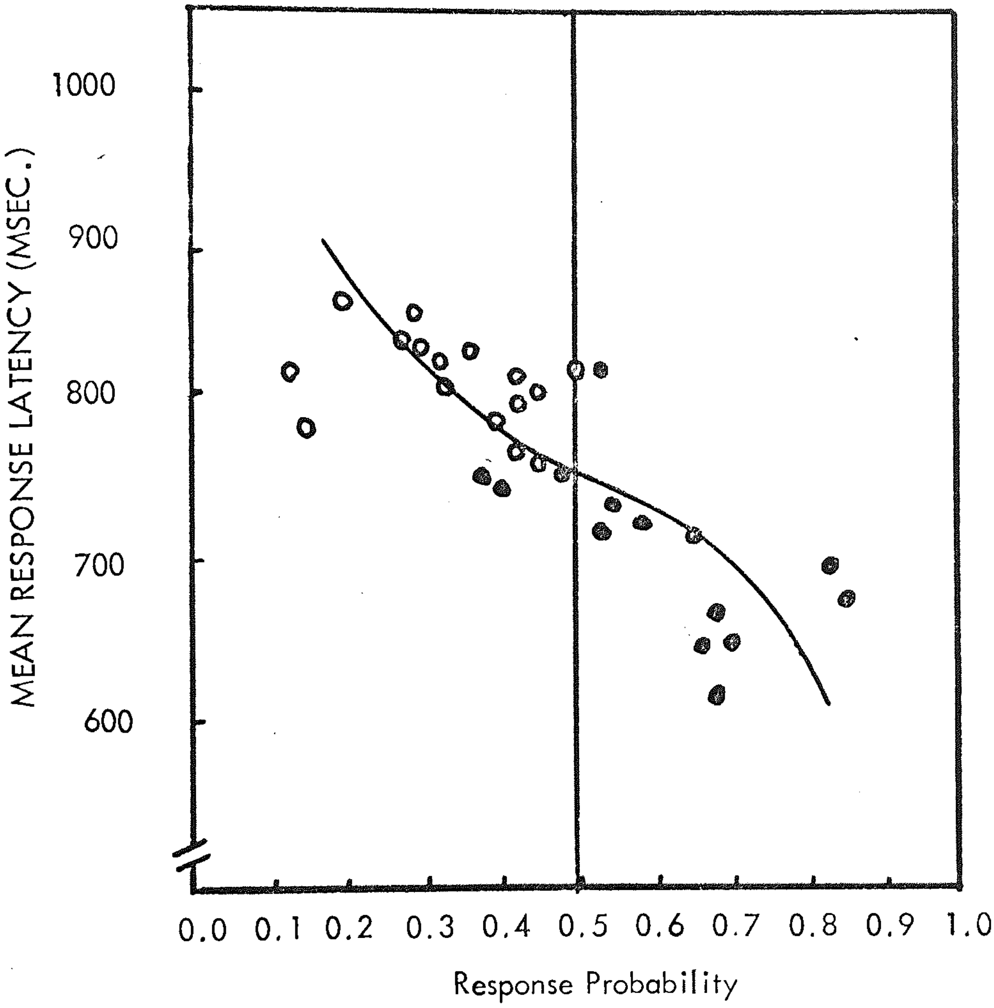


FIGURE 11.1 Latency-probability plot for individual response latencies to signals in either half of the task (correct detections denoted by closed circles and omissions by open circles).

for both halves of the task. Since there were insufficient numbers of omission responses for EEG averaging, EPs to omissions were not recorded. In order to restrict the analysis to positive (Yes) responses, therefore, only EPs averaged separately to hits and false alarms are reported. On average, about 40 response samples went into each EP.

Individual EPs for Hits and FAs are shown in Figure 11.2 for both halves of the task and for three subjects. The morphology of the EPs is generally similar to that expected for visual EPs recorded from occipital sites. The EPs were quantified by defining individual components P1, N1, P2, N2 and P3. Individual EP peaks were identified as the most positive or most negative turning points in each EP waveform, within 15 msec. windows (that is, the labels P and N were only attached to major positive and negative peaks where the inter-peak interval was greater than 15msec.). Also, only peaks having amplitudes outside the noise band of $\pm$ 2.5 uV were identified. The baseline (zero amplitude) was established by the mean EEG level in a 50 msec. period preceding the stimulus, and zero time was taken as the crossover of the calibration pulse and the baseline. The latencies of each peak were measured by hand to an accuracy of $\pm$ 2 msec. Where there was a certain ambiguity in the exact peaking time of a component, (as for P3 for subject 1 in Figure 11.2) the latencies as assessed by the author and by another rater were averaged. Generally there was only about 10 msec. difference between the peak times of the two raters.

Table 11.3 gives the mean latencies of the main EP components for both Hit and FA EPs in either half of the task. The latencies of the 'late' components P2, N2 and P3 increased significantly from the first to the second half of the task ($p < .01$ in each case; the Wilcoxon test was used in this and subsequent comparisons of latency between halves of

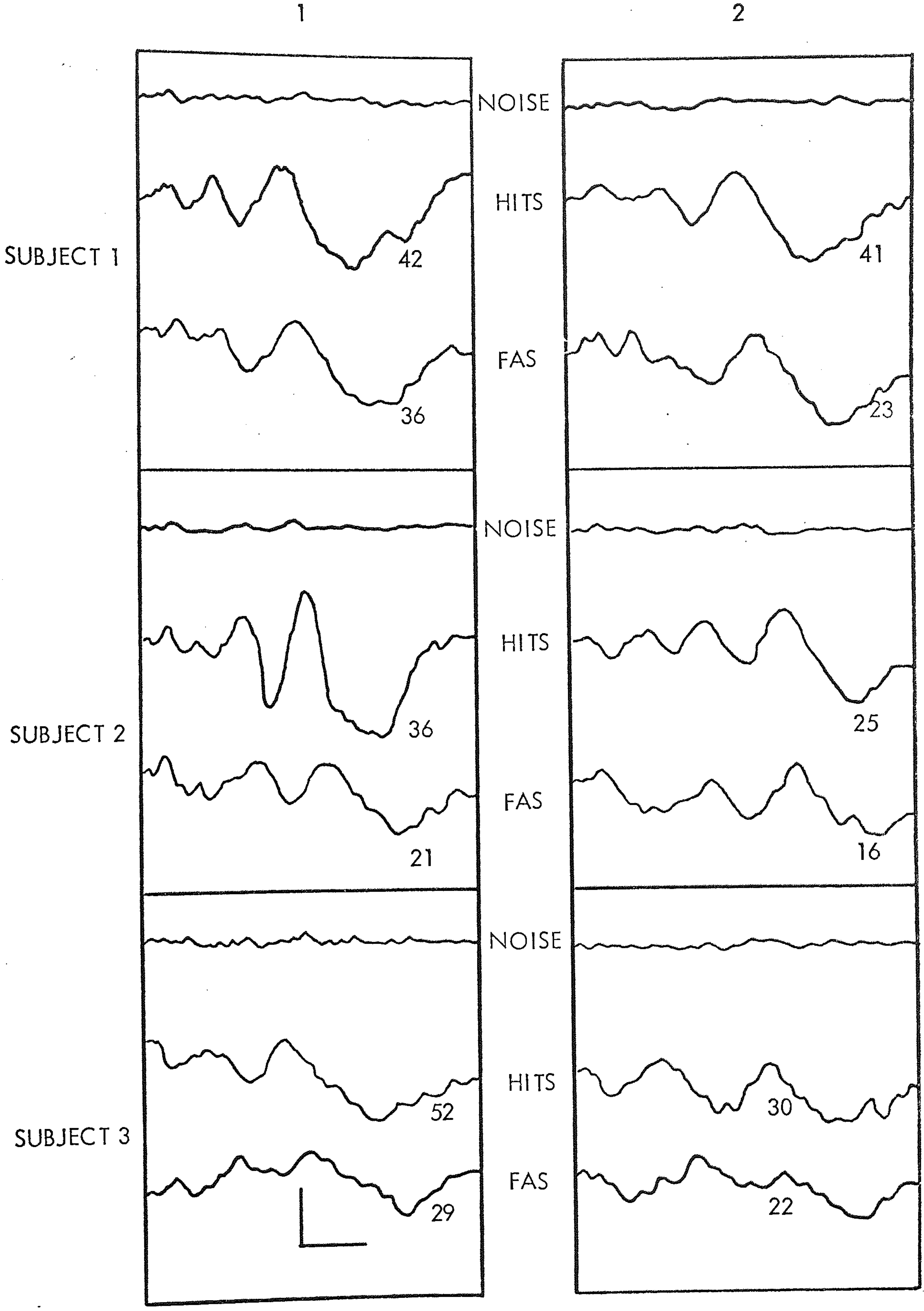


FIGURE 11.2 Occipital potentials averaged separately on correct (HITS) and incorrect (FAS) positive response trials in both the first (1) and second (2) halves of the task. The averaged potentials in the no-stimulus (NOISE) conditions before (1) and after (2) each monitoring session are also shown. (Calibration: 100 msec, 10 microvolts, negativity upwards; the numbers beside each trace refer to the number of responses in each average evoked potential).

the task and between response categories). However, there were no significant change in the latencies of the 'early' components P1 and N1.

The latencies of the late components P2, N2 and P3 in the EPs to FAs were significantly longer than the latencies of corresponding components in the EPs to Hits ($p < .01$ in all cases). This is apparent in the EP plots in Figure 11.2 as well as in Table 11.3. Again, this did not hold for the early component latencies, P1 and N1 latencies being approximately the same in Hit and FA EPs.

A further link between the EP late component latencies and decision latencies is illustrated in Figure 11.3, in which the EPs for the subjects working with the rating scale have been plotted. These EPs were averaged to Hits made with two degrees of confidence, "Certain" and "Doubtful". It is immediately apparent that Doubtful EPs have longer late component latencies than Certain EPs; in accordance with the decision theory latency model, these latency differences are accompanied by an increased mean latency for "Doubtful" responses over "Certain" responses, and an increase in the latencies of both response categories over time at work. For one subject, Certain Yes responses had a mean latency of 602 msec.,

| | HITS | | FALSE ALARMS | |
|---|---|---|---|---|
| EP Component | 1st Half | 2nd Half | 1st Half | 2nd Half |
| P1 | 75 | 78 | 70 | 73 |
| N1 | 160 | 174 | 165 | 176 |
| P2 | 218 | 251 | 249 | 261 |
| N2 | 271 | 285 | 301 | 332 |
| P3 | 341 | 360 | 412 | 424 |

TABLE 11.3 Mean evoked potential component latencies (msec.) for EPs averaged to Hits and False Alarms in the two halves of the task.

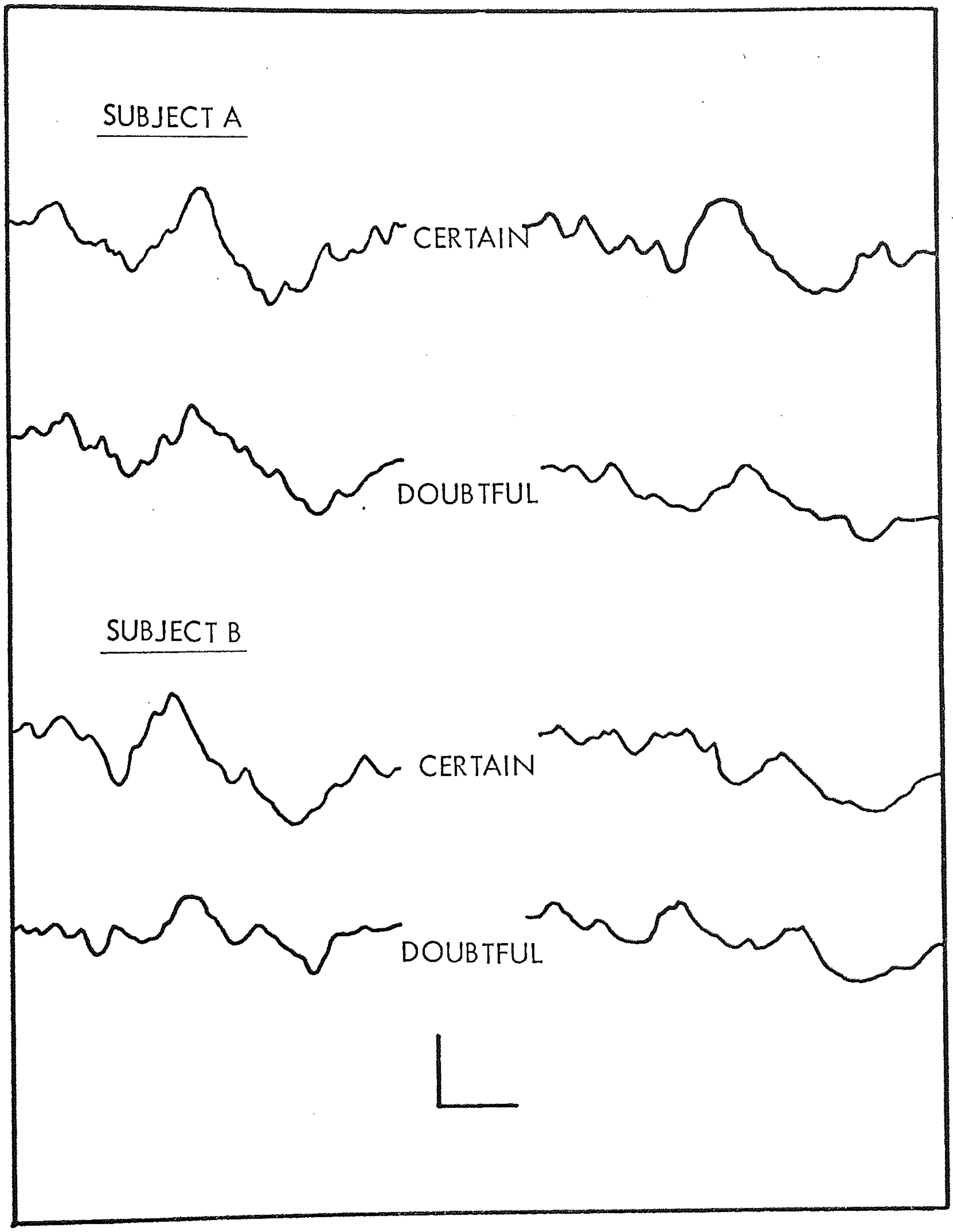


FIGURE 11.3 Averaged evoked potentials on 'Certain' and 'Doubtful' positive response trials in both the first and second halves of the task (left and right hand traces, respectively). (Calibration as in Figure 11.2).

increasing to 635 msec. in the second half of the task, while Doubtful Yes responses increased in latency from 680 msec. to 695 msec. Similar trends were observed for the other subject, and both subjects exhibited the usual finding of stable sensitivity and an increasing criterion over time on task.

### 11.2.3 Discussion

With increased time at work on a monitoring task, the mean response latencies associated with positive responses (Hits and false alarms) significantly increase, while those associated with negative responses (correct rejections and omissions) decrease. For low signal probability monitoring situations however, where there is little effect on decision criterion variation on correct rejection probability (since the criterion is set relatively high), the decrease in correct rejection latency may not be significant, as in the present experiment and in Experiment 4. The trends in the other latencies are however, consistent with the results obtained in previous experiments and in agreement with a decision theory latency model relating latency and the response criterion.

In the present study, a further relationship has been established between response latency, the criterion and late EP component latency. Changes both between and within subjects in the mean latencies of positive responses were shown to be reliably related to similar changes in the latencies of the 'late' components (P2, N2 and P3) of EPs averaged to Hits and FAs. Furthermore, the results indicated that throughout the 40-min. task, correct identifications of events as signals (Hits) were made significantly faster than incorrect identifications of non-signal events as signals (FAs) and this increase in decision time for FAs was accompanied by similar increases in the peaking times of the late components of EPs averaged to FAs over those averaged to Hits.

The selective averaging and analysis of brain electrical activity corresponding to different perceptual-response categories therefore supports the decision theory view that decision latency is a function of the relative strength of the evidence pointing towards or against the presence of a signal; variations in the relative strength of the evidence due to variations in criterion are reflected in parallel changes in decision latencies and, in the discrimination situation studied in this experiment, in the latencies of the late components of the averaged EPs. We have seen in Chapter 4 that the results of some recent studies have suggested that the amplitude of the late components, especially P3, are sensitive to variations in the decision criterion, (Paul and Sutton, 1972; Squires et al., 1975a, b). The present results are consistent with this suggestion, although we have mainly dealt with latency measures in this experiment. As is apparent in Figure 11.3, however, the EPs to responses made with high confidence have larger amplitudes as well as shorter component latencies than those for EPs averaged to responses made with a lower confidence. This result is consistent with the notion of a relationship between EP amplitude and rating criteria, as demonstrated by Squires et al. (1975a, b).

It should be noted that it is unlikely that the decision related evoked potentials obtained in the present experiment were contaminated by concomitant variations in the resolution time of post-stimulus CNV, since the bandpass of the recording system ensured that any slow wave components were attenuated and not included in the averaging process. Although some investigators have suggested that the high amplitude late EP components (e.g. P300) may represent a 'reactive return' of the CNV to baseline (Naatanen, 1975; Wilkinson and Ashby, 1974), Donchin, Tueting, Ritter, Kutas and Heffley (1975) have shown that CNV and P300 have different topographical and latency distributions, and may be considered to be relatively independent.

The observed relationship in the present experiment between EP latency and decision criteria in monitoring situations provides some support for the speculation that the form of the averaged electrical activity is different according to the time taken for the corresponding behavioural response to be evoked, that is, the average peaking time of late EP components is directly related to the time taken to make a corresponding response. Furthermore, the finding that only the late components were related to changes in response latency is significant: if the results were due to chance covariation of the response and EP latencies, one would expect some chance correlation in the early component latencies too; and it is precisely these components which have been shown not to be influenced by variations in the decision criterion (Paul and Sutton, 1972, 1973; Squires et al., 1975a, b).

Finally, it is true that the finding that the late component latencies increased with time on task does not, in isolation, provide evidence for the latency-criterion relationship, since this could be interpreted as an arousal-based physiological decrement, as have the findings of a number of other psychophysiological studies in vigilance (see 4.7.2). However, the trends in the latencies for EPs to FAs, and their relation to the Hit EP latencies provide more convincing support for the decision theory interpretation, as does the relationship between the latencies of EPs averaged to responses of different confidence level and the latencies of these responses.

## 11.3 Conclusions

The results of Experiment 6 show that the average time taken to make positive identifying responses in a monitoring situation increases with time at work, while the time taken to respond negatively either decreases or remains constant. For both types of response, incorrect responses have longer latencies than correct responses. The trends in latency for positive responses are reliably related to variations in the latencies of late components of evoked potentials averaged to corresponding response events. These results establish the decision theory interpretation of response latencies, and suggest that evoked potential latency measures may provide temporal correlates of decision making activity in monitoring situations.

# C H A P T E R   T W E L V E

## SENSITIVITY DECREMENT IN MONITORING TASKS: A TASK CLASSIFICATION ANALYSIS

## 12.1 Introduction

One of the four dimensions comprising the taxonomic system proposed in this thesis is the complexity of monitoring tasks. In Chapter 2 this dimension was identified more directly with the number of sources to be monitored in a display, since complex tasks having multiple sources may be distinguished from those combining the main monitoring task with another task. Both types of monitoring situation have been investigated in the literature (for example, see Hamilton, 1969; Hockey, 1969), but there have been few studies providing data on sensitivity and criterion trends in multisource tasks. In this chapter we shall consider certain aspects of source complexity in greater detail.

Two studies are reported in this chapter. In the first, Experiment 7, the relationship between the source complexity and speed-closure dimensions and their effects on sensitivity shifts in monitoring performance were investigated. In the second, an organizational study of a portion of the monitoring literature is reported. In this study a task classification analysis of all the experiments reporting sensitivity and criterion data was carried out.

The results of Experiments 1, 2 and 4 have shown that for both visual and auditory monitoring tasks, there is a decrement in sensitivity when the rate of stimulus presentation in the task is high. This decrement in perceptual efficiency is, however, only observed when a particular type of signal discrimination is involved, which we have termed perceptual speed. Closure tasks do not show such a decrement. It will also be recalled that the source of the difference between these two types of task was traced to the differences in the temporal sequence of the mechanisms involved in the discrimination of each type of signal. In speed tasks it was hypothesized that information is integrated over events, while in closure tasks it was assumed that the detection of a signal is made within a single event.

The relationship between the source complexity and the speed-closure dimensions becomes apparent when the effect of the addition of extra signal sources to a single-source speed task is considered. In such a task, which generally involves the detection of a change in some attribute (intensity, duration etc.) of a repetitive stimulus, a second source carrying the same stimulus may be added. If, however, the signal stimulus is only presented on one source, the signal may be discriminated by comparing the sources within a single event. On this reasonable assumption, the addition of extra sources to a speed task has the effect of 'turning it into' a closure task. Using the criterion of sensitivity decrement, this may be examined by studying the trends in perceptual sensitivity for multi-source speed and closure tasks.

The examination of performance trends in multi-source tasks is also relevant to the need, which was well expressed in Swets and Kristofferson's (1970) paper, to specify the types of monitoring display which yield a sensitivity decrement. It may be that such decrements are not obtained for multi-source tasks, although there is no direct evidence on this point. What evidence there is comes from early studies, not analysed with decision theory, which indicate that although mean performance may be reduced when more than one source has to be monitored, there is little or no decrement in multi-source tasks (Jerison and Wallis, 1957). Wiener (1964) however found that the same decrement obtained for both one-source and three-source monitoring. In reviewing these studies, Broadbent (1971) suggested that the lack of decrement observed in Jerison and Wallis's three-source task was due to a low response criterion induced by a signal rate three times as high as for the one-source task. In these studies, 'decrement' refers to the decline in the probability of detection over the monitoring period, there being little information on trends in false alarms, which would permit, in principle, an analysis in terms of sentitivity and bias. Recently, however, Milosevic (1974) carried out such an analysis, and found that while there was a reduced decrement in detections for a five-source task compared to a single-source task, there were no differences between tasks in the trends in sensitivity (d') over the monitoring period. Although Milosevic used a speed task, his results do not bear on the hypothesis linking the source complexity and speed-closure dimensions to sensitivity decrement, since he used a low event rate, which as we have seen, does not lead to a sensitivity decrement in either speed or closure tasks.

## 12.2 Experiment 7


### Summary

Two studies investigating sensitivity decrement in visual monitoring tasks are reported. In Experiment 7, the trends in sensitivity for three high-rate multi-source speed and closure tasks were compared. It was found that a significant decrement in sensitivity was obtained only for the task in which signal discrimination involved the reporting of a change in a stimulus relative to preceding stimuli. For the two tasks in which signal and non-signal features were present within a single event, no decrement occurred. These results were confirmed and extended in the second study, in which a task classification analysis of 33 experiments reporting sensitivity data was carried out. This analysis indicated that, despite some inconsistent results in the literature, task classification enables the improved specification of the types of monitoring display yielding sensitivity decrement. Such a specification can be made in terms of three of the task dimensions comprising the task classification system.

### 12.2.1 Method

In this experiment the sensitivity trends over a monitoring session for three multi-source monitoring tasks were investigated. In all three tasks stimuli were presented at a high rate, so as to maximize the chances of detecting a sensitivity decrement. Two of the tasks used were the multi-source equivalents of the (single-source) visual speed and closure tasks VS1 and VC1, which have been used in previous studies. The task combining VC1 with an extra source is referred to as the multi-source closure task (MC). The task combining VS1 with an extra source is termed the multi-source speed-closure task (MSC), because of the (hypothesized) change to a closure task with the addition of an extra stimulus source.

If the hypothesis advanced in the introduction to this chapter is supported therefore, a sensitivity decrement should not be found in either MC or MSC. However, as this would only, as it were, provide negative evidence in support of the hypothesis, a third task was used to provide for the possibility of more positive evidence. This task was designed so that, using a speed task with an extra source, the speed type of signal discrimination (the 'signal' defined relative to preceding stimuli) was retained even though another source was added. This was done by presenting the signal stimulus on both sources, so that a comparison between sources would not enable the discrimination of the signal. Under the hypothesis put forward, therefore, a significant sensitivity decrement should be observed only for this task, termed the multi-source speed task (MS). This is a rather surprising prediction, since this task, with its inherent source redundancy, appears intuitively to be the least 'difficult' of the three tasks. As Kahneman (1973) has pointed out, however, intuitive notions of task difficulty need not necessarily be related to the complexity of the demands a particular task makes on the observer. The 'objective' difficulty was the same for all three tasks, as the tasks were matched to give approximately d' values under 'alerted' conditions.

Subjects

Thirty six subjects aged 18 to 27 years participated in the experiment. They were randomly assigned to one of three equal groups, with the restriction that each group contained an equal number of male and female subjects. Each subject reported having good, uncorrected vision.

Apparatus

The apparatus and experimental system was as described previously, except that two display sources instead of one were used (each fitted with variable duration shutters as described in 6.2.2). Three tasks were used: Multi-source speed-closure task (MSC): This task was the two-source equivalent of the speed task VS1. Two circular light sources, separated by about 35 cm., had to be monitored for the occasional appearance of a slightly less intense flash on one of the sources. The sources came on intermittently every 2 secs and signals appeared at either source in a pseudo-random manner, with the restriction that an equal number of signals appeared at each source. The signal schedule, the intensity of the light flashes, the stimulus sizes, and all other features of each source were the same as for the VS1 task described preiously.
Multi-source closure task (MC): This task was a two-source version of the closure task VC1. Each source came on every 2 secs. and subjects had to detect a circular pink spot appearing at the centre of one of the sources. An equal number of signals appeared at either source.
Multi-source speed task (MS): This task was the same as MSC except that signals (slightly dimmer flashes) were presented simultaneously at both sources.

Further details of these tasks are given in section 6.2.5. The task variables signal rate (2/min.), event rate (30/min.) and task duration (45 min.) were the same for all three tasks.

Procedure

The experimental procedure was the same as described in Chapter 6, with each subject receiving extensive training and expectancy matching sessions prior to the monitoring session.

### 12.2.2 Results

The probability of detections (Hits) and False Alarms (FAs) and the indices d' and log B were calculated for each 15-min. period of monitoring. Initially these measures were computed without regard to the source on which a signal appeared.

The data for the experiment are shown in Figure 12.1 and Table 12.1. The performance trends are very similar for MSC and MC, while performance on the MSC task appears to be indicative of a sensitivity decrement over time at work. In order to test these impressions, each performance measure was analysed in a 3 x 3 (Tasks x Time Blocks) analysis of variance. The summary tables are given in Tables 12.2 to 12.5.

For d', there were significant effects for Time Blocks and the interaction between Tasks and Time Blocks. Analysis of simple effects revealed that there was a significant sensitivity decrement in the MS task but not in the MSC and MC tasks (see Table 12.2). This analysis also shows that the decrement in sensitivity was apparent only for the latter half of the monitoring period, since there was no difference

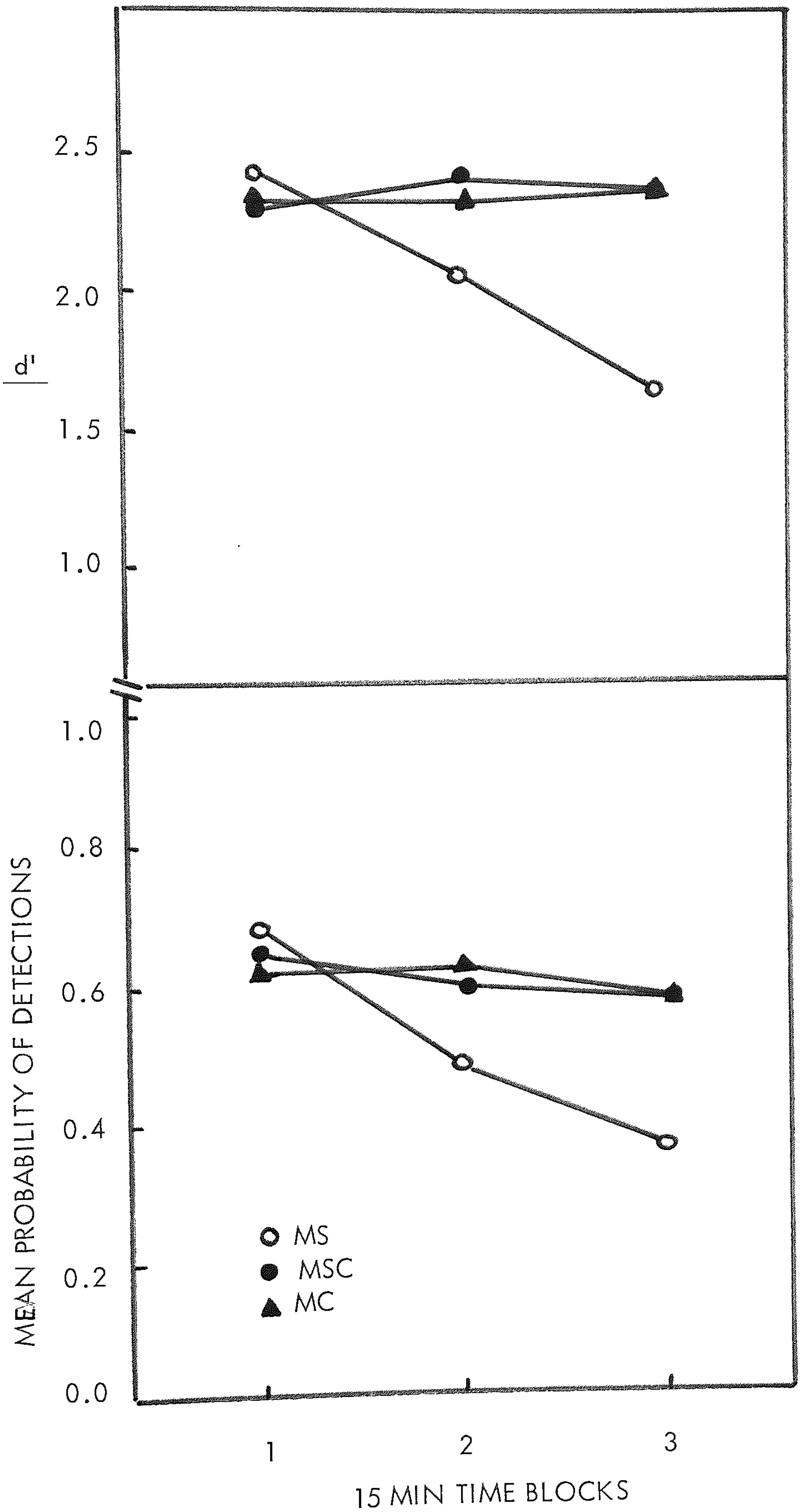


FIGURE 12.1 The mean probability of detections and mean values of d' as as a function of time blocks and for each of the three tasks (see text).

| | Median False Alarm Probability | | | Mean Log B | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 |
| Tasks | | | | | | |
| MS | .027 | .017 | .014 | .741 | .999 | 1.037 |
| MSC | .030 | .015 | .016 | .760 | .931 | 1.000 |
| MC | .021 | .022 | .015 | .860 | .855 | .998 |

TABLE 12.1 Median False Alarm probabilities and mean log B values in each time block (1, 2, 3) and for the three tasks.

| SOURCE | SS | DF | MS | F | P |
|---|---|---|---|---|---|
| A (Tasks) | 1.36045 | 2 | 0.68022 | 5.761 | .05 |
| SWG | 3.89636 | 33 | 0.11807 | | |
| B (Time Blocks) | 0.54494 | 2 | 0.27247 | 13.884 | .001 |
| AB | 1.73732 | 4 | 0.43433 | 22.132 | .001 |
| B x SWG | 1.29521 | 66 | 0.01962 | | |
| Simple effects | | | | | |
| B at A1 (MS Task) | 2.24487 | 2 | 1.12243 | 57.196 | .001 |
| B at A2 (MSC Task) | 0.03255 | 2 | 0.01628 | 0.829 | ns |
| B at A3 (MC Task) | 0.00484 | 2 | 0.00242 | 0.123 | ns |
| B x SWG | 1.29521 | 66 | 0.01962 | | |
| A at B1 | 0.07976 | 2 | 0.03988 | 0.760 | ns |
| A at B2 | 0.48674 | 2 | 0.24337 | 4.641 | .05 |
| A at B3 | 2.53127 | 2 | 1.26563 | 24.135 | .001 |
| Within cell | 5.19158 | 99 | 0.05244 | | |

TABLE 12.2 Analysis of variance of d'.

between tasks in d' values in the first 15-min. time block. Similar results were obtained for the probability of Hits, there being a greater decrement in the MS task than in the other two tasks (see Figure 12.1). Thus for both d' and the probability of Hits, performance trends were similar for the MSC and MC tasks, and different to those for the MS task.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 222.72 | 2 | 111.361 | 5.505 | .05 |
| SWG | 667.528 | 33 | 20.228 | | |
| B (Time Blocks) | 337.056 | 2 | 168.528 | 95.157 | .001 |
| AB | 274.722 | 4 | 68.681 | 38.780 | .001 |
| B x SWG | 116.889 | 66 | 1.771 | | |
| Simple effects | | | | | |
| B at A1 (MS Task) | 573.167 | 2 | 286.583 | 161.816 | .001 |
| B at A2 (MSC Task) | 24.389 | 2 | 12.194 | 6.885 | .05 |
| B at A3 (MC Task) | 14.222 | 2 | 7.111 | 4.015 | .05 |
| B x SWG | 116.889 | 66 | 1.771 | | |
| A at B1 | 17.556 | 2 | 8.778 | 1.108 | ns |
| A at B2 | 128.722 | 2 | 64.361 | 8.123 | .05 |
| A at B3 | 351.167 | 2 | 175.583 | 22.160 | .001 |
| Within cell | 784.417 | 99 | 7.923 | | |

TABLE 12.3 Analysis of variance of Hits.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.03828 | 2 | 0.01914 | 0.208 | ns |
| SWG | 3.03446 | 33 | 0.09195 | | |
| B (Time Blocks) | 1.08154 | 2 | 0.54077 | 20.897 | .001 |
| AB | 0.23249 | 4 | 0.05812 | 2.246 | ns |
| B x SWG | 1.70791 | 66 | 0.02588 | | |

TABLE 12.4 Analysis of variance of False Alarms (log transformed).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.02350 | 2 | 0.01175 | 0.095 | ns |
| SWG | 2.29452 | 33 | 0.06953 | | |
| B (Time Blocks) | 1.02862 | 2 | 0.51431 | 22.521 | .001 |
| AB | 0.16443 | 4 | 0.04111 | 1.800 | ns |
| B x SWG | 1.50722 | 66 | 0.02284 | | |

TABLE 12.5 Analysis of variance of Log B.

There were no differential trends over time in FAs or log B, as Tables 12.4 and 12.5 indicate. With increased time at work, there was a significant decrease in FAs and a significant increment in log B in all three tasks.

| | Mean Hit Probability | | Mean False Alarm Probability | |
|---|---|---|---|---|
| | Source 1 | Source 2 | Source 1 | Source 2 |
| Tasks | | | | |
| MSC | .636 | .572 | .023 | .020 |
| MC | .594 | .626 | .022 | .021 |

TABLE 12.6 Mean Hit and False Alarm probabilities associated with each signal source for the two tasks without source redundancy.

| | Single-Source (Experiment 1) | Multi-Source (Experiment 7) |
|---|---|---|
| Tasks | | |
| Speed | 1.87 (VS1) | 2.11 (MS) |
| Closure | 2.56 (VC1) | 2.35 (MC) |

TABLE 12.7 Comparison of mean d' values for single and multi-source speed and closure tasks (see text).

The data were also analysed for differences arising out of the source from which signals were detected. No differences in Hits or FAs were obtained (as indicated by chi-squared tests) between signal sources for either the MSC or MC tasks (see also Table 12.6).

Finally, the results for d' were compared with those obtained in similar conditions in Experiment 1, to examine the effects of source complexity. Table 12.7 shows mean d' values for single and dual-source tasks; for both categories of task, there were no significant differences in d' between sources (t tests), although d' was somewhat reduced with the addition of an extra source for the closure task.

### 12.2.3 Discussion

When monitoring displays with rapidly changing stimuli, subjects suffer a loss in perceptual efficiency with increased time at work, if the display involves the discrimination of signals according to the perceptual speed classification category. This finding, along with similar results obtained in Experiments 1, 2 and 4, confirms it as a characteristic of long term performance, rather than a specific or isolated phenomenon related to aspects of the tasks or apparatus used. For the specification of sensitivity decrement, therefore, the interactions of three task dimensions have to be taken into account: source complexity, the time course of events (event rate), and the perceptual speed flexibility of closure dimension of Levine, Romashko and Fleishman (1971, 1973).

The prediction that the addition of extra sources to a speed task effectively changes it into a closure task was supported by the data. Thus while a sensitivity decrement for a single source speed task was obtained in Experiments 1 and 2, no such decrement was observed if an extra source was added to the task, as in the so-called multi-source speed-closure task (MSC) in the present study. Against this weak evidence for the hypothesis, stronger evidence was provided by the results for the multi-source speed task (MS), for which there was a significant decrement in sensitivity with time at work. Thus Kahneman's (1973) suggestion that intuitive notions of task difficulty need not be reflected in task performance appears to be supported.

One objection which might be raised with regard to the interpretation of the present results concerns a feature distinguishing the MS task from the other two tasks, which has not been considered before, that is, source redundancy. In the MS task, signals were presented at both sources at once, while in the MSC and MC tasks signals appeared unpredictably and one at a time at either source. This feature might therefore be proposed as the factor underlying the finding of the differential sensitivity trends, perhaps with reference to the findings of Jerison (Jerison and Wallis, 1957) that a decrement is not observed for tasks with spatial uncertainty. However, it has been noted in 12.1 that Jerison's findings were confounded with the effects of signal probability, and, in any case, his results relate to a decrement in detections only. Furthermore, for low event rates, Milosevic (1974) did not obtain a decrement in sensitivity for either a one-source or five-source (speed) task; these results are consonant with the present results. Spatial uncertainty cannot be considered an important factor in the present study, moreover, since each task required the monitoring of only two sources. Finally, further evidence for this approach to the specification of sensitivity decrement may be sought by examining the literature and classifying all the studies which report sensitivity and bias data for monitoring tasks. As we shall see, despite some conflicting results, such an organizational study reveals that a specification of the existing data on sensitivity trends in monitoring tasks may be made in terms of the dimensions identified in the task classification system. This study is described in the following section.

## 12.3 Sensitivity Decrement in Monitoring Tasks: An Organizational Study

The results of Experiment 7 may be seen as providing a sounder basis for some of the proposals put forward in the introductory chapters regarding the specification of displays which are susceptible to a sensitivity decrement over the monitoring period. The results, by themselves, are very clear on this point, and suggest that such a specification may be made on the basis of the dimensions in the task classification system. It is therefore of interest to see whether this specification is supported by the data on sensitivity trends which have been reported in the monitoring literature. Although only four task dimensions have been considered in the task classification system, an initial classification of tasks on these dimensions might reveal discrepancies indicative of the need to consider other dimensions.

The organization of laboratory and operational monitoring tasks within a task classification framework was first reported by Levine et al., (1971), although, as noted in Chapter 2, some less rigorous attempts at systematization of the monitoring literature have been reported by Poulton (1973) and Teichner (1974). One difficulty with the integrative study reported by Levine et al., (1971) is that it did not report classificatory data relating to trends in sensitivity and bias over the monitoring period (probably because of the paucity, at the time, of studies reporting both correct and incorrect detection scores, or a bias-free measure of sensitivity). The task classification analysis reported in this section was therefore carried out to provide such data. In Chapter 5, it was decided to postpone such a classificatory study until an empirical basis for the task classification categories had been established, and until some further empirical results relating to sensitivity decrements had been obtained. These having been achieved, we may now turn to such an examination of the relevant studies.

### 12.3.1 Method

Since the beginnings of experimental research into monitoring behaviour (Mackworth, 1948, 1950), a great many studies have been reported. The majority of these studies, however, have reported performance data using only the detection probability measure, or occasionally, a measure of response latency (Davies and Tune, 1970). The number of studies reporting a relatively 'pure' (bias-free) measure of performance from a very small minority of the studies. This is fortuitous for the purposes of the present study, since somewhere in the region of 800-900 studies of monitoring behaviour exist in the literature.

Thirty three studies reporting sensitivity (d' or a similar index) data were identified through a literature search in Psychological Abstracts and Ergonomics Abstracts. Most of these were included in the final set for further analysis. Five studies were excluded for the following reasons: 1) failure to analyse sensitivity data over time, 2) inadequacy of task description, 3) use of a special testing paradigm distinguishing a study sharply from the others. Another study was rejected because only sensitivity data computed from group detection scores were reported. Twenty seven studies passing these criteria were included in the final set for classification. The tasks used in each study were classified on each of the four task dimensions, namely, source complexity, time course of events, sense modality and signal discrimination type (speed/closure). In accordance with the results of Experiment 7, a 'speed' classification was restricted to tasks in which the signal discrimination required a comparison over of stimuli between events. The cut-off between low and high-rate tasks for the time course of events dimension was arbitrarily set so that tasks with an event rate of less than 24/min. were classified as 'low-rate' tasks. Tasks with event rates of 24/min. or greater, and 'continuous' tasks were categorized as 'high-rate' tasks. Classification on the modality and source complexity dimensions does not require further explanation.

Brief descriptions of the displays used in each study were also recorded, as were the following task variables: 1) the signal duration (SD), 2) the signal rate (SR) and 3) the task duration (TD). The presence or absence of a significant decrement in sensitivity over the monitoring period was recorded for each experimental condition. Early in this classification process, it became apparent that there were no consistent differences between displays in the trends over time in measures of bias (B). Hence classificatory data are reported only for sensitivity.

### 12.3.2 Results and discussion

An initial classification of the results obtained in the twenty seven studies is presented in Table 12.8, where they are broadly classified by display modality (visual or auditory), and by presence or absence of sensitivity decrement. The last two columns in the table correspond to a classification on the signal discrimination type (speed or closure) and time course of events dimensions. The relevant values of each task variable are also tabulated.

On the whole, this classification appears to impose a degree of order on the various results reported in the literature regarding sensitivity trends. A glance at the first half of the table shows that for closure tasks, or for speed tasks with a low rate of stimulation, there is no sensitivity decrement. This is in agreement with the empirical results obtained in the experiments in this thesis.

On the other hand, the second half of the table reveals that, with three exceptions, all the displays for which a sensitivity decrement is obtained are classified as speed tasks with a high event rate. This is again in agreement with the results in this thesis. In two of the exceptions (Mackworth, 1968, Mackworth and Taylor, 1963), the display used continuous stimulation, which may bring into play the additional factor of visual fixation and pursuit. The display used in these studies, the Continuous Clock, is sufficiently different from other 'discrete' closure tasks to suggest that the closure classification assigned to it does not completely characterize its main features. The third exception, the experiment reported by Mackworth (1965a), used such a high event

(200/min.) so as to make the task virtually continuous. Subjects in this experiment also complained of visual fatigue because of the high rate of flashing used (Mackworth, 1965a, p. 422); some unsatisfactory features of this study were also noted in Chapter 5 to cast some doubt on the generality of this result. Of course, it might be the case that at such a high rate of stimulation, any type of display will yield a sensitivity decrement.

There are also some further exceptions, listed in the first half of Table 12.8, to the generalization that sensitivity decrement is restricted to tasks combining a speed signal discrimination with a high event rate. For visual displays, Colquhoun (1969) and Mackworth (1968) used high event rate speed tasks but did not obtain a sensitivity decrement. Interestingly, the study by Mackworth (1968) is the only one in which sensitivity data is reported for the Clock Test (although Loeb and Binford, 1968, report such data for an analogue of the Clock Test, in which an array of lamps in a circle was used). The lack of sensitivity decrement in this study might reflect the influence of task factors not identified within the task classification system. Colquhoun (1969), however, used a visual intensity discrimination task for which other studies have obtained a sensitivity decrement (Loeb and Binford, 1971; Experiments 1 and 2), and it is therefore more difficult to account for his results. One possible intervening factor is that this study appears to be the only one in which subjects served under more than one event rate level. Another factor in the other discrepant results may be that the arbitrary cut-off between low and high event rates may well vary in different categories of task, although there is a fair degree of agreement between studies and the experiments reported in this thesis that a sensitivity decrement is only observed for rates greater than about 20-24/min.

There are thus some exceptions to the generalization about the types of displays yielding sensitivity decrement which cannot be easily explained

| Author(s) | Task | SD | SR | TD | ER | ST |
|---|---|---|---|---|---|---|
| | | Task Variables | | | | |
| Visual displays without sensitivity decrement | | | | | | |
| Baddeley and Colquhoun (1969) | Detection of larger disc in linear display of six discs | 1.8 | v | 40 | 30 | C |
| Broadbent and Gregory (1963a)* | Detection of brighter flash on one of three flashing lights | 0.3 | 1.2 | 80 | 21 | C |
| Broadbent and Gregory (1965)* | 1. As above | 0.3 | 3.4 | 70 | 17 | C |
| | 2. As above with one light | 0.3 | 3.4 | 70 | 17 | S |
| Colquhoun (1969) | Detection of an increase in intensity of flashing light | 0.3 | v | 120 | 30 | S |
| Colquhoun and Edwards (1970) | See Baddeley and Colquhoun (1969) | 1.8 | 1.8 | 40 | 30 | C |
| Davies, Lang and Shackleton (1973) | Detection of an increase in intensity of flashing light | 0.5 | 1.2 | 40 | 12 | S |
| Guralnick and Harvey (1970)* | Detection of increase in deflection of pointer needle | - | 0.5 | 80 | 10 | S |
| Jerison, Pickett and Stenson (1965) | Detection of increase in deflection of two 'light bars' (within an event) | - | 0.3 | 80 | 30 | C |
| Mackworth (1965a) | Detection of brighter flash on one of two flashing lights | .08 | 6.0 | 30 | 40 | C |
| Mackworth (1968) | Detection of double jump of pointer (Clock Test) | - | 0.5 | 40 | 60 | S |
| Milosevic (1974)* | 1. Detection of brighter flash on one of five flashing lights | 0.8 | 0.5 | 60 | 19 | C |
| | 2. As above with one light | 0.8 | 0.5 | 60 | 19 | S |
| Williges (1973) | Detection of increase in duration of flashing light | 1.3 | - | 60 | 10 | S |
| Auditory displays without sensitivity decrement | | | | | | |
| Colquhoun (1967) | Detection of increase in intensity of continuous tone | 0.2 | - | 45 | c | C |
| Hatfield and Soderquist (1970)* | Detection of increase in intensity of 70 dB noise pulse | 0.5 | 1.0 | 90 | 24 | S |
| Loeb and Binford (1964) | As above with 60 dB pulse | 0.5 | 0.5 | 80 | 30 | S |
| Loeb and Binford (1968) | As above | 0.5 | 0.5 | 80 | 24 | S |
| Levine (1966)* | Detection of pure tone in noise | 0.3 | 1.0 | 150 | c | C |
| Milosevic (1975)* | Detection of increase in intensity of 42 dB, 700Hz tone | 0.5 | 0.5 | 60 | 17 | S |

TABLE 12.8 Continued overleaf.

| Author(s) | Task | SD | SR | TD | ER | ST |
|---|---|---|---|---|---|---|
| | | Task Variables | | | | |
| Visual displays with sensitivity decrement | | | | | | |
| Hatfield and Loeb (1968)* | Detection of increase in intensity of flashing light | 0.5 | 1.0 | 90 | 24 | S |
| Loeb and Binford (1971) | As above | 0.5 | 2.0 | 60 | 24 | S |
| Loeb and Binford (1968) | Analogue of Clock Test (see test) | v | 0.5 | 80 | 24 | S |
| Mackworth (1965a) | See previous page | .15 | 6.0 | 30 | 200 | C |
| Mackworth (1968) | Continuous Clock (see text) | .16 | 0.5 | 40 | c | C |
| Mackworth and Taylor (1963) | Continuous Clock (see text) | - | 4.0 | 60 | c | C |
| Auditory displays with sensitivity decrement | | | | | | |
| Bennedetti and Loeb (1972) | 1. Detection of increase in intensity of 60 dB noise pulse | 0.5 | 2.0 | 80 | 24 | S |
| | 2. As above | 0.5 | 2.0 | 80 | 24 | S |
| Binford and Loeb (1966)* | As above | 0.5 | 0.5 | 90 | 24 | S |
| Deaton, Tobias and Wilkinson (1971) | Detection of decrease in duration of white noise pulse | 0.5 | 7.5 | 30 | 30 | S |
| Hatfield and Loeb (1968)* | See Loeb and Binford (1964) | 0.5 | 1.0 | 90 | 24 | S |
| Hatfield and Soderquist (1969) | See Hatfield and Soderquist (1970), previous page | 0.5 | 0.5 | 90 | 24 | S |
| Loeb and Binford (1971) | See Loeb and Binford (1964) | 0.5 | 2.0 | 60 | 24 | S |
| Siddle (1972) | Detection of 2 dB increase in 1 KHz tone pulse | 1.0 | 0.5 | 50 | 30 | S |

TABLE 12.8 List of studies reporting data using the d' measure of sensitivity. Key: SD = signal duration (sec); SR = signal rate (per min); TD = task duration (min); ER = event rate (per min); ST = signal discrimination type, speed (S), or closure (C); v = variable; c = continuous; - = not specified; * = confidence ratings used.

on the basis of the four-dimensional classification system. This is perhaps to be expected, given the scope of the classification system, and a degree of unreliability between experiments due to variations in experimental procedures and subject populations, and other uncontrolled variables. In the original proposal of task classification in Chapter 1, it was pointed out that task classification alone cannot hope to furnish a comprehensive account of all the studies in a given portion of the literature. Nevertheless, it does provide a degree of consistency in any attempt to review the literature, and as in the present analysis, it brings a degree of order into a seeming multiplicity of different results. Thus while we have noted some exceptions, on the whole, the task classification analysis does enable a better specification of the types of display yielding sensitivity decrements. Table 12.8, which shows that a significant sensitivity decrement is obtained in 14 out of 33 conditions in the 27 reported studies, indicates that such a decrement is a much wider and relatively more frequent phenomenon than generally realized (Mackworth, 1970). In particular, the classificatory analysis shows that sensitivity decrements are not restricted to visual tasks, a view which is often put forward in reviews of this area (Broadbent, 1971; Mackworth, 1969, 1970).

The specification of the sensitivity decrement may be made in terms of three of the four task dimensions of the task classification system. This is further illustrated in Figure 12.2, in which the displays listed in Table 12.8 have been categorized into single cells of a four-dimensional vector matrix, as in the analysis of vector spaces. Each cell represents a combination of all four task dimensions of the classification system, while larger blocks of cells represent a combination of a subset of dimensions (for example, tasks representing a combination of the visual mode and single source categories can be represented in the central four cells of the matrix). In each cell, the symbol 'n' was recorded if the

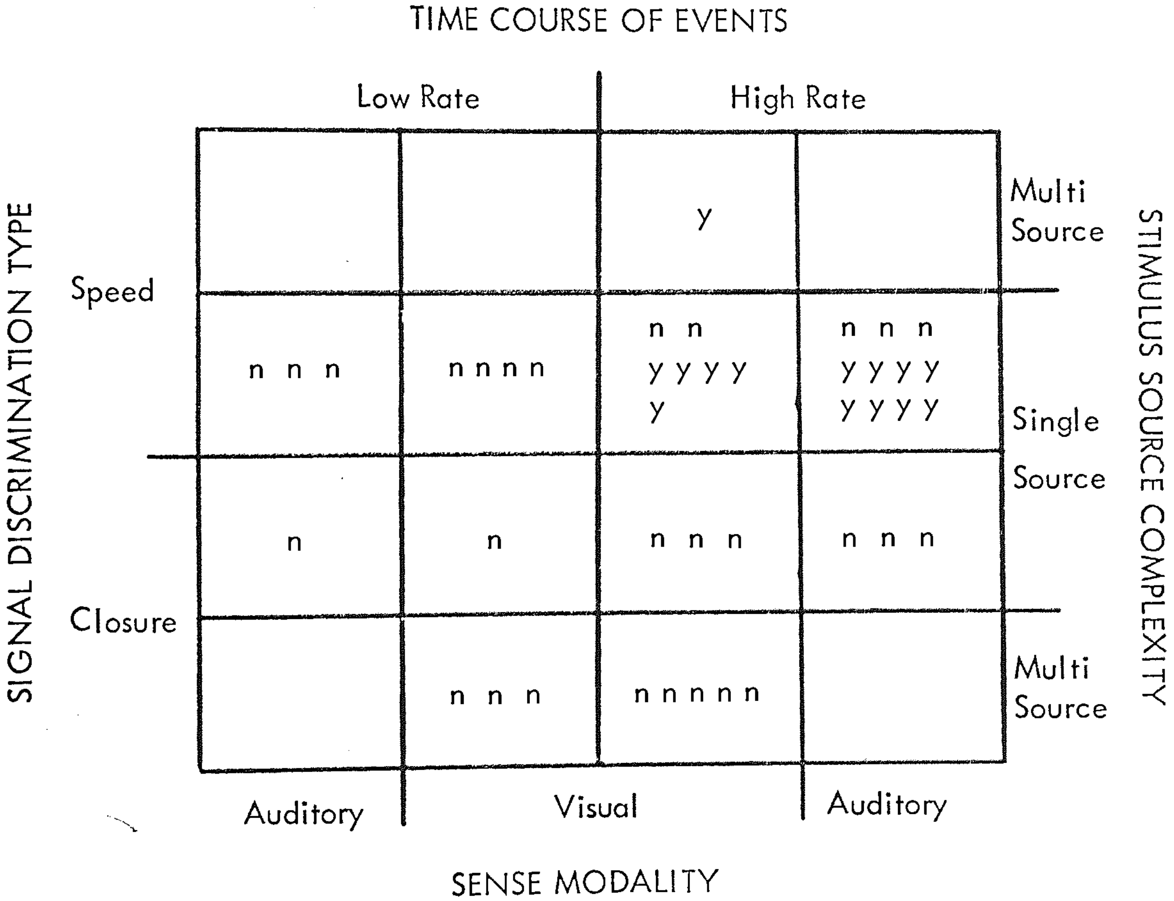


FIGURE 12.2 Classification of studies in Table 12.8 according to task category and presence (y) or absence (n) of a significant sensitivity decrement over the monitoring period (see text).

corresponding display was not associated with a sensitivity decrement, while 'y' was recorded if it was. All the listings in Table 12.8 were included in this matrix, except the two previously mentioned studies by Mackworth in which continuous visual displays were used. In addition, the relevant results from the experiments in this thesis were also recorded.

Figure 12.3 shows that there is a fairly high degree of agreement between studies regarding the types of display for which a sensitivity decrement is obtained. Most of the studies are consistent with the view that a decrement can be obtained for both visual and auditory displays, but only for displays combining the perceptual speed type of signal discrimination with a high stimulation rate. For tasks in other categories, there is unanimous agreement between the studies that there is no significant decrement in sensitivity with time on task. We may fairly conclude that the task classification analysis leads to a specification of displays with decrement that is in agreement with the empirical results obtained in this thesis.

## 12.4 Conclusions

In multisource monitoring tasks, sensitivity decrement is only obtained for tasks in which the discrimination of a signal requires the identification of a change in the stimulus relative to preceding stimuli. In tasks where information pertaining to signal and non-signal features is presented within the same event, no decrement is obtained. A task classification analysis of previous studies in the literature yields results consistent with these conclusions, and shows that a specification of the displays giving sensitivity decrements may be made in terms of three task dimensions within the task classification system.

# CHAPTER THIRTEEN

## DISCUSSION AND CONCLUSIONS

The preceding chapter concluded the main experimental part of this thesis. In this final chapter, the results are re-examined for their theoretical and practical implications. An attempt will be made to relate the results of the various experiments to each other, and, in the usual way, to the results of previous studies, with particular reference to the two main approaches to the analysis of monitoring behaviour adopted in this thesis, namely, task classification and decision theory. For convenience, the two approaches will be considered separately in two sections, although frequent references to either approach will occur in both sections.

## 13.1 Task Classification

Task classification was introduced as a general tool for the description and evaluation of task performance with different monitoring displays. In the introductory chapters of this thesis, an attempt was made to define the components of such a tool, through a consideration of the important features of different displays, and of the discrimination and decision processes involved in different types of monitoring task. This approach to task classification is related to that advanced by Fleishman (1972, 1975a, b), although, unlike the Fleishman approach, the classification categories proposed in Chapter 2 are based on task 'dimensions', rather than on task 'abilities'. However, both these approaches to taxonomy share the same goal, and seek to validate empirically the categories comprising each classification system.

A preliminary four-dimensional classification system for monitoring tasks was developed in Chapter 2 in accordance with these considerations. A large portion of the subsequent experimental work was devoted to an investigation of the reliability of these task dimensions in providing efficient descriptors of performance. In general, two classes of performance were considered: firstly, performance changes over the monitoring period, and as a function of independent variables, and secondly, the consistency of individual performances between different monitoring tasks. We shall discuss the implications of task classification for these two areas in some detail.

### 13.1.1 Performance differentiation

On the whole, the utility of task classification was greater for the assessment of performance consistency between different monitoring tasks than for the description of performance trends over the monitoring period. The assumed differences between the different task categories were not accompanied by large differences between tasks in the within-session performance trends. Thus, for example, there was little difference in detection performance between speed and closure tasks, contrary to expectations arising from the results of Levine, Romashko and Fleishman (1971). In particular, their findings that the effects of target presentation rate and sense modality on monitoring performance are a function of whether the task requires perceptual speed or flexibility of closure, was not supported by the data obtained in this thesis. Of course, this might merely be indicative of the relative similarity of the tasks employed in the experimental studies. However, the results of the correlation studies (Experiments 3 and 5) suggest that this can only be a partial explanation, since the speed and closure categories were found to describe reliably different classes of performance as far as the consistency of individual performances was concerned. The physical similarity of tasks was found not to be a significant factor in the determination of the consistency of performance in different monitoring situations. Overall, therefore, the results may be interpreted as showing that for the various combinations of task categories within the four-dimensional task classification system, performance differentiation between categories is limited to a few task categories.

On the other hand, where performance differentiation between different task categories was observed, it was found to hold consistently across several different studies. Thus it has been repeatedly noted that the effect of a high stimulation rate is to reduce monitoring efficiency over a work period, but only for tasks classified within the perceptual speed category. This relation holds for both visual and auditory tasks, and is apparent both when 'discrete' (accuracy) and 'continuous' (latency) measures are used to assess performance. The emergence of the same type of effect in several experiments and in an organizational study of the literature serves to confirm its reliability and generality over the whole spectrum of monitoring tasks. Levine, Romashko and Fleishman (1971) originally suggested that new, previously obscured functional relationships could be revealed by using a task classification approach, and the results obtained here are in general agreement with this suggestion.

A task classification analysis thus appears to lead to improved generalizations of the effects of independent variables on performance, and of the specification of those types of task to which a particular class of performance is restricted. With regard to the latter, we have discussed in detail in the preceding chapter the importance of the speed-closure dimension for the specification of displays yielding a sensitivity decrement. Thus the answer to the query, "what types of monitoring display exhibit a decrement in sensitivity over the work period?", which was posed in Chapter 5, and originally by Swets and Kristofferson (1970), may be answered very succinctly with reference to the task classification system. As Figure 13.1 shows, sensitivity decrements are restricted to tasks represented by the hatched area; that is, either visual or auditory speed tasks with a high event rate, and multi-source, visual speed tasks. As noted in Chapter 12, traditional theories of monitoring behaviour which do not identify task-specific factors, and other theories incorporating task-related constructs such as 'coupling', do not provide the

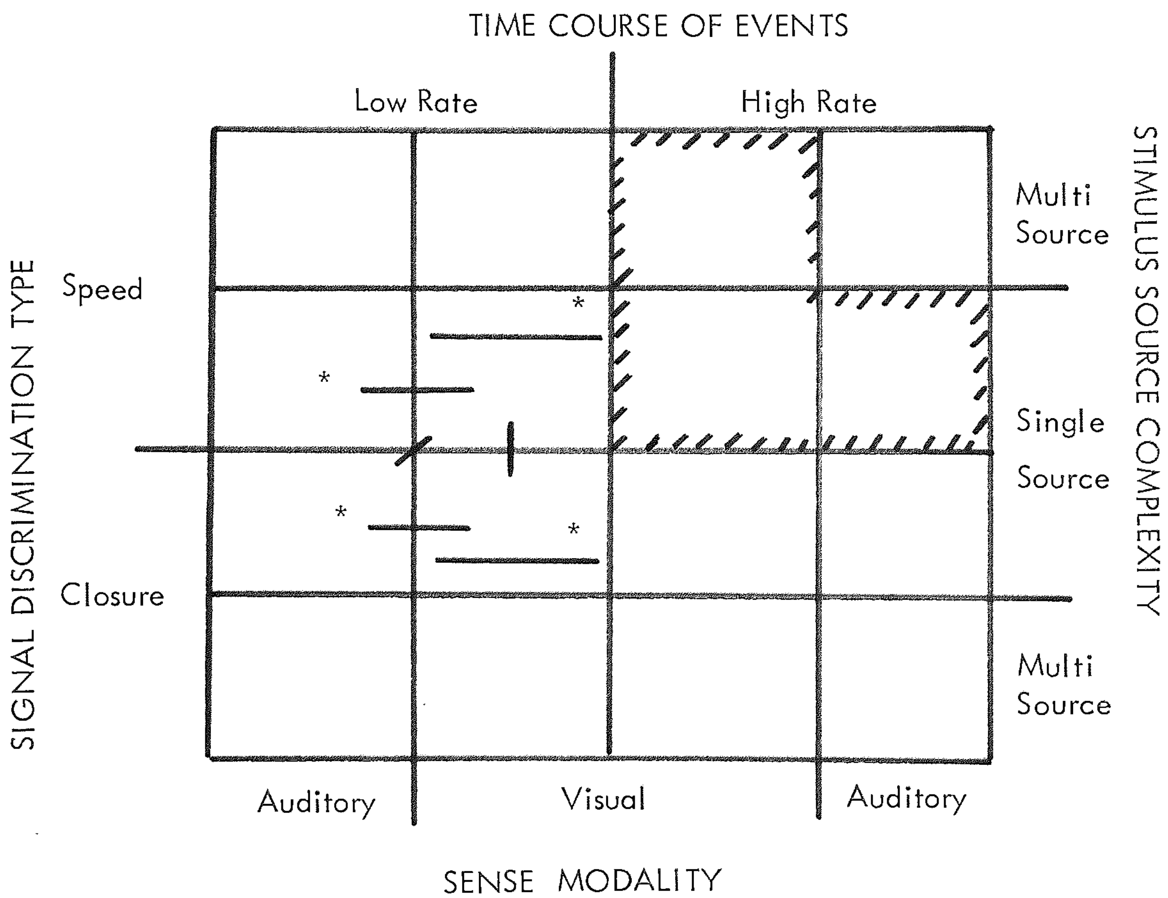


FIGURE 13.1 The domain of displays for which a sensitivity decrement over the monitoring period is obtained (indicated by hatched area). The lines within and across cells indicate the degree of consistency in performance (d') between task categories (the length of each line is proportional to the corresponding correlation coefficient; * significantly greater than 0, $p < .05$).

same predictive capacity for the specification of sensitivity decrement as do the task categories identified on the task classification system.

Thus although the differentiation in monitoring performance was limited to a few classification categories in the studies reported here, the results nevertheless endorse the value of the task classification approach. The results might be indicative only of the relative 'homogeneity' of the performance areas considered, as compared with other, more 'heterogeneous' tasks, for example, tasks involving retention and retrieval processes. However, Levine, Romashko and Fleishman (1971) have pointed out that task homogeneity provides a more stringent test of the task classification approach, since, if performance differentiation is obtained for more homogeneous tasks, then performance differentiation of an area with more heterogeneous tasks can be readily expected. Levine, Greenbaum and Notkin (1973) have reported such differentiation of performance functions for a heterogeneous research area, that of the effects of alchohol on human performance. They found that there were marked differences in the effects of alchohol depending on the category of task performed. In general, their study was an attempt to structure the existing data to provide more reliable generalizations across studies and new functional relationships between independent variables and performance. Although performance differentiation was limited to a few task categories and a few classes of performance in the studies reported in this thesis, the results may nevertheless be interpreted in the same general way as providing support for the task classification approach.

As far as performance changes over a monitoring period are concerned, therefore, task classification enables a clearer distinction to be made between tasks in which criterion shifts only are obtained, and those where both sensitivity and criterion shifts are obtained. As noted previously, however, task classification was also examined in relation to changes in individual performances between tasks; this aspect of monitoring behaviour is given further consideration in the next section.

### 13.1.2 Consistency of individual differences in monitoring performance

One of the early and generally prevalent views held about monitoring behaviour was that individual differences in performance are almost wholly task specific (Baker, 1963; Buckner, Harabedian and McGrath, 1960). This view was based, at the time, on the results of a few studies indicating a generally low correlation in performance between visual and auditory monitoring tasks. Since that time, there have been further investigations of inter-task correlations in monitoring performance, and while the results have not been completely consistent, they have pointed out that the early view of complete task specificity paints too bleak a picture of the situation. It has been previously noted that the reasons for the apparent task specificity had not been examined in a systematic way, and that a closer examination of the various tasks used in the literature might lead to a resolution of this problem.

If such an examination is carried out, a very general view emerges: this is that those tasks which impose similar types of demands on the monitor are more likely to share common performance variance than tasks which make differing demands. Two experiments were especially designed to investigate whether performance consistency is related to task demands as identified within the task classification system. The results from these two studies were very clearly in favour of this general hypothesis: when tasks are matched on either category of the speed-closure dimension, individual differences in performance are highly consistent from task to task, whether performances within (Experiment 3) or across (Experiment 5) modalities are considered. On the other hand, for task pairs classified either across the speed-closure dimension only, or across both the speed-closure and the modality dimensions, the correlation in performance between tasks is very low. These results are illustrated in Figure 13.2(a) and (b) for correlation coefficients for the d' and detection probability measures respectively. Both figures display the same upward trend in correlation values as more 'compatible' (relative to the two task dimensions) task pairs are considered. In Figure 13.2(a) the first plotted point shows that for tasks classified across two classification dimensions, the obtained correlation in performance is very low and almost zero. A somewhat higher but still non-significant correlation is obtained if tasks are matched on one dimension (modality) but not the other (speed-closure), as indicated by the second plotted point. As the next two points show, significant correlations are obtained only if tasks are equated on the speed-closure dimension. Finally, for tasks equated across both dimensions, the highest correlations are obtained. For such task pairs, the obtained correlations are only slightly lower than the task reliabilities (see Experiment 8, Appendix B).

These results therefore clearly show that the early view of complete task specificity of individual differences is no longer tenable. On the other hand, this does not imply an acceptance of the opposing view, proposed by Tyler, Waag and Halcomb (1972), that monitoring performance is non-specific and mediated by a common predominant 'vigilance' factor. Performance consistency is closely dependent on the type of signal to be monitored; if this is the same in two different displays, then performance with these two displays is likely to be correlated, even if the displays are presented to different senses, or if they differ in other ways. Of course performances may not be correlated if the displays differ on some important dimension not identified in a task classification system. This is a matter for empirical investigation of the relative 'importance' of task dimensions which have not been considered, such as

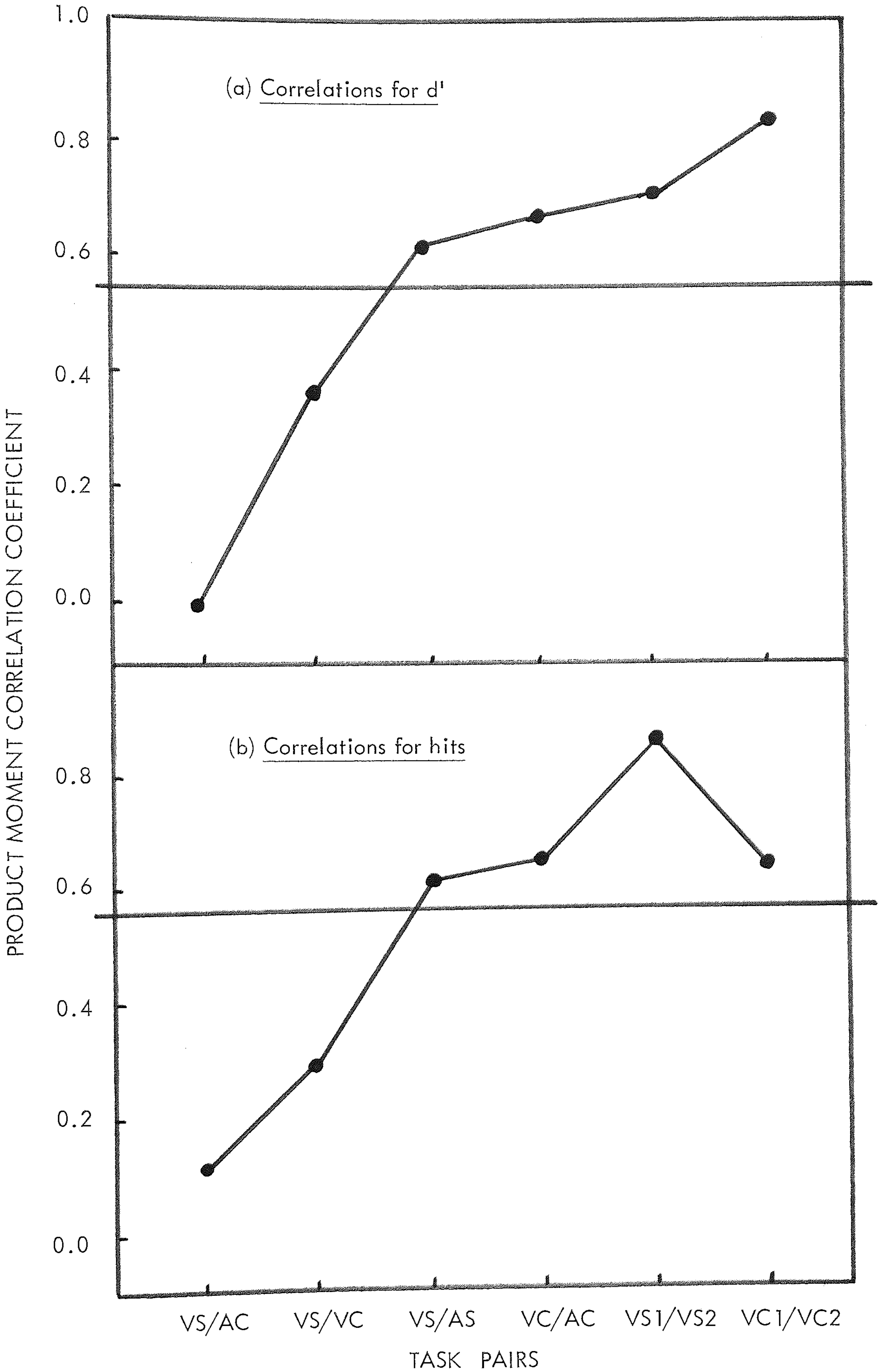


FIGURE 13.2 Product moment correlations between different task pairs for two measures of performance (V = Visual; A = Auditory; S = Speed; C = Closure; correlations above the mid-lines are significantly greater than 0, $p < .05$).

visual search, signal duration, and so on. At the same time, it must be emphasized that the superficial similarity or dissimilarity of tasks per se does not significantly affect performance consistency (Parasuraman, 1976a). It seems clear, therefore, that the consistency of individual performances is largely determined by the similarity of the demands of various tasks, as identified within a task classification system.

It has been suggested that task demands as defined in particular by the speed and closure categories provide the principal determinant of performance consistency. However, the results of Experiment 5 also pointed to the involvement of modality specific factors. No doubt as other factors are considered, such as visual search, the divergence between visual and auditory tasks becomes greater. Thus it is not sufficient to merely state that visual and auditory tasks impose differing demands on the subject. In some cases they may do so, but this may be due to additional features of each type of task, such as visual search or binaural detection (Eijkman and Vendrik, 1965; see also Jeffress, 1971). In other cases, such as for the simpler visual tasks of Experiment 5, there may not be any substantial differences between sensory modalities.

The empirical figures plotted in Figure 13.2 may be taken as a representative function for the relation between the degree of performance consistency between two tasks and the 'compatibility' of the tasks, as assessed on certain task dimensions. Thus, if one were to carry out a number of studies similar to Experiments 3 and 5 for several different task pairs, and plot the obtained correlation coefficients, one might obtain a similar function if the task pairs were arranged accordingly (a movement to the right along the horizontal axis representing an increase in 'compatibility'). This function would then describe the changeover from low to high performance consistency as task compatibility is increased; one might accordingly account for two apparently conflicting results by differentiating the corresponding task pairs along the axis of 'compatibility'.

This approach to performance consistency in monitoring tasks suggests that much of the debate about the existence of a common 'vigilance factor' (see Davies and Tune, 1970, pp. 33-37) has, for the most part, missed the point. The main argument in this debate rests on the assumption that 'vigilance' is a unitary process which is the principal determinant of performance in a number of prolonged tasks termed 'vigilance' or 'monitoring' tasks. Thus if a correlation in performance between two tasks is observed, then both tasks require 'vigilance'; if on the other hand, a low correlation is observed, one task does not require vigilance, or the 'vigilance' hypotehsis is incorrect. This is essentially the argument put forward by, among others, Buckner et al., (1960) and Tyler et al., (1972). However, such an 'all-or-none' hypothesis in terms of whether a task requires vigilance or not provides very little explanatory power for many of the results in the literature, and particularly for the results for speed and closure tasks obtained in this thesis.

### 13.1.3 Implications for a taxonomy of monitoring tasks

We have seen that the results of the experiments reported in this thesis clearly point to the importance of the perceptual speed-flexibility of closure task dimension for a taxonomy of monitoring tasks. The results of Experiments 1 and 4 in particular indicate that there is a high degree of consistency in group performance levels if monitoring tasks are categorized on this dimension. However, the correlation studies in Experiments 3 and 5 provide the more crucial evidence regarding the consistency of individual performances, and as such clearly show that the speed and closure categories describe reliably different classes of monitoring behaviour. The speed and closure categories are therefore strong candidates for inclusion in a taxonomy of monitoring tasks, in accordance with the original speculations of Levine, Romashko and Fleishman (1971) and Theologus and Fleishman (1971).

The speed-closure dimension represents one component in a four-dimensional task classification system which was proposed as an initial exploratory tool for the analysis of monitoring behaviour. Of the other three 'dimensions', the time course of events has also been shown to exert a significant influence on monitoring performance. Although the complete range of tasks represented by this dimension (from low stimulation rate through high stimulation rate to continuous tasks) was not examined, the results for low and high rate tasks were consistent enough to permit generalization to continuous tasks. As noted previously, this dimension is important from the point of view of the specification of monitoring displays yielding a sensitivity decrement; there is some evidence that continuous tasks also show a sensitivity decrement (Mackworth, 1968; Mackworth and Taylor, 1963). One of the problems associated with the analysis of performance on such tasks is the difficulty of deriving reliable indices of sensitivity. Mackworth (1965b) demonstrated that a sensitivity decrement was observed for such tasks by assuming an arbitrary 'decision interval', thus facilitating the derivation of the sensitivity index d'. As Luce and Green (1974) have pointed out, this is not a particularly satisfactory method of analysis. An urgent requirement for further research into monitoring behaviour is thus the development of detection models for the analysis of continuous displays, which may be more typical of 'real-life' tasks than 'discrete displays'.

The distinction between discrete and continuous tasks is thus also an important one from the point of view of a taxonomy of monitoring tasks, and which merits further investigation. The same is true of the stimulus source complexity dimension, which was considered in Experiment 7. This experiment demonstrated the interaction between this dimension and the speed-closure dimension in relation to performance trends over a monitoring period. The degree of source complexity in the displays used in this experiment was not very great. This was a deliberate choice; for as source complexity is increased, a visual search factor may come into play. This is another candidate for further research into task classification and monitoring performance. Visual search is an important component of many inspection and monitoring jobs, and thus the application of a task classification analysis to such tasks would need to consider this as a task dimension.

A final point concerning the representativeness of the tasks used in the experiments reported in this thesis concerns the signal duration factor. All the displays used in the experiments reported in this thesis employed short-duration or 'transient' stimuli. These correspond to 'limited-hold' tasks, in Broadbent's (1968) terminology. A number of 'unlimited hold' tasks, where the signal is present until detected, have also been reported in the literature (see 2.2). On the whole, however, tasks with transient signals are more representative of the tasks reported in the monitoring and inspection literature, and probably, of many operational tasks, such as radar and sonar displays. Poulton (1973) has also pointed out that laboratory monitoring tasks with brief signals are representative of inspection tasks which involve visual search.

## 13.2 Decision Processes in Monitoring Behaviour

Having discussed the implications of task classification for the evaluation of task performance with different monitoring displays, we may now re-examine the experimental results from the point of view of the decision processes involved in monitoring situations. It has been noted that a consideration of decision processes leads to an analysis of monitoring performance within the broad framework of statistical decision theory. In general, the results obtained in the experimental studies provide strong support for this approach to the analysis of monitoring performance.

This approach to the analysis of monitoring behaviour represents a 'broad' rather than a rigorous application of the theory of signal detectability (TSD), in which there is no commitment, a priori, to a particular type of detection model. Thus, it is not necessary to assert many of the strong assumptions implicit in TSD, such as those relating to the Gaussian representation of stimuli and the choice of responses according to the optimization of expected value. Instead, it need only be assumed that there is some randomness in the representation of stimuli in the observer (the 'evidence'), and between the evidence and the various response states; it may then be investigated whether performance changes in monitoring situations are associated with a change in the form of the evidence (sensitivity), or with a change in the manner of the allocation of response states to responses (criterion). The results indicate that either or both of these changes may underly the observed variations in monitoring performance, according to the type of task used and to other features of the monitoring situation. In this section we shall consider the evidence for criterion and sensitivity shifts in monitoring behaviour and discuss some of the theoretical interpretations of these changes.

### 13.2.1 Criterion shifts

The principal finding arising out of the application of TSD to the analysis of monitoring behaviour has been in the re-interpretation of the traditional 'vigilance decrement'. Such an analysis has indicated that in most monitoring tasks there is no real decrement in efficiency, but a gradual increase over the work period in the criterial level of response of the observer. The results have shown that the increment in the criterion is relatively independent of the type of display monitored. It has also been demonstrated that a new interpretation of latency data, using a conceptual model extended from TSD, also provides support for the notion of an increment in the criterion during the monitoring period. The association between the traditional 'vigilance decrement' and the increase in the response criterion thus emerges as the most general characteristic of monitoring behaviour, and, as such, may be said to be firmly established.

Less well established, however, is the view that the response criteria corresponding to different categories of response show an unequal change over the monitoring period. This view stems mainly from the findings of Broadbent and Gregory (1963a, 1965), who found that while strict response criteria increase over the run, risky criteria remain stationary; the implication is that there is an expansion in the scale of the 'evidence' used by the observer with elapsed time (Broadbent, 1971). At the other extreme, Milosevic (1974, 1975) has reported time-related increases in both strict and risky criteria, for both visual and auditory monitoring tasks. In the middle ground are the results of Levine (1966) and Loeb and Binford (1964), who reported that although all criteria increase with time at work, the increase in lax criteria is usually less than that for strict criteria. The results obtained in this thesis are generally in agreement with these latter studies, but do not resolve the discrepancies between the studies. Criterion increments may depend on the particular values of criteria 'chosen' by observers and on the signal probability, and these are usually different between studies; thus what are termed 'lax' criteria in one study may correspond to 'medium' or 'strict' criteria in another, where the overall B values are lower. The lack of a sufficient number of studies reporting rating data, and the inherent unreliability in attaching a meaning to absolute values of B prevent a strict analysis of the data to see whether this is so. There is a general indication, however, that Broadbent and Gregory's study, which is the only one reporting no increase in lax criteria, had generally lower values of B than in the other studies. Thus an increase in the criterion may be observed only if the criterion values are high, or relatively 'strict' to begin with.

In earlier chapters, we have entertained a number of different interpretations of the criterion increment in monitoring tasks. These include the expectancy approach, the expected value interpretation of Williges (1973, 1976), an interpretation based on subjective gains and losses (Broadbent, 1971), and Smith's (1968) cost model. In general, the results may be interpreted as providing support for the expectancy approach, although the experiments were not specifically concerned with critically distinguishing between these theoretical positions.

Under the expectancy approach, the criterion increment is assumed to be due to a continual re-adjustment of the criterion by self feedback of detection information. The observer is assumed to begin the session by under-responding to signals, and does so increasingly as he revises his criterion upwards; the results from most of the experiments in this thesis which show that the overall response probability p(Yes) was less than the signal probability p, and became lower still with time, are consistent with this interpretation. On the other hand, as Broadbent (1971) has pointed out, the expectancy approach faces the difficulty that in rating tasks it must be assumed that only the confident responses are taken into account in the self feedback loop, but there appears to be little justification for this assumption.

The expectancy approach does, however, receive greater support in the results of a series of experiments by Baddeley and Colquhoun (1969; Colquhoun and Colquhoun, 1964, 1967) and by Williges (1969, 1971, 1973). These were reviewed in detail in Chapter 5, and so we will not re-consider them here, except to note that Williges (1973, 1976) has gone beyond the expectancy approach by interpreting B shifts in vigilance in terms of the expected value model of TSD. He suggested that since the mean values of B approach the optimum B (as determined by the maximization of expected value) toward the end of the work period, the monitor's performance actually represents an attempt to reach optimal behaviour; this was termed the 'vigilance increment' (Williges, 1976).

The present results also show that the <u>mean</u> values of B approach the optimum value of B as the monitoring period proceeds. However, an acceptance of Williges' interpretation must be tempered by three considerations. Firstly, in Williges' experiments as they are reported, no validation of the equal-variance TSD model on which the empirical and ideal B values are dependent, is presented. It has been noted previously that the confidence limits on B values in low signal probability tasks may be very large, so that while the <u>trends</u> in the criterion may be usefully assessed using this index, attaching a meaning to the <u>absolute</u> B values is rather tenuous. Mackworth and Taylor (1963) and Taylor (1967) therefore do not recommend the use of B at all. Secondly, it is almost a truism that observers do not usually make decisions based on criteria as extreme as those predicted by expected value models. Hence, for $p \lessgtr 0.5$, we would expect that $B \lessgtr B_{opt}$: the corresponding increase or decrease in B over the work period can then be explained in terms of expectancy alone (the probability matching model of Thomas and Legge, 1970 , also predicts that observed values of B are not as extreme as the optimum values; see 4.2.2). Thirdly, the ideal observer would take into account both the *a priori* signal probability and the payoff matrix in using an optimum value B, but Williges' results themselves (Williges, 1971) indicate that the magnitude and direction of the changes in B need not follow those predicted by the expected value model (see also Levine, 1966).

It therefore does not appear necessary to assume that the monitor's behaviour represents an attempt to reach optimal behaviour. This is consistent with the interpretation of TSD within the framework of decision theory, in which reference to an expected value model of decision making need not be made. Nevertheless, the results of Williges, when interpreted in a more general way, do point to the importance of expectancy mechanisms in the interpretation of the criterion increment in monitoring tasks.

It seems fairly clear that criterion adjustment is a feature of long term performance, and future developments in this area of research might profit from an analysis of monitoring performance using detection models which incorporate criterion variability, such as the so-called 'error-correcting' models (see Kac, 1962). Recently there has been a revival of interest in detection models which do not assume a 'static' criterion (as does TSD), and it would seem natural to extend such learning detection models to the analysis of criterion increment in monitoring situations (Parasuraman, 1976b, in preparation; Thomas, 1973).

### 13.2.2 Interpretation of sensitivity decrements

Although variations in performance in monitoring tasks are principally associated with criterion shifts, changes in sensitivity may also occur in certain situations. In this section we shall therefore discuss the various theoretical interpretations of the sensitivity decrement in relation to the results of the experiments and those of previous investigators.

Previous explanations of reduced efficiency in monitoring tasks have made use of several general theoretical constructs, such as 'coupling' (Loeb and Binford, 1968, 1971), 'neural habituation' (Mackworth, 1969) and 'observing responses' (Jerison, 1967b, 1970). On the basis of any of these constructs, it may be postulated that auditory monitoring tasks are less susceptible to performance degradation than visual tasks. At first this seems a reasonable proposition; however, as noted in Chapter 12, there are several examples of auditory tasks for which sensitivity decrements over the work period are obtained (see Table 12.8). The view that sensitivity decrement is restricted to visual tasks (Broadbent, 1971, p. 99; Mackworth, 1970, p. 41) does not therefore tell the whole story, and the finding that a decrement may also be obtained for auditory tasks provides general evidence against these theories. However, some further, more substantial objections to the theories may also be considered.

<u>Coupling</u>

Under the coupling hypothesis, sensitivity decrement is assumed to be restricted to loosely-coupled (usually visual) tasks, since in such tasks efficiency would be more likely to be affected by changes in the orientation of the peripheral receptor mechanisms (Loeb and Binford, 1968). The coupling theory would appear to be embarrassed by the finding that auditory tasks, which may be assumed to be closely coupled, may also lead to a decrement in efficiency over the work period. Moreover, in Chapter 3 it was noted that the coupling hypothesis could not account for some of the data relating to cross-modal correlations in sensitivity; a number of studies have shown that it is more likely for performance in an auditory task to be correlated with that in a <u>loosely-coupled</u> visual task than with that in a closely-coupled visual task. The coupling hypothesis therefore suffers from a certain ambiguity in that an auditory task can be said to be either closely or loosely coupled. The same ambiguity applies to the interpretation of sensitivity decrement in terms of coupling, since opposite sets of results may be explained: 1) that visual (loosely coupled) tasks yield a decrement while auditory (closely coupled) tasks do not (Loeb and Binford, 1968), or 2) that visual (loosely coupled) tasks do not yield a decrement while closely coupled (visual or auditory) tasks do (Deaton, Tobias and Wilkinson, 1971; Hatfield and Loeb, 1968).

The explanation for the first result is the one mentioned previously. The explanation for the second type of result, as proposed by Deaton et al. and Hatfield and Loeb, is, that for closely coupled (usually auditory) tasks, there is less opportunity to attend to irrelevant stimulation and thus maintain efficiency than for loosely coupled (visual tasks). The coupling theory manifestly loses credibility in being able to interpret opposing results, and it is rather odd that one of the investigators in this area (M. Loeb) should be associated with both types of coupling interpretation.

Finally, the coupling hypothesis cannot adequately deal with all the empirical data obtained in this thesis in relation to task classification and sensitivity decrement. Coupling may thus be treated as a general feature of monitoring tasks but does not provide a strong theoretical framework for the interpretation of reduced efficiency in monitoring tasks.

Neural habituation

J.F. Mackworth's (1969) habituation theory of vigilance postulates that the habituation of the neural responses to the repetitive background events ('unwanted signals') of a monitoring task might underly the decrease in sensitivity over the monitoring period. Mackworth's theory is largely speculative; it rests on the assumption of a casual link between the evoked response decrement and performance decrement in monitoring tasks; it further assumes that a common habituation mechanism underlies the decrement in both of these cases.

The evidence for a habituation mechanism for performance sensitivity presumably derives from Mackworth's own results that in the particular visual task (Continuous Clock) she has used, there is an approximately exponential decay in sensitivity over time; however this appears to be a highly specific result confined to continuous tasks requiring constant visual fixation or visual pursuit. Furthermore, while the criteria for the phenomenon of habituation (see Thompson and Spencer, 1966) may be met, with difficulty, for the evoked response decrement (but see Ritter, Vaughan and Costa, 1968), there does not seem to be any justification for viewing monitoring performance decrement within a strict habituation paradigm. Moreover, there has been no demonstration of a link between the two types of decrement.

It has been previously noted that recent research in evoked responses suggest that selective averaging of brain electrical activity according to different behavioural events may provide correlates of decision making processes; while the evoked potential does provide a more central index of stimulus linked physiological activity (and hence, of potential habituation phenomena), the results of Experiment 6 and other recent research suggest that evoked potentials may be more sensitive to fluctuations in 'cognitive' decision making activity than to variations in the observing process _per se_ (Sutton and Tueting, 1975; Donchin, Kubovey, Kutas, Johnson and Herning, 1973).

Finally, the largely suggestive habituation theory of Mackworth does not provide any explanatory power for the results of this thesis or of previous studies in relation to the different types of task giving sensitivity decrements.

Jerison's observing response model

Jerison (1967b, 1970; Jerison, Pickett and Stenson, 1965) has put forward a theoretical description of monitoring behaviour in terms of the types of observation of monitoring displays. He suggested that observing behaviour is comprised of three main classes of observing response: alert or 'optimal' observing, blurred or 'sub-optimal' observing, and 'distracted' observing, which corresponds to no observing of the display. As far as visual displays are concerned, these classes of observing behaviour may be grasped intuitively. However, their meaning is less obvious where auditory displays are concerned. Jerison considered the 'distracted' mode of observing to represent a 'decision not to observe'; if this is taken to include the decision to inhibit the peripheral response to the stimulus, then such an interpretation is difficult for auditory displays, since as it was noted as early as Chapter 2 that the evidence suggests that voluntary control cannot be exerted over the peripheral cochlear response to auditory stimuli (Picton, Hillyard, Galambos and Schiff, 1971). To be fair however, it is not clear whether Jerison intended his model to include the observation of auditory displays; his own experimental work has been concerned almost exclusively with visual monitoring tasks.

Jerison proposed that reduced efficiency in monitoring tasks might be due to an increase in periods of inefficient (blurred and distracted) observing with time at work. As far as the results for the probability of detection are concerned, the available data are consistent with

Jerison's theory. From the point of view of changes in sensitivity, however, several objections to the theory may be raised. Although Jerison himself would probably counter such objections by arguing against the validity of the TSD metric d' to describe sensitivity in monitoring tasks (Jerison, 1967a), they can be formulated without reference to d' and B. The essence of the argument is as follows (see also Broadbent, 1971). Jerison's three-way observing model implies that the relative degree of randomness in internal representation (evidence) of signals and non-signals will change with elapsed time. If this takes the form of increased variability in signal variance due to an increasing frequency of 'distractions', then the operating characteristic (OC) will be markedly skewed and become more so with time. This is illustrated in Figure 13.3(a). Such a trend in OC skew was not observed in the experimental studies carried out, and is also inconsistent with the results of previous studies where the OC has been plotted (Loeb and Binford, 1964; Milosevic, 1975, personal communication). As far as sensitivity decrement is concerned, therefore, Jerison's theory is not supported by the available evidence.

This is not to imply that the theory cannot be applied to other situations. It appears to be particularly relevant to visual tasks only, and might therefore be useful in the analysis of performance of complex visual displays; although here again the indication is that theories based on changing patterns of observation between components of the display (Hamilton, 1969; Hockey, 1969) provide greater explanatory power than 'all-or-none' observation theories.

It therefore transpires that theories based on the concepts of coupling, neural habituation and observing responses cannot adequately interpret the finding of sensitivity decrement in monitoring tasks; although a modified version of observing response theory that can account for auditory sensitivity decrement may provide the basis for the interpretation of sensitivity decrement. Some alternative theoretical approaches are therefore now considered, but in a slight digression from the main argument, another aspect of OC skew is considered first. This also has some bearing on the interpretation of sensitivity decrement.

Sensitivity decrement and OC skew

Following on suggestions by Green and Swets (1966) and Taylor (1967), Mackworth (1970) proposed that OC skew in a monitoring situation might reflect the degree of uncertainty about some aspect of the signal to be monitored. Theoretically, if there are two types of signal, one of which is 'known exactly', while only limited information is available on the other signal, the OC for the latter type of signal will exhibit a greater degree of variability in the signal distribution as compared to the noise distribution (Green and Swets, 1966; Taylor, 1967). Thus in a monitoring task, if it is assumed that the observer gets to know the signal characteristics better as the task proceeds, there will be a decrease in the OC skew with time on task; the OC slope should accordingly approach unity at the end of the session or in later sessions. This is illustrated in Figure 13.3(b).

However none of the OCs obtained in Experiments 2 and 4 conform to this pattern; in Experiment 2 it was of particular interest to investigate whether such a learning effect was present, and if the obtained sensitivity decrement in the first session would be abolished in the second session. However, the same decrement was obtained in the second session, and there was no systematic trend in OC skew. Figure 13.3(c) illustrates the typical OCs obtained under conditions of sensitivity decrement.

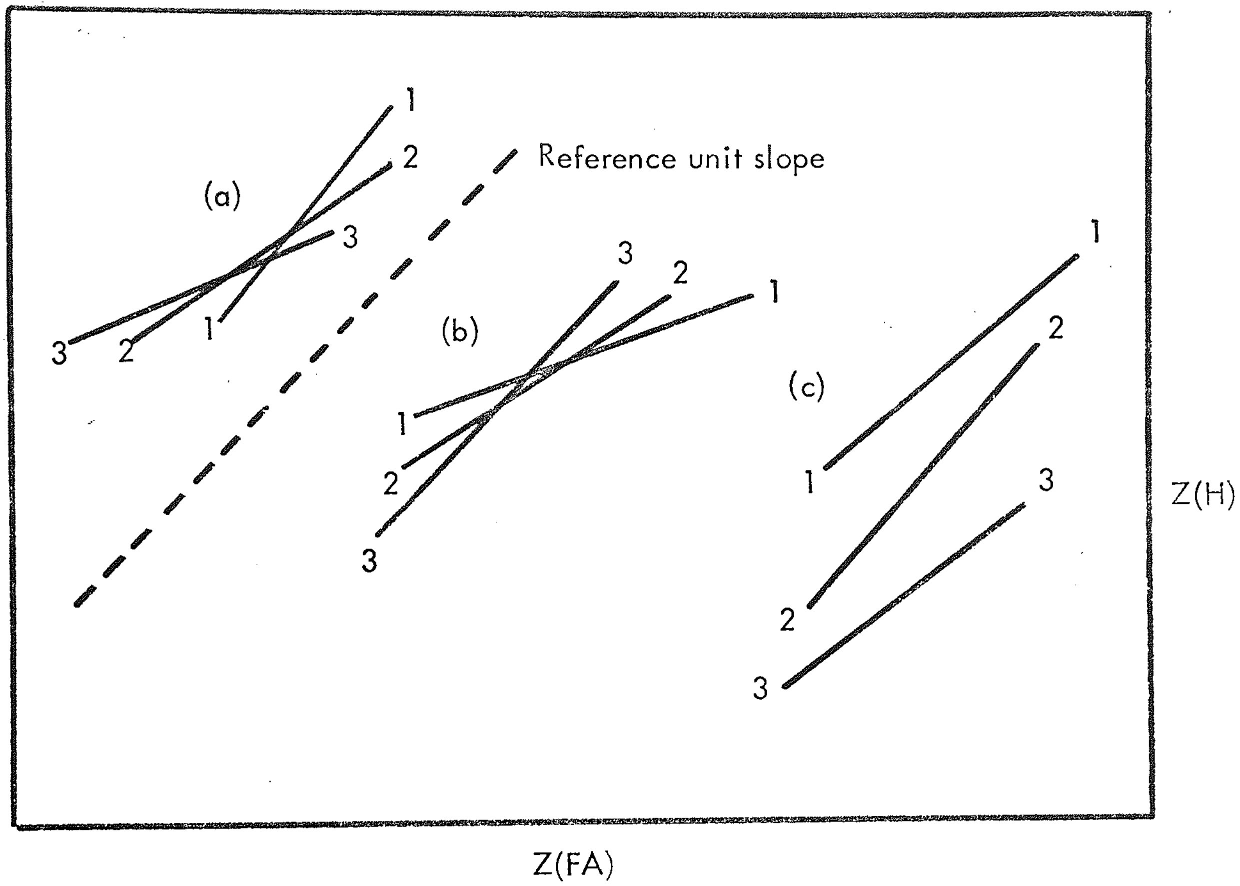


FIGURE 13.3 Operating characteristics derived from the theoretical observations of Jerison (a) and Mackworth (b), and from typical data from the rating experiments 2 and 4 (c). Numbers refer to successive time blocks in a monitoring session.

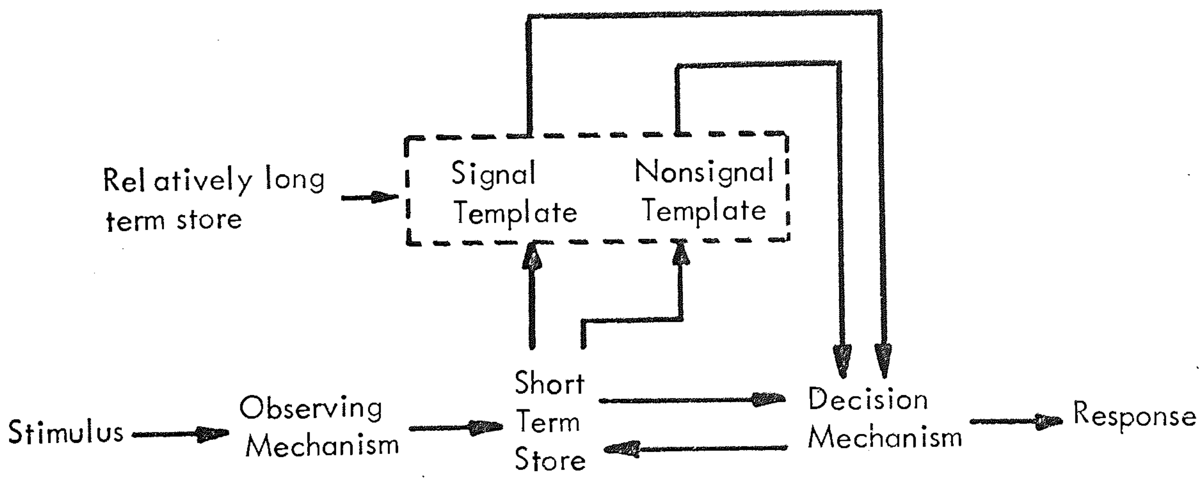


FIGURE 13.4 Information flow diagram of monitoring performance (after Davies and Tune, 1970, p.327).

Memory data-limits and time pressure

Given the negative evidence for the coupling, habituation and observing response approaches to sensitivity decrement, some alternative approaches are considered here. The theories discussed thus far appear to be consistent with some of the data, but cannot interpret data obtained in other monitoring situations. Thus while it appears that failures of observation can account fairly successfully for the occurrence of a sensitivity decrement in continuous tasks requiring visual fixation, such as the Continuous Clock (Mackworth, 1968), it is clear that observing response theory cannot be extended to other types of sensitivity decrement. In particular it cannot explain why a decrement is obtained only for tasks requiring perceptual speed and not flexibility of closure. It is therefore of interest to re-examine the features of these tasks more closely.

In the speed task, the observer must detect a change in some feature of a repetitive stimulus; the information required to make such a discrimination is therefore not present at one point in time; a characteristic of speed tasks is that the distinction between the signal and non-signal stimuli has to be remembered. In the closure task, on the other hand, a single 'stimulus' may contain both the signal and non-signal stimulus features, and these must be discriminated; however the information required for the discrimination is present within a single event at the same instant in time.

These differences between speed and closure tasks immediately suggest two things: firstly, there is a greater dependence on memory in the speed task since information has to be integrated over events. Secondly, while the signal in the closure task is, in a sense, specified uniquely (e.g. a 1KHz tone, a pink circle), the signal in the speed task is only specified relative to another stimulus presented previously (a brighter light than standard, a shorter duration tone than standard). The second distinction between tasks need not always apply when multi-source monitoring tasks are considered (see Chapter 12), but the first distinction always holds.

As we have noted previously, Norman and Bobrow (1975) have pointed out that many discrimination tasks, including monitoring tasks, may be said to be data-limited tasks, in which performance is limited by the relative quality of the input data (see also 2.2). Norman and Bobrow also distinguished between signal data and memory data limits on performance; in the former, performance is more directly dependent on the quality of the signal (as in the detection of a weak signal in noise), while in the latter, performance is affected by limitations in the stored representations of input data (as in delayed comparison tasks). A speed task may be expecially susceptible to performance degradation, therefore, because in addition to limitations on performance due to the quality of the input data, there also may be added demands on memory storage which are difficult to meet in such tasks.

It is thus conceivable that the greater dependence on memory in speed tasks, coupled with the demands of a high rate of stimulation, might underlie the observed decrement in sensitivity in such tasks and not in closure tasks. Under this hypothesis therefore, there is a reduction in the input to the decision making stage in the TSD model; in closure tasks it may be assumed that the sensory evidence required for the discrimination of a signal is not affected by disturbances in the memory store, since information relating to signal and non-signal features are presented together when a detection is required.

In terms of Davies and Tune's (1970) model of vigilance performance (shown in Figure 13.4), in closure tasks, the observing mechanism and the memory store have relatively little effect on the efficiency with which signals are discriminated, because the signal and non-signal features and their 'templates' are activated on each presentation of the signal. In the speed task however, there is a greater dependence on memory for the signal, so that any disturbances in this, which might occur with high stimulation rates, will result in a decrement in sensitivity.

This type of interpretation places a greater emphasis on memory mechanisms than is usual in monitoring research. As Davies and Tune (1970) have pointed out, most investigators have not given much consideration to memory in the interpretation of monitoring performance. Johnston, Howell and Williges (1969) and Williges (1971) have however, suggested that in addition to the two general mechanisms which have been considered in relation to monitoring behaviour, namely attention and decision-making, a factor related to memory also needs to be taken into account. They did not specify the nature of this factor in detail, but proposed that tasks with a high 'memory load' would be especially susceptible to performance deterioration. To demonstrate this, Johnston et al. tested subjects on two monitoring tasks involving the discrimination of alphanumeric stimuli presented in cells of an 8 x 8 matrix. In one task, signals were defined as deletions of some stimuli with respect to the previous display, while in the other task signals were added to the display. It was the former display which was hypothesized to impose the greater load on memory. The results were in support of the hypothesis, in that a significant decrement in d' was obtained for the 'deletions' task but not for the 'additions' task.

Hence the interpretation of the sensitivity decrement in terms of the memory load factor appears to be a plausible explanation for the available results, despite the lack of other supporting data from studies investigating memory factors in vigilance, and the relatively unsophisticated nature of this conceptualization of memory load. The theory may, however, face some difficulty arising out of a consideration of the effects of signal and event rate on sensitivity decrement. If the sensitivity decrement is assumed to be related to reductions in the output of a memory store, then an increase in the signal rate should serve to arrest this process, and hence the decrement. In a related context, Dornic (1967) has suggested that the 'memory trace' for signals is built up and consolidated by the occurrence of signals, while non-signals act as a source of interference which disrupts the memory trace; the degree of consolidation and disruption is proportional to the frequency of intervening signal and non-signal stimuli. Given this assumption, an increase in event rate, which increases the number of intervening non-signals, would lead to a reduction in efficiency over the monitoring period. This is what has been observed for speed tasks with high event rates. However, this interpretation also predicts that an increase in the signal rate should lead to a reduction or removal of the sensitivity decrement, since there will be less disruption of the memory trace. Furthermore, if there is an equal increase in both the signal and event rates, Dornic's theory predicts that there will be no change in performance, since the number of intervening non-signals will remain the same (this is equivalent to the decision theory view that if conditional signal probability is held constant, there is no change in performance). However, as the results of the experiments and those of Loeb and Binford (1968) show, sensitivity decrement is not affected significantly by signal rate. It is not evident whether this poses a difficulty for the theory, since Dornic's proposals do not make clear whether it is the number of intervening stimuli or the intervening time interval length which is the important factor in determining the effects of memory on performance. Presumably, both factors are important, but there is little evidence relating to their independent effects on monitoring performance. It is therefore difficult to evaluate the implications of Dornic's trace-interference theory for the interpretation of sensitivity decrement in terms of memory data-limitations on performance.

Another apparent difficulty with a memory data-limits theory may be considered by briefly examining the features of speed tasks and comparing them with those of the delayed-comparison tasks which are often used in recognition memory research. These tasks share certain features, since in both cases a 'signal' has to be discriminated by comparing it to a previously presented standard; although unlike most monitoring tasks, more than two standard and comparison stimuli are used in delayed-comparison tasks. In such tasks it is usually found that d' decreases as the delay interval between the standard and comparison stimulus increases; and a number of models of the decay in the memory trace for the standard stimulus may be proposed to account for this result (see for example, Kinchla and Smyzer, 1967). At first glance, this result appears to be the opposite of that obtained in Experiments 1 and 4 with regard to the effects of ER (which is the inverse of the interstimulus or delay interval) on d'. However the reduction in d' with ER in these Experiments is due mainly to the decrement over time in the high ER condition, which appears only after about 15 to 30 minutes of monitoring. In the first 15-min. time block, or for 'alerted' detection, there is no effect of ER on d'. Thus there is no discrepancy between the effects of ER in monitoring and delayed-comparison tasks. Furthermore, in delayed-comparison tasks the decrease in d' as the delay interval increases is much less marked when there is only one standard stimulus (Aiken and Lau, 1966).

The memory data-limits theory may also be related to Kahneman's (1973) conceptions of effort and time pressure. Kahneman proposed that effort is 'mobilized' in response to the demands of a task, and that there is a fixed allocation of effort for each task; time pressure, which is inherent in the structure of the task, therefore increases the effort demands of the task. Kahneman also suggested that "the investment of less than this standard (effort) causes a deterioration in performance, but in most tasks it is impossible to completely eliminate errors by a voluntary increase in effort .... Time pressure is a particularly important determinant of mementary effort. Tasks that impose a heavy load on memory necessarily impose severe time pressure" (Kahneman, 1973, p.27). If increased event rate is identified with increased time pressure, then a relationship between Kahneman's theory of effort and the memory load theory may be forged. It remains a loose relationship, but a consideration of sensitivity decrements from the point of view of the effort theory might help to resolve some of the difficulties faced by the memory load theory.

Conclusions on sensitivity decrement

The foregoing considerations have made clear that 'unitary' theories of monitoring behaviour cannot adequately deal with a majority of the available data on monitoring performance. Neither the observing response and coupling models nor the habituation theory of Mackworth (1969) are able to provide a consistent interpretation of the data relating to sensitivity decrement. On the other hand, a theoretical approach based on an identification of the memory demands of various monitoring tasks appears to be more promising, despite some difficulties with the theory as it stands. Further work needs to be done on the effects of practice and other variables on the sensitivity decrement within a task classification framework, in order to see whether the difficulties are major ones, in which case the whole theory may have to be abandoned, or are more minor problems which may be resolved as fresh experimental evidence is gathered. The appeal of this theoretical approach is that it is not only in agreement with the empirical results obtained in this thesis, but also with the specification of tasks giving a sensitivity decrement which may be drawn by analysing all the reported data within a task classification framework.

### 13.2.3 Analysis of response latencies

There are two distinctive features associated with the treatment of response latencies in this thesis. Firstly, the need to examine not only the response times associated with correct identifications of target signals, but also of other response categories has been emphasized. Secondly, a conceptual model was outlined whereby response latency data may be interpreted within the framework of decision theory (Parasuraman and Davies, 1976). Previous investigations of response latency in monitoring tasks have not been associated with either of these features (see Buck, 1966). The investigations in this thesis have been primarily concerned with response latencies in tasks with 'transient' signals and not with the so-called 'hold' tasks (see 2.2), where signals remain present until identified. In general, response latencies for 'hold' signals have been interpreted in other ways, as in the probe measurement of 'mental vigilance' (Haider, 1963), or in the assessment of changing patterns of observation of complex displays (Hockey, 1969).

Buck (1966) has also made the distinction between 'transient' and 'hold' tasks, and incorporated the results with these tasks into a theoretical model of response latency in monitoring tasks. According to the model, the onset of the detection rate decrement is determined by the relationship of the initial 'vigilance level' to a critical level below which performance degradation occurs. The critical level is set by parameters such as signal intensity and duration. If the critical level is low, as in 'hold' tasks, the decline in 'perceptual vigilance' results in an increment in detection latency, but so long as the vigilance level exceeds the (low) critical level, there is no decrement in detection rate. A decline in detection rate and in increment in latency only occur when the critical level is high, that is, when the signal duration is short or the signal is otherwise 'weak'.

Buck's model provides a reasonable conceptual framework for the interpretation of detection and latency data in 'hold' and 'transient' tasks; and the correlation between these two measures which is usually found lends some support to the notion of a common central process of 'perceptual vigilance'. However, the model does not make any provision for false detections, nor does it consider the latencies associated with other response categories. If, as in Buck's model, temporal changes in monitoring performance are assumed to be associated with a decline in arousal or 'perceptual vigilance', the latencies of <u>all</u> response categories might be expected to increase with time at work. The results reported here show that this is not the case; specifically, the latencies of both correct and incorrect rejections of target signals either <u>decrease</u> or remain stable over the work period. These results are not compatible with the notion of a decline in perceptual vigilance, but are consistent with the decision theory view that the criterion becomes increasingly stringent as the task proceeds. Response latency may be postulated to be inversely related to the strength of the evidence received by the observer <u>relative</u> to the chosen criterial level of evidence which is required before a response is elicited. In TSD terms, the further the received evidence is from the criterion, the greater the relative strength of the evidence, and hence the shorter the decision latency. An example from everyday life of the predictions of this model is that it takes less time to decide to place a bet if the odds offered are, say, 100 to 1 than if they are only 2 to 1.

The decision theory latency model proposed in this thesis considered the effects of variations in both the criterion and in sensitivity. On the whole, the results were in harmony with the predictions of the model, although there were some difficulties of interpretation when shifts in both the criterion and sensitivity are obtained (see 7.2.4). The analysis of latency data also raises some unresolved problems relating to the reliability of measures of bias, in particular the problem of whether a change in sensitivity is accompanied by a change in the criterial likelihood ratio B or a change in the criterial cut-off c. The results show that, for the event rate variable at least, it is c rather than B which is kept constant when sensitivity is decreased; this has also been suggested by Hardy and Legge (1968) and McNicol (1972), and, implicitly, by Gesheider, Wright and Evans (1968), but has been disputed by Broadbent (1971), who proposes that it is B rather than c which remains constant. The results do not settle this issue, but indicate that the effects on bias may well be different depending on the variable which brings about the change in sensitivity; thus a set of parameters rather than a single parameter might be required to specify a change in bias. As was noted in the review of bias measures in 4.4.2, a satisfactory clarification of these issues may only be possible when more data within the context of sensitivity shifts and isobias functions are made available. The same conclusion holds for a re-analysis of sensitivity and criterion data in binaural detection reported by Treisman (1972), whose paper indicated the need for considering both c and B in evaluating sensitivity shifts.

Finally, we may briefly consider whether the decision theory model is adequate as it stands, or requires some modifications. It was seen in Chapter 4 that in fixed sample models, the variations in the time taken to gather the evidence are assumed not to significantly influence the overall response latency; response time is assumed to be mainly comprised of the time taken to decide between response alternatives. Broadbent (1973, 1975, 1976), on the other hand, has suggested that, as in a random-walk model, the response latency may be related to the time taken to gather the evidence, which itself depends on the probability of the signal and the criterion level. This type of model may be fitted to some of the data obtained in this thesis, but in its simplest form, the random walk model faces the problem that it predicts that correct and incorrect responses have equal latencies (but see Link and Heath, 1975), and the results clearly falsify this prediction. Some further reasons for the suitability of fixed sample models for monitoring tasks were outlined in 4.6. It was also noted, in Experiment 4, that confident responses are shorter in latency than responses of low confidence level; this result is again consistent with a fixed sample decision theory model.

Experiment 4 also demonstrated that there is less movement over the monitoring period in lax criteria. Broadbent (1975) therefore suggested that there should be similarly less movement in latencies associated with intermediate confidence responses, if the decision theory model is valid; if a change in latency were obtained, then one would have to assume that variations in input sampling time also affect the overall response time. Although this proposal rests on the assumption that lax criteria do remain stable (which has previously been queried), the results for auditory monitoring from Experiment 4 are not compatible with this suggestion. For the visual tasks used in Experiment 1, this could not be examined directly because confidence ratings were not taken. However, one method of analysis for binary response tasks is to divide the response latency distributions into assumed rating categories (as in the construction of latency based operating characteristics; see Norman and Wickelgren, 1969). Such an analysis showed that there were no changes in the latencies of responses 'just clearing' the criterion with time on task.

It does not appear necessary, therefore, to consider an input sampling stage for the analysis of response latency data in the types of monitoring situation studied in this thesis. However, it is fairly clear, particularly from the suggestions of Broadbent (1971, 1973) that such a stage has to be considered in tasks where signals are not limited in duration or are not 'weak'; here a two-process theory, combining a sequential sampling stage with a subsequent decision stage, as suggested by Broadbent (1973, pp. 27 - 30), might be appropriate. Such alternative models might also be necessary for the analysis of latencies in recognition memory tasks, where attempts to interpret latency data in terms of strength theories extended from TSD have been relatively unsuccessful (see Corballis, 1975, for a review). For data-limited monitoring tasks however, a decision theory model of the type proposed in this thesis appears adequate. Of course, the general decision theory analysis was restricted to these tasks for the analysis of detection measures as well. It may therefore be concluded that both detection and latency data for data-limited tasks may be interpreted within the same decision theory framework.

## 13.3 Implications and Suggestions for Further Research

A number of possible departure points for further research have been suggested in this chapter. In this section some of these, and some other implications of the results, will be briefly considered.

As far as a taxonomy of monitoring tasks is concerned, the results show that the task classification system outlined in Chapter 2 suffices to cover a fairly wide range of tasks. It may, however, be considered as a starting point for the development of a more comprehensive taxonomy of monitoring tasks. In particular, Craig and Colquhoun (1975) have pointed out that further research into monitoring performance with multi-attribute signals is required if improved generalizations to operational performance are to be made. It would appear therefore that further developments in this field would profit from an examination of this and other task dimensions, such as visual search and signal duration. One might then go on to incorporate tracking, encoding and other more complex functions in the basic monitoring situation and study these within a taxonomic framework. This might provide a much needed bridge between the studies on simple monitoring and those on time-shared and other more complex tasks.

From the point of view of the specification of the effects of independent variables, the results carry the implication that further research into the effects of other variables on performance, especially those that might be termed 'soft' variables (e.g. motivational and training variables) may also benefit from a taxonomic approach. For instance, it has been demonstrated in this thesis that criterion shifts alone are obtained in some monitoring tasks, while other tasks lead to sensitivity shifts as well. It would therefore be especially valuable from an applied point of view if a taxonomy were used to derive training principles for optimizing response strategies and arresting performance degradation in operational monitoring tasks.

The results of this thesis also suggest that an end to the long and arduous search for reliable correlates of monitoring performance may be in sight. Such attempts at deriving selection devices for monitoring and inspection tasks have been singularly unsuccessful (see Davies and Shackleton, 1975; Wiener, 1975b), and investigators have often been forced to conclude that the best selection device is a short form of the task itself (Baker, 1963; Buckner et al., 1960). The results of

Experiment 3 and 5 and of some recent research (reviewed in Chapter 3) suggest that the situation may not be as bad as this. The categorization of tasks according to the consistency of individual differences, which has been demonstrated in this thesis, represents a first step towards the derivation of more reliable selection procedures, and this might be followed up in further research.

As far as the applications of decision theory are concerned, this thesis has made clear that it provides a reliable method of analysis for a number of different monitoring tasks. Further research is required, however, into the range of applicability of the theory, and, in particular, into the types of tasks for which a rigorous or 'loose' form of the theory is appropriate. It has already been noted that decision theory faces a number of problems in the analysis of performance in multisource and otherwise more 'complex' tasks. The same might be expected of operational monitoring tasks; although, in some applied fields such as industrial inspection (Drury and Fox, 1975; Sheehan and Drury, 1971), medical diagnostics (Swets, 1972) and insurance risk assessment (Baker and Schunck, 1975), the basic theory has proved to be extremely useful and robust.

Finally, on a more theoretical level, it has been previously noted that there is a need for the development of learning detection models in relation to monitoring tasks (Parasuraman, 1976b). The criterion increment may be further analysed using such models, since, unlike TSD, they incorporate the notion of criterion variation. A number of specific points for further research also emerge as a result of the demonstration of the relation between response latencies and decision criteria; the same applies to the results arising out of the analysis of evoked potentials in monitoring.

## 13.4 General Conclusions

The results of the studies reported in this thesis lead to the conclusion that task classification and decision theory provide a useful framework for the evaluation of several aspects of monitoring behaviour in different monitoring tasks. In general, task classification enables a more reliable specification of the range of tasks to which particular classes of performance are restricted. In this thesis, such improved specifications have been demonstrated for 1) the tasks in which reliable decrements in monitoring efficiency occur and 2) the tasks for which individual differences in performance are consistently maintained.

It may also be concluded that improved generalizations may be made regarding the effects of independent variables on performance by classifying tasks within a taxonomic framework, although only a few, yet crucial, interactions between independent variables and task categories emerged in this thesis. These results therefore suggest that theoretical interpretations of behaviour may be refined and selected according to their capacity for generalizing predictions to different tasks. Using this approach, it may be concluded that interpretations based on expectancy mechanisms and memory data limitations provide the most parsimonious explanation of several aspects of monitoring performance.

Decision theory emerges as the most reliable method for the analysis of performance changes in monitoring tasks. Using this analysis either in the general terms of statistical decision theory, or in the more specific terms of the theory of signal detectability, it is possible to distinguish performance changes in monitoring situations where there is a variation in perceptual efficiency from those in which there is only a change in the response strategies employed by subjects. Task classification also permits a specification of the type and range of monitoring situations in which either or both these types of performance change occur.

Two other conclusions relating to task classification and decision theory may be noted. Firstly, task classification serves to point out that a decision theory analysis may only be applicable for tasks in which the available sensory information is limited or unreliable, that is, in the so-called data-limited tasks. In continuous, multiple-choice and other more complex and 'real-life' tasks, more sophisticated elaborations of decision theory may be required. Secondly, support for decision theory is usually sought in the analysis of measures of performance accuracy. The results in this thesis show that the analysis of response latency measures using a model extended from the theory of signal detectability also provides support for the decision theory approach to monitoring behaviour. Thus decision theory enables, within limits, the interpretation of both accuracy and latency performance measures within the same theoretical framework.

Finally, perhaps the most general conclusion of the studies reported in this thesis is that the twin task classification-decision theory approach may be fruitfully applied in other behavioural situations. Some of these include the research areas of the effects of noise on performance, task characteristics in auditory recognition tasks, and the effects of alchohol and other drugs on task performance (see also Fleishman, 1975b). As in monitoring behaviour, task classification may provide a suitable basis for the analysis of performance across different behavioural or task categories, while decision theory, in one of its several manifestations, may provide a suitable method for the analysis of performance itself.

REFERENCES


ABRAHAMSON I.G. and LEVITT H. (1969) Statistical analysis of data from experiments in human signal detection. Journal of Mathematical Psychology, 6, 391-417.

ADAMS J.A. (1956) Vigilance in the detection of low-intensity visual stimuli. Journal of Experimental Psychology, 52, 204-208.

ADAMS J.A., HUMES J.M. and STENSON H.H. (1962) Monitoring of complex displays: III Effects of repeated sessions on human vigilance. Human Factors, 4, 149-158.

AIKEN E.G. and LAU A.W. (1966) Memory for the pitch of a tone. Perception and Psychophysics, 1, 231-233.

ALEXANDER D.A. (1973) Attention dysfunction in senile dementia. Psychological Reports, 32, 229-230.

ALLUISI E.A. (1967) Methodology in the use of synthetic tasks to assess complex performance. Human Factors, 9, 375-384.

ALTHAM P.M.E. (1973) A non-parametric measure of signal discriminability. British Journal of Mathematical and Statistical Psychology, 26, 1-12.

ANDERSON N.H. (1974) Algebraic models in perception. In Cartarette E.C. and Friedman M.P. (eds.), Handbook of Perception Volume II. New York: Academic Press.

ANDERSON R.P., HALCOMB C.G. and DOYLE R.B. (1973) The measurement of attentional deficits. Exceptional Children, 39, 534-539.

ANNETT J. and DUNCAN K.D. (1967) Task analysis and training design. Occupational Psychology, 41, 211-221.

ANNETT J. and PATERSON L. (1967) Training for auditory detection. Acta Psychologica, 27, 420-426.

ATKINSON R.C. (1963) A variable sensitivity theory of signal detection. Psychological Review, 70, 91-106.

ATKINSON R.C. and KINCHLA R.A. (1965) A learning model for forced-choice detection experiments. British Journal of Mathematical and Statistical Psychology, 18, 183-206.

AUDLEY R.J. (1970) Choosing. Bulletin of the British Psychological Society, 23, 177-191.

AUDLEY R.J. (1973) Some observations on theories of choice reaction time: Tutorial survey. In Kornblum S. (Ed.), Attention and Performance IV. New York: Academic Press.

AUDLEY R.J. and MERCER A. (1968) The relation between decision time and relative response frequency in a blue-green discrimination. British Journal of Mathematical and Statistical Psychology, 21, 183-192.

AUDLEY R.J. and PIKE A.R. (1965) Some alternative stochastic models of choice. British Journal of Mathematical and Statistical Psychology, 18, 207-225.

BADDELEY A.D. and COLQUHOUN W.P. (1969) Signal probability and vigilance: A reappraisal of the 'signal rate' effect. British Journal of Psychology, 60, 169-178.

BADDELEY A.D., HATTER J.E., SCOTT D. and SNASHALL A. (1970) Memory and time of day. Quarterly Journal of Experimental Psychology, 22, 605-609.

BAKAN P. (1959) Extraversion-introversion and improvement in an auditory vigilance task. British Journal of Psychology, 50, 325-332.

BAKER C.H. (1959) Towards a theory of vigilance. Canadian Journal of Psychology, 13, 35-42.

BAKER C.H. (1963) Consistency of performance in two visual vigilance tasks. In Buckner D.N. and McGrath J.J. (Eds.), Vigilance: A Symposium. New York: McGraw-Hill.

BAKER E.M. and SCHUCK J.R. (1975) Theoretical note: Use of signal detection theory to clarify problems of evaluating performance in industry. Organizational Behavior and Human Performance, 13, 307-317.

BAKER R.A. and WARE J.R. (1966) The relationship between vigilance and monotonous work. Ergonomics, 9, 109-114.

BENNEDETTI L.H. and LOEB M. (1972) A comparison of auditory monitoring performance in blind subjects with that of sighted subjects in light and dark. Perception and Psychophysics, 11, 10-16.

BERGUM B.O. (1966) A taxonomic analysis of continuous performance. Perceptual and Motor Skills, 23, 47-54.

BINFORD J.R. and LOEB M. (1966) Changes within and over repeated sessions in criterion and effective sensitivity in an auditory vigilance task. Journal of Experimental Psychology, 72, 339-345.

BLACKWELL H.R. (1963) Neural theories of simple visual detection. Journal of the Optical Society of America, 53, 129-160.

BLACKWELL P.J. and BELT J.A. (1971) Effect of differential levels of ambient noise on vigilance performance. Perceptual and Motor Skills, 32, 734.

BLAKE M.J.F. (1967) Time of day effects on performance in a range of tasks. Psychonomic Science, 9, 349-350.

BROADBENT D.E. (1950) The twenty dials test under quiet conditions (MRC APU Report 130/50).

BROADBENT D.E. (1951) The twenty dials and twenty lights tests under noise conditions (MRC APU Report 160/51).

BROADBENT D.E. (1954) Some effects of noise on visual performance. Quarterly Journal of Experimental Psychology, 6, 1-5.

BROADBENT D.E. (1958) Perception and Communication. London: Pergamon.

BROADBENT D.E. (1963) Possibilities and difficulties in the concept of arousal. In Buckner D.N. and McGrath J.J. (Eds.), Vigilance: A Symposium. New York: McGraw-Hill.

BROADBENT D.E. (1965) A reformulation of the Yerkes-Dodson Law. British Journal of Mathematical and Statistical Psychology, 18, 145-157.

BROADBENT D.E. (1967) Word-frequency effect and response bias. Psychological Review, 74, 1-15.

BROADBENT D.E. (1971) Decision and Stress. London: Academic Press.

BROADBENT D.E. (1973) In Defence of Empirical Psychology. London: Methuen.

BROADBENT D.E. (1975, 1976) Personal communication to D.R. Davies.

BROADBENT D.E. and GREGORY M. (1963a) Vigilance considered as a statistical decision. British Journal of Psychology, 54, 309-323.

BROADBENT D.E. and GREGORY M. (1963b) Division of attention and the decision theory of signal detection. Proceedings of the Royal Society B, 158, 221-231.

BROADBENT D.E. and GREGORY M. (1965) Effects of noise and of signal rate upon vigilance analysed by means of decision theory. Human Factors, 7, 155-162.

BROADBENT D.E. and GREGORY M. (1967) Perception of emotionally toned words. Nature, 215, 581-584.

BUCK L. (1966) Reaction time as a measure of perceptual vigilance. Psychological Bulletin, 65, 291-308.

BUCKNER D.N. (1963) An individual-difference approach to explaining vigilance performance. In Buckner D.N. and McGrath J.J. (Eds.), Vigilance: A Symposium. New York: McGraw-Hill.

BUCKNER D.N., HARABEDIAN A. and MCGRATH J.J. (1960) A study of individual differences in vigilance performance (Human Factors Research Inc. Report TR-2, Los Angeles).

BUCKNER D.N. and MCGRATH J.J. (1963a) (Eds.), Vigilance: A Symposium. New York: McGraw-Hill.

BUCKNER D.N. and MCGRATH J.J. (1963b) A comparison of performance on single and dual sensory mode vigilance tasks. In Buckner D.N. and McGrath J.J. (Eds.). Vigilance: A Symposium. New York: McGraw-Hill.

CARTARETTE E.C, FRIEDMAN M.P. and COSMIDES R. (1965) Reaction-time distributions in the detection of weak signals in noise. Journal of the Acoustical Society of America, 38, 531-542.

CLARK D.L., BUTLER R.A. and ROSNER B.S. (1970) Are evoked responses necessary?: A reply to Donchin and Sutton. Communications in Behavioral Biology, 5, 105-110.

CLARK W.C. and MEHL L. (1973) Signal detection theory procedures are not equivalent when thermal stimuli are judged. Journal of Experimental Psychology, 97, 148-153.

COLES M.G.H. and GALE A. (1971) Physiological reactivity as a predictor of performance in a vigilance task. Psychophysiology, 8, 594-599.

COLQUHOUN W.P. (1961) The effect of unwanted signals on performance in a vigilance task. Ergonomics, 4, 41-52.

COLQUHOUN W.P. (1966) The effects of unwanted signals on performance in a vigilance task: A reply to Jerison. Ergonomics, 9, 417-419.

COLQUHOUN W.P. (1967) Sonar target detection as a decision process. Journal of Applied Psychology, 51, 187-190.

COLQUHOUN W.P. (1969) Effects of raised ambient temperature and event rate on vigilance performance. Aerospace Medicine, 40, 413-417.

COLQUHOUN W.P. (1975) Evaluation of auditory, visual and dual-mode displays for prolonged sonar monitoring in repeated sessions. Human Factors, 17, 425-437.

COLQUHOUN W.P. and BADDELEY A.D. (1964) The role of pre-test expectancy in vigilance decrement. Journal of Experimental Psychology, 68, 156-160.

COLQUHOUN W.P. and BADDELEY A.D. (1967) Influence of signal probability during pretraining on vigilance decrement. Journal of Experimental Psychology, 73, 153-155.

COLQUHOUN W.P., BLAKE M.J.F. and EDWARDS R.S. (1968) Experimental studies of shift workI: A comparison of 'rotating' and 'stabilized' 4 hour shift systems. Ergonomics, 11, 437-453.

COLQUHOUN W.P. and EDWARDS R.S. (1970) Practice effects on a visual vigilance task with and without search. Human Factors, 12, 537-546.

COLQUHOUN W.P. and GOLDMAN R.F. (1972) Vigilance under induced hyperthemia. Ergonomics, 15, 621-632.

CORBALLIS M.C. (1975) Access to memory: An analysis of recognition times. In Rabbitt P.M.A. and Dornic S. (Eds.), Attention and Performance V. London: Academic Press.

CORCORAN D.W.J. (1965) Personality and the inverted-U relation. British Journal of Psychology, 56, 267-273.

COTTERMAN T.E. (1959) Task classification: An approach to partially ordering information on human learning (WADC TN-58-374, Wright-Patterson Air Force Base, Ohio).

COULES J. and AVERY D.L. (1966) Human performance and basal skin conductance in a vigilance-type task with and without knowledge of results. Perceptual and Motor Skills, 23, 1295-1302.

CRAIG A. and COLQUHOUN W.P. (1975) Vigilance: A review. In Drury C.G. and Fox J.G. (Eds.), Human Reliability in Quality Control. London: Taylor and Francis.

CREELMAN C.F. and DONALDSON W. (1968) ROC curves for discrimination of linear extent. Journal of Experimental Psychology, 77, 514-516.

DANIEL R.S. (1967) Alpha and theta EEG in vigilance. Perceptual and Motor Skills, 25, 697-703.

DARDANO J.F. (1969) Observing behaviour of mental patients under a fixed interval schedule. Psychological Reports, 24, 635-653.

DAVENPORT W.G. (1969) Vibrotactile vigilance: Effects of costs and values of signals. Perception and Psychophysics, 5, 25-28.

DAVIES D.R. (1968) Physiological and psychological effects of exposure to high intensity noise. Applied Acoustics, 1, 215-233.

DAVIES D.R. (1970) Monotony and work. Science Journal, 6, 26-30.

DAVIES D.R. and HOCKEY G.R.J. (1966) The effects of noise and doubling the signal frequency on individual differences in visual vigilance performance. British Journal of Psychology, 57, 381-389.

DAVIES D.R., HOCKEY G.R.J. and TAYLOR A. (1969) Varied auditory stimulation, temperament differences and vigilance performance. British Journal of Psychology, 60, 453-457.

DAVIES D.R., LANG L. and SHACKLETON V.J. (1973) The effects of music and task difficulty on performance at a visual vigilance task. British Journal of Psychology, 64, 383-389.

DAVIES D.R. and PARASURAMAN R. (1976) Cortical evoked potentials and vigilance: A decision theory analysis. Paper to be presented at the Symposium on Vigilance: Relationships between Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy.

DAVIES D.R. and SHACKLETON V.J. (1975) Psychology and Work. London: Methuen.

DAVIES D.R. and TREACHER A.C.C. (1968) The effects of increasing the signal frequency on heart rate changes during vigilance performance. (Unpublished report, University of Leicester).

DAVIES D.R. and TUNE G.S. (1970) Human Vigilance Performance. London: Staples Press.

DEATON M., TOBIAS J.S. and WILKINSON R.T. (1971) The effect of sleep deprivation on signal detection parameters. Quarterly Journal of Experimental Psychology, 23, 449-451.

DEESE J. (1955) Some problems in the theory of vigilance. Psychological Review, 62, 359-368.

DONCHIN E., KUBOVY M., KUTAS M., JOHNSON R. and HERNING R. (1973) Graded changes in evoked response (P300) amplitude as a function of cognitive activity. Perception and Psychophysics, 14, 319-324.

DONCHIN E. and LINDSLEY D.B. (1966) Averaged evoked potentials and reaction times to visual stimuli. Electroencephalography and Clinical Neurophysiology, 20, 217-223.

DONCHIN E. and LINSDLEY D.B. (1969) (Eds.), Average Evoked Potentials: Methods Results and Evaluations. Washington DC: NASA SP-191.

DONCHIN E. and SUTTON S. (1970) The 'psychological significance' of evoked responses: A comment on Clark, Butler and Rosner. Communications in Behavioral Biology, 5, 111-114.

DONCHIN E., TUETING P., RITTER W., KUTAS M. and HEFFLEY E. (1975) On the independence of the CNV and P300 components of the human averaged evoked potential. Electroencephalography and Clinical Neurophysiology, 38, 449-461.

DORFMAN D.D. and ALF E. Jnr. (1968) Maximum likelihood estimation of parameters of signal detection theory - a direct solution. Psychometrika, 33, 117-124.

DORFMAN D.D. and ALF E. Jnr. (1969) Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals - rating method data. Journal of Mathematical Psychology, 6, 487-496.

DORNIC S. (1967) Expectancy of signals and the memory trace. Studia Psychologica, 9, 87-91.

DRURY C.G. and FOX J.G. (1975) (Eds.), Human Reliability in Quality Control. London: Taylor and Francis.

DUFFY E. (1962) Activation and Behavior. New York: Wiley.

DUSOIR A.E. (1974) Thomas and Legge's matching hypothesis for detection and recognition tasks: Two tests. Perception and Psychophysics, 16, 466-470.

EASON R.G., BEARDSHALL A. and JAFFEE S. (1965) Performance and physiological indicants of activation in a vigilance situation. Perceptual and Motor Skills, 20, 3-13.

EDWARDS W., LINDMANN H. and PHILLIPS L.D. (1965) Emerging technologies for making decisions. In Newcomb T.M. (Ed.), New Directions in Psychology Vol. 2. New York: Holt, Rinehart and Winston.

EGAN J.P. (1975) Signal Detection Theory and ROC Analysis. New York: Academic Press.

EGAN J.P. and CLARKE F.R. (1966) Psychophysics and signal detection. In Sidowski J.B. (Ed.), Experimental Methods and Instrumentation in Psychology. New York: McGraw-Hill.

EGAN J.P., GREENBERG G.L. and SCHULMAN A.I. (1961a) Operating characteristics, signal detectability and the method of free response. Journal of the Acoustical Society of America, 37, 1102-1109.

EGAN J.P., SHULMAN A.I. and GREENBERG G.Z. (1959) Operating characteristics determined by binary decisions and by ratings. Journal of the Acoustical Society of America, 31, 768-773.

EGAN J.P., SCHULMAN A.I. and GREENBERG G.Z. (1961b) Memory for waveform and time uncertainty in auditory detection. Journal of the Acoustical Society of America, 33, 779-781.

EGETH H. and SMITH E.E. (1967) On the nature of errors in a choice reaction time task. Psychonomic Science, 8, 345-346.

EIJKMAN E. and VENDRIK A.J.H. (1965) Can a sensory system be specified by its internal noise? Journal of the Acoustical Society of America, 37, 1102-1109.

ELLIOT E. (1960) Perception and alertness. Ergonomics, 3, 357-364.

EMMERICH D.S. (1968) ROCs obtained with two signal intensities presented in random order, and a comparison between Yes-No and rating ROCs. Perception and Psychophysics, 3, 35-40.

EMMERICH D.S., GRAY J.L., WATSON C.S. and TANIS D.C. (1972) Response latency, confidence and ROCs in auditory signal detection. Perception and Psychophysics, 11, 65-72.

EYSENCK H.J. (1959) The Maudsley Personality Inventory. London: University of London Press.

EYSENCK H.J. and EYSENCK S.B.G. (1964) Eysenck Personality Inventory. San Diego: Educational and Testing Service.

FARINA A.J. Jnr. (1969) Development of a taxonomy of human performance: A review of descriptive schemes for human task behavior (American Institutes for Research - TR-2, Washington DC).

FARINA A.J. Jnr. and WHEATON G.R. (1971) Development of a taxonomy of human performance: The task characteristics approach to performance prediction (American Institutes for Research TR-7, Washington DC).

FITTS P.M. (1962) Factors in complex skill training. In Glaser R. (Ed.), Training Research and Education. Pittsburgh: University of Pittsburgh Press.

FLEISHMAN E.A. (1967a) Performance assessment based on an empirically derived task taxonomy. Human Factors, 9, 349-366.

FLEISHMAN E.A. (1967b) Development of a behavior taxonomy for describing human tasks: A correlational experimental approach. Journal of Applied Psychology, 51, 1-10.

FLEISHMAN E.A. (1972) On the relation between abilities, learning and human performance. American Psychologist, 27, 1017-1032.

FLEISHMAN E.A. (1975a) Toward a taxonomy of human performance. American Psychologist, 30, 1127-1149.

FLEISHMAN E.A. (1975b) Taxonomic problems in human performance research. In Singleton W.T. and Spurgeon P. (Eds.), Measurement of Human Resources. London: Taylor and Francis.

FRANKMANN J.P. and ADAMS J.A. (1962) Theories of vigilance. Psychological Bulletin, 257-272.

GAGNÉ R.M. (1962) Human functions in systems. In Gagné R.M. (Ed.), Psychological Principles in System Development. New York: Holt, Rinehart and Winston.

GALANTER E. (1974) Psychological decision mechanisms and perception. In Cartarette E.C. and Friedman M.P. (Eds.), Handbook of Perception Vol. II. New York: Academic Press.

GARNER W.R. (1974) Attention: The processing of multiple sources of information. In Cartarette E.C. and Friedman M.P. (Eds.), Handbook of Perception Vol. II. New York: Academic Press.

GAUSSIN J. (1972) Observers detecting a signal in two multiple observation tasks. Acta Psychologica, 36, 253-274.

GESHEIDER G.A., WRIGHT J.H. and EVANS M.B. (1968) Reaction time in the detection of vibrotactile signals. Journal of Experimental Psychology, 77, 501-504.

GREEN D.M. (1964) General prediction relating yes-no and forced choice results. Journal of the Acoustical Society of America, 36, 1042.

GREEN D.M. and LUCE R.D. (1971) Detection of auditory signals presented at random times III. Perception and Psychophysics, 9, 257-268.

GREEN D.M. and MOSES F.L. (1966) On the equivalence of two recognition measures of short-term memory. Psychological Bulletin, 66, 228-234.

GREEN D.M. and SWETS J.A. (1966) Signal Detection Theory and Psychophysics. New York: Wiley (2nd Edition, 1973).

GREENHOUSE S.W. and GEISSER S. (1959) On methods in the analysis of profile data. Psychometrika, 24, 95-112.

GREY D.R. and MORGAN B.J.T. (1972) Some aspects of ROC curve-fitting: Normal and logistic models. Journal of Mathematical Psychology, 9, 128-139.

GRIER J.B. (1971) Non-parametric indexes for sensitivity and bias: Computing formulas. Psychological Bulletin, 75, 424-429.

GROLL E. (1966) Central nervous system and peripheral activation variables during vigilance performance. Z. exp. Ang. Psychol., 13, 248-264. (German).

GRUBER A. (1964) Sensory alternation and performance in a vigilance task. Human Factors, 6, 3-12.

GUILFORD J.P. (1967) The Nature of Human Intelligence. New York: McGraw-Hill.

GUNN W.J. and LOEB M. (1967) Correlation of performance in detecting visual and auditory signals. American Journal of Psychology, 80, 236-242.

GURALNICK M.J. (1973) Effects of event rate and signal difficulty on observing responses and detection measures. Journal of Experimental Psychology, 99, 261-265.

GURALNICK M.J. and HARVEY K.G. (1970) Response requirement and performance in a visual vigilance task. Psychonomic Science, 20, 215-217.

HAIDER M. (1963) Experimental work on continued attention and mental vigilance during a monotonous task. Z. exp. Ang. Psychol., 10, 1-18 (German).

HAIDER M., SPONG P. and LINDSLEY D.B. (1964) Attention, vigilance and cortical evoked potentials in humans. Science, 145, 180-182.

HAMILTON P. (1969) Selective attention in multisource monitoring tasks. Journal of Experimental Psychology, 82, 34-37.

HAMMERTON M. and ALTHAM P.M.E. (1971) A non-parametric alternative to d'. Nature, 234, 487-488.

HARDY G.R. and LEGGE D. (1968) Cross-modal induction of changes in sensory thresholds. Quarterly Journal of Experimental Psychology, 20, 20-29.

HARTLEY L.R., OLSSON R. and INGLEBY J.D. (1973) Visual assistance in an auditory detection task. Quoted in Poulton (1973).

HATFIELD J.L. and LOEB M. (1968) Sense mode coupling in a vigilance task. Perception and Psychophysics, 4, 29-36.

HATFIELD J.L. and SODERQUIST D.R. (1969) Practice effects and signal detection indices in an auditory vigilance task. Journal of the Acoustical Society of America, 46, 1458-1463.

HATFIELD J.L. and SODERQUIST D.R. (1970) Coupling effects and performance in vigilance tasks. Human Factors, 12, 351-359.

HAYS W.L. (1963) Statistics. London: Holt, Rinehart and Winston.

HENMON V.A.C. (1911) The relation of the time of a judgement to its accuracy. Psychological Review, 18, 186-201.

HERON A. (1956) A two-part personality inventory for use as a research criterion. British Journal of Psychology, 47, 243-251.

HILLYARD S.A. and PICTON T.W. (1974) Human auditory evoked potentials. II: Effects of attention. Electroencephalography and Clinical Neurophysiology, 36, 191-199.

HOCKEY G.R.J. (1969) Environmental stress and attention to complex visual displays. (Unpublished PhD. Thesis, University of Cambridge).

HOCKEY G.R.J. and COLQUHOUN W.P. (1972) Diurnal variation in human performance: A review. In Colquhoun W.P. (Ed.), Aspects of Human Efficiency: Diurnal Rhythm and Loss of Sleep. London: English University Press.

HILLYARD S., SQUIRES K.C., BAUER J.W. and LINDSAY P.H. (1971) Evoked potential correlates of auditory signal detection. Science, 172, 1357-1360.

HODOS W. (1970) Nonparametric index of response bias for use in detection and recognition experiments. Psychological Bulletin, 74, 351-354.

HOLLAND J.G. (1963) Human vigilance. In Buckner D.N. and McGrath J.J. (Eds.), Vigilance: A Symposium. New York: McGraw-Hill.

HOWELL W.C., JOHNSTON W.A. and GOLDSTEIN I.L. (1966) Complex monitoring and its relation to the classical problem of vigilance. Organizational Behavior and Human Performance, 1, 129-150.

INGHAM J.G. (1970) Individual differences in signal detection. Acta Psychologica, 34, 39-50.

INGLEBY J.D. (1968) Decision-making processes in human perception and memory. (Unpublished Ph.D. Thesis, University of Cambridge).

INGLEBY J.D. (1973) The separation of bias and sensitivity in multiple-choice tasks. Perception, 2, 295-305.

JEFFRESS L.A. (1971) Binaural phenomena. In Tobias J.V. (Ed.) Modern Foundations of Auditory Theory. New York: Academic Press.

JENKINS H.M. (1958) The effects of signal rate on performance in visual monitoring. American Journal of Psychology, 71, 647-661.

JERISON H.J. (1967a) Signal detection theory in the analysis of human vigilance. Human Factors, 9, 285-288.

JERISON H.J. (1967b) Activation and long-term performance. Acta Psychologica, 27, 373-389.

JERISON H.J. (1970) Vigilance: A paradigm and some physiological speculations. Acta Psychologica, 33, 367-380.

JERISON H.J. and PICKETT R.M. (1964) Vigilance: The importance of the elicited observing rate. Science, 143, 970-971.

JERISON H.J., PICKETT R.M. and STENSON H.H. (1965) The elicited observing rate and decision process in vigilance. Human Factors, 7, 107-138.

JERISON H.J. and WALLIS R.A. (1957) Experiments on vigilance: II One-clock and three-clock monitoring. (USAF WADC TR 57-206).

JOHNSTON W.A., HOWELL W.C. and WILLIGES R.G. (1969) The components of complex monitoring. Organizational Behavior and Human Performance, 4, 112-124.

KAC M. (1962) A note on learning signal detection. IRE Transactions on Information Theory, 8, 126-128.

KAHNEMAN D. (1973) Attention and Effort. Englewood Cliffs, N.J.: Prentice-Hall.

KAPPAUF W.E. and POWE W.E. (1959) Performance decrement at an audio-visual checking task. Journal of Experimental Psychology, 57, 49-56.

KARLIN L. (1970) Cognition, preparation and sensory evoked potentials. Psychological Bulletin, 73, 122-136.

KEISTER M.E. and McLAUGLIN R.J. (1972) Vigilance performance related to introversion-extraversion and caffeine. Journal of Experimental Research in Personality, 6, 5-11.

KENNEDY R.S. (1971) A comparison of performance on visual and auditory vigilance tasks. Human Factors, 13, 93-98.

KEPPEL G. (1973) Design and Analysis: A Researchers Handbook. Englewood Cliffs, N.J.: Prentice-Hall.

KINCHLA R. and SMYZER F. (1967) A diffusion model of perceptual memory. Perception and Psychophysics, 2, 219-229.

KIRK R.E. (1968) Experimental Design: Procedures for the Behavioral Sciences. Belmont, Calif: Brooks-Cole.

KRANTZ D.H. (1969) Threshold theories of signal detection. Psychological Review, 76, 308-324.

LACEY J.I. (1967) Somatic response patterning and stress: Some revisions of activation theory. In Appley M.H. and Trumbull R. (Eds.), Psychological Stress. New York: Appleton-Century Crofts.

LAMING D.R.J. (1968) Information Theory of Choice-Reaction Times. London: Academic Press.

LAMING D.R.J. (1973) Mathematical Psychology. London: Academic Press.

LARKIN W.D. (1965) Rating scales in detection experiments. Journal of the Acoustical Society of America, 37, 748-749.

LEMMON V.W. (1927) The relation of reaction time to measures of intelligence, memory and learning. Archives of Psychology, 94, 51-63.

LESHOWITZ B. (1969) Comparison of ROC curves from one- and two-interval rating scale procedures. Journal of the Acoustical Society of America, 46, 399-402.

LEVINE J.M. (1966) The effects of values and costs on the detection and identification of signals in auditory vigilance. Human Factors, 8, 525-537.

LEVINE J.M., GREENBAUM G.D. and NOTKIN E.R. (1973) The effects of alchohol on human performance: A classification and integration of research findings. (American Institutes for Research, Washington DC).

LEVINE J.M., ROMASHKO T. and FLEISHMAN E.A. (1971) Development of a taxonomy of human performance: Evaluation of an abilities classification system for integrating and generalizing research findings. (American Institutes for Research TR-12, Washington DC).

LEVINE J.M., ROMASHKO T. and FLEISHMAN E.A. (1973) Evaluation of an abilities classification system for integrating and generalizing human performance research findings: An application to vigilance tasks. Journal of Applied Psychology, 58, 149-157.

LEVINE J.M. and TEICHNER W.H. (1971) Development of a taxonomy of human performance: An information-theoretic approach. (American Institutes for Research TR-9, Washington DC).

LEWIS J., BADDELEY A.D., BONHAM K.G. and LOVETT D. (1970) Traffic pollution and mental efficiency. Nature, 225, 95-96.

LINK S.W. and HEATH R.A. (1975) A sequential theory of psychological discrimination. Psychometrika, 40, 77-105.

LOEB M. and BINFORD J.R. (1964) Vigilance for auditory intensity changes as a function of preliminary feedback and confidence level. Human Factors, 7, 445-458.

LOEB M. and BINFORD J.R. (1968) Variation in performance on auditory and visual monitoring tasks as a function of signal and stimulus frequencies. Perception and Psychophysics, 4, 361-366.

LOEB M. and BINFORD J.R. (1971) Modality, difficulty and coupling in vigilance behavior. American Journal of Psychology, 84, 529-541.

LUCE R.D. (1959) Individual Choice Behavior. New York: Wiley.

LUCE R.D. (1963a) A threshold theory for simple detection experiments. Psychological Review, 70, 61-79.

LUCE R.D. (1963b) Detection and recognition. In Luce R.D., Bush R.R. and Galanter E. (Eds.), Handbook of Mathematical Psychology Vol. 1. New York: Wiley.

LUCE R.D. (1966) A model for detection in temporally unstructured experiments with a Poisson distribution of signal presentations. Journal of Mathematical Psychology, 3, 48-64.

LUCE R.D. and GREEN D.M. (1972) A neural timing theory for response times and the psychophysics of intensity. Psychological Review, 79, 14-57.

LUCE R.D. and GREEN D.M. (1974) Detection, discrimination and recognition. In Cartarette E.C. and Friedman M.P. (Eds.), Handbook of Perception Vol. II. New York: Academic Press.

MACKWORTH J.F. (1965a) Deterioration of signal detectability during a vigilance task as a function of background event rate. Psychonomic Science, 3, 421-422.

MACKWORTH J.F. (1965b) Decision interval and signal detectability during a vigilance task. Canadian Journal of Psychology, 19, 111-117.

MACKWORTH J.F. (1968) Effects of signal rate on performance in two kinds of vigilance task. Human Factors, 10, 11-18.

MACKWORTH J.F. (1969) Vigilance and Habituation: A Neuropsychological Approach. Harmondsworth, Middlesex: Penguin.

MACKWORTH J.F. (1970) Vigilance and Attention: A Signal Detection Approach. Harmondsworth, Middlesex: Penguin.

MACKWORTH J.F. and TAYLOR M.M. (1963) The d' measure of signal detectability in vigilance-like situations. Canadian Journal of Psychology, 17, 302-325.

MACKWORTH N.H. (1948) The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology, 1, 6-21.

MACKWORTH N.H. (1950) Researches on the measurement of human performance. (MRC Special Report 268, London: HMSO).

MACKWORTH N.H. (1957) Some factors affecting vigilance. Advancement of Science, 53, 389-393.

MARKOWITZ J. and SWETS J.A. (1967) Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. Perception and Psychophysics, 2, 91-100.

MCGILL W.J. (1967) Neural counting mechanisms and energy detection in audition. Journal of Mathematical Psychology, 4, 351-376.

MCGRATH J.J. (1963) Some problems of definition and criteria in the study of vigilance performance. In Buckner D.N. and McGrath J.J. (Eds.), Vigilance: A Symposium. New York: McGraw-Hill.

MCNICOL D. (1972) A Primer of Signal Detection Theory. London: Allen and Unwin.

MELTON A.W. and BRIGGS G.E. (1960) Engineering psychology. Annual Review of Psychology, 11, 71-98.

MILLER R.B. (1967) Task taxonomy: Science or technology? In Singleton W.T., Easterby R.S. and Whitfield D. (Eds.), The Human Operator in Complex Systems. London: Taylor and Francis.

MILLER R.B. (1971) Development of a taxonomy of human performance: A user-oriented approach. (American Institutes for Research TR-6, Washington DC).

METZGER K.R., WARM J.S. and SENTER R.J. (1974) Effects of background event rate and artificial signals on vigilance performance. Perceptual and Motor Skills, 38, 1175-1181.

MILOSEVIC S. (1974) Effect of time and space uncertainty on a vigilance task. Perception and Psychophysics, 15, 331-334.

MILOSEVIC S. (1975) Changes in detection measures and skin resistance during an auditory vigilance task. Ergonomics, 18, 1-18.

MOORE S.F. and GROSS S.J. (1973) Influence of critical signal regularity, stimulus event matrix and cognitive style on vigilance performance. Journal of Experimental Psychology, 99, 137-139.

MORAY N. and O'BRIEN T. (1967) Signal-detection theory applied to selective listening. Journal of the Acoustical Society of America, 42, 765-772.

NAATANEN R. (1975) Selective attention and evoked potentials in humans: A critical review. Biological Psychology, 2, 237-307.

NACHMIAS J. (1968) Effects of presentation probability and number of response alternatives on simple visual detection. Perception and Psychophysics, 3, 151-155.

NACHMIAS J. (1972) Signal detection theory and its application to problems in vision. In Jameson D. and Hurvich L.M. (Eds.), Handbook of Sensory Physiology Vol. VII/4. Berlin: Springer-Verlag.

NACHMIAS J. and STEINMAN R.M. (1963) Study of absolute visual detection by the rating scale method. Journal of the Optical Society of America, 53, 1206-1213.

NORMAN D.A. (1964) Sensory thresholds, response biases, and the neural quantum theory. Journal of Mathematical Psychology, 1, 88-120.

NORMAN D.A. and BOBROW D.G. (1975) On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.

NORMAN D.A. and WICKELGREN W.A. (1969) Strength theory of decision rules and latency in retrieval from short-term memory. Journal of Mathematical Psychology, 6, 192-208.

OGILVIE J.C. and CREELMAN C.D. (1968) Maximum likelihood estimation of receiver operating characteristic curve parameters. Journal of Mathematical Psychology, 5, 377-391.

OPTON E.M., Jnr. (1964) Electroencephalographic correlates of performance lapses on an attention task in young and older men. Dissertation Abstracts, 25, 3115-3116.

PARASURAMAN R. (1975a) Response bias and physiological reactivity. Journal of Psychology, 91, 309-313.

PARASURAMAN R. (1975b) Response latencies in visual monitoring. Perceptual and Motor Skills, 40, 636.

PARASURAMAN R. (1976a) Consistency of individual differences in human performance: An abilities classification analysis. Journal of Applied Psychology, 60 (in press).

PARASURAMAN R. (1976b) A learning model for unstructured detection tasks. In preparation.

PARASURAMAN R. and DAVIES D.R. (1975) Response and evoked potential latencies associated with commission errors in visual monitoring. Perception and Psychophysics, 17, 465-468.

PARASURAMAN R. and DAVIES D.R. (1976) Decision theory analysis of response latencies in vigilance. Journal of Experimental Psychology: Human Perception and Performance, 2 (in press).

PARDUCCI A. and SANDUSKY A.J. (1970) Limits on the applicability of signal detection theories. Perception and Psychophysics, 7, 63-64.

PARKS T.E. (1966) Signal-detectability theory of recognition memory performance. Psychological Review, 73, 44-58.

PAUL D.D. and SUTTON S. (1972) Evoked potential correlates of response criterion in auditory signal detection. Science, 177, 362-364.

PAUL D.D. and SUTTON S. (1973) Evoked potential correlates of psycho-physical judgements: The threshold problem. A new reply to Clark, Butler and Rosner. Behavioral Biology, 9, 421-433.

PICTON T.W. and HILLYARD S.A. (1974) Human auditory evoked potentials: II Effects of attention. Electroencephalography and Clinical Neurophysiology, 36, 191-200.

PICTON T.W., HILLYARD S.A., GALAMBOS R. and SCHIFF M. (1971) Human auditory attention: A central or peripheral process? Science, 173, 351-353.

PIKE A.R. (1968) Latency and relative frequency of response in psychophysical discrimination. British Journal of Mathematical and Statistical Psychology, 21, 161-182.

POLLACK I. and NORMAN D.A. (1964) A nonparametric analysis of recognition experiments. Psychonomic Science, 1, 125-126.

POLLACK I., NORMAN D.A. and GALANTER E. (1964) An efficient non-parametric analysis of recognition memory. Psychonomic Science, 1, 327-328.

POPE L.T. and MCKECHNIE D.F. (1963) Correlation between visual and auditory vigilance performance. (USAF, AMRL TR-63-57).

POULTON E. C. (1960) The optimal perceptual load in a paced auditory inspection task. British Journal of Psychology, 51, 127-139.

POULTON E.C. (1970) Environment and Human Efficiency. Springfield, Illinois: Thomas.

POULTON E.C. (1973) The effect of fatigue on inspection work. Applied Ergonomics, 4, 73-83.

POULTON E.C. and EDWARDS R.S. (1974) Interactions and range effects in experiments on pairs of stresses: Mild heat and low-frequency noise. Journal of Experimental Psychology, 102, 621-628.

PRICE R.L. and SMITH D.B.D. (1974) The P3(00) wave of the averaged evoked potential: A bibliography. Physiological Psychology, 2, 387-391.

RICHARDSON J.T.E. (1972) Nonparametric indexes of sensitivity and response bias. Psychological Bulletin, 78, 429-432.

RITTER W., SIMSON R. and VAUGHAN H.G., Jnr. (1972) Association cortex potentials and reaction time in auditory discrimination. Electroencephalography and Clinical Neurophysiology, 33, 547-555.

RITTER W. and VAUGHAN H.G., Jnr. (1969) Averaged evoked responses in vigilance and discrimination: A reassessment. Science, 164, 326-328.

RITTER W., VAUGHAN H.G., Jnr. and COSTA L.D. (1968) Orienting and habituation to auditory stimuli: A study of short term changes in average evoked responses. Electroencephalography and Clinical Neurophysiology, 25, 550-556.

ROHRBAUGH J.W., DONCHIN E. and ERIKSEN C.W. (1974) Decision making and the P300 component of the cortical evoked response. Perception and Psychophysics, 15, 368-374.

SAKITT B. (1973) Indices of discriminability. Nature, 241, 133-134.

SAYERS B. McA., BEAGLEY H.A. and HENSHALL W.R. (1974) The mechanism of auditory evoked responses. Nature, 247, 481-483.

SCHULMAN A.I. and GREENBERG G.Z. (1970) Operating characteristics and a priori probability of the signal. Perception and Psychophysics, 8, 317-320.

SCHULMAN A.I. and MITCHELL R.R. (1966) Operating characteristics from yes-no and forced-choice procedures. Journal of the Acoustical Society of America, 40, 473-477.

SHEEHAN J.J. and DRURY C.G. (1971) The analysis of industrial inspection. Applied Ergonomics, 2, 74-78.

SIDDLE D.A.T. (1972) Vigilance decrement and speed of habituation of the GSR component of the orienting response. British Journal of Psychology, 63, 191-194.

SIEGEL S. (1956) Nonparametric Methods for the Behavioral Sciences. New York: McGraw-Hill.

SIMPSON A.J. (1967) Signal detection and vigilance. (Unpublished Ph.D. thesis, University of Reading).

SIMPSON A.J. and FITTER M.J. (1973) What is the best measure of detectability? Psychological Bulletin, 80, 481-488.

SIPOS I. (1970) Vigilance performance in verbal signals. Studia Psychologica, 12, 29-33.

SMITH P.T. (1968) Cost, discriminability and response bias. British Journal of Mathematical and Statistical Psychology, 21, 35-60.

SMITH R.L. (1966) Monotony and Motivation: A Theory of Vigilance. Los Angeles: Dunlop and Associates, Inc.

SMITH R.L. and LUCACCINI L.F. (1969) Vigilance research: Its application to industrial problems. Human Factors, 11, 149-156.

SPERLING G. (1960) The information available in brief visual presentations. Psychological Monographs, 74 (Whole No. 498).

SQUIRES K.C., SQUIRES N.K. and HILLYARD S.A. (1975a) Vertex evoked potentials in a rating-scale detection task: Relation to signal probability. Behavioral Biology, 13, 21-34.

SQUIRES K.C., SQUIRES N.K. and HILLYARD S.A. (1975b) Decision-related cortical evoked potentials during an auditory signal detection task with cued observation intervals. Journal of Experimental Psychology: Human Perception and Performance, 1, 268-279.

STERN R.M. (1966) Performance and physiological arousal during two vigilance tasks varying in signal presentation rate. Perceptual and Motor Skills, 23, 691-700.

SUTTON S., BRAREN M., ZUBIN J. and JOHN E.R. (1965) Evoked potential correlates of stimulus uncertainty. Science, 150, 1187-1188.

SUTTON S. and TUETING P. (1975) The sensitivity of the evoked potential to psychological variables. In Venables P.H. and Christie M.J. (Eds.), Research in Psychophysiology. London: Wiley.

SVERKO B. (1968) Intermodal correlations in vigilance performance. Proceedings of the 16th International Congress of Applied Psychology, Amsterdam.

SWETS J.A. (1959) Indices of signal detectability obtained with various psychophysical procedures. Journal of the Acoustical Society of America, 31, 511-513.

SWETS J.A. (1972) Signal detection in medical diagnosis. In Jacquez J.A. (Ed.), Computer Diagnosis and Diagnostic Methods. Springfield, Illinois: Thomas.

SWETS J.A. (1973) The relative operating characteristic in psychology. Science, 182, 990-1000.

SWETS J.A. and KRISTOFFERSON A.B. (1970) Attention. Annual Review of Psychology, 21, 339-366.

SWETS J.A., TANNER W.P., Jnr. and BRIDSALL T.G. (1961) Decision processes in perception. Psychological Review, 68, 301-340.

TANNER W.P. and SWETS J.A. (1954) A decision-making theory of visual detection. Psychological Bulletin, 61, 401-409.

TAUB H.A. and OSBORNE H. (1968) Effects of signal and stimulus rates on vigilance performance. Journal of Applied Psychology, 52, 133-138.

TAYLOR M.M. (1965) Detectability measures in vigilance: Comment on a paper by Wiener, Poock and Steele. Perceptual and Motor Skills, 20, 1217-1221.

TAYLOR M.M. (1967) Detectability theory and the interpretation of vigilance data. Acta Psychologica, 27, 390-399.

TEICHNER W.H. (1972) Prediction of human performance. Proceedings of the Annual Reliability and Maintainability Symposium, San Francisco, California. (IEEE Catalog Number 72CHO 577-7R, pp. 420-430).

TEICHNER W.H. (1974) The detection of a simple visual signal as a function of time on watch. Human Factors, 16, 339-353.

TEICHNER W.H. and OLSEN D. (1969) Predicting performance in space environments. (NASA Report CR-1370, Washington DC).

TEICHNER W.H. and WHITEHEAD J. (1971) Development of a taxonomy of human performance: Evaluation of a task classification system for generalizing research findings from a data base. (American Institutes for Research TR-8, Washington DC).

THEOLOGUS G.C. and FLEISHMAN E.A. (1971) Development of a taxonomy of human performance: Validation study for classifying human tasks. (American Institutes for Research TR-10, Washington DC).

THEOLOGUS G.C., ROMASHKO T and FLEISHMAN E.A. (1970) Development of a taxonomy of human performance: A feasibility study of ability dimensions for classifying human tasks. (American Institutes for Research TR-5, Washington DC).

THOMAS E.A.C. (1973) On a class of additive learning models: Error correcting and probability matching. Journal of Mathematical Psychology, 10, 241-264.

THOMAS E.A.C. (1975) Criterion adjustment and probability matching. Perception and Psychophysics, 18, 158-162.

THOMAS E.A.C. and LEGGE D. (1970) Probability matching as a basis for detection and recognition decisions. Psychological Review, 77, 65-72.

THOMAS E.A.C. and MYERS J.L. (1972) Implications of latency data for threshold and nonthreshold models of signal detection. Journal of Mathematical Psychology, 9, 263-285.

THOMPSON R.F. and SPENCER W.A. (1966) Habituation: A model phenomenon for the study of neuronal substrates of behavior. Psychological Review, 73, 16-43.

TREISMAN A. and GEFFEN G. (1967) Selective attention: Perception or response? Quarterly Journal of Experimental Psychology, 19, 1-17.

TREISMAN M. (1972) Detection of binaural tone stimuli: Time sharing or criterion change? Journal of the Acoustical Society of America, 51, 625-630.

TYLER D.M. and HALCOMB C.G. (1974) Monitoring performance with a time shared encoding task. Perceptual and Motor Skills, 38, 383-386.

TYLER D.M., WAAG W.L. and HALCOMB C.G. (1972) Monitoring performance across sense modes: An individual differences approach. Human Factors, 14, 539-549.

UTTAL W.R. (1967) Evoked brain potentials: Signs or codes? Perspectives in Biology and Medicine, 10, 627-639.

VAUGHAN H.G., Jnr. and RITTER W. (1970) The sources of auditory evoked responses recorded from the human scalp. Electroencephalography and Clinical Neurophysiology, 28, 360-367.

VAUGHAN H.G., Jnr. and RITTER W. (1973) Physiologic approaches to the analysis of attention and performance. In Kornblum S. (Ed.), Attention and Performance IV. London: Academic Press.

VERSCHOOR A.M. and VAN WIERENGEN P.C. (1970) Vigilance performance and skin conductance. Acta Psychologica, 33, 394-401.

VICKERS D., NETTELBECK T. and WILLSON R.J. (1972) Perceptual indices of performance: The measurement of 'inspection time' and 'noise' in the visual system. Perception, 1, 263-295.

WAGENAAR W.A. (1969) Note on the construction of digram-balanced Latin squares. Psychological Bulletin, 72, 384-386.

WALLACK P.M. and ADAMS S.K. (1969) The utility of signal detection theory in the analysis of industrial inspection accuracy. American Institution of Industrial Engineers Transactions, 1, 33-44.

WARE J.R., SIPOWICZ R.R. and BAKER R.A. (1961) Auditory vigilance in repeated sessions. Perceptual and Motor Skills, 13, 127-129.

WATSON C.S., RILLING M.E. and BOURBON W.T. (1964) Receiver-operating characteristics determined by a mechanical analog to the rating scale. Journal of the Acoustical Society of America, 36, 283-288.

WEINTRAUB D.J. and HAKE H.W. (1962) Visual discrimination, an interpretation in terms of detectability theory. Journal of the Optical Society of America, 52, 1179-1184.

WEITZEL C. and DOBSON R. (1974) Sensitivity to visual Gaussian noise patterns as determined by binary and rating procedures. Perceptual and Motor Skills, 38, 147-154.

WELFORD A.T. (1972) Arousal, channel-capacity and decision . Nature, 194, 365-366.

WELFORD A.T. (1971) What is the basis of choice-reaction time? Ergonomics, 14, 679-693.

WHEATON G.R. (1968) Development of a taxonomy of human performance: A review of classificatory systems relating to tasks and performance. (American Institutes for Research TR-1, Washington DC).

WHEATON G.R., SHAFFER E.J., MIRABELLA A. and FLEISHMAN E.A. (1973) Methods for predicting job-ability requirements: 1. Ability requirements as a function of changes in the characteristics of a auditory signal identification task. (American Institutes for Research, Washington DC).

WHITTENBURG J.A., ROSS S. and ANDREWS T.G. (1956) Sustained perceptual efficiency as measured by the Mackworth 'Clock' Test. Perceptual and Motor Skills, 6, 109-116.

WICKELGREN W.A. (1968) Testing two-state theories with operating characteristics and a posteriori probabilities. Psychological Bulletin, 69, 126-131.

WIENER E.L. (1964) Multiple channel monitoring. Ergonomics, 7, 453-460.

WIENER E.L. (1973) Adaptive measurement of vigilance decrement. Ergonomics, 16, 353-363.

WIENER E.L. (1975a) On simultaneous monitoring and tracking. Journal of Applied Psychology, 60, 100-105.

WIENER E.L. (1975b) Individual and group differences in inspection. In Drury C.G. and Fox J.G. (Eds.), Human Reliability in Quality Control. London: Taylor and Francis.

WILDING J.M. (1970) The relation between decision-time and accuracy in the identification of visual stimuli. (Unpublished Ph.D. thesis, University of London).

WILKINSON R.T. (1961) Comparison of paced, unpaced, irregular and continuous displays in watchkeeping. Ergonomics, 4, 259-267.

WILKINSON R.T. (1968) Sleep deprivation: Performance tests for partial and selective sleep deprivation. In Abt, L.A. and Riess B.F. (Eds.), Progress in Clinical Psychology Vol. 7. New York: Grune.

WILKINSON R.T. and ASHBY S. (1974) Selective attention, contingent negative variation and the evoked potential. Biological Psychology, 1, 137-139.

WILKINSON R.T. and HAINES E. (1970) Evoked response correlates of expectancy during vigilance. Acta Psychologica, 33, 402-413.

WILKINSON R.T., MORLOCK H.C. and WILLIAMS H.L. (1966) Evoked cortical response during vigilance. Psychonomic Science, 4, 221-222.

WILLIGES R.C. (1969) Within session criterion changes compared to an ideal observer criterion in a visual monitoring task. Journal of Experimental Psychology, 81, 61-66.

WILLIGES R.C. (1971) The role of payoffs and signal ratios on criterion changes during a monitoring task. Human Factors, 13, 261-267.

WILLIGES R.C. (1973) Manipulating the response criterion in visual monitoring. Human Factors, 15, 179-185.

WILLIGES R.C. (1976) The vigilance increment: An ideal observer hypothesis. Paper presented at the Symposium on Monitoring Behavior and Supervisory Control, Berchtesgaden, W. Germany.

WILLIGES R.C. and NORTH R.A. (1972) Knowledge of results and decision making performance in visual monitoring. Organizational Behavior and Human Performance, 10, 44-57.

WINER B.J. (1962) Statistical Principles in Experimental Design. London: McGraw-Hill.

WYATT S. and LANGDON J.N. (1932) Inspection processes in industry. (Industrial Health Research Board Report 63, London: HMSO).

## APPENDIX A GLOSSARY OF SYMBOLS AND MAJOR ABBREVIATIONS

| | |
|---|---|
| AC | Auditory closure (task) |
| ANOVA | Analysis of variance |
| AS | Auditory speed (task) |
| B, $\beta$ | Likelihood ratio at criterion |
| $B_O$, $\beta_{Opt}$ | Optimum likelihood ratio on expected value model |
| C | (Flexibility of) Closure |
| CRL | Correct rejection latency |
| c | Criterion; Criterion cut-off value |
| cm | Centimetre |
| DL | Detection latency |
| d | Difference between the means of the signal and noise distributions |
| $d_a$ | Sensitivity index equivalent to d scaled by the root mean square of the signal and noise variances |
| d' | Sensitivity index equivalent to d and $d_a$ when the signal and noise variances are equal |
| dB | Decibel |
| $\Delta\sigma$ | Difference between signal and noise standard deviations |
| EEG | Electroencephalograph |
| EP | Evoked potential |
| ER | Event rate |
| FA | False alarm |
| FAL | False alarm latency |
| f(x/n) | Probability density function of the evidence (x) given noise (n) alone |
| f(x/s) | Probability density function of the evidence given signal plus noise(s) |
| H, Hit | Correct detection |
| Hz | Hertz (cycles per second) |

| | |
|---|---|
| MC | Multi-source closure (task) |
| ML | Omission latency |
| MS | Multi-source speed (task) |
| MSC | Multi-source speed-closure (task) |
| m | Metre |
| $\mu$V | Microvolt |
| N | Noise present (No) response |
| n | Noise |
| OC | Operating characteristic |
| P3, P300 | Late positive component of evoked potential |
| p | Probability |
| p(FA),p(S/n) | Probability of a false alarm |
| p(H),p(S/s) | Probability of a correct detection (hit) |
| S | Signal present (Yes) response; (Perceptual) Speed |
| SP | Signal probability |
| s | Signal |
| $\sigma n$ | Standard deviation of noise distribution |
| $\sigma s$ | Standard deviation of signal distribution |
| TSD | Theory of Signal Detectability |
| VC | Visual closure (task) |
| VS | Visual speed (task) |
| x | Evidence variable |
| y | Ordinate of probability density function |
| Z(FA),Z(S/n) | Normalized value of p(FA), p(S/n) |
| Z(H),Z(S/s) | Normalized value of p(H), P(S/s) |
| z | Normalized value of sample relative to population mean |

## APPENDIX B RELIABILITY OF PERFORMANCE ON VISUAL SPEED AND CLOSURE TASKS (EXPERIMENT 8)

### Outline

In order to provide a reference point for the results of the main experiments, especially Experiments 3 and 5, an experiment was carried out to examine the reliability of performance on two speed and closure monitoring tasks. Previous studies have generally indicated that the variability in performance on the same task is reliable between sessions separated by a few days or weeks (Buckner, Harabedian and McGrath, 1960), although there are very few studies which report reliability figures for decision theory metrics, or for different types of task.

An investigation of inter-session performance reliability also bears on aspects of the demand characteristics of different tasks and on the adequacy of the training procedures used. In all of the experiments reported in this thesis, a rather extensive training of subjects was undertaken so as to minimize learning effects during monitoring sessions. By comparing performance levels on repeated sessions, such effects, if any, may be evaluated.

Performance reliability was examined by testing subjects on three sessions on three separate days. The first two sessions were separated by a day (approximately 24 hours), and the third session was held at the same time of day a week after the first session. Thus estimates of day to day and week to week reliabilities were obtained.

The two tasks used in Experiment 1 were selected for reliability testing; a visual speed task (VS1) and a visual closure task (VC2). These tasks were also used in Experiment 3. (See 6.2.5 for task descriptions). All features of the tasks and the apparatus were as described previously, except that task duration was 30 minutes. The same experimental procedure as described in 6.3 was employed.

Twenty male subjects, aged 17 to 26 years, were employed. They were randomly assigned to two groups, each of which worked with either the speed or the closure task. Each subject served in three sessions. The subjects serving in the present study had not participated in experiments on detection or monitoring behaviour before. However, the same training regime was used for these subjects as in the previous experiments. Further characteristics of the subjects are listed in Table 6.2.

Results and Discussion

Mean performance levels were computed for each half (15-min. time block) of the monitoring session, for the measures p(H), P(FA), d' and B. These are displayed in Table B.1 for performance on either task. The usual trends over time blocks are apparent, and there is some indication of an improvement in performance over sessions. Each performance metric was analysed in a 2 x 3 x 2 analysis of variance (Tasks x Sessions x Time Blocks), with repeated measures on the latter two factors. The summary tables for these analyses are shown in Tables B.2 to B.5.

Inspection of these tables reveals that there was no reliable change in performance over sessions. A slight improvement in mean performance is apparent from Table B.1, especially between Session 1 and Session 2, but this was not statistically significant. The only significant effect obtained was for the Time Blocks factor; the proportion of Hits and False Alarms declined significantly with Time Blocks, while B increased. Furthermore, there was a significant triple interaction between Tasks, Sessions and Time Blocks for the p(FA) measure. A firm interpretation

| Performance Measure | Speed Task | | | | | | Closure Task | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SS1 | | SS2 | | SS3 | | SS1 | | SS2 | | SS3 | |
| | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| p(H) | .78 | .75 | .82 | .77 | .81 | .71 | .80 | .71 | .76 | .72 | .81 | .77 |
| p(FA) | .050 | .041 | .047 | .039 | .052 | .033 | .049 | .025 | .039 | .029 | .048 | .042 |
| d' | 2.52 | 2.52 | 2.73 | 2.62 | 2.63 | 2.58 | 2.57 | 2.58 | 2.58 | 2.56 | 2.63 | 2.57 |
| B | 3.5 | 4.5 | 4.0 | 4.6 | 3.4 | 8.0 | 3.5 | 6.7 | 4.9 | 7.5 | 3.0 | 4.3 |

TABLE B.1 Mean performance levels in successive time blocks (1, 2) and sessions (SS) for the speed and closure tasks.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.00913 | 1 | 0.00913 | 0.0303 | ns |
| SWG | 5.35626 | 18 | 0.29757 | | |
| B (Sessions) | 0.13897 | 2 | 0.06948 | 2.8698 | ns (.10) |
| AB | 0.13090 | 2 | 0.06545 | 2.7032 | ns (.10) |
| B x SWG | 0.87163 | 36 | 0.02421 | | |
| C (Time Blocks) | 0.04641 | 1 | 0.04641 | 4.3048 | ns (.10) |
| AC | 0.00675 | 1 | 0.00675 | 0.6261 | ns |
| C x SWG | 0.19407 | 18 | 0.01078 | | |
| BC | 0.03401 | 2 | 0.01701 | 1.0717 | ns |
| ABC | 0.01689 | 2 | 0.00845 | 0.5323 | ns |
| BC x SWG | 0.57126 | 36 | 0.01587 | | |

TABLE B.2 Analysis of variance for d'.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.8333 | 1 | 0.8333 | 0.0698 | ns |
| SWG | 214.8000 | 18 | 11.9333 | | |
| B (Sessions) | 0.6167 | 2 | 0.3083 | 0.1745 | ns |
| AB | 9.1167 | 2 | 4.5583 | 2.5802 | ns (.10) |
| B x SWG | 63.6000 | 36 | 1.7667 | | |
| C (Time Blocks) | 22.5333 | 1 | 22.5333 | 38.7516 | .001 |
| AC | 0.0000 | 1 | 0.0000 | 0.0000 | ns |
| C x SWG | 10.4667 | 18 | 0.5815 | | |
| BC | 1.0167 | 2 | 0.5083 | 1.0558 | ns |
| ABC | 3.6500 | 2 | 1.8250 | 3.7904 | ns (.10) |
| BC x SWG | 17.3333 | 36 | 0.4815 | | |

TABLE B.3 Analysis of variance for Hits.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.03104 | 1 | 0.03104 | 0.1348 | ns |
| SWG | 4.14447 | 18 | 0.23025 | | |
| B (Sessions) | 0.07932 | 2 | 0.03966 | 0.8423 | ns |
| AB | 0.26582 | 2 | 0.13291 | 2.8228 | ns (.10) |
| B x SWG | 1.69506 | 36 | 0.04709 | | |
| C (Time Blocks) | 0.76321 | 1 | 0.76321 | 34.9524 | .001 |
| AC | 0.00007 | 1 | 0.00007 | 0.0031 | ns |
| C x SWG | 0.39304 | 18 | 0.02184 | | |
| BC | 0.06185 | 2 | 0.03092 | 2.0058 | ns |
| ABC | 0.18509 | 2 | 0.09254 | 6.0027 | .05 |
| BC x SWG | 0.55500 | 36 | 0.01542 | | |

TABLE B.4 Analysis of variance for False Alarms (log transformed).

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.09633 | 1 | 0.09633 | 0.4036 | ns |
| SWG | 4.29588 | 18 | 0.23866 | | |
| B (Sessions) | 0.03493 | 2 | 0.01746 | 0.3614 | ns |
| AB | 0.19763 | 2 | 0.09881 | 2.0449 | ns |
| B x SWG | 1.73958 | 36 | 0.04832 | | |
| C (Time Blocks) | 0.64533 | 1 | 0.64533 | 37.1435 | .001 |
| AC | 0.01633 | 1 | 0.01633 | 0.9401 | ns |
| C x SWG | 0.31273 | 18 | 0.01737 | | |
| BC | 0.01803 | 2 | 0.00901 | 0.5265 | ns |
| ABC | 0.09133 | 2 | 0.04566 | 2.6676 | ns (.10) |
| BC x SWG | 0.61625 | 36 | 0.01712 | | |

TABLE B.5 Analysis of variance of log (B+1).

cannot be given to this interaction, but it appears to be related to the differential decrement in False Alarms in the speed and the closure tasks.

Table B.6 lists the day-to-day ($R_{dd}$) and week-to-week ($R_{ww}$) reliability figures for each task, using the mean values of the measures of Hits, False Alarms, d' and B. For the speed task, both $R_{dd}$ and $R_{ww}$ are high and significantly greater than zero, for all four performance measures. These coefficients compare favourably with the correlation coefficients obtained in Experiment 3 for performance consistencies across similarly classified tasks. On average, the task reliabilities obtained here are higher than the coefficients for intramodal performances in Experiment 3. However, as Table B.6 indicates, the day-to-day reliabilities for performance on the closure task are substantially lower than the other coefficients, with three out of the four correlations being non-significant. Examination of the raw data revealed that this was due to the extreme variability in performance of one subject. In Session 1, this subject gave the best performance (in terms of any performance metric), but was the poorest monitor in Session 2, apparently adopting a much stricter criterion. In Session 3, this subject again had the best overall performance. Post-experimental interview did not reveal any reasons for the poor performance in Session 2. This result does not greatly detract from the general finding of high task reliabilities, since if the performance data of the idiosyncratic subject were removed from consideration, consistently high reliabilities are obtained for each performance measure (.97, 61, 68, and .83 for Hits, false alarms, d' and B, respectively).

Performance reliabilities for corresponding halves of each monitoring session were also high, as Table B.7 shows, although $R_{dd}$ values for the closure task were again low. All but three of the other values were statistically significant.

| | SPEED TASK | | CLOSURE TASK | |
|---|---|---|---|---|
| | $R_{dd}$ | $R_{ww}$ | $R_{dd}$ | $R_{ww}$ |
| Hits | .96 | .77 | .43+ | .88 |
| False Alarms | .82 | .84 | .17+ | .69 |
| d' | .88 | .77 | .59 | .98 |
| B | .83 | .69 | .35+ | .87 |

TABLE B.6 Day-to-day ($R_{dd}$) and week-to-week ($R_{ww}$) performance reliabilities (Spearman rank correlation coefficients) for speed and closure tasks (+ not significant at .05 level).

| | SPEED TASK | | | | CLOSURE TASK | | | |
|---|---|---|---|---|---|---|---|---|
| | $R_{dd}$ | | $R_{ww}$ | | $R_{dd}$ | | $R_{ww}$ | |
| | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| Hits | .86 | .69 | .92 | .75 | .31+ | .37+ | .80 | .81 |
| False Alarms | .93 | .49+ | .97 | .40+ | .26+ | .23+ | .55 | .67 |
| d' | .71 | .77 | .68 | .76 | .55 | .53+ | .88 | .88 |
| B | .85 | .77 | .96 | .58 | .19+ | .56 | .56 | .83 |

TABLE B.7 Performance reliabilities in corresponding halves (1 and 2) of each monitoring session (+ not significant at .05 level).

| | EXPERIMENTS | | | | | |
|---|---|---|---|---|---|---|
| Tasks | P | 1 | 3 | 4* | 5 | M |
| VS1 | | 2.48 | 2.53,2.52 | | 2.60,2.66 | 2.56 |
| VS2 | 2.50 | | 2.54 | | | 2.52 |
| VC1 | 2.53 | | 2.47,2.48 | | 2.57 | 2.51 |
| VC2 | | 2.51 | 2.46 | | | 2.49 |
| AS | | | | 2.56 | 2.61 | 2.59 |
| AC | | | | 2.44 | 2.59,2.56 | 2.53 |

TABLE B.8 Mean d' values for visual and auditory speed and closure tasks in Experiments 1 to 5 (P = Pilot study to Experiment 1; see 6.2.5 for task descriptions; *values of da).

These results therefore demonstrate that performance reliabilities for sessions separated either by a day or a week are consistently high for both speed and closure tasks. The results are therefore in agreement with previous studies (Buckner et al., 1960), but show additionally that sensitivity d', and criterion B were also reliably correlated across sessions. There was also no evidence of practice effects in the present study. Practice or learning effects are occasionally noted in monitoring tasks, as in Experiment 2 (see also Mackworth, 1970), but the extensive training and expectancy matching procedures employed probably contributed to their absence in the present study.

Finally, an indication of performance reliability in the various tasks employed can be gained by examining the mean performance levels obtained in different experiments for the same task. This is possible since Experiments 1 to 5 shared at least one condition in which the task and all relevant independent variable values were the same. As Table B.8 shows, the mean d' values are fairly stable within tasks, the variability between experiments being within the limits ($\pm$ 10%) to be expected due to random experimental error.

# A P P E N D I X C

## SUMMARY TABLES FOR ANALYSES OF VARIANCE

Note

Summary tables for Experiments 1 to 5 are given in Sub-appendices C1 to C5, respectively.

ANOVA = Analysis of variance

SWG = Subjects within groups

SP = Signal probability

ER = Event rate

SS = Sum of squares

MS = Mean square

DF = Degrees of freedom

ns = not significant

Appendix C1 : Experiment 1

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.0007 | 1 | 0.0007 | 0.0027 | ns |
| B (Event Rate) | 0.2683 | 2 | 0.1342 | 0.5093 | ns |
| AB | 0.0410 | 2 | 0.0205 | 0.0778 | ns |
| SWG | 14.2229 | 54 | 0.2634 | | |
| C (Time Blocks) | 4.1153 | 2 | 2.0577 | 21.9711 | .001 |
| AC | 0.0285 | 2 | 0.0143 | 0.1527 | ns |
| BC | 0.0128 | 4 | 0.0032 | 0.0342 | ns |
| ABC | 0.2890 | 4 | 0.0723 | 0.7720 | ns |
| C x SWG | 10.1146 | 108 | 0.0937 | | |

TABLE C1.1 ER ANOVA for False Alarms (log transformed) in Conditions I, II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.1051 | 1 | 0.1051 | 0.9800 | ns |
| B (Signal Prob.) | 4.2642 | 2 | 2.1321 | 19.8800 | .001 |
| AB | 0.3444 | 2 | 0.1722 | 1.6057 | ns |
| SWG | 5.7910 | 54 | 0.1072 | | |
| C (Time Blocks) | 2.2642 | 2 | 1.1321 | 28.9478 | .001 |
| AC | 0.0290 | 2 | 0.0145 | 0.3708 | ns |
| BC | 0.0040 | 4 | 0.0010 | 0.0256 | ns |
| ABC | 0.0835 | 4 | 0.0209 | 0.5344 | ns |
| C x SWG | 4.2237 | 108 | 0.0392 | | |

TABLE C1.2 SP ANOVA for False Alarms (log transformed) in Conditions II, IV and V.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.0108 | 1 | 0.0108 | 0.0451 | ns |
| B (Signal Prob.) | 0.2448 | 1 | 0.2448 | 1.0197 | ns |
| AB | 0.0028 | 1 | 0.0028 | 0.0117 | ns |
| SWG | 8.6429 | 36 | 0.2401 | | |
| C (Time Blocks) | 0.1953 | 2 | 0.0976 | 0.7867 | ns |
| AC | 0.0467 | 2 | 0.0233 | 0.1881 | ns |
| BC | 0.0078 | 2 | 0.0039 | 0.0313 | ns |
| ABC | 0.0190 | 2 | 0.0095 | 0.0764 | ns |
| C x SWG | 8.9356 | 72 | 0.1241 | | |

TABLE C1.3 SP ANOVA for d' in Conditions 11 and IV.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.0227 | 1 | 0.0227 | 0.1152 | ns |
| B (Event Rate) | 1.5976 | 2 | 0.7988 | 3.6841 | .05 |
| AB | 0.0840 | 2 | 0.0420 | 0.1937 | ns |
| SWG | 11.7085 | 54 | 0.2168 | | |
| | | | | | |
| C (Time Blocks) | 3.3955 | 2 | 1.6978 | 19.2681 | .001 |
| AC | 0.0002 | 2 | 0.0001 | 0.0011 | ns |
| BC | 0.0683 | 4 | 0.0171 | 0.1941 | ns |
| ABC | 0.0644 | 4 | 0.0161 | 0.1827 | ns |
| C x SWG | 9.5163 | 108 | 0.0881 | | |

TABLE C1.4 ER ANOVA for log (B+1) in Conditions I, II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 3392.03 | 1 | 3392.03 | 0.3268 | ns |
| B (Signal Prob.) | 52668.30 | 1 | 52668.30 | 5.0740 | .05 |
| AB | 8636.03 | 1 | 8636.03 | 0.8320 | ns |
| SWG | 373682.93 | 36 | 10380.08 | | |
| | | | | | |
| C (Time Blocks) | 121661.15 | 2 | 60830.58 | 33.4872 | .001 |
| AC | 1129.22 | 2 | 564.61 | 0.3108 | ns |
| BC | 1028.15 | 2 | 514.08 | 0.2830 | ns |
| ABC | 151.22 | 2 | 75.61 | 0.0416 | ns |
| C x SWG | 130790.27 | 72 | 1816.53 | | |

TABLE C1.6 SP ANOVA for Detection Latency in Conditions II and IV.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 68640.83 | 1 | 68640.83 | 5.5074 | .05 |
| B (Event Rate) | 60840.03 | 1 | 60840.03 | 4.8815 | .05 |
| AB | 51667.5 | 1 | 51667.50 | 4.1455 | .05 |
| SWG | 448685.33 | 36 | | | |
| | | | | | |
| C (Time Blocks) | 121925.07 | 2 | 60962.53 | 45.0825 | .001 |
| AC | 97.87 | 2 | 48.93 | 0.0362 | ns |
| BC | 1787.47 | 2 | 893.73 | 0.6609 | ns |
| ABC | 744.80 | 2 | 372.40 | 0.2754 | ns |
| C x SWG | 97361.47 | 72 | 1352.24 | | |
| **Simple effects** | | | | | |
| B at A1 (Speed Task) | 112321.00 | 1 | 112321.00 | 18.0390 | .001 |
| B at A2 (Closure Task) | 186.50 | 1 | 186.50 | 0.0150 | ns |

TABLE C1.5 ER ANOVA for Detection Latency in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 3785.63 | 1 | 3785.63 | 0.1932 | ns |
| B (Event Rate) | 34.13 | 1 | 34.13 | 0.0017 | ns |
| AB | 1080.00 | 1 | 1080.00 | 0.0551 | ns |
| SWG | 705309.40 | 36 | 19591.93 | | |
| C (Time Blocks) | 18442.82 | 2 | 9221.41 | 4.0028 | .05 |
| AC | 355.72 | 2 | 177.86 | 0.0772 | ns |
| BC | 2298.62 | 2 | 1149.31 | 0.4989 | ns |
| ABC | 948.05 | 2 | 474.03 | 0.2058 | ns |
| C x SWG | 165870.72 | 72 | 2303.76 | | |

TABLE C1.7 ER ANOVA for False Alarm Latency in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 31557.63 | 1 | 31557.63 | 1.6952 | ns |
| B (Signal Prob.) | 139400.83 | 1 | 139400.83 | 7.4882 | .01 |
| AB | 6931.20 | 1 | 6931.20 | 0.3723 | ns |
| SWG | 670177.53 | 36 | 18616.04 | | |
| C (Time Blocks) | 33859.40 | 2 | 16929.70 | 8.6355 | .01 |
| AC | 16.07 | 2 | 8.03 | 0.0041 | ns |
| BC | 0.87 | 2 | 0.43 | 0.0002 | ns |
| ABC | 810.60 | 2 | 405.30 | 0.2067 | ns |
| C x SWG | 141155.07 | 72 | 1960.49 | | |

TABLE C1.8 SP ANOVA for False Alarm Latency in Conditions II and IV.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 10982.53 | 1 | 10982.53 | 0.8959 | ns |
| B (Event Rate) | 6020.85 | 1 | 6020.85 | 0.4912 | ns |
| AB | 3286.53 | 1 | 3286.53 | 0.2681 | ns |
| SWG | 441305.13 | 36 | 12258.48 | | |
| C (Time Blocks) | 23329.40 | 2 | 13664.70 | 16.7647 | .001 |
| AC | 468.87 | 2 | 234.43 | 0.2876 | ns |
| BC | 1056.07 | 2 | 528.03 | 0.6478 | ns |
| ABC | 1236.07 | 2 | 618.03 | 0.7582 | ns |
| C x SWG | 58686.27 | 72 | 815.08 | | |

TABLE C1.9 ER ANOVA for Correct Rejection Latency in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 17812.03 | 1 | 17812.03 | 1.1301 | ns |
| B (Signal Prob.) | 19969.20 | 1 | 19969.20 | 1.2669 | ns |
| AB | 821.63 | 1 | 821.63 | 0.0521 | ns |
| SWG | 367424.60 | 36 | 15761.79 | | |
| C (Time Blocks) | 31182.92 | 2 | 15591.46 | 19.8221 | .001 |
| AC | 2773.52 | 2 | 1366.76 | 1.7376 | ns |
| BC | 16.25 | 2 | 8.13 | 0.0103 | ns |
| ABC | 18.32 | 2 | 9.16 | 0.0116 | ns |
| C x SWG | 56633.00 | 72 | 766.57 | | |

TABLE C1.10 SP ANOVA for Correct Rejection Latency in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 89216.53 | 1 | 89216.53 | 5.7611 | .05 |
| B (Event Rate) | 82163.33 | 1 | 82163.33 | 5.3056 | .05 |
| AB | 72717.63 | 1 | 72717.63 | 4.6957 | .05 |
| SWG | 557500.20 | 36 | 15486.12 | | |
| C (Time Blocks) | 36666.05 | 2 | 18333.03 | 18.3537 | .001 |
| AC | 3514.12 | 2 | 1757.06 | 1.7590 | ns |
| BC | 3859.12 | 2 | 1929.56 | 1.9317 | ns |
| ABC | 3601.72 | 2 | 1800.86 | 1.8029 | ns |
| C x SWG | 71919.00 | 72 | 998.88 | | |

Simple effects

| | | | | | |
|---|---|---|---|---|---|
| B at A1 (Speed Task) | 154737.0 | 1 | 152737.0 | 9.9920 | .01 |
| B at A2 (Closure Task) | 144.0 | 1 | 144.0 | 0.0093 | ns |

TABLE C1.11 ER ANOVA for Omission Latency in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 3080.53 | 1 | 3080.53 | 0.3083 | ns |
| B (Signal Prob.) | 4737.63 | 1 | 4737.63 | 0.4741 | ns |
| AB | 700.83 | 1 | 700.83 | 0.0701 | ns |
| SWG | 359760.47 | 36 | 9993.35 | | |
| C (Time Blocks) | 28635.22 | 2 | 14317.61 | 17.6047 | .001 |
| AC | 1988.32 | 2 | 994.16 | 1.2224 | ns |
| BC | 1886.62 | 2 | 943.31 | 1.1599 | ns |
| ABC | 1691.52 | 2 | 845.75 | 1.0399 | ns |
| C x SWG | 58556.33 | 72 | 813.28 | | |

TABLE C1.12 SP ANOVA for Omission Latency in Conditions II and IV.

Appendix C2 : Experiment 2

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.02567 | 1 | 0.02567 | 0.2247 | ns |
| SWG | 0.68545 | 6 | 0.11424 | | |
| B (Sessions) | 0.13125 | 1 | 0.13125 | 4.8138 | ns (.10) |
| AB | 0.01367 | 1 | 0.01367 | 0.5013 | ns |
| B x SWG | 0.16360 | 6 | 0.02727 | | |
| C (Time Blocks) | 0.20325 | 2 | 0.10163 | 9.2394 | .05 |
| AC | 0.04549 | 2 | 0.02274 | 2.0677 | ns |
| C x SWG | 0.13199 | 12 | 0.01100 | | |
| BC | 0.00040 | 2 | 0.00020 | 0.0186 | ns |
| ABC | 0.03159 | 2 | 0.01579 | 1.4518 | ns |
| BC x SWG | 0.13054 | 12 | 0.01088 | | |

TABLE C2.1 ANOVA for log B1.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.03000 | 1 | 0.03000 | 0.2333 | ns |
| SWG | 0.77163 | 6 | 0.12861 | | |
| B (Sessions) | 0.13081 | 1 | 0.13081 | 10.1351 | .05 |
| AB | 0.00378 | 1 | 0.00368 | 0.2861 | ns |
| B x SWG | 0.07708 | 6 | 0.01285 | | |
| C (Time Blocks) | 0.34287 | 2 | 0.17143 | 12.3395 | .05 |
| AC | 0.03815 | 2 | 0.01908 | 1.3730 | ns |
| C x SWG | 0.16672 | 12 | 0.01389 | | |
| BC | 0.02062 | 2 | 0.01031 | 1.3873 | ns |
| ABC | 0.01995 | 2 | 0.00998 | 1.3424 | ns |
| BC x SWG | 0.08917 | 12 | 0.00743 | | |

TABLE C2.2 ANOVA for log B2

| SOURCE | SS | DF | MS | F | p |
| --- | --- | --- | --- | --- | --- |
| A (Tasks) | 0.19508 | 1 | 0.19508 | 1.3890 | ns |
| SWG | 0.84266 | 6 | 0.14044 | | |
| B (Sessions) | 0.04563 | 1 | 0.04563 | 3.9382 | ns (.10) |
| AB | 0.05201 | 1 | 0.05201 | 4.8883 | ns (.10) |
| B x SWG | 0.06953 | 6 | 0.01159 | | |
| C (Time Blocks) | 0.26608 | 2 | 0.13304 | 8.7928 | .05 |
| AC | 0.02389 | 2 | 0.01194 | 0.7894 | ns |
| C x SWG | 0.18157 | 12 | 0.01513 | | |
| BC | 0.00283 | 2 | 0.00142 | 0.2272 | ns |
| ABC | 0.03140 | 2 | 0.01570 | 2.5224 | ns |
| BC x SWG | 0.07470 | 12 | 0.00623 | | |

TABLE C2.3 ANOVA for log B3.

| SOURCE | SS | DF | MS | F | p |
| --- | --- | --- | --- | --- | --- |
| A (Tasks) | 0.01110 | 1 | 0.01110 | 0.1518 | ns |
| SWG | 0.43886 | 6 | 0.07314 | | |
| B (Sessions) | 0.03255 | 1 | 0.03255 | 0.4095 | ns |
| AB | 0.00350 | 1 | 0.00350 | 0.0441 | ns |
| B x SWG | 0.47690 | 6 | 0.07948 | | |
| C (Time Blocks) | 0.11220 | 2 | 0.05610 | 1.1484 | ns |
| AC | 0.05408 | 2 | 0.02704 | 0.5535 | ns |
| C x SWG | 0.58625 | 12 | 0.04885 | | |
| BC | 0.03168 | 2 | 0.01584 | 0.2484 | ns |
| ABC | 0.01180 | 2 | 0.00590 | 0.0926 | ns |
| BC x SWG | 0.76512 | 12 | 0.06376 | | |

TABLE C2.4 ANOVA for $\sigma_s/\sigma_n$.

Appendix C3 : Experiment 3

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.6000 | 1 | 0.6000 | 0.1200 | ns |
| SWG | 40.0000 | 8 | 5.0000 | | |
| B (Tasks) | 0.600 | 1 | 0.6000 | 1.4400 | ns |
| AB | 0.0667 | 1 | 0.0667 | 0.1600 | ns |
| B x SWG | 3.3333 | 8 | 0.4167 | | |
| C (Time Blocks) | 30.6333 | 2 | 15.3167 | 15.3167 | .01 |
| AC | 0.7000 | 2 | 0.3500 | 0.3500 | ns |
| C x SWG | 16.0000 | 16 | 1.0000 | | |
| BC | 2.1000 | 2 | 1.0500 | 3.6000 | ns |
| ABC | 1.2333 | 2 | 0.6167 | 2.1143 | ns |
| C x SWG | 4.6667 | 16 | 0.2917 | | |

TABLE C3.1 ANOVA for Hits in Group 1.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.00193 | 1 | 0.00193 | 0.0212 | ns |
| SWG | 0.72537 | 8 | 0.09067 | | |
| B (Tasks) | 0.00006 | 1 | 0.00006 | 0.0073 | ns |
| AB | 0.02017 | 1 | 0.02017 | 2.4454 | ns |
| B x SWG | 0.06597 | 8 | 0.00825 | | |
| C (Time Blocks) | 1.60309 | 2 | 0.80155 | 32.4720 | .001 |
| AC | 0.00826 | 2 | 0.00423 | 0.1674 | ns |
| C x SWG | 0.39495 | 16 | 0.02468 | | |
| BC | 0.03367 | 2 | 0.01684 | 1.7506 | ns |
| ABC | 0.02956 | 2 | 0.01478 | 1.5371 | ns |
| BC x SWG | 0.15387 | 16 | 0.00962 | | |

TABLE C3.2 ANOVA for False Alarms (log transformed) in Group 1.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.06017 | 1 | 0.06017 | 1.5973 | ns |
| SWG | 0.30133 | 8 | 0.03767 | | |
| B (Tasks) | 0.00150 | 1 | 0.00150 | 0.1071 | ns |
| AB | 0.00150 | 1 | 0.00150 | 0.1071 | ns |
| B x SWG | 0.11200 | 8 | 0.01400 | | |
| C (Time Blocks) | 0.07600 | 2 | 0.03800 | 0.9175 | ns |
| AC | 0.10133 | 2 | 0.05067 | 1.2233 | ns |
| C x SWG | 0.66267 | 16 | 0.04147 | | |
| BC | 0.11200 | 2 | 0.05600 | 3.2000 | ns |
| ABC | 0.02800 | 2 | 0.01400 | 0.8000 | ns |
| BC x SWG | 0.28000 | 16 | 0.01750 | | |

TABLE C3.3 ANOVA for d' in Group 1.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.00032 | 1 | 0.00032 | 0.0019 | ns |
| SWG | 1.37484 | 8 | 0.17186 | | |
| B (Tasks) | 0.00641 | 1 | 0.00641 | 0.7113 | ns |
| AB | 0.01291 | 1 | 0.01291 | 1.4330 | ns |
| B x SWG | 0.07205 | 8 | 0.00901 | | |
| C (Time Blocks) | 1.96892 | 2 | 0.98446 | 45.8769 | .001 |
| AC | 0.00520 | 2 | 0.00260 | 0.1212 | ns |
| C x SWG | 0.34334 | 16 | 0.02146 | | |
| BC | 0.01130 | 2 | 0.00565 | 0.8088 | ns |
| ABC | 0.01002 | 2 | 0.00501 | 0.7122 | ns |
| BC x SWG | 0.11181 | 16 | 0.00699 | | |

TABLE C3.4 ANOVA for log B in Group 1.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 4001.67 | 1 | 4001.67 | 0.3038 | ns |
| SWG | 105382.40 | 8 | 13172.80 | | |
| B (Tasks) | 147.27 | 1 | 147.27 | 0.0548 | ns |
| AB | 8.07 | 1 | 8.07 | 0.0030 | ns |
| B x SWG | 21482.00 | 8 | 2685.25 | | |
| C (Time Blocks) | 51840.83 | 2 | 25920.42 | 26.6500 | .001 |
| AC | 935.83 | 2 | 467.92 | 0.4811 | ns |
| C x SWG | 15562.00 | 16 | 972.63 | | |
| BC | 250.23 | 2 | 125.12 | 0.1048 | ns |
| ABC | 3297.63 | 2 | 1648.82 | 1.3817 | ns |
| BC x SWG | 19092.80 | 16 | 1193.30 | | |

TABLE C3.5 ANOVA for Detection Latency in Group 1.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.0667 | 1 | 0.0667 | 0.0119 | ns |
| SWG | 45.0000 | 8 | 5.6250 | | |
| B (Tasks) | 0.2667 | 1 | 0.2667 | 0.2038 | ns |
| AB | 0.2667 | 1 | 0.2667 | 0.2038 | ns |
| B x SWG | 10.4667 | 8 | 1.3083 | | |
| C (Time Blocks) | 31.3000 | 2 | 15.6500 | 15.6500 | .001 |
| AC | 2.0333 | 2 | 1.0167 | 1.0167 | ns |
| C x SWG | 16.0000 | 16 | 1.0000 | | |
| BC | 1.0333 | 2 | 0.5167 | 1.0164 | ns |
| ABC | 0.8333 | 2 | 0.4167 | 0.8197 | ns |
| BC x SWG | 8.1333 | 16 | 0.5083 | | |

TABLE C3.6 ANOVA for Hits in Group 2.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.01411 | 1 | 0.01411 | 0.0822 | ns |
| SWG | 1.37342 | 8 | 0.17168 | | |
| B (Tasks) | 0.00241 | 1 | 0.00241 | 0.0843 | ns |
| AB | 0.00600 | 1 | 0.00600 | 0.2102 | ns |
| B x SWG | 0.22833 | 8 | 0.02854 | | |
| C (Time Blocks) | 0.89834 | 2 | 0.44917 | 10.3273 | .05 |
| AC | 0.00842 | 2 | 0.00421 | 0.0968 | ns |
| C x SWG | 0.69590 | 16 | 0.04349 | | |
| BC | 0.00690 | 2 | 0.00345 | 0.3193 | ns |
| ABC | 0.03141 | 2 | 0.01571 | 0.6337 | ns |
| BC x SWG | 0.36955 | 16 | 0.02479 | | |

TABLE C3.7 ANOVA for False Alarms (log transformed) in Group 2.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.01667 | 1 | 0.01667 | 0.1177 | ns |
| SWG | 1.13267 | 8 | 0.14158 | | |
| B (Tasks) | 0.00000 | 1 | 0.00000 | 0.0000 | ns |
| AB | 0.04267 | 1 | 0.04267 | 2.9942 | ns |
| B x SWG | 0.11400 | 8 | 0.01425 | | |
| C (Time Blocks) | 0.04433 | 2 | 0.02217 | 0.1623 | ns |
| AC | 0.03033 | 2 | 0.01517 | 0.1110 | ns |
| C x SWG | 2.18533 | 16 | 0.13658 | | |
| BC | 0.03900 | 2 | 0.01950 | 0.7222 | ns |
| ABC | 0.00233 | 2 | 0.00117 | 0.0432 | ns |
| BC x SWG | 0.43200 | 16 | 0.02700 | | |

TABLE C3.8 ANOVA for d' in Group 2.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.01204 | 1 | 0.01204 | 0.0473 | ns |
| SWG | 2.03720 | 8 | 0.25465 | | |
| B (Tasks) | 0.00000 | 1 | 0.00000 | 0.0000 | ns |
| AB | 0.00028 | 1 | 0.00028 | 0.0049 | ns |
| B x SWG | 0.45927 | 8 | 0.05741 | | |
| C (Time Blocks) | 1.29016 | 2 | 0.64508 | 16.9102 | .01 |
| AC | 0.02084 | 2 | 0.01042 | 0.2732 | ns |
| C x SWG | 0.61036 | 16 | 0.03815 | | |
| BC | 0.00604 | 2 | 0.00302 | 0.0928 | ns |
| ABC | 0.03852 | 2 | 0.01926 | 0.5918 | ns |
| BC x SWG | 0.52073 | 16 | 0.03255 | | |

TABLE C3.9 ANOVA for log B in Group 2.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 236.02 | 1 | 236.02 | 0.0197 | ns |
| SWG | 95656.33 | 8 | 11957.04 | | |
| B (Tasks) | 18903.75 | 1 | 18903.75 | 6.6223 | .05 |
| AB | 4.82 | 1 | 4.82 | 0.0017 | ns |
| B x SWG | 22836.60 | 8 | 2854.58 | | |
| C (Time Blocks) | 36099.70 | 2 | 18049.85 | 31.3377 | .001 |
| AC | 481.63 | 2 | 240.82 | 0.4181 | ns |
| C x SWG | 9215.67 | 16 | 575.98 | | |
| BC | 745.90 | 2 | 372.95 | 0.5256 | ns |
| ABC | 991.03 | 2 | 495.52 | 0.6983 | ns |
| BC x SWG | 11353.40 | 16 | 709.59 | | |

TABLE C3.10 ANOVA for Detection Latency in Group 2.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.1500 | 1 | 0.1500 | 0.0291 | ns |
| SWG | 41.2000 | 8 | 5.1500 | | |
| B (Tasks) | 1.3500 | 1 | 1.3500 | 0.4880 | ns |
| AB | 0.0167 | 1 | 0.0167 | 0.0060 | ns |
| B x SWG | 22.1333 | 8 | 2.7667 | | |
| C (Time Blocks) | 9.4333 | 2 | 4.7167 | 4.1665 | .10 |
| AC | 2.7000 | 2 | 1.3500 | 1.1868 | ns |
| C x SWG | 18.2000 | 16 | 1.1375 | | |
| BC | 3.7000 | 2 | 1.8500 | 1.9389 | ns |
| ABC | 0.0333 | 2 | 0.0167 | 0.0175 | ns |
| BC x SWG | 15.2667 | 16 | 0.9542 | | |

TABLE C3.11 ANOVA for Hits in Group 3.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.00028 | 1 | 0.00028 | 0.0095 | ns |
| SWG | 0.23832 | 8 | 0.02979 | | |
| B (Tasks) | 0.00002 | 1 | 0.00002 | 0.0009 | ns |
| AB | 0.00140 | 1 | 0.00140 | 0.0804 | ns |
| B x SWG | 0.13947 | 8 | 0.01743 | | |
| C (Time Blocks) | 0.25470 | 2 | 0.12735 | 15.2929 | .01 |
| AC | 0.01512 | 2 | 0.00655 | 0.9080 | ns |
| C x SWG | 0.13324 | 16 | 0.00833 | | |
| BC | 0.01227 | 2 | 0.00614 | 0.6378 | ns |
| ABC | 0.01310 | 2 | 0.00655 | 0.6812 | ns |
| BC x SWG | 0.15389 | 16 | 0.00962 | | |

TABLE C3.12 ANOVA for False Alarms (log transformed) in Group 3.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.03267 | 1 | 0.03267 | 0.1252 | ns |
| SWG | 2.08733 | 8 | 0.26092 | | |
| B (Tasks) | 0.02400 | 1 | 0.02400 | 0.1982 | ns |
| AB | 0.00067 | 1 | 0.00067 | 0.0055 | ns |
| B x SWG | 0.96867 | 8 | 0.12108 | | |
| C (Time Blocks) | 0.04900 | 2 | 0.02450 | 0.3467 | ns |
| AC | 0.05033 | 2 | 0.02517 | 0.3561 | ns |
| C x SWG | 1.13067 | 16 | 0.07067 | | |
| BC | 0.15100 | 2 | 0.07550 | 2.2152 | ns |
| ABC | 0.02033 | 2 | 0.01017 | 0.2983 | ns |
| BC x SWG | 0.54533 | 16 | 0.03408 | | |

TABLE C3.13 ANOVA for d' in Group 3

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.00067 | 1 | 0.00067 | 0.0084 | ns |
| SWG | 0.63779 | 8 | 0.07972 | | |
| B (Tasks) | 0.00963 | 1 | 0.00963 | 0.6317 | ns |
| AB | 0.00113 | 1 | 0.00113 | 0.0739 | ns |
| B x SWG | 0.12191 | 8 | 0.01524 | | |
| C (Time Blocks) | 0.53356 | 2 | 0.26678 | 15.7995 | .01 |
| AC | 0.04140 | 2 | 0.02120 | 1.2260 | ns |
| C x SWG | 0.27017 | 16 | 0.01689 | | |
| BC | 0.02266 | 2 | 0.01133 | 0.6815 | ns |
| ABC | 0.01862 | 2 | 0.00931 | 0.5600 | ns |
| BC x SWG | 0.26605 | 16 | 0.01663 | | |

TABLE C3.14 ANOVA for log B in Group 3

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 4932.27 | 1 | 2932.27 | 0.3012 | ns |
| SWG | 131015.67 | 8 | 16376.96 | | |
| B (Tasks) | 8213.40 | 1 | 8213.40 | 1.5252 | ns |
| AB | 187.23 | 1 | 187.23 | 0.0348 | ns |
| B x SWG | 43082.33 | 8 | 5385.29 | | |
| C (Time Blocks) | 36093.23 | 2 | 18046.62 | 31.9900 | .001 |
| AC | 1562.63 | 2 | 781.32 | 1.3850 | ns |
| C x SWG | 9026.13 | 16 | 564.13 | | |
| BC | 191.10 | 2 | 95.55 | 0.1413 | ns |
| ABC | 834.63 | 2 | 417.32 | 0.6171 | ns |
| BC x SWG | 10820.27 | 16 | 676.27 | | |

TABLE C3.15 ANOVA for Detection Latency in Group 3

## APPENDIX C4 : Experiment 4

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.18881 | 1 | 0.18881 | 1.6439 | ns |
| B (Signal Probability) | 0.11781 | 1 | 0.11781 | 1.0258 | ns |
| AB | 0.03745 | 1 | 0.03745 | 0.3261 | ns |
| SWG | 4.13477 | 36 | 0.11485 | | |
| C (Time Blocks) | 0.01827 | 2 | 0.00914 | 0.7808 | ns |
| AC | 0.00990 | 2 | 0.00495 | 0.4231 | ns |
| BC | 0.00153 | 2 | 0.00077 | 0.0655 | ns |
| ABC | 0.00080 | 2 | 0.00040 | 0.0343 | ns |
| C x SWG | 0.84243 | 72 | 0.01170 | | |

TABLE C4.1 SP ANOVA for $d_a$ in Conditions I and II.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.01951 | 1 | 0.01951 | 0.2406 | ns |
| B (Event Rate) | 0.00954 | 1 | 0.00954 | 0.1177 | ns |
| AB | 0.28130 | 1 | 0.28130 | 3.4701 | ns |
| SWG | 2.91828 | 36 | 0.08106 | | |
| C (Time Blocks) | 0.43956 | 2 | 0.21978 | 14.7217 | .01 |
| AC | 0.09640 | 2 | 0.04820 | 3.2284 | ns |
| BC | 0.12606 | 2 | 0.06303 | 4.2220 | .05 |
| ABC | 0.10722 | 2 | 0.05361 | 3.5910 | ns |
| C x SWG | 1.07489 | 72 | 0.01493 | | |

TABLE C4.2 ER ANOVA for log B1 in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.21590 | 1 | 0.21590 | 4.5804 | .05 |
| B (Signal Probability) | 0.25669 | 1 | 0.25669 | 5.4457 | .05 |
| AB | 0.00547 | 1 | 0.00547 | 0.1160 | ns |
| SWG | 1.69688 | 36 | 0.04714 | | |
| C (Time Blocks) | 0.58981 | 2 | 0.29490 | 31.0095 | .001 |
| AC | 0.06293 | 2 | 0.03146 | 3.3081 | ns |
| BC | 0.04824 | 2 | 0.02412 | 2.5363 | ns |
| ABC | 0.01178 | 2 | 0.00589 | 0.6193 | ns |
| C x SWG | 0.68472 | 72 | 0.00951 | | |

TABLE C4.3 SP ANOVA for log B1 in Conditions I and II.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.01200 | 1 | 0.01200 | 0.1879 | ns |
| B (Event Rate) | 0.14700 | 1 | 0.14700 | 2.3015 | ns |
| AB | 0.00000 | 1 | 0.00000 | 0.0000 | ns |
| SWG | 2.29933 | 36 | 0.06387 | | |
| C (Time Blocks) | 0.22173 | 2 | 0.11086 | 9.6574 | .05 |
| AC | 0.07850 | 2 | 0.03925 | 3.4191 | ns |
| BC | 0.16958 | 2 | 0.08479 | 7.3861 | .01 |
| ABC | 0.06753 | 2 | 0.03376 | 2.9412 | ns |
| C x SWG | 0.82653 | 72 | 0.01148 | | |

TABLE C4.4 ER ANOVA for log B2 in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.02523 | 1 | 0.02523 | 0.2995 | ns |
| B (Signal Probability) | 0.63365 | 1 | 0.63365 | 7.5216 | .01 |
| AB | 0.00261 | 1 | 0.00261 | 0.0310 | ns |
| SWG | 3.03282 | 36 | 0.08425 | | |
| C (Time Blocks) | 0.81498 | 2 | 0.40749 | 33.3137 | .001 |
| AC | 0.02981 | 2 | 0.01490 | 1.2183 | ns |
| BC | 0.00386 | 2 | 0.00193 | 0.1579 | ns |
| ABC | 0.03365 | 2 | 0.01683 | 1.3756 | ns |
| C x SWG | 0.88070 | 72 | 0.01223 | | |

TABLE C4.5 SP ANOVA for log B2 in Conditions I and II.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.97560 | 1 | 0.97560 | 10.2370 | .01 |
| B (Event Rate) | 1.46523 | 1 | 1.46523 | 15.3747 | .001 |
| AB | 0.07803 | 1 | 0.09530 | 0.8188 | ns |
| SWG | 3.43085 | 36 | 0.09530 | | |
| C (Time Blocks) | 0.16094 | 2 | 0.08047 | 5.7804 | .05 |
| AC | 0.00931 | 2 | 0.00466 | 0.3345 | ns |
| BC | 0.07369 | 2 | 0.03684 | 2.6466 | ns |
| ABC | 0.11792 | 2 | 0.05896 | 4.2353 | .05 |
| C x SWG | 1.00229 | 72 | 0.01392 | | |

TABLE C4.6 ER ANOVA for log B3 in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 0.00954 | 1 | 0.00954 | 0.0911 | ns |
| B (Signal Probability) | 1.43664 | 1 | 1.43664 | 13.7239 | .001 |
| AB | 0.64974 | 1 | 0.64974 | 6.2068 | .05 |
| SWG | 3.76854 | 36 | 0.10468 | | |
| C (Time Blocks) | 0.28408 | 2 | 0.14204 | 3.2993 | ns |
| AC | 0.06701 | 2 | 0.03351 | 0.7786 | ns |
| BC | 0.02017 | 2 | 0.01009 | 0.2344 | ns |
| ABC | 0.01622 | 2 | 0.00811 | 0.1884 | ns |
| C x SWG | 3.09888 | 72 | 0.04304 | | |

TABLE C4.7 SP ANOVA for log B3 (lax criterion) in Conditions I and II.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 211848.03 | 1 | 211848.03 | 26.4126 | .001 |
| B (Event Rate) | 28520.83 | 1 | 28520.83 | 3.5559 | ns |
| AB | 63572.03 | 1 | 63572.03 | 7.9260 | .01 |
| SWG | 288745.47 | 36 | 8020.71 | | |
| C (Time Blocks) | 48418.85 | 2 | 24209.43 | 41.5734 | .001 |
| AC | 37.32 | 2 | 18.66 | 0.0320 | ns |
| BC | 2359.62 | 2 | 1179.81 | 2.0260 | ns |
| ABC | 1833.82 | 2 | 916.91 | 1.5746 | ns |
| C x SWG | 41927.73 | 72 | 582.33 | | |
| Simple effects | | | | | |
| B at A1 (Speed Task) | 88627.00 | 1 | 88627.00 | 11.0498 | .01 |
| B at A2 (Closure Task) | 3466.00 | 1 | 3466.00 | 0.4321 | ns |

TABLE C4.8 ER ANOVA for Detection Latency in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 22742.53 | 1 | 22742.53 | 2.6026 | ns |
| B (Signal Probability) | 120967.50 | 1 | 120967.50 | 13.9388 | .001 |
| AB | 3286.53 | 1 | 3286.53 | 0.3787 | ns |
| SWG | 312424.47 | 36 | 8678.46 | | |
| C (Time Blocks) | 19150.47 | 2 | 9575.23 | 18.9542 | .001 |
| AC | 2666.67 | 2 | 1333.33 | 2.6393 | ns |
| BC | 1288.60 | 2 | 643.30 | 1.2734 | ns |
| ABC | 428.87 | 2 | 214.43 | 0.4245 | ns |
| C x SWG | 36372.73 | 72 | 505.18 | | |

TABLE C4.9 SP ANOVA for Detection Latency in Conditions I and II.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 440.83 | 1 | 440.83 | 0.0217 | ns |
| B (Event Rate) | 6453.33 | 1 | 6453.33 | 0.3175 | ns |
| AB | 9612.30 | 1 | 9612.30 | 0.4726 | ns |
| SWG | 732079.00 | 36 | 20335.53 | | |
| C (Time Blocks) | 47947.22 | 2 | 23973.61 | 11.9938 | .01 |
| AC | 1351.12 | 2 | 675.56 | 0.3380 | ns |
| BC | 8207.82 | 2 | 4103.91 | 2.0532 | ns |
| ABC | 286.25 | 2 | 143.13 | 0.0716 | ns |
| C x SWG | 143915.60 | 72 | 1998.83 | | |

TABLE C4.10 ER ANOVA for False Alarm Latency in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 13824.53 | 1 | 13824.53 | 0.8066 | ns |
| B (Signal Probability) | 66176.03 | 1 | 66176.03 | 3.8610 | ns (.10) |
| AB | 2.13 | 1 | 2.13 | 0.0001 | ns |
| SWG | 617023.13 | 36 | 17139.53 | | |
| C (Time Blocks) | 69213.02 | 2 | 34606.51 | 15.2478 | .001 |
| AC | 5970.32 | 2 | 2985.16 | 1.3176 | ns |
| BC | 3415.52 | 2 | 1707.76 | 0.7538 | ns |
| ABC | 1543.82 | 2 | 771.91 | 0.3407 | ns |
| C x SWG | 163122.67 | 72 | 2265.59 | | |

TABLE C4.11 SP ANOVA for False Alarm Latency in Conditions I and II.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 3808.13 | 1 | 3808.13 | 0.2250 | ns |
| B (Event Rate) | 2520.83 | 1 | 2520.83 | 0.1490 | ns |
| AB | 1672.53 | 1 | 1672.53 | 0.0988 | ns |
| SWG | 609247.13 | 36 | 16923.53 | | |
| C (Time Blocks) | 775.72 | 2 | 387.85 | 0.4806 | ns |
| AC | 1729.32 | 2 | 864.66 | 1.0715 | ns |
| BC | 2078.12 | 2 | 1039.06 | 1.2876 | ns |
| ABC | 440.52 | 2 | 220.26 | 0.2729 | ns |
| C x SWG | 5810.67 | 72 | 806.97 | | |

TABLE C4.12 ER ANOVA for Correct Rejection Latency in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 7744.13 | 1 | 7744.13 | 0.3403 | ns |
| B (Signal Probability) | 11564.03 | 1 | 11564.03 | 0.5081 | ns |
| AB | 4514.13 | 1 | 4514.13 | 0.1984 | ns |
| SWG | 819292.20 | 36 | 22758.12 | | |
| C (Time Blocks) | 3245.07 | 2 | 1622.53 | 1.8655 | ns |
| AC | 2213.07 | 2 | 1106.53 | 1.2709 | ns |
| BC | 305.87 | 2 | 152.93 | 0.1756 | ns |
| ABC | 3636.27 | 2 | 1818.13 | 2.0881 | ns |
| C x SWG | 62690.40 | 72 | 870.70 | | |

TABLE C4.13 SP ANOVA for Correct Rejection Latency in Conditions I and II.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 10697.46 | 1 | 10697.46 | 0.7775 | ns |
| B (Event Rate) | 76457.01 | 1 | 76456.01 | 5.5571 | .05 |
| AB | 21094.01 | 1 | 21094.01 | 1.5332 | ns |
| SWG | 495300.03 | 36 | 13758.33 | | |
| C (Time Blocks) | 7746.52 | 2 | 3873.26 | 4.2842 | .05 |
| AC | 1335.62 | 2 | 667.81 | 0.7387 | ns |
| BC | 387.92 | 2 | 193.96 | 0.2145 | ns |
| ABC | 730.02 | 2 | 365.01 | 0.4037 | ns |
| C x SWG | 65093.27 | 72 | 904.07 | | |

TABLE C4.14 ER ANOVA for Omission Latency in Conditions II and III.

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Tasks) | 20228.87 | 1 | 20228.87 | 1.9598 | ns |
| B (Signal Probability) | 153940.03 | 1 | 153940.03 | 14.9142 | .001 |
| AB | 10083.33 | 1 | 10083.33 | 0.9769 | ns |
| SWG | 371582.47 | 36 | 10321.74 | | |
| C (Time Blocks) | 11873.52 | 2 | 5936.76 | 13.9697 | .001 |
| AC | 650.22 | 2 | 325.11 | 0.7650 | ns |
| BC | 850.52 | 2 | 425.26 | 1.0007 | ns |
| ABC | 157.62 | 2 | 78.81 | 0.1854 | ns |
| C x SWG | 30598.13 | 72 | 424.97 | | |

TABLE C4.15 SP ANOVA for Omission Latency in Conditions I and II.

## Appendix C5 : Experiment 5

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.4167 | 1 | 0.4167 | 0.0846 | ns |
| SWG | 39.4000 | 8 | 4.9250 | | |
| B (Tasks) | 2.0167 | 1 | 2.0167 | 1.7664 | ns |
| AB | 1.3500 | 1 | 1.3500 | 1.1825 | ns |
| B x SWG | 9.1333 | 8 | 1.1417 | | |
| C (Time Blocks) | 49.7333 | 2 | 24.8667 | 39.7867 | .001 |
| AC | 0.0933 | 2 | 0.4667 | 0.7467 | ns |
| C x SWG | 10.0000 | 16 | 0.6250 | | |
| BC | 0.1333 | 2 | 0.0667 | 0.2105 | ns |
| ABC | 2.2170 | 2 | 1.1085 | 3.5002 | ns |
| BC x SWG | 5.0667 | 16 | 0.3167 | | |

TABLE C5.1 ANOVA for Hits in Group 1

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.02440 | 1 | 0.02440 | 0.1654 | ns |
| SWG | 1.18031 | 8 | 0.14754 | | |
| B (Tasks) | 0.09520 | 1 | 0.09520 | 5.8035 | .05 |
| AB | 0.00888 | 1 | 0.00888 | 0.5414 | ns |
| B x SWG | 0.13123 | 8 | 0.01640 | | |
| C (Time Blocks) | 2.73187 | 2 | 1.36594 | 61.7744 | .001 |
| AC | 0.00444 | 2 | 0.00222 | 0.1005 | ns |
| C x SWG | 0.35379 | 16 | 0.02211 | | |
| BC | 0.02994 | 2 | 0.01497 | 0.7726 | ns |
| ABC | 0.02822 | 2 | 0.01411 | 0.7282 | ns |
| BC x SWG | 0.31007 | 16 | 0.01938 | | |

TABLE C5.2 ANOVA for False Alarms (log transformed) in Group 1

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.12604 | 1 | 0.12604 | 0.3841 | ns |
| SWG | 2.62543 | 8 | 0.32818 | | |
| B (Tasks) | 0.00204 | 1 | 0.00204 | 0.0248 | ns |
| AB | 0.04108 | 1 | 0.04108 | 0.4997 | ns |
| B x SWG | 0.65773 | 8 | 0.08222 | | |
| C (Time Blocks) | 0.05187 | 2 | 0.02594 | 0.5127 | ns |
| AC | 0.05574 | 2 | 0.02787 | 0.5510 | ns |
| C x SWG | 0.80929 | 16 | 0.05058 | | |
| BC | 0.10184 | 2 | 0.05092 | 2.3469 | ns |
| ABC | 0.03710 | 2 | 0.01855 | 0.8550 | ns |
| BC x SWG | 0.34715 | 16 | 0.02169 | | |

TABLE C5.3 ANOVA for d' in Group 1

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.00913 | 1 | 0.00913 | 0.0755 | ns |
| SWG | 0.96761 | 8 | 0.12095 | | |
| B (Tasks) | 0.06403 | 1 | 0.06403 | 2.9503 | ns |
| AB | 0.01176 | 1 | 0.01176 | 0.5419 | ns |
| B x SWG | 0.17361 | 8 | 0.0217 | | |
| C (Time Blocks) | 2.30604 | 2 | 1.15302 | 66.8935 | .001 |
| AC | 0.02370 | 2 | 0.01185 | 0.6876 | ns |
| C x SWG | 0.27579 | 16 | 0.01724 | | |
| BC | 0.00026 | 2 | 0.00013 | 0.0079 | ns |
| ABC | 0.01519 | 2 | 0.00760 | 0.4532 | ns |
| BC x SWG | 0.26815 | 16 | 0.01676 | | |

TABLE C5.4 ANOVA for log (B+1) in Group 1

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 2.0167 | 1 | 2.0167 | 0.2895 | ns |
| SWG | 55.7333 | 8 | 6.9667 | | |
| B (Tasks) | 12.1500 | 1 | 12.1500 | 6.9429 | .05 |
| AB | 0.0167 | 1 | 0.0167 | 0.0095 | ns |
| B x SWG | 14.0000 | 8 | 1.7500 | | |
| C (Time Blocks) | 15.8333 | 2 | 7.9167 | 4.5455 | .10 |
| AC | 0.6333 | 2 | 0.3167 | 0.1818 | ns |
| C x SWG | 27.8667 | 16 | 1.7417 | | |
| BC | 1.3000 | 2 | 0.6500 | 1.5294 | ns |
| ABC | 2.2333 | 2 | 1.1167 | 2.6275 | ns |
| BC x SWG | 6.8000 | 16 | 0.4250 | | |

TABLE C5.5 ANOVA for Hits in Group 2

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.00024 | 1 | 0.00024 | 0.0009 | ns |
| SWG | 2.12183 | 8 | 0.26523 | | |
| B (Tasks) | 0.67416 | 1 | 0.67416 | 6.4494 | .05 |
| AB | 0.22083 | 1 | 0.22083 | 2.1126 | ns |
| B x SWG | 0.83625 | 8 | 0.10453 | | |
| C (Time Blocks) | 1.00153 | 2 | 0.50077 | 7.8817 | .05 |
| AC | 0.12757 | 2 | 0.06379 | 1.0039 | ns |
| C x SWG | 1.01657 | 16 | 0.06354 | | |
| BC | 0.05383 | 2 | 0.02692 | 0.9921 | ns |
| ABC | 0.10696 | 2 | 0.05348 | 1.9713 | ns |
| BC x SWG | 0.43407 | 16 | 0.02713 | | |

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.17713 | 1 | 0.17713 | 0.4176 | ns |
| SWG | 3.39348 | 8 | 0.42419 | | |
| B (Tasks) | 0.00600 | 1 | 0.00600 | 0.0690 | ns |
| AB | 0.05046 | 1 | 0.05046 | 0.5806 | ns |
| B x SWG | 0.69531 | 8 | 0.08691 | | |
| C (Time Blocks) | 0.01159 | 2 | 0.00580 | 0.3022 | ns |
| AC | 0.00196 | 2 | 0.00098 | 0.0512 | ns |
| C x SWG | 0.30678 | 16 | 0.01914 | | |
| BC | 0.03823 | 2 | 0.01912 | 1.0003 | ns |
| ABC | 0.01675 | 2 | 0.00838 | 0.4385 | ns |
| BC x SWG | 0.30575 | 16 | 0.01911 | | |

TABLE C5.7 ANOVA for d' in Group 2

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.00104 | 1 | 0.00104 | 0.0062 | ns |
| SWG | 1.34286 | 8 | 0.16786 | | |
| B (Tasks) | 0.46288 | 1 | 0.46288 | 6.3532 | .05 |
| AB | 0.21004 | 1 | 0.21004 | 2.8829 | ns |
| B x SWG | 0.58286 | 8 | 0.07286 | | |
| C (Time Blocks) | 0.81433 | 2 | 0.40717 | 5.7867 | .05 |
| AC | 0.12000 | 2 | 0.06000 | 0.8528 | ns |
| C x SWG | 1.12580 | 16 | 0.07036 | | |
| BC | 0.06634 | 2 | 0.03317 | 1.3408 | ns |
| ABC | 0.04908 | 2 | 0.02454 | 0.9920 | ns |
| BC x SWG | 0.39584 | 16 | 0.02474 | | |

TABLE C5.8 ANOVA for log (B+1) in Group 2

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.0000 | 1 | 0.0000 | 0.0000 | ns |
| SWG | 38.7333 | 8 | 4.8417 | | |
| B (Tasks) | 5.4000 | 1 | 5.4000 | 1.3935 | ns |
| AB | 0.2667 | 1 | 0.2667 | 0.0688 | ns |
| B x SWG | 31.0000 | 8 | 3.8750 | | |
| C (Time Blocks) | 13.1633 | 2 | 6.8167 | 8.2211 | .05 |
| AC | 2.1000 | 2 | 1.0500 | 1.2663 | ns |
| C x SWG | 13.2667 | 16 | 0.8292 | | |
| BC | 0.7000 | 2 | 0.3500 | 0.6222 | ns |
| ABC | 0.6333 | 2 | 0.3167 | 0.5630 | ns |
| BC x SWG | 9.0000 | 16 | 0.5625 | | |

TABLE C5.9 ANOVA for Hits in Group 3

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.25350 | 1 | 0.25350 | 0.6210 | ns |
| SWG | 3.26593 | 8 | 0.40824 | | |
| B (Tasks) | 0.58806 | 1 | 0.58806 | 2.8925 | ns |
| AB | 0.24833 | 1 | 0.24833 | 1.2214 | ns |
| B x SWG | 1.62645 | 8 | 0.20331 | | |
| C (Time Blocks) | 0.85969 | 2 | 0.42985 | 13.4877 | .01 |
| AC | 0.00556 | 2 | 0.00278 | 0.0872 | ns |
| C x SWG | 0.50991 | 16 | 0.03187 | | |
| BC | 0.00688 | 2 | 0.00344 | 0.1119 | ns |
| ABC | 0.01777 | 2 | 0.00889 | 0.2890 | ns |
| BC x SWG | 0.49191 | 16 | 0.03074 | | |

TABLE C5.10 ANOVA for False Alarms (log transformed) in Group 3

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.26136 | 1 | 0.26136 | 1.7523 | ns |
| SWG | 1.19320 | 8 | 0.14915 | | |
| B (Tasks) | 0.15811 | 1 | 0.15811 | 0.8044 | ns |
| AB | 0.05046 | 1 | 0.05046 | 0.2567 | ns |
| B x SWG | 1.57247 | 8 | 0.19656 | | |
| C (Time Blocks) | 0.02804 | 2 | 0.01402 | 0.8128 | ns |
| AC | 0.02779 | 2 | 0.01390 | 0.8059 | ns |
| C x SWG | 0.27588 | 16 | 0.01725 | | |
| BC | 0.04834 | 2 | 0.02417 | 1.1983 | ns |
| ABC | 0.00469 | 2 | 0.00235 | 0.1163 | ns |
| BC x SWG | 0.32273 | 16 | 0.02017 | | |

TABLE C5.11 ANOVA for d' in Group 3

| SOURCE | SS | DF | MS | F | p |
|---|---|---|---|---|---|
| A (Order) | 0.11094 | 1 | 0.11094 | 0.4945 | ns |
| SWG | 1.79489 | 8 | 0.22436 | | |
| B (Tasks) | 0.27473 | 1 | 0.27473 | 2.4633 | ns |
| AB | 0.06403 | 1 | 0.06403 | 0.5741 | ns |
| B x SWG | 0.89221 | 8 | 0.11153 | | |
| C (Time Blocks) | 0.74825 | 2 | 0.37413 | 12.6526 | .01 |
| AC | 0.01204 | 2 | 0.00602 | 0.2036 | ns |
| C x SWG | 0.43711 | 16 | 0.02957 | | |
| BC | 0.01457 | 2 | 0.00729 | 0.4100 | ns |
| ABC | 0.05937 | 2 | 0.02969 | 1.6702 | ns |
| BC x SWG | 0.28439 | 16 | 0.01777 | | |

TABLE C5.12 ANOVA for log (B+1) in Group 3

# APPENDIX D A COMPARISON OF SINGLE, BINARY AND RATING RESPONSE MODES

In the experiments reported in the main part of the thesis, three different response procedures have been employed. In some studies (Experiments 3, 5, 7 and 8) the 'conventional' single mode of response was used, where subjects were required to respond 'Detect' only when they thought they had detected a signal. Experiments 1 and 6 used the binary mode of response, with subjects being required to respond either 'Yes' or 'No' to both signal and non-signal events. This procedure is equivalent to the Yes/No paradigm in signal detection experiments. Finally, in two studies (Experiments 2 and 4), a four category rating scale was used. Although these studies were concerned with the investigation of different aspects of monitoring performance, they were designed so that some of them shared at least one experimental condition in common, enabling post-hoc comparisons of performance levels across response modes.

A comparison of single and binary response modes may be made by examining the results of Experiments 1 and 3, while Experiments 4 and 5 may be compared for differences between the single and rating modes. Lastly a comparison of the binary and rating modes may be achieved by comparing Experiments 1 and 2.

Comparative performance data for the three response modes obtained in the results of Experiments 1 to 5 are given in Table D.1, for the single-binary, single-rating, and binary-rating comparisons. The original data were also examined for possible differences between response modes in the performance trends over time, but since no discriminable differential trends were observed, only mean performance levels averaged across time blocks are given in Table D.1.

A. SINGLE - BINARY MODE COMPARISON
(EXPERIMENTS 1 AND 3)

| | Visual Speed Task | | Visual Closure Task | |
|---|---|---|---|---|
| | Single | Binary | Single | Binary |
| P(H) | .751 | .716 | .731 | .740 |
| p(FA) | .040 | .037 | .040 | .037 |
| d' | 2.53 | 2.48 | 2.46 | 2.51 |
| Log B | .587 | .687 | .618 | .682 |
| DL | 657 | 689 | 677 | 683 |

B. SINGLE - RATING MODE COMPARISON **
(EXPERIMENTS 4 AND 5)

| | Auditory Speed Task | | Auditory Closure Task | |
|---|---|---|---|---|
| | Single | Rating | Single | Rating |
| P(H) | .847 | .760 | .791 | .782 |
| P(FA) | .074 | .043* | .051 | .032 |
| d' | 2.61 | 2.56 | 2.59 | 2.44 |

C. BINARY - RATING MODE COMPARISON **
(EXPERIMENTS 1 AND 2)

| | Visual Speed Task | | Visual Closure Task | |
|---|---|---|---|---|
| | Binary | Rating | Binary | Rating |
| P(H) | .436 | .661 | .704 | .750 |
| P(FA) | .027 | .032 | .032 | .047 |
| d' | 1.87 | 2.17 | 2.56 | 2.43 |

TABLE D.1. Mean performance levels in corresponding experimental conditions for single, binary and rating response modes. (* Difference between response modes significant at .05 level; all other differences are non-significant; ** For rating scale metrics, mean values corresponding to the lax criterion have been given, while values of da are given for sensitivity.)

For the single and binary response modes, mean performance levels are remarkably stable across response modes for performance on both the visual speed and the closure task. The largest difference in mean performance between modes was of the order of 17% for log B (speed task). However, statistical comparisons of mean performance levels on the single and binary key tasks using Mann-Whitney U tests (Siegel, 1956) revealed no significant differences. Thus, along with similar previous results (Davies, Lang and Shackleton, 1973; Guralnick and Harvey, 1970; Whittenburg, Ross and Andrews, 1956), these results establish that the addition of a second response key to the single key that is usually employed in monitoring tasks does not exert any significant effect on performance.

No significant differences in the mean probability of Hits and in d' between the single and rating response modes were observed. However, it was found that the mean probability of False Alarms was significantly higher, for the speed task, with the single response mode than with the rating mode ($p < .05$). These results are, if anything, the opposite to that expected, since it has been previously reported that the addition of a rating scale serves to increase the False Alarm rate (Guralnick and Harvey, 1970; Loeb and Binford, 1964). It is also surprising that although there were no significant differences, the mean probability of Hits was also lower for the rating response mode. However, the important result relates to the finding of no significant differences for the sensitivity measure between response modes.

Finally, no significant differences between modes were obtained for the binary-rating comparison of the results of Experiments 1 and 2. This is apparent in Table D.1, which also indicates that somewhat lower sensitivity values were obtained in this case since the comparisons were made for a high event rate of 30/min.

These results therefore indicate that the different response procedures used to obtain monitoring performance indices are equivalent, for both visual and auditory speed and closure tasks. There is some evidence that the rating method is not equivalent to the single key response mode as far the proportion of False Alarms are concerned. However, mean sensitivity values are very similar for all three response modes, and performance levels are almost identical for the single and binary response modes.

## APPENDIX E SEX DIFFERENCES IN MONITORING PERFORMANCE

Davies and Tune (1970) have pointed out that in many investigations of monitoring performance in which both male and female subjects are employed, researchers often assume that there are no sex differences in performance. In this thesis, male and female subjects participated in two experiments (4 and 7), and thus it is of interest to examine whether there were any sex differences in performance in these studies.

The raw performance scores in Experiments 4 and 7 were partitioned by sex and examined for possible differences between males and females in both the mean level of performance and in the performance trends over time. Since no substantial sex differences in monitoring performance in different time blocks were apparent, only mean performance data, averaged over the monitoring period, are reported here. Table E.1 lists the mean values of sensitivity and criterion for male and female subjects in Experiments 4 and 7. Male and female performance scores in each condition were compared using independent Mann-Whitney U tests. On the whole, no reliable differences in performance attributable to sex were obtained. In three cases, an effect of sex was obtained, but the small size of the effects and the low sample size (N = 5 in Experiment 4, N = 6 in Experiment 7) indicate that these are probably chance results arising out of the low power of the tests. When male and female subjects were compared without regard to experimental condition (N = 30 in Experiment 4, N = 18 in Experiment 7), no significant differences due to sex were obtained (see 'Mean' column in Table E.1). Thus, although male subjects had slightly larger d', log B scores than female subjects (overall, they detected slightly more signals and male slightly fewer false alarms), there were no reliable differences in performance between male and females.

| Experiment 4 | | Speed Task I | II | III | Closure Task I | II | III | Mean |
|---|---|---|---|---|---|---|---|---|
| $d_a$ | Males | 2.49 | 2.57 | 2.01 * | 2.46 | 2.45 | 2.46 | 2.41 |
| | Females | 2.44 | 2.55 | 1.80 | 2.37 | 2.44 | 2.52 | 2.35 |
| $\log B^1$ | Males | 1.17 * | 1.00 | 1.06 | 1.17 | 1.05 | 1.06 | 1.08 |
| | Females | 1.00 | 0.90 | 0.98 | 1.07 | 0.88 | 1.02 | 0.96 |

| Experiment 7 | | MS Task | MSC Task | MC Task | Mean |
|---|---|---|---|---|---|
| d' | Males | 2.14 | 2.28 | 2.35 | 2.26 |
| | Females | 2.07 | 2.41 | 2.35 | 2.28 |
| log B | Males | 0.99 | 0.81 * | 0.92 | 0.91 |
| | Females | 0.85 | 0.99 | 0.87 | 0.90 |

TABLE E.1 The mean values of d' and log B for male and female subjects in Experiments 4 and 7 (* difference between males and females significant, $p < .05$; 1 values for the central or medium criterion).

These results are consistent with previous studies in which performance differences between male and female subjects have been reported (see Davies and Tune, 1970). These studies, like the present brief report, have generally not treated sex as the primary variable of interest, and have used relatively small sample sizes. Recently, however, Waag, Halcomb and Tyler (1973) carried out a study specifically aimed at examining the main effects of sex on monitoring performance. They used a large sample size (N = 220) and a visual monitoring task of the type used by Jerison and Pickett (1964). Although it was found that male subjects detected more signals and made fewer false alarms, the magnitude of the effects were very small; sex differences accounted for only 4% of the variance in detections and less than 1% in false alarms. Waag et al., (1973) did not report sensitivity and bias data, but their results for detections and false alarms were the same as in the two experiments reported here.

We may conclude that sex differences were not a significant source of variance in the experiments reported in this thesis.

# APPENDIX F  ESTIMATION OF HIT AND FALSE ALARM PROBABILITIES

In signal detection and monitoring tasks, performance may be such that no false alarms or no omissions are elicited during a given detection period. Alternatively, one may wish to obtain performance metrics for small sub-sections of the detection period during which such 'perfect' performance is obtained, even though the mean rate of errors may be high. These two cases pose problems for the estimation of the operating Hit and False Alarm (FA) probabilities, which might be required for the computation of sensitivity and bias indices.

The experimental technique of estimating the probability of a response involves measuring the relative frequency of that response over a number of trials. Relative frequency tends 'in probability' to the actual response probability as the number of trials becomes large. If a given detection period is therefore 'blank' with respect to FAs or omissions, some other technique must be used to estimate probability since relative frequency is zero.

The estimation technique presented here considers the likelihood of no response (of any type) occurring within a certain detection period. If the 'true' probability of a response is P(R) the probability of not obtaining a response in a single trial is

$$1 - P(R)$$

Assuming that this probability is fixed, and that successive trials are independent, the probability of no responses in n trials is

$$(1 - P(R))^n$$

The likelihood of this can be tested by setting

$$(1 - P(R))^n \leqslant \tfrac{1}{2}$$

$$\text{or, Min } (P(R)) = 1 - \frac{1}{2^{1/n}}$$

This formula can thus be to estimate Hit and FA probabilities in 'blank' periods. For Hits, n is set to the number of signal presentations, and Hit probability is taken as 1 - P(R). For FAs, n is set to the number of non-signals, and p(FA) is taken as P(R).

## APPENDIX G ESTIMATION OF LUMINANCE VALUES USING AN EXPOSURE METER

In the visual tasks used in the several experiments reported in this thesis, standardization of the luminance of visual sources was often necessary to preserve the equality of experimental conditions for all subjects. In the later experiments this was always done with the aid of a spot photometer, with which the luminance of both direct and background sources could be obtained. For Experiments 1, 2 and 3, however, each of which employed visual tasks, a suitable photometer was not available at the time of investigation. A technique was thus developed whereby approximate luminance values could be obtained using a simple exposure meter of the type used for indoor and outdoor photography.

The spectral response of the CIE (International Commission on Illumination) 'standard observer' describes a hump shaped inverted-U function, with the response falling off to zero at either end of the visible frequency spectrum. This is illustrated in Figure G.1. The spectral response of the standard Selenium photocell used in exposure meters is not the same as the CIE standard observer response. However we can utilize the fact that the difference in the responses is a fairly linear function of frequency to estimate the luminance of broad-band sources. Both incandescent and flourescent sources have, typically, fairly broad spectra. By combining the response characteristics of a Selenium cell for flourescent and incandescent illumination with the typical output characteristics for such sources, an empirical equation giving intensity levels in log ft.L. for exposure readings was derived. The equation and its use is described below Figure G.1.

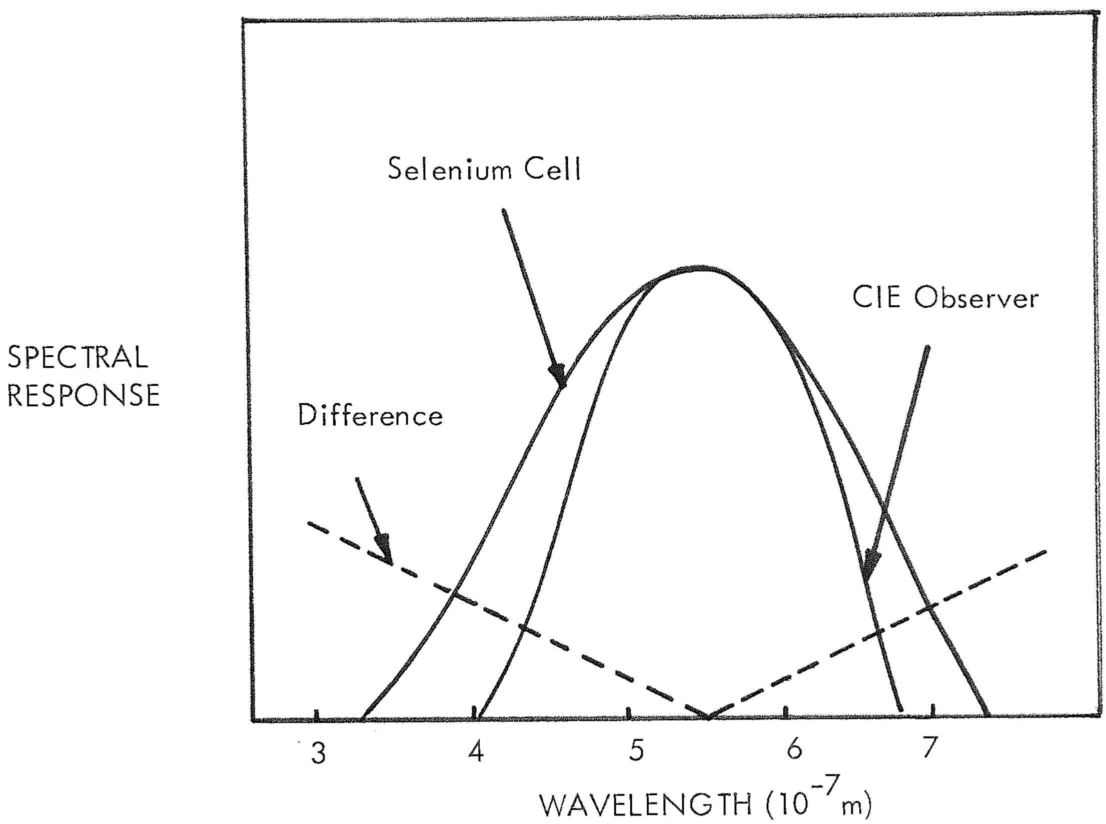


FIGURE G.1 Spectral response of selenium cell and CIE observer.

---

$$\text{Log}_{10}\,\text{Intensity} = K + 0.6\,\text{Log}_2\,F_{stop} - \text{Log}_{10}\,t_e$$

Notes:

1. Intensity is in log foot lamberts, with $t_e$ (exposure time) in milliseconds.
2. K is 1.6 for an incandescent source, and 1.93 for a flourescent source.
3. The above equation holds for ASA 1000 (DIN 21). For other sensitivity (ASA) ranges, increase or decrease readings by $\log_{10}100 - \log_{10}ASA$.
4. The equation holds only for selenium cell exposure meters and for reasonably broad band sources.

---

## APPENDIX H LIST OF PUBLICATIONS

Note

Portions of the experimental work of this thesis are also reported in five of the following papers. The number(s) given at the end of each reference indicate the associated experiment(s).


1. DAVIES D.R. and PARASURAMAN R. (1976) Cortical evoked potentials and vigilance: A decision theory analysis. Symposium on Vigilance: Relationships between Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy (in press) (6).

2. PARASURAMAN R. (1975a) Response bias and physiological reactivity. Journal of Psychology, 91, 309-313.

3. PARASURAMAN R. (1975b) Response latencies in visual monitoring. Perceptual and Motor Skills, 40, 636.

4. PARASURAMAN R. (1976) Consistency of individual differences in human performance: An abilities classification analysis. Journal of Applied Psychology, 60 (in press) (3).

5. PARASURAMAN R. and DAVIES D.R. (1975) Response and evoked potential latencies associated with commission errors in visual monitoring. Perception and Psychophysics, 17, 465-468. (6).

6. PARASURAMAN R. and DAVIES D.R. (1976a) Decision theory analysis of response latencies in vigilance. Journal of Experimental Psychology: Human Perception and Performance, 2 (in press) (1).

7. PARASURAMAN R. and DAVIES D.R. (1976b) A taxonomic analysis of vigilance performance. Symposium on Vigilance: Relationships between Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy (in press) (3, 5, 7).